



Published in final edited form as:

*Neuroimage*. 2021 November ; 243: 118482. doi:10.1016/j.neuroimage.2021.118482.

## Deep learning-based parameter estimation in fetal diffusion-weighted MRI

Davood Karimi<sup>a,\*</sup>, Camilo Jaimes<sup>a</sup>, Fedel Machado-Rivas<sup>a</sup>, Lana Vasung<sup>b</sup>, Shadab Khan<sup>a</sup>, Simon K. Warfield<sup>a</sup>, Ali Gholipour<sup>a</sup>

<sup>a</sup>Computational Radiology Laboratory (CRL), Department of Radiology, Boston Children's Hospital, and Harvard Medical School, USA

<sup>b</sup>Department of Pediatrics at Boston Children's Hospital, and Harvard Medical School, Boston, Massachusetts, USA

### Abstract

Diffusion-weighted magnetic resonance imaging (DW-MRI) of fetal brain is challenged by frequent fetal motion and signal to noise ratio that is much lower than non-fetal imaging. As a result, accurate and robust parameter estimation in fetal DW-MRI remains an open problem. Recently, deep learning techniques have been successfully used for DW-MRI parameter estimation in non-fetal subjects. However, none of those prior works has addressed the fetal brain because obtaining reliable fetal training data is challenging. To address this problem, in this work we propose a novel methodology that utilizes fetal scans as well as scans from prematurely-born infants. High-quality newborn scans are used to estimate accurate maps of the parameter of interest. These parameter maps are then used to generate DW-MRI data that match the measurement scheme and noise distribution that are characteristic of fetal data. In order to demonstrate the effectiveness and reliability of the proposed data generation pipeline, we used the generated data to train a convolutional neural network (CNN) to estimate color fractional anisotropy (CFA). We evaluated the trained CNN on independent sets of fetal data in terms of reconstruction accuracy, precision, and expert assessment of reconstruction quality. Results showed significantly lower reconstruction error ( $n = 100$ ,  $p < 0.001$ ) and higher reconstruction precision ( $n = 20$ ,  $p < 0.001$ ) for the proposed machine learning pipeline compared with standard estimation methods. Expert assessments on 20 fetal test scans showed significantly better overall reconstruction quality ( $p < 0.001$ ) and more accurate reconstruction of 11 regions of interest ( $p < 0.001$ ) with the proposed method.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\*Corresponding author. davood.karimi@childrens.harvard.edu (D. Karimi).

Credit authorship contribution statement

**Davood Karimi:** Conceptualization, Methodology, Software, Writing – original draft, Formal analysis, Investigation, Writing – review & editing. **Camilo Jaimes:** Validation, Formal analysis. **Fedel Machado-Rivas:** Validation, Formal analysis, Visualization. **Lana Vasung:** Validation, Formal analysis. **Shadab Khan:** Conceptualization, Methodology, Software. **Simon K. Warfield:** Supervision, Writing – original draft. **Ali Gholipour:** Supervision, Writing – original draft, Project administration, Funding acquisition.

Code availability:

All codes as well as some trained models for this manuscript are publicly available at <https://github.com/bchimagine>.

## Keywords

Fetal MRI; diffusion weighted imaging; color fractional anisotropy; deep learning

---

## 1. Introduction

In utero fetal brain imaging can provide unique insights into brain development and disorders before birth. In this regard, diffusion-weighted magnetic resonance imaging (DW-MRI) is a powerful non-invasive tool that can reveal detailed information about brain microstructure and connectivity (Basser and Pierpaoli (2011), Johansen-Berg and Behrens (2013)). Over the past two decades, many studies have used in utero DW-MRI to study normal and abnormal fetal brain development. Early works used a small number of DW measurements to compute parameters such as apparent diffusion coefficient (ADC) and fractional anisotropy (FA) (Righini et al. (2003), Bui et al. (2006), Baldoli et al. (2002)). Those works demonstrated the potential of DW-MRI for studying fetal cerebral white matter maturation and degeneration. With improved imaging and image reconstruction techniques, more recent studies have used diffusion tensor imaging (DTI) and more complex models to produce detailed pictures of fetal brain microstructure and connectivity (Khan et al. (2019), Deprez et al. (2019), Kasprian et al. (2008)).

Fetal DW-MRI faces challenges that distinguish it from non-fetal imaging: 1) the fetal and maternal motions can be significant, 2) the signal to noise ratio (SNR) is very low due to the fetal head being embedded in the mother's body, 3) imaging artifacts and geometric distortions can be significant, and 4) the number of measurements is small because scan times have to be short to minimize maternal discomfort (Deprez et al. (2019), Gholipour et al. (2014)). There have been significant progress in faster acquisition methods and more accurate fetal head tracking and slice-to-volume registration (SVR) techniques (Jiang et al. (2009), Oubel et al. (2012), Marami et al. (2017)). However, existing methods fail to fully compensate for the motion and artifacts. Therefore, the quality of data are usually not sufficient for accurate and robust estimation of parameters of interest. Moreover, previous studies have used voxel-wise least squares (LS)-based methods for parameter estimation (Zanin et al. (2011), Jakab et al. (2015), Marami et al. (2016)). As a result, the accuracy and robustness of parameter estimation in fetal DW-MRI lags far behind non-fetal imaging.

Figure 1 shows example scans from a fetal and a pre-term newborn subject of the same gestational age and their corresponding reconstructed color fractional anisotropy (CFA) images, estimated using the weighted linear least squares (WLLS) method for tensor estimation (Koay et al., 2006). The newborn scan is from the developing Human Connectome Project (dHCP) dataset (Bastiani, 2019). The dHCP subjects were imaged at 280 gradient directions. For this dataset, we estimated the SNR to be 15–20 dB. The fetal DWI volume was created from a scan acquired in our institution and processed using our motion correction and SVR pipeline (Khan et al. (2019)). In this manuscript, a DWI “volume” refers to a 4D array, where the last dimension corresponds to different diffusion-sensitizing gradients. In our fetal scans we are restricted to 24–48 measurements and the

SNR is typically in the range 5–16 dB. Consequently, the estimated fetal CFA is noisy and lacks much of the details that can be seen in the newborn CFA image of the same age.

The LS methods are based on models of diffusion signal and noise that can be unrealistic for fetal data. Moreover, since they estimate the parameter of interest on a voxel-wise basis, they fail to exploit the spatial regularity of the parameters. This shortcoming is more important in fetal DW-MRI because of low signal quality. Machine learning methods have the potential to overcome these limitations: 1) Rather than assuming a known model of signal or noise, they can learn the model from data. 2) They can exploit the spatial regularities by learning to estimate the parameter of interest based on the measurements in a neighborhood of a voxel.

Many studies have used machine learning for parameter estimation in DW-MRI. Classical machine learning methods were used to estimate white matter permeability, for fiber orientation estimation, and tractography (Neher et al. (2017), Nedjati-Gilani et al. (2017), Schultz (2012)). Recently, deep learning has received increasing attention. Golkov (2016) showed that deep learning could reduce the required number of measurements for estimating certain diffusion parameters by a factor of 12. Other studies used deep learning to estimate FA, generalized FA, mean diffusivity, neurite orientation dispersion index, and kurtosis (Gibbons (2019), Ye et al. (2019), Aliotta et al. (2019)). Several recent works have estimated diffusion tensor and fiber orientation distribution with deep learning (Karimi et al. (2021b), Tian et al. (2020), Koppers and Merhof (2016), Lin (2019)).

The above studies have used high-quality MRI datasets (Ye et al. (2019), Tian et al. (2020)) or ex-vivo brain scans and histological dissection (Nat, 2019). Such data are impossible or very costly to obtain for fetal DW-MRI. One may use fetal scans and the corresponding parameter maps estimated using standard methods. However, such parameter maps would be inaccurate. Alternatively, one may use newborn data and parameter maps. However, the measurement scheme and noise distribution are different between newborn and fetal data. For example, a lower diffusion strength is usually used in fetal imaging, which renders newborn data useless for fetal applications. Hence, because of this central limitation in obtaining *reliable training data with accurate ground-truth*, no previous work has employed machine learning for fetal DW-MRI.

Our goal is to propose a solution to this limitation, thereby enabling the use of machine learning for fetal DW-MRI. Our proposed method uses both fetal scans and high-quality pre-term newborn scans. Although neither fetal nor newborn scans on their own are adequate, we show that a combination of the two can be used to generate reliable training data. We demonstrate the effectiveness of the proposed approach by using the generated data for estimation of CFA, which is widely used in studying fetal brain micro-architecture and organization (Wakana et al. (2004), Khan et al. (2019)). We train a convolutional neural network (CNN) to estimate CFA from the generated data and compare our method with standard methods, both quantitatively and in terms of expert evaluations.

## 2. Materials and methods

### 2.1. Training data generation

Our proposed method generates reliable training data by synergistically combining the best that fetal and newborn scans offer: 1) parameter maps estimated from newborn scans, and 2) measurement scheme and noise distribution from fetal scans. Figure 2 shows our data generation pipeline. We explain this pipeline in detail below.

**2.1.1. Processing of fetal scans**—For this step we used scans of  $n = 20$  fetuses with gestational age of  $30.9 \pm 5.0$  weeks (range: [23.4,38.0]). We processed these fetal scans using our motion tracking and SVR pipeline (Marami et al., 2016). From the reconstructed DW volumes, we estimated the noise probability distribution function (PDF). In non-fetal DW-MRI, measurement noise can be accurately modeled as a Rician or non-central chi-square distribution (Canales-Rodríguez et al. (2015), Dietrich et al. (2008)). In fetal scans the noise is stronger and includes residual (uncorrected) motion errors. Since we are unaware of the true distribution of noise in the reconstructed fetal volumes, we use kernel density estimation (KDE) (Murphy, 2012) to estimate the noise PDF from data. Since the noise is known to be signal-dependent, we estimate the noise PDF for signal levels from 0.01 to 0.99 in steps of 0.01.

Figure 3 shows our approach to estimating the noise PDF. Standard SVR methods create DW volumes via registering the acquired slices to a common space and using a point spread function to assign the measurements in each slice pixel to the right voxels in the volume. All fetal scans used in this work were single-shell scans at  $b = 500$ . Depending on the available scan time and maternal discomfort, in our fetal scans we usually acquire measurements at 24–48 diffusion directions. The measurements in one voxel of an SVR-reconstructed DW-MRI volume are shown in Figure 3(a). The variation in the measurements along a certain diffusion gradient direction represent noise. We consider all the measurements,  $X = \{x_i\}_{i=1}^n$  along the same direction, with a tolerance of  $5^\circ$ , and estimate the PDF for the diffusion signal corresponding to a mean signal of  $\text{mean}(X)$  as:

$$p(x) = \frac{1}{n} \sum_{i=1}^n \mathcal{K}_h(x - x_i), \quad (1)$$

where  $\mathcal{K}_h$  is a normalized Gaussian kernel with a standard deviation of  $h = 1$ . Example estimated PDFs are shown in Figure 3(b). Furthermore, we stored the gradient tables for these 20 fetal scans.

**2.1.2. Processing of newborn scans**—There were 82 pre-term (i.e., gestational age < 38 weeks at the time of scan) newborns in the dHCP dataset (Bastiani, 2019). The gestational age of these subjects was  $35.4 \pm 2.0$  weeks (range: [29.3,38.0]). We estimated the diffusion tensor and CFA images for these scans. Each scan in this dataset included measurements in three different shells ( $b = 400, 1000, 2600$ ). Following the standard practice (Jones et al., 1999), we used the 88 measurements in the  $b = 1000$  shell from this dataset for estimating the tensor and CFA images. Note that the  $b$  value used in

this estimation would not restrict the b values of the fetal scans used at test time. This was because from the newborn scans processed in this stage we used only the estimated parameters, and not the measurements. For non-fetal diffusion tensor imaging a b value of approximately 1000 is commonly used, whereas for fetal imaging lower b values in the range [500,700] are commonly used (Kasprian et al. (2008), Khan et al. (2019)).

**2.1.3. Generating training data**—We used the fetal and newborn scans processed as described above to generate our training data. We performed the following steps to generate a training data sample, i.e., a DW-MRI volume and its corresponding CFA image:

1. Randomly select one of the fetal DW volumes,  $V^k$ .
2. Randomly select one of the newborn subjects. We use the reconstructed tensor and CFA images for this subject, which we denote as  $\text{Tensor}^k$  and  $\text{CFA}^k$ , respectively.
3. We use  $\text{CFA}^k$  as our target parameter map. We need to synthesize a fetal-style DW-MRI volume that matches this parameter map. We initialize an empty volume,  $\text{DW}^k$ , with the same size as  $\text{CFA}^k$  in the voxel space to hold the synthesized diffusion data. For each voxel in  $\text{CFA}^k$ :
  - a. Select the tensor,  $D$ , from the same voxel in  $\text{Tensor}^k$  and a gradient table,  $G$ , from a voxel in  $V^k$ .
  - b. Estimate the mean signal in each direction  $g \in G$  using the standard diffusion tensor model  $m = S S_0 = \exp(-bg^T Dg)$ , where  $S$  is the diffusion-weighted signal, and  $S_0$  is the baseline non-diffusion-weighted signal. Moreover,  $b$  is the diffusion strength, which as we have mentioned above is equal to  $b = 500$  in all fetal DW-MRI data in this work.
  - c. The measurement  $m$  generated in the above step is noise-free. To obtain noisy measurements, we add noise to  $m$  by sampling the estimated PDF corresponding to  $m$ . We do this using the inversion method (Casella and Berger, 2021). Denote with  $p_m(x)$  the probability distribution for a signal mean  $m$ . The inversion method first computes the cumulative probability distribution  $c_m(x)$ . Then, one can generate a uniformly distributed random number  $u \in [0, 1]$  and compute  $\hat{m} = c_m^{-1}(u)$  as the noisy version of  $m$ .
  - d. Repeat the above steps for all  $g \in G$  and place the generated noisy diffusion signal in the corresponding voxel in  $\text{DW}^k$ .

Note that the above process does not require spatial registration of fetal and newborn scans into a common space. The generated data will spatially match the high-quality newborn scans. One can obtain a dataset of size  $K$ ,  $\mathcal{D}_{\text{train}} = \{\text{DW}^k, \text{CFA}^k\}_{k=1}^K$ , by repeating the steps described above  $K$  times. Each sample in this dataset will consist of a diffusion-weighted volume and its corresponding CFA image. The random selection of  $V^k$ ,  $\text{Tensor}^k$ , and  $\text{CFA}^k$  and the random noise added to the synthesized diffusion signal ensure that each data sample

is unique. For this work, we synthesized a total of 600 data samples, i.e.,  $K = 600$ . We used 500 of these samples for training and validation and used the remaining 100 samples for test. For proper validation, it was essential to ensure that there was no overlap between training and test sets. Specifically in our setting, it was important to ensure that newborn and fetal scans used in the generation of training, validation, and test sets had no overlap. Therefore, we split the data into training, validation, and test sets at random on a *patient-wise* basis. In other words, the newborn and fetal scans that were used to generate the 100 test samples were completely separate from the newborn and fetal scans that were used to generate the 500 training samples. This ensured that the training, validation, and test data were completely independent. Moreover, we applied our trained method on an independent set of 20 fetal test scans, described in more detail below, and evaluated it in terms of reconstruction precision as well as in terms of expert neuroanatomist assessments. These additional 20 fetal test scans had not been used in the generation of our training data in any way.

The data synthesized with our pipeline satisfy the requirements that we desire: 1) the synthesized DW-MRI volumes,  $DW^k \in \mathcal{D}_{\text{train}}$ , follow the noise distribution and measurement scheme (i.e., gradient table) of fetal scans, and 2) the corresponding CFA images,  $CFA^k \in \mathcal{D}_{\text{train}}$ , are accurate target parameter maps that are reconstructed from high-quality newborn scans. Hence, we expect that this data should be effective for reliable training of deep learning models for parameter estimation in fetal DW-MRI.

## 2.2. Proposed machine learning methods

**2.2.1. Resampling of diffusion measurements**—In general, the number and direction of diffusion-weighted gradients are different between different scans. In fetal imaging, because of the SVR process used to generate the DW-MRI volumes, the number and directions of measurements are generally different between different voxels of the same volume. Therefore, as the first step, measurements obtained using different gradient tables need to be transformed into a unified basis or grid in  $q$  space. This is necessary for the machine learning model to be applicable to scans acquired with different gradient tables. There are two common ways of achieving this goal: 1) representation in a spherical harmonics basis (Lin, 2019), and 2) interpolation onto a spherical grid (Karimi et al., 2021a). In our recent work we found that, for fODF estimation, the interpolation method was slightly better than representation in spherical harmonics bases. Therefore, here we use the interpolation approach by considering a uniform spherical grid of size 200 as in (Karimi et al., 2021a).

We used the Fibonacci spiral sphere method (González, 2010) to construct the spherical grid. We represent this grid using the set of unit vectors from the origin to the grid points,  $U = \{u_i\}_{i=1}^{200}$ . Consider the DW measurements in a voxel,  $\{s(q_j, b)\}_{j=1}^n$ . In this study we only work with single-shell measurements, hence  $b$  is constant.  $q_j$  is the unit vector indicating the direction of diffusion gradient for the  $j^{\text{th}}$  measurement. We resample the measured signal  $s$  onto the sphere  $U$  using weighted averaging as follows:



$$s(u^i) = \sum_{\Omega_{ij}} w_{ij} s(q_j), \quad w_{ij} \propto \frac{1}{\angle(q_j, u_i) + \epsilon}, \quad (2)$$

where  $\Omega_{ij}$  is the set of  $M$  closest measurement directions,  $q_j$ , to the direction under consideration,  $u_i$ . We set  $M = 5$  in our experiments.  $\angle(q_j, u_i)$  denotes the angle between  $q_j$  and  $u_i$ , and  $\epsilon = 0.1$  rad is meant to avoid division by very small numbers. The weights  $w_{ij}$  are normalized to sum to one.

**2.2.2. CNN architecture**—The CNN architecture that we used in this work is an encoder-decoder architecture in the spirit of UNet (Çiçek et al., 2016). Additional dense connections and skip connections have been added, similar to DenseNet (Huang et al., 2017) and 3D UNet++ (Zhou et al., 2018). A schematic representation of the architecture is shown in Figure 4. The network accepts 3D patches of size  $48^3$  voxels from the diffusion volume as input and predicts the CFA for that patch. As we have explained above, upon resampling of the diffusion measurements, the measurements in each voxel are of size 200. Therefore, the CNN input has 200 channels. The CNN output has three channels, for the three (RGB) components of the CFA image.

The number of feature maps in the first stage of the network was set to 12, which was the largest possible on our GPU memory. During training, we sampled patches from random locations in the training DW-MRI volumes and corresponding locations in the target CFA images. These were used to compute the loss and update the network weights using the Adam optimizer (Kingma and Ba, 2014). As training loss, we used the  $\ell_2$  norm between the predicted and ground truth CFA. We used an initial learning rate of  $10^{-3}$ , which we reduced by 0.5 every time the validation loss did not decrease after a training epoch. Once the network was trained, on a test volume we used a sliding window processing with a stride of 16 voxels along each dimension to estimate CFA for a DW-MRI volume of arbitrary size.

### 2.3. Evaluation methods and criteria

We compared our trained model with the following three methods:

- WLLS reconstruction (Koay et al., 2006). This method estimates the diffusion tensor,  $D$ , in a voxel by considering residuals of the form  $\log(s_i/s_{0i}) - bq_i^T Dq_i$  and weights that are proportional to the diffusion signal. This method is commonly used for estimating the diffusion tensor parameters in non-fetal and fetal DW-MRI (Marami et al. (2017), Khan et al. (2019)).
- nonlinear least squares (NLS) reconstruction (Koay et al., 2006). This method estimates  $D$  by fitting the diffusion signal without log-transformation.
- RESTORE (robust estimation of tensors by outlier rejection) (Chang et al., 2005). This method iteratively identifies outlier measurements and excludes them from the fitting process.

Our evaluations and comparisons are based on three different criteria, described below.

**2.3.1. Reconstruction accuracy.**—As mentioned in Section 2.1 above, we left 100 of the synthesized DW-MRI volumes for test. For these volumes, we had high-quality target CFA images. Therefore, we could compute the reconstruction error as the difference between the estimated and target CFA. We compared different methods on these test volumes in terms of the root mean square of the reconstruction error (RMSE).

**2.3.2. Reconstruction precision.**—For this evaluation, we used scans of 20 test fetuses. These fetuses were different from the 20 fetuses used in the generation of our training data described in Section 2.1.1. The scans of these fetuses had not been used in generating the training data or in CNN training in any way. The gestational age of these fetuses was  $31.5 \pm 5.3$  weeks (range: [24.0,38.7]). For fetal scans we did not have high-quality reference CFA images. Therefore, on these scans we assessed the “reconstruction precision”, which we defined as the inverse of variance (Murphy, 2012). Each of these fetal scans consisted of 36–60 sets of diffusion measurements. For each fetus, we bootstrap-selected subsets of 24 measurements and reconstructed the CFA image. We used the same bootstrap-selected sets for all compared methods. We computed precision as the inverse of CFA variance across bootstraps in each voxel (Murphy, 2012). To compute a single value of precision for each fetal scan and each reconstruction method, we averaged the precision across all voxels.

**2.3.3. Expert assessments.**—Three experts with extensive experience in fetal neuroanatomy assessed CFA images of the 20 test fetuses. These were the 20 independent test fetuses, mentioned above in Section 2.3.2, that had not been used in the generation of our training data. These assessments included two separate and independent parts.

**1. Overall assessment.:** We asked one of the expert neuroanatomists to rate the accuracy and fidelity of the reconstructions based on her knowledge of fetal neuroanatomy on a 1–5 scale with 5 being the best quality.

**2. Detailed assessment.:** A more detailed assessment was conducted by two other experts: a board certified pediatric neuroradiologist with expertise in fetal neuroimaging and a post-doctoral research fellow with expertise in developmental neuroanatomy. To compare the quality of the reconstructions, these two experts evaluated multiple structures: cerebral cortex, posterior fossa [middle cerebellar peduncle (MCP), transverse pontine fibers (TPF), descending pyramidal tracts in brainstem], large white matter tracts [corpus callosum (CC), inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), supratentorial corticospinal tract (CST)], small white matter tracts [fornix, superior longitudinal fasciculus (SLF), cingulum], and deep white matter (corresponding to the internal and external capsules). To perform this evaluation, for each subject, the experts opened the CFA images reconstructed by WLLS and the proposed method side-by-side and scrolled through them to view different structures in detail. To compare the reconstructions, they used a 5-point scale ranging from  $-2$  to  $+2$ , analogous to that used in (Conklin et al., 2019). Briefly, in this scoring system  $-2$  strongly favored the WLLS reconstruction,  $+2$  strongly favored the CNN reconstruction, and 0 meant no significant difference. To determine superiority, the experts evaluated all the above-mentioned regions of interest by



considering: (a) conspicuity of the structure, (b) correspondence with known neuroanatomy, and (c) accuracy of orientation of the primary eigenvector as represented by standard RGB CFA display. In addition to the above-mentioned regions, the presence of artifacts was also evaluated using the same scale.

In order to make our best efforts to achieve a fair comparison, we named the CFA images reconstructed with WLLS and our method as “CFA 01” and “CFA 02” at random before presenting them to the three experts for evaluation. The experts were told that the names were randomized, but they were blind to the naming order.

### 3. Results and Discussion

#### 3.1. Reconstruction accuracy

On the 100 test images used in this evaluation, the reconstruction RMSE for the proposed method was  $0.0379 \pm 0.0030$ . For WLLS, NLS, and RESTORE, it was  $0.0807 \pm 0.0034$ ,  $0.0805 \pm 0.0036$ , and  $0.0800 \pm 0.0034$ , respectively. On all 100 test images the RMSE achieved by the proposed method was lower than WLLS, NLS, and RESTORE. We performed paired t-tests to assess the statistical significance of the differences. The tests showed that the difference between our method and the other three methods was statistically significant ( $p < 0.001$ ). However, the differences between the other three methods were statistically insignificant ( $p > 0.5$ ). Figure 5 shows CFA images reconstructed with different methods on selected test volumes along with the reference images. The images reconstructed with the proposed method are much closer to the reference image. Whereas the images reconstructed with the three competing methods are noisy and lack some of the important details, the images reconstructed with the proposed method display all of the detail that are present in the reference image.

#### 3.2. Reconstruction precision

Figure 6 displays example images from the experiments to assess reconstruction precision. It shows CFA images reconstructed with our proposed method and the competing methods from three different bootstrap-selected measurements of a fetal test scan. The reconstruction precision on the 20 independent fetal test scans for the proposed method was  $297.8 \pm 19.0$ . For WLLS, NLS, and RESTORE, the reconstruction precision was, respectively,  $41.1 \pm 5.50$ ,  $41.4 \pm 5.73$ , and  $40.2 \pm 5.70$ . On all 20 test images, the reconstruction precision for our proposed method was higher than the other three methods. Paired t-tests showed that the precision of our method was significantly higher than each of the other three methods ( $p < 0.001$ ), but the differences between the three competing methods were not significant ( $p > 0.5$ ).

#### 3.3. Expert assessments

**3.3.1. Overall assessment**—Figure 7 displays CFA images reconstructed by WLLS and the proposed method on five fetal test scans of different gestational ages. These five scans were from among the 20 independent fetal test scans. The figure also shows the scores assigned by our expert neuroanatomist. The score assigned to CFA images reconstructed with WLLS and the proposed method were, respectively,  $1.30 \pm 0.46$  and  $3.85 \pm 0.79$ . The

score received by our proposed method was higher than WLLS on all 20 fetal test scans. We performed a Wilcoxon signed-rank test on these scores and found that the difference between our method and WLLS was statistically significant ( $p < 0.001$ ).

**3.3.2. Detailed assessment**—As we mentioned above, the detailed assessment by two independent experts focused on neuroanatomic concordance, sharpness, and orientation of 12 different regions of interest as well as on reconstruction artifacts. Figure 8 shows a summary of the results of this assessment, separately for the 12 regions of interest and for the presence of artifacts. The scores for both experts and all 20 subjects have been combined for generating this plot. The averaged scores from both experts indicated superior quality for the images reconstructed with the proposed method for all structures ( $p < 0.001$ ), except for the deep white matter where they were comparable ( $p = 0.652$ ). Scores were highest for cortex, small tracts, and artifact evaluation. There was high interrater agreement between the two raters (average ICC 0.777, 95% CI 0.436 – 0.912,  $p = 0.001$ ).

In terms of computational time, the proposed method estimated the CFA for an entire fetal DW-MRI volume in approximately five minutes. Most of the computational time of our method was spent on the signal resampling step (Equation (2)). The time taken by the CNN computation was less than 10 seconds for a fetal brain. Compared with our method, WLLS needed almost twice the computational time, approximately 10 minutes, for a fetal brain.

The rationale behind our proposed method is that such a machine learning-based technique can learn to predict the parameter of interest from fetal diffusion signal that suffers from strong noise and motion. Moreover, because our model is a CNN, it can exploit the signal from neighboring voxels to achieve a more robust estimation. Since we cannot control these factors independently, it is hard to judge which of them contribute more to the superiority of our method compared with the other techniques. To investigate the effect of motion, we estimated the average magnitude of fetal head motion between subsequent slice acquisitions based on the slice-to-volume registration transforms in our SVR pipeline. Figure 9(a)–(b) shows the reconstruction precision for the proposed method and WLLS as a function of motion for the 20 fetal test scans. While the reconstruction precision for the proposed method remained almost constant, the precision of WLLS showed a tendency to decrease with larger motion. Figure 9(c) shows the average of the two experts' scores from the detailed assessment approach versus the magnitude of motion. As we explained in Section 2.3.3, this was a relative score that was assigned by comparing the proposed method and WLLS, with positive scores favoring the proposed method. Figure 9(c) shows that the advantage of the proposed method over WLLS in terms of expert scores increases with stronger motion. Overall, this analysis indicates that the advantage of the proposed method over WLLS increases with stronger fetal head motion.

Our proposed machine learning-based method estimates the parameter of interest without directly solving an inverse problem. The inquisitive reader may wonder about the potential shortcomings and pitfalls of the proposed method. For example, in Figure 5 the results obtained with the proposed method look less noisy than the reference results. The reader may wonder if the proposed method might learn to infer outputs that are not supported by data. We have performed extensive experiments to ensure that our method is sound and

reliable. In Figure 10 we show an example of these experiments. In this experiment, we smeared or added random noise to a small patch of a DWI volume. We then applied our method on the smeared or noisy volume. In the figure, we show the original image and the images in which the patch has been smeared or added noise to. The location of the patch has been marked with a red square on the image of the original volume. We have shown the estimated CFA image as well as a zoomed-in section of the CFA image for each case. As can be seen in these examples, our method displays the desired behavior: the CFA image estimated for the patch that has been smeared or added noise to shows the corresponding effect. Therefore, our method does not infer outputs that are not supported by the data. The sharper and less noisy reconstructions of our method compared with the reference (Figure 5) is due to inherently different approaches that they follow. The reference results are obtained on a voxel-wise basis. In other words, the reference parameters in each voxel are estimated based on the diffusion signal in that voxel alone. Our method, on the other hand, by seeing a large number of noisy data samples during training, can learn to partially factor out the effect of noise. Moreover, because our model is a CNN, it has the ability to learn the spatial patterns and use the data in neighboring voxels for more accurate and less noisy estimation.

### 3.4. Limitations and future work

Deep learning models, in general, benefit from larger training datasets. In this work we used 20 fetal scans and 82 newborn scans in generating our training data. The imbalance in the number of fetal and newborn scans (20 versus 82) was in part intentional, because the fetal and newborn scans were used for different purposes in our data generation pipeline. Specifically, the fetal scans were used to estimate the noise and to sample the gradient tables, whereas the newborn scans were used to generate the parameter maps. Because of the high inter-subject variability in the structures of the parameter maps, we decided to use a larger number of newborn scans than fetal scans. This allowed our CNN model to see a wider range of variability in the spatial patterns of CFA during training. Even though our results on independent test data are very good, future works may achieve better results by using larger and more diverse datasets. As an example, most of the newborn scans that we used in this work were closer to 38 weeks of age. From the 82 newborn scans that we used to synthesize our training data, only 15 of them were younger than 34 weeks and the youngest was 29.3 weeks. Given this limitation, the fact that our method works well on fetal test scans as young as 24 weeks is very promising (Figure 7). Nonetheless, a training dataset with a larger number of younger newborn scans may further improve the performance of the proposed method on younger fetuses.

Another factor that can be further investigated in future works is the effect of heterogeneity in the diffusion data, such as heterogeneity in multi-center data. The 20 fetal scans that were used in generating our training data were acquired using Siemens Skyra ( $n = 18$ ) and Prisma ( $n = 2$ ) scanners. The other 20 fetal scans that were used for testing our method were acquired using Siemens Skyra ( $n = 15$ ), Prisma ( $n = 3$ ), and Trio ( $n = 2$ ) scanners. These were all 3T scanners. The scans were acquired at three different sites in Massachusetts. We did not observe any differences in the performance of our proposed method on scans obtained with these three scanners. The variability in multi-center data can be higher than the variability in the data used in this study. Therefore, a complete investigation of this issue

will require more extensive experiments. Nonetheless, our results are promising with regard to the generalizability of our method to multi-center data.

#### 4. Conclusions

DW-MRI continues to play a prominent role in shaping our understanding of fetal brain development and disorders and their impacts on cognitive development and disabilities later in life. Therefore, accurate and robust parameter estimation in fetal DW-MRI is of crucial importance. In this work, we presented the first successful application of deep learning for parameter estimation in fetal DW-MRI. Our main contribution was a methodology that enabled us to generate large amounts of reliable training data. Our quantitative and qualitative evaluations demonstrated the superiority of the proposed deep learning method to the standard estimation methods. While we focused on CFA estimation to demonstrate the effectiveness of our proposed methodology, our methods can be extended and adapted for estimating other parameters. For other diffusion tensor imaging parameters, for example, this can be accomplished by simply replacing CFA with the parameter of interest (e.g., fractional anisotropy or mean diffusivity) in the methods described above and other obvious necessary changes such as changing the number of CNN's output channels.

#### Acknowledgment

Research reported in this publication was supported in part by the National Institute of Neurological Disorders and Stroke, the National Institute of Biomedical Imaging and Bioengineering, and the National Library of Medicine of the National Institutes of Health (NIH) under award numbers R01NS106030, R01EB031849, R01EB019483, and R01LM013608; in part by the Office of the Director of the NIH under Award Number S10OD0250111; and in part by a Technological Innovations in Neuroscience Award from the McKnight Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Lana Vasung was supported by the Ralph Schlager Fellowship of Harvard University. Camilo Jaimes is supported by the scholarship of the American Roentgen Ray Society.

#### Data availability:

Part of the data is from the developing Human Connectome Project (dHCP) dataset. This dataset is publicly available at: <http://www.developingconnectome.org/project/data-release-user-guide/>. Part of the data includes in-house fetal diffusion MRI, which cannot be shared with the public at this time.

#### References

- Aliotta E, Nourzadeh H, Sanders J, Muller D, Ennis DB, 2019. Highly accelerated, model-free diffusion tensor mri reconstruction using neural networks. *Medical physics* 46, 1581–1591. [PubMed: 30677141]
- Baldoli C, Righini A, Parazzini C, Scotti G, Triulzi F, 2002. Demonstration of acute ischemic lesions in the fetal brain by diffusion magnetic resonance imaging. *Annals of neurology* 52, 243–246. [PubMed: 12210800]
- Basser PJ, Pierpaoli C, 2011. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *Journal of magnetic resonance* 213, 560–570. [PubMed: 22152371]
- Bastiani M, et al. , 2019. Automated processing pipeline for neonatal diffusion mri in the developing human connectome project. *NeuroImage* 185, 750–763. [PubMed: 29852283]

- Bui T, Daire JL, Chalard F, Zaccaria I, Alberti C, Elmaleh M, Garel C, Luton D, Blanc N, Sebag G, 2006. Microstructural development of human brain assessed in utero by diffusion tensor imaging. *Pediatric radiology* 36, 1133–1140. [PubMed: 16960686]
- Canales-Rodríguez EJ, Daducci A, Sotiropoulos SN, et al. , 2015. Spherical deconvolution of multichannel diffusion mri data with non-gaussian noise models and spatial regularization. *PLoS one* 10, e0138910. [PubMed: 26470024]
- Casella G, Berger RL, 2021. *Statistical inference*. Cengage Learning.
- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 424–432.
- Chang LC, Jones DK, Pierpaoli C, 2005. Restore: robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 53, 1088–1095.
- Conklin J, Longo M, Cauley S, Setsompop K, González R, Schaefer P, Kirsch J, Rapalino O, Huang S, 2019. Validation of highly accelerated wave-caipi swi compared with conventional swi and t2\* -weighted gradient recalled-echo for routine clinical brain mri at 3t. *American Journal of Neuroradiology* 40, 2073–2080. [PubMed: 31727749]
- Deprez M, Price A, Christiaens D, Estrin GL, Cordero-Grande L, Hutter J, Daducci A, Tournier JD, Rutherford M, Counsell SJ, et al. , 2019. Higher order spherical harmonics reconstruction of fetal diffusion mri with intensity correction. *IEEE transactions on medical imaging* 39, 1104–1113. [PubMed: 31562073]
- Dietrich O, Raya JG, Reeder SB, Ingrisch M, et al. , 2008. Influence of multichannel combination, parallel imaging and other reconstruction techniques on mri noise characteristics. *Magnetic resonance imaging* 26, 754–762. [PubMed: 18440746]
- Gholipour A, Estroff JA, Barnewolt CE, Robertson RL, Grant PE, Gagoski B, Warfield SK, Afacan O, Connolly SA, Neil JJ, et al. , 2014. Fetal mri: a technical update with educational aspirations. *Concepts in Magnetic Resonance Part A* 43, 237–266.
- Gibbons EK, et al. , 2019. Simultaneous noddi and gfa parameter map generation from subsampled q-space imaging using deep learning. *Magnetic resonance in medicine* 81, 2399–2411. [PubMed: 30426558]
- Golkov V, et al. , 2016. q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE transactions on medical imaging* 35, 1344–1351. [PubMed: 27071165]
- González Á, 2010. Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. *Mathematical Geosciences* 42, 49.
- Huang G, Liu Z, Weinberger KQ, van der Maaten L, 2017. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 3.
- Jakab A, Kasprian G, Schwartz E, Gruber GM, Mitter C, Prayer D, Schöpf V, Langs G, 2015. Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage* 111, 277–288. [PubMed: 25725467]
- Jiang S, Xue H, Counsell S, Anjari M, Allsop J, Rutherford M, Rueckert D, Hajnal JV, 2009. Diffusion tensor imaging (dti) of the brain in moving subjects: application to in-utero fetal and ex-utero studies. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 62, 645–655.
- Johansen-Berg H, Behrens TE, 2013. *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press.
- Jones DK, Horsfield MA, Simmons A, 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 42, 515–525.
- Kainz B, Steinberger M, Wein W, Kuklisova-Murgasova M, Malamateniou C, Keraudren K, Torsney-Weir T, Rutherford M, Aljabar P, Hajnal JV, et al. , 2015. Fast volume reconstruction from motion corrupted stacks of 2d slices. *IEEE transactions on medical imaging* 34, 1901–1913. [PubMed: 25807565]

- Karimi D, Vasung L, Jaimes C, Machado-Rivas F, Khan S, Warfield S, Gholipour A, 2021. A machine learning-based method for estimating the number and orientations of major fascicles in diffusion-weighted magnetic resonance imaging. *Medical Image Analysis* 102129. [PubMed: 34182203]
- Karimi D, Vasung L, Jaimes C, Machado-Rivas F, Warfield SK, Gholipour A, 2021. Learning to estimate the fiber orientation distribution function from diffusion-weighted mri. *NeuroImage* 239, 118316. [PubMed: 34182101]
- Kasprian G, Brugger PC, Weber M, Krssák M, Krampf E, Herold C, Prayer D, 2008. In utero tractography of fetal white matter development. *Neuroimage* 43, 213–224. [PubMed: 18694838]
- Khan S, Vasung L, Marami B, Rollins CK, Afacan O, Ortinau CM, Yang E, Warfield SK, Gholipour A, 2019. Fetal brain growth portrayed by a spatiotemporal diffusion tensor mri atlas computed from in utero images. *NeuroImage* 185, 593–608. [PubMed: 30172006]
- Kingma DP, Ba J, 2014. Adam: A method for stochastic optimization. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Koay CG, Chang LC, Carew JD, Pierpaoli C, Basser PJ, 2006. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *Journal of magnetic resonance* 182, 115–125. [PubMed: 16828568]
- Koppers S, Merhof D, 2016. Direct estimation of fiber orientations using deep learning in diffusion imaging. In: *International Workshop on Machine Learning in Medical Imaging*. Springer, pp. 53–60.
- Lin Z, et al. , 2019. Fast learning of fiber orientation distribution function for mr tractography using convolutional neural network. *Medical physics* 46, 3101–3116. [PubMed: 31009085]
- Marami B, Salehi SSM, Afacan O, Scherrer B, Rollins CK, Yang E, Estroff JA, Warfield SK, Gholipour A, 2017. Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *NeuroImage* 156, 475–488. [PubMed: 28433624]
- Marami B, Scherrer B, Afacan O, Erem B, Warfield SK, Gholipour A, 2016. Motion-robust diffusion-weighted brain mri reconstruction through slice-level registration-based motion tracking. *IEEE transactions on medical imaging* 35, 2258–2269. [PubMed: 27834639]
- Murphy KP, 2012. *Machine learning: a probabilistic perspective*.
- Nat V, et al. , 2019. Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted mri. *Magnetic resonance imaging* 62, 220–227. [PubMed: 31323317]
- Nedjati-Gilani GL, Schneider T, Hall MG, Cawley N, Hill I, Ciccarelli O, Drobnjak I, Wheeler-Kingshott CAG, Alexander DC, 2017. Machine learning based compartment models with permeability for white matter microstructure imaging. *NeuroImage* 150, 119–135. [PubMed: 28188915]
- Neher PF, Côté MA, Houde JC, Descoteaux M, Maier-Hein KH, 2017. Fiber tractography using machine learning. *Neuroimage* 158, 417–429. [PubMed: 28716716]
- Oubel E, Koob M, Studholme C, Dietemann JL, Rousseau F, 2012. Reconstruction of scattered data in fetal diffusion mri. *Medical image analysis* 16, 28–37. [PubMed: 21636311]
- Righini A, Bianchini E, Parazzini C, Gementi P, Ramenghi L, Baldoli C, Nicolini U, Mosca F, Triulzi F, 2003. Apparent diffusion coefficient determination in normal fetal brain: a prenatal mr imaging study. *American journal of neuroradiology* 24, 799–804. [PubMed: 12748074]
- Schultz T, 2012. Learning a reliable estimate of the number of fiber directions in diffusion mri. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 493–500.
- Tian Q, Bilgic B, Fan Q, Liao C, Ngamsombat C, Hu Y, Witzel T, Setsompop K, Polimeni JR, Huang SY, 2020. Deepdti: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage* 219, 117017. [PubMed: 32504817]
- Wakana S, Jiang H, Nagae-Poetscher LM, Van Zijl PC, Mori S, 2004. Fiber tract-based atlas of human white matter anatomy. *Radiology* 230, 77–87. [PubMed: 14645885]
- Ye C, Li X, Chen J, 2019. A deep network for tissue microstructure estimation using modified lstm units. *Medical image analysis* 55, 49–64. [PubMed: 31022640]



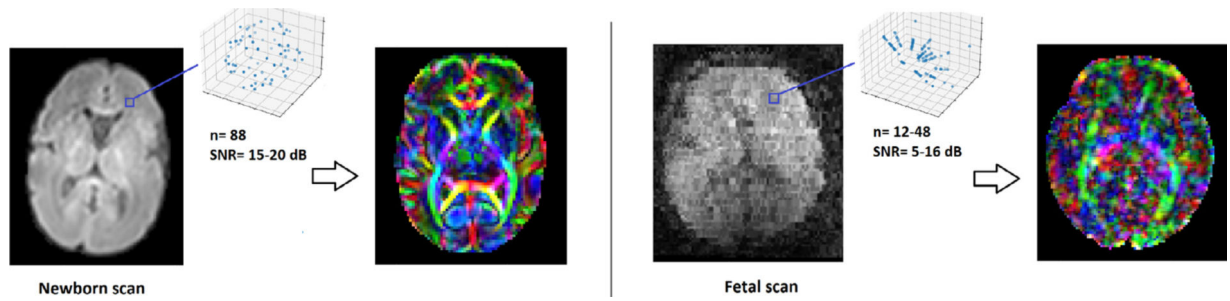
- Zanin E, Ranjeva JP, Confort-Gouny S, Guye M, Denis D, Cozzone PJ, Girard N, 2011. White matter maturation of normal human fetal brain. an in vivo diffusion tensor tractography study. *Brain and behavior* 1, 95–108. [PubMed: 22399089]
- Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J, 2018. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, pp. 3–11.

Author Manuscript

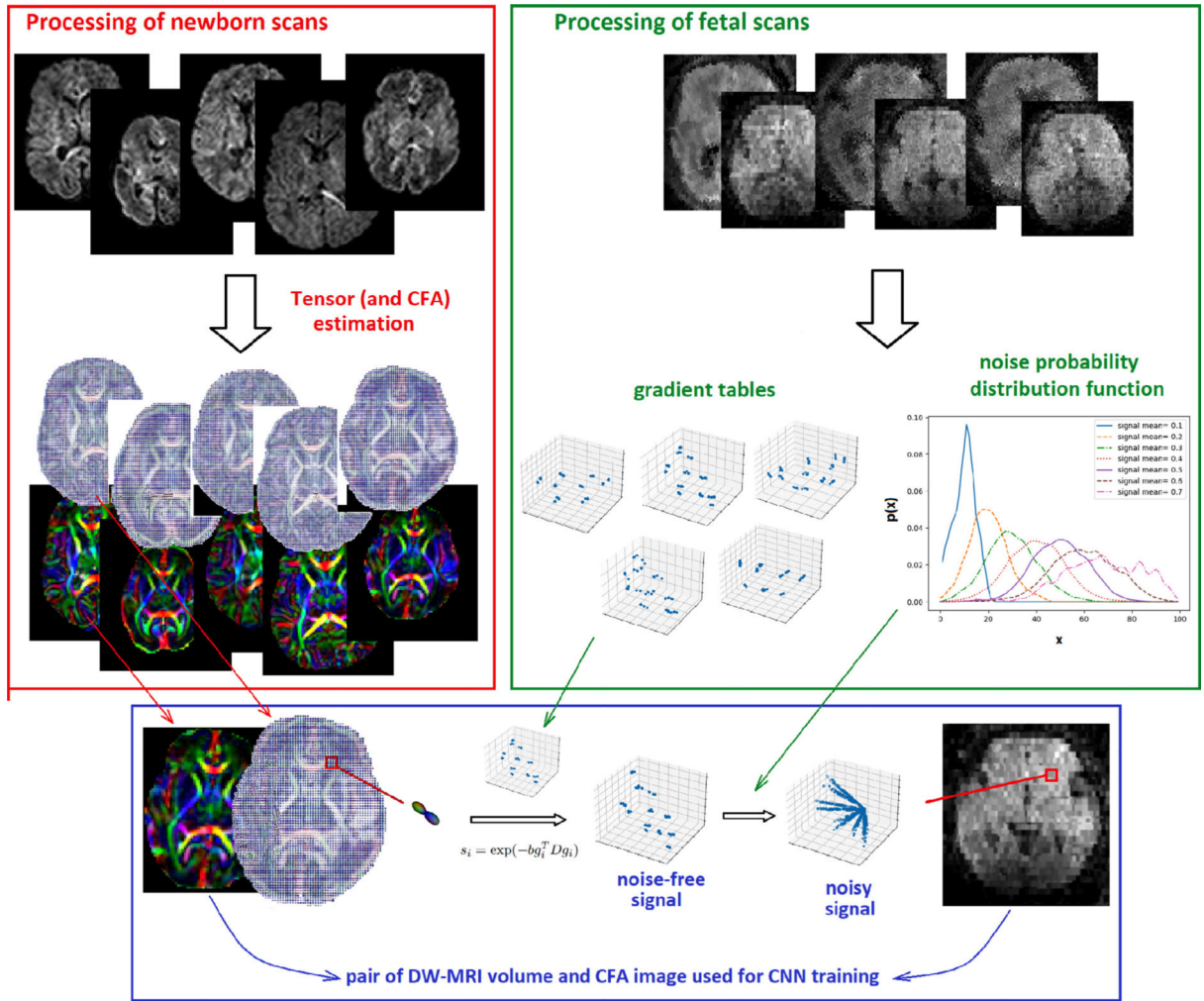
Author Manuscript

Author Manuscript

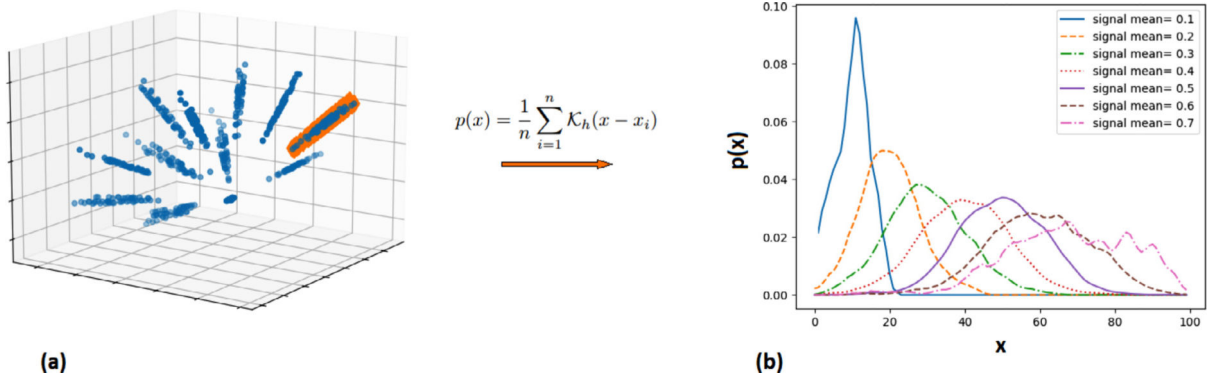
Author Manuscript



**Fig. 1.** DW-MRI scans and CFA reconstructions for a newborn subject and a fetal subject of the same gestational age (34 weeks). The number of diffusion-sensitized measurements for this fetal scan was 36; the number of data points in each voxel is larger than the number of measurements because SVR methods use a point-spread function with a width larger than one voxel (Kainz et al. (2015), Marami et al. (2017)).



**Fig. 2.** The proposed pipeline for synthesizing fetal DW-MRI data and corresponding CFA images. Top left pane: We estimate accurate CFA and tensor images for a set of high-quality pre-term newborn scans. Top right pane: We analyze a set of fetal DW-MRI volumes, reconstructed using our SVR pipeline (Marami et al., 2016) in order to estimate the noise probability distribution. We also store gradient tables for voxels in these volumes. Bottom pane: We use tensor images estimated from the newborn scans to synthesize DW-MRI data that follow the acquisition scheme and noise distribution of fetal scans.

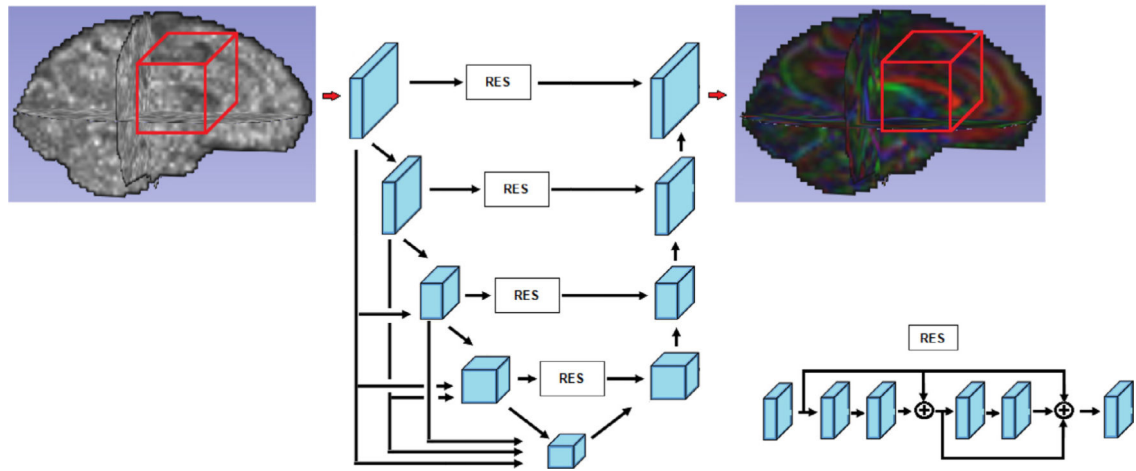


(a)

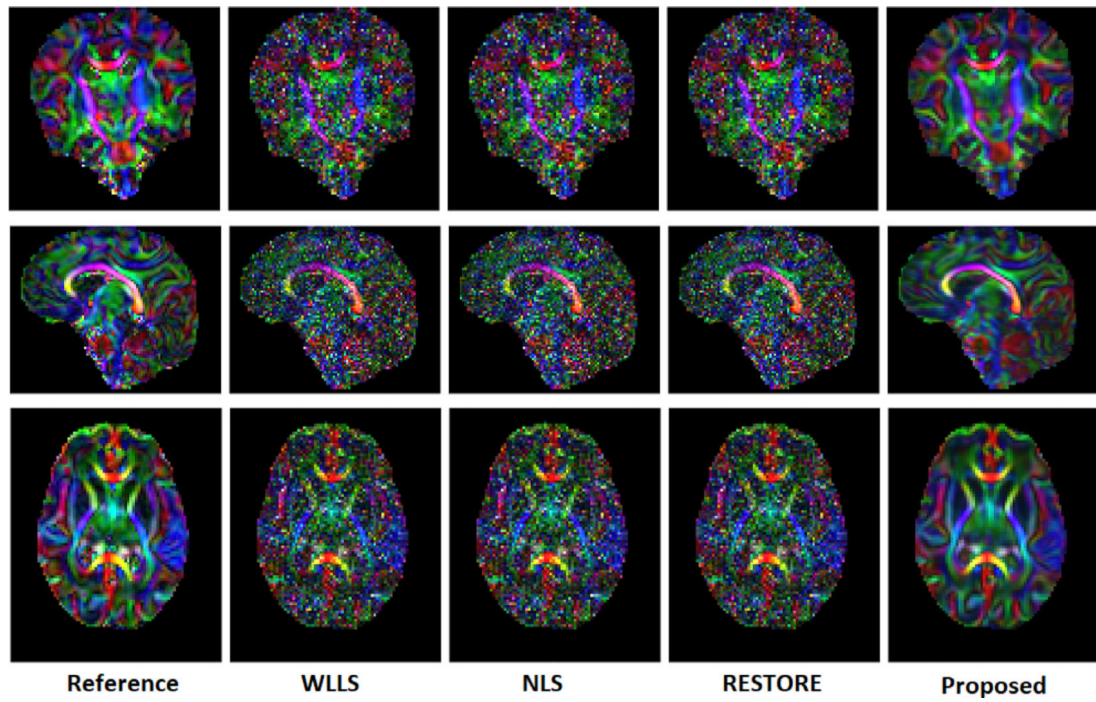
(b)

**Fig. 3.**

(a) Estimation of noise PDF. We consider measurements that fall along the same diffusion gradient direction in a voxel. We estimate the noise PDF using kernel density estimation, conditioned on the signal mean. (b) Example noise PDFs for several signal mean values.

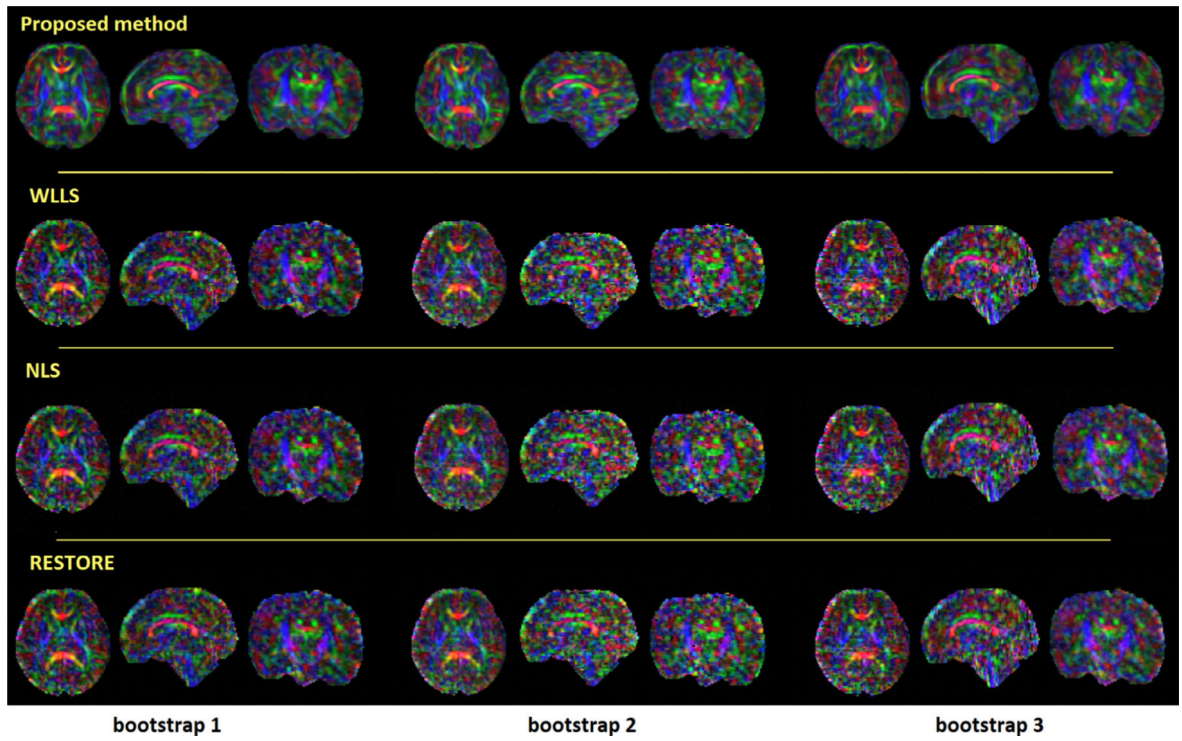


**Fig. 4.** The network architecture used to predict the CFA image from a DW-MRI volume. All convolutional layers are followed by ReLU activation. The lower right section of the figure shows the residual module (RES) with short and long skip connections.

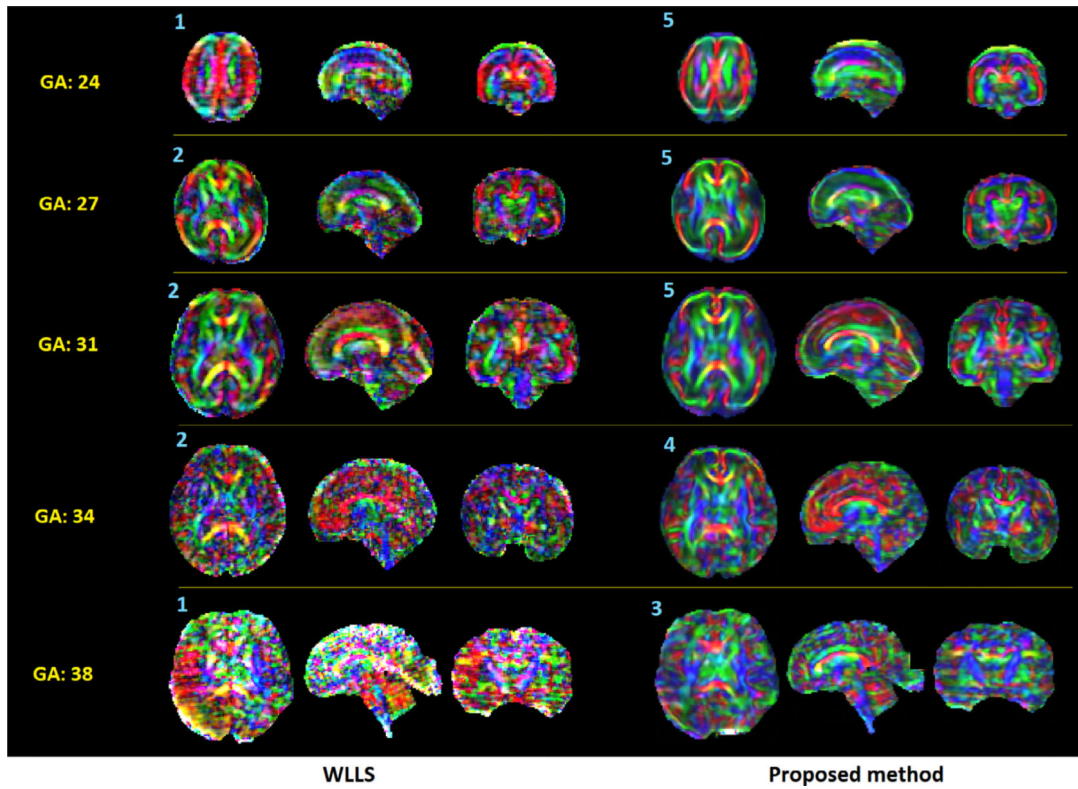


**Fig. 5.**  
Comparison of different methods on example test DW-MRI volumes from the data synthesized in Section 2.1.

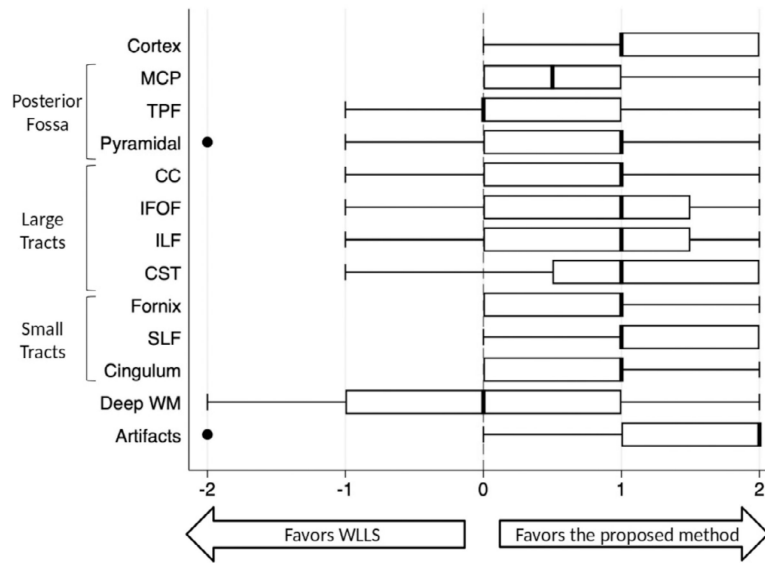




**Fig. 6.** CFA images reconstructed by different methods from three different bootstrap-selected measurements of a fetal test scan.



**Fig. 7.** CFA images reconstructed with WLLS and the proposed method on five fetal test scans of different gestational ages (GA). The gestational age for each subject is shown on the left side of the figure. The scores assigned by our expert neuroanatomist are shown on the top left corner of each image.



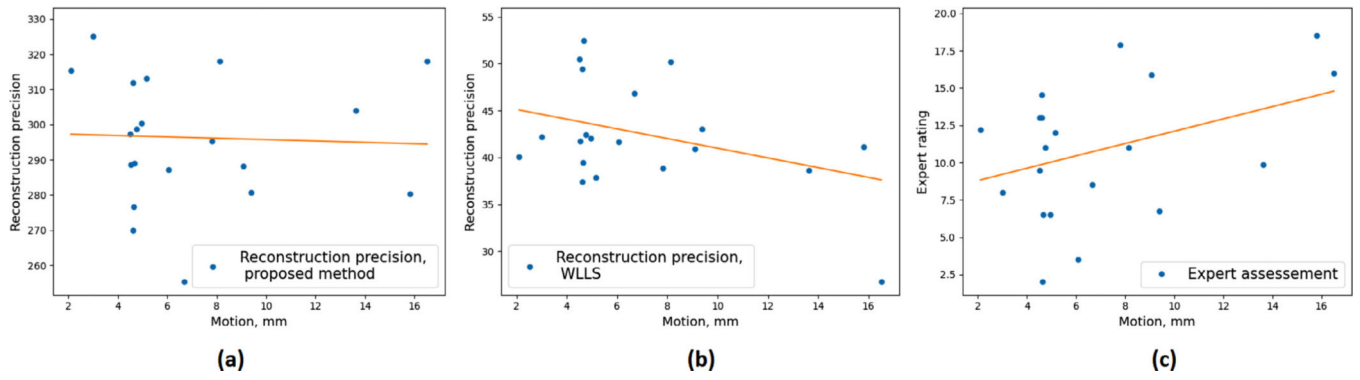
**Fig. 8.** Detailed assessment of neuroanatomic concordance, sharpness, and orientation of 12 different regions of interest for the CFA images reconstructed with WLLS and the proposed method on 20 fetal test scans.

Author Manuscript

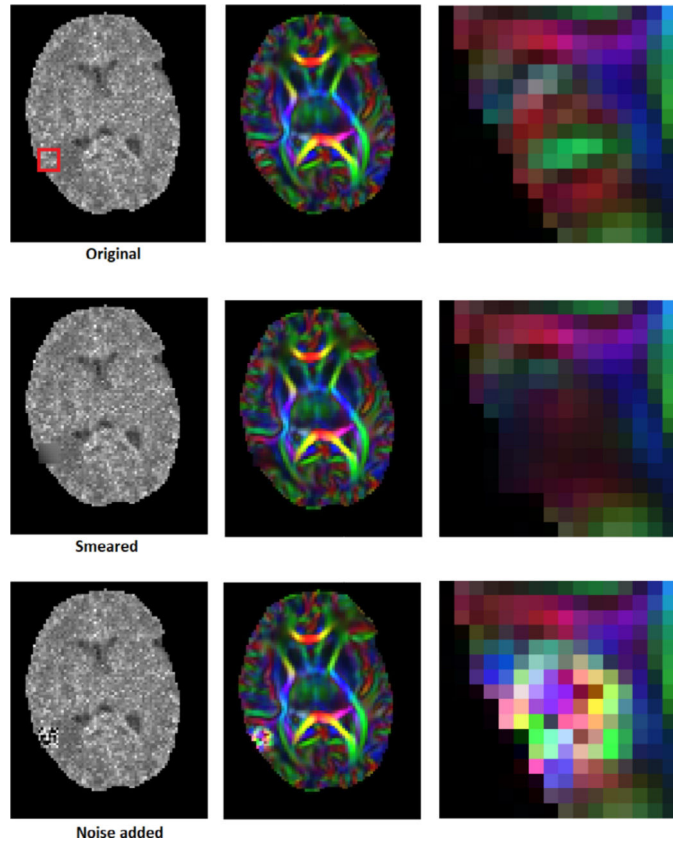
Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 9.** (a) and (b) Reconstruction precision for, respectively, the proposed method and WLLS as a function of average fetus head motion between slice acquisitions. (c) Average of the two experts' scores from the detailed assessment described in Section 2.3.3 as a function of fetal head motion.



**Fig. 10.**

The results of a sanity test to ensure that the method does not learn to produce (i.e., inpaint) output values that are not supported by the data. A slice of the original input DW volume and the corresponding CFA parameter map is shown in the first row. We smeared or added random noise to a small patch of a the volume and applied our method on the smeared or noisy volume. The red square on the original volume marks the location of the patch. We have shown the estimated CFA image (middle column) as well as a zoomed-in section of the CFA image (right column) for each case. Results show that the CFA values estimated for the patch that has been smeared or added noise to displays the corresponding effect.