LOGO

Compensation of Motion-induced Signal Loss in *in vivo* DTI of the Human Heart Using GAN-Based Super Angular Resolution

Yunlong He, Lihui Wang*, Feng Yang, Patrick Clarysse, Yuanjie Zheng, Zeyu Deng, Marc Robini, and Yuemin Zhu*

Abstract—**Diffusion tensor imaging (DTI) has become an indispensable modality for non-invasively characterizing fiber architectures of the human heart. However, clinical applications of *in vivo* cardiac DTI remain extremely limited due to large signal loss caused by motions. We propose a novel method that can compensate for motion-induced signal loss in *in vivo* cardiac DTI. This is achieved by training a deep learning network to produce higher angular-resolution diffusion weighted (DW) images from acquired lower angular-resolution DW images. The proposed network is constructed based on a generative adversarial network (GAN) combined with L2 norm to produce realistic cardiac DW images. Besides, angular consistency between DW images with the same angular resolution is employed as an additional regularization term to improve the performance of signal loss compensation. To train and test the proposed network, *in vivo* cardiac DTI datasets are used in combination with *ex vivo* cardiac DTI datasets to imitate plenty of motion-induced DW images with lower angular resolution and their ground-truth DW images with higher angular resolution. Experimental results showed that the produced higher angular-resolution images enable significant improvement of the estimation accuracy of diffusion tensors, fractional anisotropy (FA) and mean diffusivity (MD). Moreover, fiber orientations are better characterized. The proposed method thus provides an interesting post-processing way to compensate for motion-induced signal loss in *in vivo* cardiac DTI.**

Index Terms—**Cardiac tissue, diffusion tensor imaging, superresolution, generative adversarial networks.**

I. INTRODUCTION

DIFFUSION tensor imaging (DTI) noninvasively probes diffusion properties and fiber structures of human tissues by measuring the diffusion of water molecules in them. In cardiac imaging, for example, water diffusion process in each voxel can be described by a diffusion tensor estimated from a set of diffusion-weighted (DW) signals acquired in various diffusion gradient directions [1]–[3]. The major eigenvector corresponding to the largest eigenvalue of the diffusion tensor defines the direction of water diffusion that

is least restricted in the muscle fiber tracts and thus reflects the principal orientation of the fibers inside that voxel . This makes it possible to depict the fiber architecture of human myocardium and infer the microstructural changes related to cardiovascular diseases, e.g., myocardial infarction [4], hypertrophic cardiomyopathy [5], etc.

Cardiac DTI has been widely used to characterize the myocardial fiber architectures of the heart and evaluate the histological arrangement of the myocardial fibers [6]–[11]. These studies have shown that fibers in the normal myocardium are orderly arranged in a crossing helical pattern. In patients, DTI has also been considered a useful tool to describe the disease-induced fiber disarray [4], [7]. The fiber integrity after myocardial infarction has often been quantified using fractional anisotropy (FA) and mean diffusivity (MD) [9]–[11]. All these findings suggest the potential of DTI for early diagnosis and treatment of cardiovascular diseases.

However, the applications of *in vivo* DTI of the human heart remain limited in clinical practice. The main reason lies in the fact that DTI essentially measures the microscopic motion of water molecules and thus is very sensitive to subject motion. Respiratory and heartbeat motions that occur during imaging induce large signal loss in DW images and thereby degrade the accuracy of estimated diffusion tensors, which potentially leads to incorrect clinical diagnoses.

To overcome this problem, researchers have made efforts to design new DTI sequences that can reduce the effects caused by motion during imaging. For example, Edelman *et al*. [6] proposed a stimulated echo acquisition mode (STEAM) sequence where diffusion signals were acquired based on the spatial position and shape of the myocardium over the entire heartbeat cycle so that the effects of heartbeat motion were minimized. Gamper *et al*. [12] designed a spin echo (SE) based sequence to decrease the sensitivity of diffusion-encoding to cardiac motion while maintaining a high signal-to-noise ratio (SNR). Besides, there exist a variety of other sequences that can be seen as extensions of STEAM or SE and have improved performance of *in vivo* DTI of the human heart [13]–[17]. Despite the power of these imaging sequences, most of them are time consuming and have to be used in combination with dedicated strategies (e.g., navigator gating and breath holding), resulting in a relatively long data acquisition time. Unfortunately, prolonging the acquisition time is often prohibitive in clinical routine because it compromises patient comfort and increases the risk of inducing motion artifacts. Therefore, for most sequences, the reduction of motion effects often comes at the expense of acquiring a limited amount of data or else the SNR is relatively poor. This makes these sequences remain challenging to use in practice.

In recent years, researchers began to develop post-processing methods capable of recovering motion-induced signal loss in clinically acquired DTI data. In 2011, Pai *et al*. [18] introduced a combined principal component analysis and temporal maximum intensity projection (PCATMIP) method to restore motion-induced signal losses in porcine liver DTI scans. Later, Wei *et al*. [19] proposed a wavelet-based image fusion and principal component analysis (WIFPCA) method to retrieve myocardial fiber structures in free-breathing DTI acquisitions of the human heart. Recently, deep learning (DL) especially the convolutional neural network (CNN),

which has become the method of choice for processing and analyzing medical images [20], [21], has been applied to compensation of signal loss in *in vivo* cardiac DW images [22]. These studies demonstrated that it is feasible to use clinically available acquisitions combined with dedicated post-processing methods to achieve signal loss compensation in *in vivo* cardiac DTI. A common idea used in most of the studies is to repeatedly acquire DW images at multiple time points, and then a post-processing algorithm is used to recover motion-induced signal loss by exploiting the diffusion information from multiple acquisitions over time. However, up to now, there exists little work on the use of information from images acquired in various diffusion gradient directions to deal with signal loss in *in vivo* DW images.

In fact, in addition to collecting images at multiple time points, it is also desirable in many cases to acquire DW images in more diffusion gradient directions. Theoretically, a diffusion tensor can be mathematically calculated from six DW signals acquired in six diffusion gradient directions [1]. In practice, however, the number of diffusion gradient directions, which defines the angular resolution of DTI data, has an important influence on the estimation of diffusion tensors [23]–[25]. Many researchers have concluded that a higher angular resolution (e.g., more than 30 diffusion gradient directions) is required to achieve more accurate diffusion tensor estimation [24]. Moreover, some DTI studies of both human brain [25] and heart [26] have shown that when the acquisition time is sufficient, the accuracy of diffusion tensor increases as the angular resolution. Seen in this light, increasing the angular resolution in our case would be an interesting alternative to compensate for motion-induced signal loss in estimating diffusion tensors. For the time being, available angular resolution in *in vivo* cardiac DTI is quite limited (typically no more than 12 diffusion gradient directions) [16], [27], [28]. This motivates the requirement for developing post-processing methods to improve the angular resolution of DTI data.

In this paper, we propose a novel super angular-resolution method to compensate for motion-induced signal loss in *in vivo* cardiac DTI. The proposed method consists of training a DL algorithm that can produce higher angular-resolution DW images from lower angular-resolution DW images corrupted by motion. Instead of using traditional CNN-based models [21] that only minimize Euclidean distance between produced and ground-truth images, we construct a generative adversarial network (GAN) [29] where two competing sub-networks can be trained iteratively to minimize errors at a higher level, e.g., urges the produced images to be perceptually similar to the ground-truth images with higher angular resolution. Meanwhile, the similarity between the produced image and its corresponding ground truth in the pixel space is guaranteed by combining adversarial learning with L2 norm. Moreover, in order to enhance the angular correlation of produced DW images, we incorporate an additional regularized term into the final objective function to narrow the angular inconsistency not only between produced images and ground-truth images, but also between image subsets with the same angular resolution.

The rest of this paper is structured as follows. In Section II, we describe the details of the proposed method. The data acquisition and processing, experimental setup, evaluation metrics and results are presented in Section III. Section IV provides the discussion, followed by the conclusion in Section V.

II. METHODS

A. Problem Formulation

The goal of the proposed method is to train a DL network that can increase the angular resolution of DW images for signal loss compensation in *in vivo* cardiac DTI. More specifically, given a set of lower angular-resolution *in vivo* DW images L: $\{l_1, l_2, \cdots, l_n\}$ that have been acquired in n diffusion gradient directions $\{d_1, d_2, \cdots, d_n\}$ and a set of higher angular-resolution DW images H: $\{h_1, h_2, \cdots, h_n, h_{n+1}, \cdots, h_m\}$ associated with m diffusion gradient directions $\{d_1, d_2, \cdots, d_n, d_{n+1}, \cdots, d_m\}$, we assume that the directions for L is a subset of directions for H. For clarity but without loss of generality, we set $m = 2n$ in our study. With the DL network, we try to learn the mapping G from L to H. Our final goal is to recover from L a set of images $G(L)$: $\{g_1, g_2, \cdots, g_m\}$ that are as similar as possible to the ground-truth higher-resolution images H.

B. GAN-based Super Angular Resolution

In the field of deep learning, a natural idea to learn the mapping G is to use a CNN-based super-resolution network [21]. However, the traditional CNN-based model that only minimizes the Euclidean distance between output and ground truth tends to neglect the information between neighboring pixels and thus causes blurring in the output image. Therefore, we adopted a generative adversarial network (GAN) [29] that utilizes an adversarial learning strategy to overcome the blurring problem and produce DW images that are perceptually similar to real DW images.

Fig. 1 illustrates an overview of the proposed GAN-based super angular-resolution network. It contains two sub-networks that are trained to compete with each other, a generator G and a discriminator D. The generator G is constructed to take existing lower angular-resolution DW images L as input and output higher angular-resolution DW images $G(L)$. D is essentially a discriminative model that takes both L and a set of higher angular-resolution DW images as input, where the higher angular-resolution images can be either $G(L)$ or ground-truth images H. It tries to determine whether the input is real pair $\{L, H\}$ or fake pair $\{L, G(L)\}$. During training the D, the network parameters are updated by minimizing the following loss:

$$\mathcal{L}_{GAN}(D) = -\mathbb{E}_L[\log(1 - D(L, G(L)))] \\ - \mathbb{E}_{L,H}[\log D(L, H)] \tag{1}$$

where the D minimizes $D(L, G(L))$ while maximizing $D(L, H)$ in order to improve its ability to classify real pair $\{L, H\}$ and fake pair $\{L, G(L)\}$. During training G, the network parameters are updated by minimizing:

$$\mathcal{L}_{GAN}(G) = \mathbb{E}_L[\log(1 - D(L, G(L)))] \tag{2}$$

where D is served as an additional network that has been updated using Eq. (1), then G minimizes $\log(1 - D(L, G(L)))$ to generate images more difficult to be classified by D. In general, Eqs. (1) and (2) are integrated into a min-max function expressed as:

$$\min_G \max_D \mathcal{L}_{GAN}(G, D) = \mathbb{E}_{L,H}[\log D(L, H)] + \\ \mathbb{E}_L[\log(1 - D(L, G(L)))] \tag{3}$$

where G attempts to minimize this loss function in opposition to D that tries to maximize it.

The adversarial learning strategy in Eq. (3) guides the network to produce DW images $G(L)$ perceptually similar to real images H, but it cannot guarantee the similarity between produced and ground-truth images in pixel space. Many previous studies have found that combining the adversarial loss with a traditional loss (e.g, L1 or L2 norm) is beneficial for improving the performance of image generation/translation [30], [31]. Therefore, in our method, we combine Eq. (3) with L2 norm $\mathcal{L}_{sim}$ to improve the similarity between the produced and ground-truth DW images. The $\mathcal{L}_{sim}$ here can be expressed as:

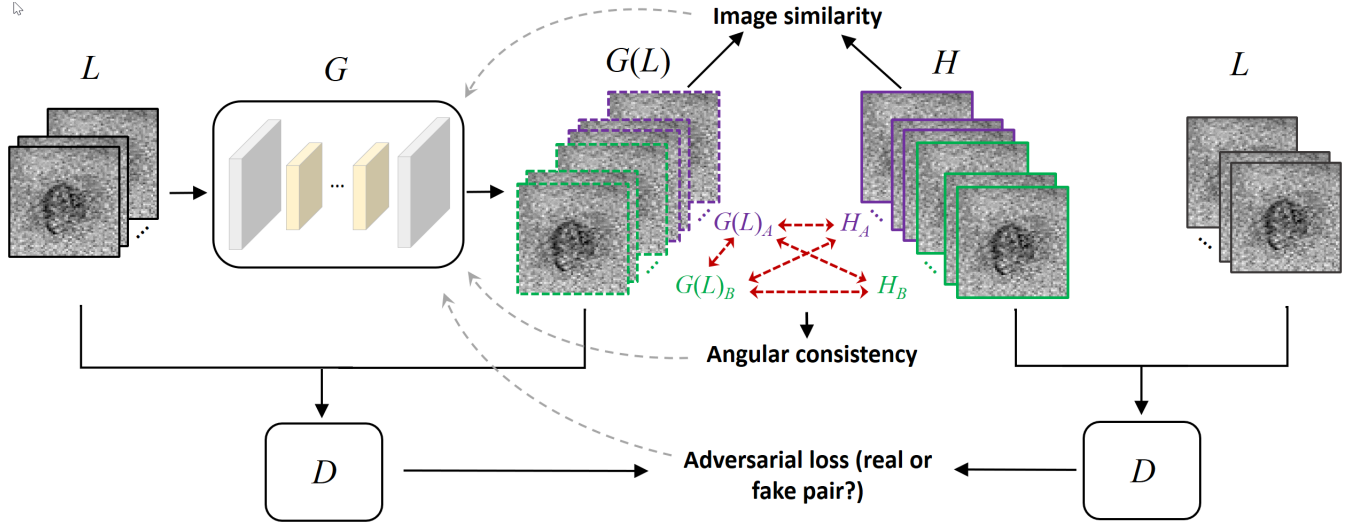$$\mathcal{L}_{sim} = \mathbb{E}_{H,L}[\|H - G(L)\|^2] \tag{4}$$

Fig. 1. Schematic illustration of the proposed GAN-based super angular-resolution network. The generator G learns the mapping from lower angular-resolution DW images L to higher angular-resolution DW images H based on three main constraints: image similarity, angular consistency, and adversarial learning strategy. The gray dashed arrows denote the back propagation to adjust the network parameters for updating G, and the red two-way solid arrows represent the angular consistency between DW image sets.

where $\|\cdot\|^2$ denotes a Euclidean distance function. For the DW images associated with multiple diffusion gradient directions, Eq. (4) can be equivalent to minimizing $\frac{1}{2n}\sum_{i=1}^{2n}\|h_i - g_i\|^2$.

C. Regularized Learning with Angular Consistency

As mentioned before, the ground-truth DW images H consist of $2n$ DW images associated with diffusion gradient directions $\{d_1, d_2, \cdots, d_{2n}\}$. A direction sampling algorithm [32] can be used to split the $2n$ directions into two subsets and each subset contains n directions uniformly distributed on a sphere. Then we can obtain two subsets of images H_A: $\{h_1, h_2, \cdots, h_n\}$ and H_B: $\{h_{n+1}, h_{n+2}, \cdots, h_{2n}\}$ according to the two subsets of directions. Similarly, we specify that the network's outputs $G(L)$ also consist of two subsets $G(L)_A$ and $G(L)_B$ which have the same angular resolution as their targets H_A and H_B, respectively. To further improve the performance of the proposed network for producing realistic DW images, we propose to incorporate an additional regularization term into the final objective function to enforce the angular correlation between these image sets. The underlying assumption behind this idea is that diffusion tensors estimated from the image sets with the same angular resolution should be consistent with each other, although their diffusion gradient directions are different.

Specifically, we denote by $\mathcal{M}(G(L))$ and $\mathcal{M}(H)$ two diffusion tensor maps estimated from the produced images $G(L)$ and their ground-truth images H, respectively, then the angular correlation between the two image sets can be enhanced by minimizing a consistency loss defined as:

$$\mathcal{L}_\varphi = \frac{1}{N}\sum_{s=1}^{N}\|\mathcal{M}(H)_s - \mathcal{M}(G(L))_s\|_F^2 \tag{5}$$

where N is the total number of diffusion tensors in each diffusion tensor map and s a spatial location of the tensor map. $\mathcal{M}(G(L))_s$ and $\mathcal{M}(H)_s$ denote sth 3×3 diffusion tensor matrices in $\mathcal{M}(G(L))$ and $\mathcal{M}(H)$, respectively, and $\|\cdot\|_F$ denotes the Frobenius norm. Besides, in order to enhance the consistency between image subsets with the same angular resolution, a similar idea is to minimize the

following function:

$$\mathcal{L}_\eta = \frac{1}{N}\sum_{s=1}^{N}\Big(\|\mathcal{M}(H_A)_s - \mathcal{M}(G(L)_A)_s\|_F^2 + \\ \|\mathcal{M}(H_B)_s - \mathcal{M}(G(L)_B)_s\|_F^2\Big) \tag{6}$$

where $\mathcal{M}(H_A)_s$, $\mathcal{M}(G(L)_A)_s$, $\mathcal{M}(H_B)_s$ and $\mathcal{M}(G(L)_B)_s$ are sth 3×3 diffusion tensor matrices in the diffusion tensor maps estimated from the image subsets H_A, $G(L)_A$, H_B and $G(L)_B$, respectively. Eq. (5) intends to reduce the angular difference between the higher angular-resolution DW images $G(L)$ and H, and Eq. (6) encourages the produced image subsets $G(L)_A$ and $G(L)_B$ are consistent with their targets H_A and H_B, respectively.

In theory, minimizing Eq. (5) combined with Eq. (6) is sufficient to ensure angular consistency between produced and ground-truth images. However, it is worth noting that the angular consistency also exists in DW images with different diffusion gradient directions. There should be high angular consistency among at least three image sets like $G(L)_A$, $G(L)_B$ and H_A (see Fig. 1). Therefore, we further improve the first term in Eq. (6) as follows:

$$\frac{1}{N}\sum_{s=1}^{N}\Big(\|\mathcal{M}(G(L)_A)_s + \mathcal{M}(G(L)_B)_s - 2\mathcal{M}(H_A)_s\|_F^2\Big) \tag{7}$$

In fact, Eq. (7) establishes a 3-cycle-consistency [33] strategy for minimizing the differences between $\mathcal{M}(G(L)_A)_s$, $\mathcal{M}(G(L)_B)_s$ and $\mathcal{M}(G(L)_B)_s$. Minimizing Eq. (7) ensures that the differences between any two of the three diffusion tensor maps would be minimized. Similarly, by changing the second term in Eq. (6), a new loss function can be written as:

$$\mathcal{L}_\eta' = \frac{1}{N}\sum_{s=1}^{N}\Big(\|\mathcal{M}(G(L)_A)_s\mathcal{M}(G(L)_B)_s - 2\mathcal{M}(H_A)_s\|_F^2 + \\ \|\mathcal{M}(G(L)_A)_s + \mathcal{M}(G(L)_B)_s - 2\mathcal{M}(H_B)_s\|_F^2\Big) \tag{8}$$

The final regularized term based on angular consistency is:

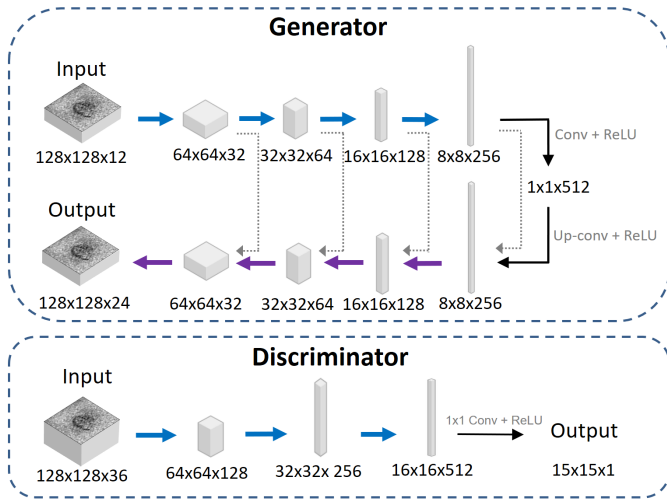$$\mathcal{L}_{cont} = \mathcal{L}_\varphi + \mathcal{L}_\eta' \tag{9}$$

Fig. 2. The detailed network architecture of the proposed generator and discriminator. The blue arrow denotes repeated 4 blocks of 4x4 convolutional layers followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with stride **2** for downsampling. The purple arrow denotes a 2x2 deconvolution layer with two 4x4 convolution layers followed by a ReLU. The gray dotted line denotes skip connections between encoder and decoder.

By combining Eq. (3), Eq. (4) and Eq. (9), we obtain our full objective function:

$$\mathcal{L}_{obj} = \arg \min_{G} \max_{D} \alpha \mathcal{L}_{GAN}(G, D) + \beta \mathcal{L}_{sim} + \lambda \mathcal{L}_{cont} \quad (10)$$

where α, β and λ are parameters for controlling the trade-off between adversarial learning, image similarity and angular consistency, respectively.

D. Network Architecture and Training Process

Fig. 2 provides the detailed architecture of the proposed generator and discriminator used in the present network. We adjusted the architecture to learn the mapping from 12 diffusion gradient directions to 24 diffusion gradient directions for super angular-resolution in *in vivo* cardiac DTI. The generator was constructed based on the Batch normalization modules [34] and U-Net [35] as encoder-decoder that fuses the features from shallow and deep layers through multi-path information to optimize the generator quickly. The generator's inputs and outputs are DW images of size $128 \times 128 \times 12$ and $128 \times 128 \times 24$, respectively, where their last channels correspond to the angular resolutions. A skip connection strategy [30] was used to connect the information between encoder and its corresponding decoder directly instead of letting the information flow pass through a mass of unrelated layers. The discriminator's inputs are $128 \times 128 \times 36$ DW images containing both the generator's input and higher angular-resolution output (or ground-truth images). Its output is a 15×15 image map where each pixel value is a probability that represents how believable the corresponding patch of the input image is. All convolutions are c $\times 4 \times 4 \times$f filters applied with stride 2, where c and f are the number of channels for the previous and next layers, respectively.

At the beginning of the training stage, we initially update parameters to train D. Then, the trained D was used to calculate an adversarial loss based on Eq. (9). Meanwhile, the images $G(L)$ and its ground-truth images H were employed to compute the similarity loss and angular consistency loss according to Eqs. (4) and (9), respectively. After that, the parameters for G were updated by combining all the computed losses. The images produced from the trained G can, in turn, be regarded as "fake" images to update

parameters for D. Finally, the capability of both G and D was increased gradually by repeating this process iteratively until all the real data $\{L, H\}$ and fake data $\{L, G(L)\}$ have no great influence on updating D.

III. EXPERIMENTS AND RESULTS

We conducted a number of experiments using human cardiac DTI datasets to evaluate the performance of the proposed method. In practice, DTI of the *in vivo* human heart is quite difficult to perform with up to 24 diffusion gradient directions. In order to train and test the proposed super angular-resolution network, in the experiments, *in vivo* cardiac DTI datasets were used in combination with *ex vivo* cardiac DTI datasets to imitate large amounts of lower angular-resolution motion-induced DW images (in 12 directions) and their corresponding ground-truth higher angular-resolution DW images (in 24 directions). In this study, the hearts used to acquire the cardiac DTI datasets were all obtained and processed in compliance with French legal and ethical guidelines. More details about the acquisition process of these hearts were described in [36] and in [19], [37]. All the experiments were carried out following the principles outlined in the Declaration of Helsinki [38].

In the following, we first describe the acquisition of original *in vivo* and *ex vivo* cardiac DTI datasets. Subsequently, the method including three pre-processing steps for imitating lower and higher angular-resolution DW images are introduced. After presenting the experimental setting and evaluation metrics, the experimental results are reported.

A. Datasets

Ex Vivo Cardiac Datasets: Six *ex vivo* human hearts from four infants and two adults were obtained and processed in compliance with French legal and ethical guidelines. The infant DTI datasets were acquired in clinical conditions with Siemens 3T MRI Magnetom Verio. The imaging parameters are the following: TE = 74 ms, TR = 7900 ms, FOV = 144×144 mm^2, slice thickness = 1.4 mm, in-plane resolution = 2 mm, slice spacing = 1.4 mm, slice duration = 123.2 ms, number of slices = 35, slice size = 104×104, diffusion sensitivity b = 700 s/mm^2, number of diffusion gradient directions = 192, 64 and 12. In each direction, MRI scans were acquired 6 times for noise reduction. The adult DTI datasets were acquired using Siemens 3T MRI Magnetom Prisma with parameters: TE = 71 ms, TR = 9600 ms, FOV = 177×177 mm^2, slice thickness=1.5 mm, slice spacing = 1.5 mm, slice duration = 123.1 ms, number of slices = 70, slice size = 122×122, diffusion sensitivity b = 700 s/mm^2, and number of diffusion gradient directions = 192. MRI scans were performed 3 times in each direction.

In Vivo Cardiac Datasets: Six healthy volunteers were recruited to acquire *in vivo* cardiac DW images with a 1.5 T clinical scanner (MAGNETOM Avanto, Siemens AG) at the Neuro-Cardiology Hospital of Lyon in France. All subjects gave informed consent to the institutional review board-approved study protocol prior to participation. The acquisition parameters are as follow: TE/TR = 51/100 ms, b-value = 200 s/mm^2, spatial resolution = $2.63 \times 2.63 \times 6$ mm^3, matrix size = 90×160, FOV = 420×236 mm^3, bandwidth = 1302 Hz, acceleration rate = 2, partial Fourier = 6/8. For each heart, 10 slices (slice thickness = 6 mm) were acquired in the short-axis view without any interslice gap. An individual mask was obtained to detect the myocardial region. The DW images were acquired in 10 different time points. For each time point, 13 DW images including 12 images associated with 12 diffusion gradient directions and one b$_0$ image (without diffusion weighting) were acquired.

B. Data Pre-processing

1) Sampling of Ex Vivo Cardiac DTI Data: We used the *ex vivo* cardiac DW images acquired in 192 diffusion gradient directions to sample data for training the mapping from low to high angular-resolution DW images. Specifically, for the 192 diffusion gradient directions, we sub-sampled 24 diffusion gradient directions from them using a q-space sampling algorithm called iterative maximum overlap construction (IMOC) [32], such that 24 directions are uniformly distributed on a sphere. By using the same algorithm, 12 directions uniformly distributed on a sphere were sub-sampled from the 24 directions. Then, two sets of DW images associated with the obtained 12 directions and 24 directions, respectively, were extracted from the 192 *ex vivo* DW images, which were served as the network's input and ground-truth, respectively.

However, training the network on a fixed set of 24 diffusion gradient directions (including the subset with 12 directions) may not be practical enough because the diffusion gradient directions used for a given angular resolution may be varied with different DTI machines that adopt different sampling schemes in practice. To design a network that can be used for different imaging machines or schemes, we further generated plenty of DW image sets as network's training data, where the associated 24 diffusion gradient directions are different from one set to another. It was shown in [39] that any independent set of diffusion gradient directions (uniformly spread on the same unit sphere) can be regarded as a rigid rotation of another set of directions with the same angular resolution. Subsequently, we rotated the fixed set of 24 directions 192 times on the sphere with the rotation angles determined by the 192 directions. Then, for each rotated direction, we extracted its nearest neighbor in the 192 directions based on Euclidean distance between directions so that we can obtain 192 sets of varied 24 directions. After that, from each set of the 24 directions, 12 directions were sub-sampled using the IMOC algorithm. Finally, 192 pairs of DW image sets were extracted from the *ex vivo* DW images according to the obtained sets of 24 directions and their corresponding 12 directions.

2) Direction Normalization: By definition, our network processes DW image set without using any directional information. Therefore, it is necessary to design a strategy that matches the diffusion gradient directions of input and output DW images. To achieve this, we first used the IMOC algorithm to generate a set of 24 diffusion gradient directions and its subset with 12 directions as two fixed sets of directions assigned to the output and input DW images, respectively. Then, each voxel in DW images associated with any new set of 12 or 24 diffusion gradient directions was normalized into the fixed set of directions using the equation below:

$$S_i' = S_0 e^{(-b \vec{g}_i^{\mathrm{T}} \underline{\mathbf{D}} \vec{g}_i)}, \tag{11}$$

where $\underline{\mathbf{D}}$ denotes the diffusion tensor computed from the un-normalized DW images for that voxel, $\vec{g}_i = (x_i, y_i, z_i)$ denotes the unit vector that designates the ith diffusion gradient direction in the fixed set with $i = 1, 2, \cdots n (n \geq 6)$, S_i' is its normalized value and b is the diffusion weighting factor [2]. Eq. (11) is the well-known Stejskal-Tanner equation [40]. The use of this equation ensures that the diffusion tensor, b-value and S_0 are kept constant during the normalization.

3) Imitation of In Vivo Motion-Induced DTI Data: After normalizing the *ex vivo* DW images, a method was used to model motion-induced effects and imitate motion-induced DW images. More precisely, given a set of *in vivo* DW images acquired in 12 diffusion gradient directions, the motion estimation algorithm PCATMIP was used to calculate mean motion-induced signal loss for each DW image. Then, a value x was computed by normalizing

the mean signal loss of all images to 0-1 for measuring the degree of signal loss. We defined the signal decay in the presence of motion as $S_m = S_i \cdot y$, where S_i and y represent the signal attenuation terms caused by molecular motion (diffusion) and cardiac motion, respectively. For each signal intensity S_m, its value of S_i can be estimated using PCATMIP algorithm so that the value of y can be obtained. After that, we modeled the relationship between the signal loss degree x and the motion-induced attenuation term y as an exponential function defined by $y = A \exp(Bx + C)$, where A, B and C are the estimation coefficients. This model was solved by the regression estimation using DW images over six *in vivo* human hearts. The obtained model provides a mean signal loss value given an arbitrary degree x. Besides, based on the value of x, we roughly divided DW images into three categories with minor, moderate, and huge signal losses. This allows us to determine the mean numbers of images for different category of signal loss. Furthermore, for each category of signal loss, the proportion of lower and higher pixel values for one image were also calculated. Based on these results, realistic motion-induced signal loss for each pixel was generated and then added to *ex vivo* DW images for imitating *in vivo* DTI data with signal loss.

C. Experimental Setting

To conduct the experiments, DTI datasets of six *ex vivo* human hearts were used to imitate lower and higher angular-resolution DW images. For each dataset, we selected 10 of the 35 slices (from the middle third of the heart) of the *ex vivo* DW images to imitate the number of slices often used in *in vivo* setting. A myocardium mask for each slice was obtained by thresholding the apparent diffusion coefficient (ADC) images calculated from the DW images using an upper threshold of $2 \times 10^{-3} mm^2/s$ (following the method in [36]). With the myocardium mask, the myocardial area of each image was extracted as region of interest (ROI). For each slice, 192 pairs of DW image sets were generated using the three data pre-processing steps detailed in Section III-B. Each pair of sets was regarded as a training sample consisting of a set of motion-induced 12 DW images and a set of 24 ground-truth DW images without motion effects. That is, for each heart, a total of $192 \times 10 = 1920$ training samples for the 10 slices were generated.

We first trained and tested the proposed network on the imitated DW images of six human hearts. For each heart, we randomly selected 1400 samples from the generated 1920 training samples for training and used the remaining 520 samples for testing. Subsequently, we tested the proposed network directly on the original *in vivo* cardiac DTI datasets for further validation. To this end, 6000 samples were randomly selected from the training samples of all the six human hearts (1000 samples per heart) to train the proposed network. According to our experimental experiences, we set parameters $\alpha = 1$, $\beta = 1$ and $\lambda = 50$ to achieve an optimal balance for minimizing the loss function. All data pre-processing steps were performed using MATLAB (MathWorks, Inc., Natick, MA, USA). The network algorithm was run on a Linux cluster with a Matrox GPU and 32 GB of system memory (RAM).

D. Evaluation Metrics

Several metrics were employed to evaluate the performance of the proposed network in both qualitative and quantitative manners. First, diffusion tensors were estimated from the DW images to be evaluated using the least squares estimation. Then, for each estimated diffusion tensor, we calculated the FA and MD that describe the degree of anisotropy and the overall diffusion level of water molecules, respectively. Finally, main fiber orientations represented by azimuth angle and elevation angle were calculated using the primary eigenvector of the diffusion tensor.
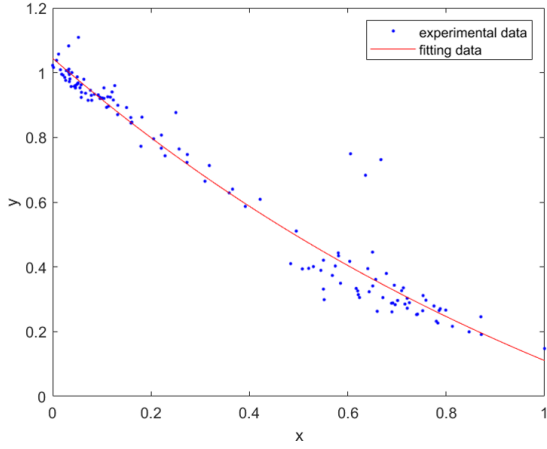
Fig. 3. Motion-induced attenuation term as a function of signal loss degree over the six human hearts. For each blue point (the experimental data), the x value denotes the degree of signal loss and the y value the mean signal loss computed from all pixels averaged inside each of image on the myocardium area.
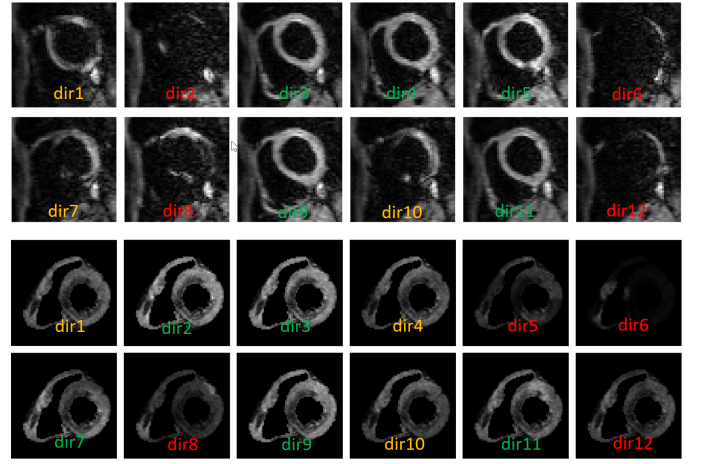

Fig. 4. An example of imitated motion-induced DW images for 12 directions. Top two rows: *In vivo* cardiac DW images. Bottom two rows: motion-induced DW images obtained by adding the estimated signal loss to *ex vivo* DW images. The direction labels superimposed on the image with green, yellow and red colors designate minor, moderate and huge signal losses, respectively.

For experiments performed on imitated DW images, we employed a similar method to that in [41] to measure estimation errors in terms of the above metrics. Specifically, the *ex vivo* DW images acquired in 192 directions were averaged over all time points and used as reference images. The errors in estimating diffusion tensors were measured by computing the Frobenius distance (FD) between estimated tensors and reference tensors (from the reference images). The errors in estimating FA, MD, azimuth angle and elevation angle were calculated by computing the mean absolute error (MAE) between estimated values and reference values (computed from reference tensor). These estimation errors were first computed for each myocardial voxel and then the mean of all errors (ME) was computed. The standard deviation of estimation errors (ErrSD) was also computed using:

$$ErrSD = \left(\frac{1}{N_{mv}} \sum_{i=1}^{N_{dt}} (Err_i - ME)^2 \right)^{\frac{1}{2}} \qquad (12)$$

where N_{mv} is the number of myocardial voxels, Err_i represents the estimation error (FD for the diffusion tensor and MAE for other metrics) at the ith voxel. For experiments directly performed on *in vivo* cardiac DTI datasets, the values of FA, MD, azimuth angle and elevation angle from the resulting DW images were used for comparison and analysis.

E. Results

In this section, we provide experimental results in three main parts. First, we show the results of imitating the motion-induced DTI data used for training the proposed network. Second, we give some results of analyzing the performance of the proposed network based on adverserval learning and angular consistency loss. Finally, we present the detailed evaluation results of the proposed network for the compensation of motion-induced signal loss in *in vivo* cardiac DTI.

Imitated Motion-induced DW Images: Fig. 3 plots the estimated relationship between the signal loss degree x and the mean signal loss value y over six human hearts. The results were obtained by fitting y as an exponential function of x using regression estimation (as mentioned in Section III-B.3). The fitting parameters of this model were RMAE = 0.038 and R-square = 0.983. As we can see, the fitted results roughly follow the behavior of monotonically decreasing exponential function. This fitted model provides a mean

value to generate realistic motion-induced signal loss.

Fig. 4 shows an example of motion-induced DW images imitated by adding the realistic signal loss to *ex vivo* DW images. Both the *in vivo* and original *ex vivo* DW images were acquired in 12 diffusion gradient directions. The mean numbers of images in the 12 directions over all *in vivo* human heart were 5, 3, and 4, respectively for minor, moderate and huge signal losses. For minor signal-loss images, the mean proportion of higher pixel values h_v ($h_v \geq 2\mathcal{R}/3$, where $\mathcal{R}$ is the maximum pixel value in the image) per image are 86%, and this proportion is 51% and 6% for moderate and huge signal losses, respectively. As observed, for a set of 12 diffusion gradient directions, the method imitates highly realistic motion-induced DW images.

Impact of Adversarial Learning and Angular Consistency: Fig. 5 gives the qualitative results of the network trained with and without the adversarial loss $\mathcal{L}_{GAN}(G, D)$. The results of training without adversarial learning were achieved by setting $\alpha = 0$, $\beta = 1$ and $\lambda = 0$, which means the case that the network only minimizes the similarity between the produced and ground-truth DW images using L2 norm. The images produced by running the network with adversarial loss were obtained by changing α into 1 for training. We can observe that without using the adversarial loss (Fig. 5, top row), the intensities of the produced DW signals are blurred and visually unrealistic. In contrast, the produced images using the adversarial learning combined with the L2 norm (Fig. 5, middle row) present more variations and are perceptually closer to the ground truth.

Table I reports the ME $\pm$ ErrSD of the diffusion tensors, FA, MD, elevation and azimuth angles for the motion-induced DW images, motion-compensated DW images produced by the proposed network

TABLE I

PERFORMANCE OF SIGNAL LOSS COMPENSATION USING THE PROPOSED NETWORK WITH AND WITHOUT ADVERSARIAL LEARNING (ADV.). THE NUMBERS DESIGNATE THE ME $\pm$ ERRSD VALUES.

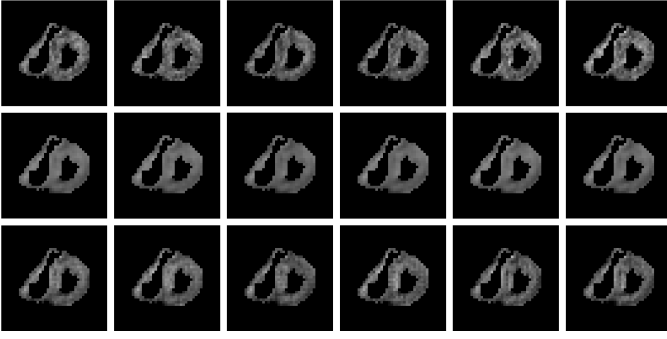|	Motion-induced	Motion-compensated	
		Without Adv.	With Adv.
Diffusion tensor	0.07 ± 0.16	0.07 ± 0.14	0.05 ± 0.10
FA	0.01 ± 0.03	0.01 ± 0.03	0.01 ± 0.02
MD	0.03 ± 0.05	0.02 ± 0.05	0.01 ± 0.03
Elevation	8.91 ± 26.26	8.21 ± 25.67	7.84 ± 25.11
Azimuth	5.78 ± 18.13	5.61 ± 17.99	5.12 ± 17.10

Fig. 5. Visual comparison of the cardiac DW images produced by our network without (middle row) and with (bottom row) adversarial learning, and ground-truth DW images (top row). Each column corresponds to one of six diffusion gradient directions.

TABLE II
COMPARISON BETWEEN WITH AND WITHOUT ANGULAR CONSISTENCY (AC) WHEN USING THE PROPOSED NETWORK FOR THE ESTIMATION ERRORS (ME ± ERRSD) OF DIFFUSION TENSOR AND FIBER ORIENTATION.

	With AC	Without AC
Diffusion tensor	0.05 ± 0.10	0.04 ± 0.08
Elevation	7.84 ± 25.11	7.74 ± 24.60
Azimuth	5.12 ± 17.10	4.67 ± 16.20

with and without adversarial learning on imitated DW images of an infant human heart. The unit of MD is 10^{-3} mm^2/s. As observed, the ME of all the metrics was decreased after performing signal loss compensation using the proposed network. When using the network without adversarial learning, the ME of these metrics was slightly but not significantly reduced. The mean errors of elevation angles and azimuth angles for the network with adversarial learning were decreased by $0.57°$ and $0.62°$, respectively, compared with the network without adversarial learning.

Table II summarizes the ME $\pm$ ErrSD of diffusion tensors, azimuth and elevation angles for the proposed method with and without using the consistency loss (Eq. (5) and (8), or Eq. (10)). We performed the training without angular consistency by setting $\lambda = 0$ and kept the optimal parameters for the others. From Table II, we can see that, compared with the network without angular consistency loss, the errors in estimating the elevation and azimuth angles are further decreased by $0.60°$ and $0.53°$, respectively, and that the diffusion tensors are decreased by 0.01×10^{-3} mm^2/s. This indicates the interest of enhancing angular consistency for improving the effect of signal loss compensation.

Performance of Signal Loss Compensation: Fig. 6 gives the 3D visualization of diffusion tensor fields derived from imitated motion-induced DW images, signal loss-compensated DW images, and reference DW images for one slice of the left ventricle of an adult human heart. The three zoom views represent the same region of the left ventricle (LV) containing fiber structures from epicardium to endocardium. The motion-induced DW images used here correspond to the images shown in Fig. 4. From Fig. 6(b), it can be observed that the shapes of most tensor cuboids for the motion-induced images associated with 12 diffusion gradient directions are not very different from each other, and that their major axes (main orientations of tensor cuboids) are almost identical. After performing the signal loss compensation, these cuboids (Fig. 6(c)) became more anisotropic and closer to the reference tensors (Fig. 6(a)). By further observing the zoom views of Fig. 6(c), we can find that the main orientations of
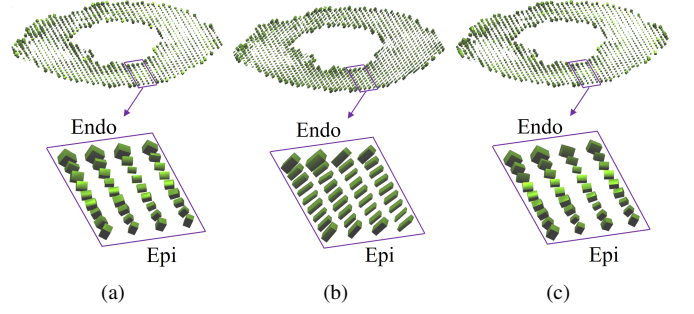

Fig. 6. 3D visualization of diffusion tensor fields using superquadric glyphs derived from (a) reference DW images, (b) imitated motion-induced DW images, and (c) signal loss-compensated DW images obtained using the proposed network. (Epi: epicardium. Endo: endocardium.)
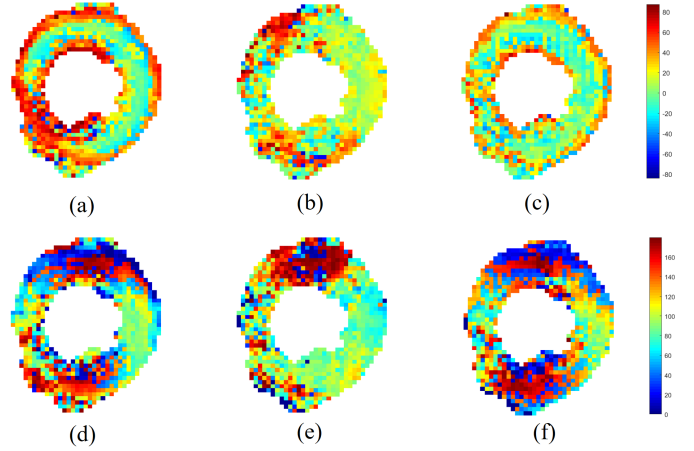

Fig. 7. Elevation angle maps (from (a) to (c)) and azimuth angle maps (from (d) to (f)). (a) and (d) Angle maps derived from reference DW images. (b) and (e) Angle maps derived from DW images corrupted by motion. (c) and (f) Angle maps derived from the signal loss-compensated DW images after using the proposed network.

tensor cuboids linearly vary from epicardium to endocardium. This variation roughly reflects the helical structure of normal myocardial fibers of the human heart.

Fig. 7 shows the elevation and azimuth angle maps calculated from imitated motion-induced DW images with and without signal loss compensation. Here we take the LV myocardium as our region of interest. From Fig. 7(a), it is observed that most elevation angles in the LV myocardium vary from about $60°$ at the endocardium to about $10°$ at the mid-wall, and back to about $50°$ at the epicardium. Also, as observed in Fig. 7(d), there is an approximately continuous circular variation for the azimuth angles. When observing clockwisely around the LV, the azimuth angles vary periodically from $0°$ to $180°$, and areas having nearly similar color appear twice. The variations of elevation and azimuth angles in the reference maps reflect the cross-helical structure of cardiac fibers. Such variations are difficult to be found in the elevation and azimuth angles maps derived from the motion-induced DW images (Fig. 7(b) and (d)). After using the proposed signal loss-compensation method, regular variation patterns of elevation angles on the LV were largely recovered except for the red area (Fig. 7(a)). The variation pattern of azimuth angles was nearly completely recovered.

Table III sumarizes the ME $\pm$ ErrSD of elevation and azimuth obtained before and after using the proposed network on the imitated motion-induced DW images of six human hearts. Clearly, for all the
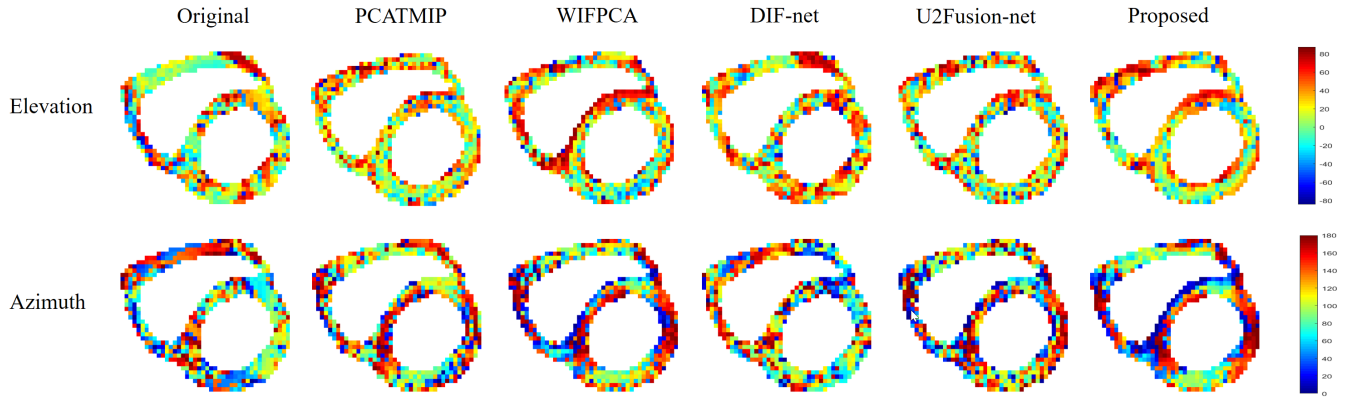
Fig. 8. Elevation angle maps (top row) and azimuth angle maps (bottom row) computed from the original *in vivo* cardiac DW images (first column on the left) and the DW images after signal loss compensation using the proposed network and state-of-the art ones (from the second to the last columns).

TABLE III
ESTIMATION ERRORS (ME ± ERRSD) OF FIBER ORIENTATION COMPUTED FROM THE IMITATED MOTION-INDUCED AND SIGNAL LOSS-COMPENSATED DW IMAGES FOR THE SIX HUMAN HEARTS (H*i* DENOTE THE DATASET OF THE *i*TH HUMAN HEART).

	Motion-induced		Signal loss-compensated	
	Elevation	Azimuth	Elevation	Azimuth
H1	5.02 ± 16.41	7.65 ± 24.66	3.01 ± 13.25	4.67 ± 19.04
H2	5.48 ± 16.98	8.66 ± 26.06	4.28 ± 14.56	7.06 ± 23.07
H3	5.04 ± 15.48	7.85 ± 23.66	3.68 ± 13.12	5.99 ± 21.43
H4	7.11 ± 19.64	10.09 ± 26.96	4.67 ± 16.20	7.74 ± 24.60
H5	9.63 ± 22.31	16.24 ± 35.95	9.14 ± 21.65	12.97 ± 32.62
H6	9.73 ± 20.67	15.18 ± 32.02	5.41 ± 15.91	8.89 ± 26.41
mean	7.35 ± 18.58	11.43 ± 28.13	$\mathbf{5.70 \pm 16.45}$	$\mathbf{8.55 \pm 24.53}$

TABLE IV
COMPARISON RESULTS OF THE PROPOSED METHOD WITH STATE-OF-THE ART ONES. THE NUMBERS DESIGNATE THE ME ± ERRSD VALUES.

	Elevation	Azimuth
Motion-induced	7.11 ± 19.64	10.09 ± 26.96
PCATMIP [18]	6.84 ± 17.34	9.40 ± 26.04
WIFPCA [19]	6.52 ± 16.86	9.07 ± 25.07
DIF-net [42]	8.52 ± 20.37	10.40 ± 28.14
U2fusion-net [43]	6.41 ± 19.25	9.42 ± 27.49
Proposed	$\mathbf{5.70 \pm 16.45}$	$\mathbf{8.55 \pm 24.53}$

TABLE V
MEAN±SD OF FA AND MD IN THE LV OF SIX *in vivo* HUMAN HEART. THE UNIT OF MD IS 10^{-3} MM^2/S.

	FA±SD	MD±SD
Original	0.63 ± 0.14	1.68 ± 0.24
PCATMIP	0.51 ± 0.12	0.85 ± 0.11
WIFPCA	0.46 ± 0.10	0.73 ± 0.09
DIF-net	1.52 ± 0.67	2.40 ± 0.18
U2fusion-net	0.61 ± 0.16	1.42 ± 0.12
Proposed	$\mathbf{0.42 \pm 0.08}$	$\mathbf{0.65 \pm 0.09}$

six hearts, the proposed network significantly reduced the errors of estimating fiber angles. Besides, the proposed network also reduced the variation of estimation errors (see the values of ErrSD). The mean values over all the hearts (last row of Table III) show that the elevation and azimuth MEs were decreased by $1.65°$ and $2.88°$, respectively, and that these ErrSD values were $2.13°$ and $3.60°$, respectively.

Table IV provides a comparison of the proposed network with state-of-the art ones on the imitated data, where the ME and ErrSD values over all of the hearts were reported. The PCATMIP and WIFPCA were proposed specifically for compensation of motion-induced signal loss in *in vivo* cardiac DTI, and they relied on combining the DW images from multiple time points. The DIF-net and U2fusion-net were two representative CNN-based methods developed for image fusion tasks. Here motion-induced signal loss were added to the *ex vivo* DW images acquired at multiple time points (without signal averages) to imitate *in vivo* DW images and then we run these four methods on the imitated motion-induced images acquired at multiple time points. For the proposed method, we simply averaged the images over all time points and then generated motion-induced data for training the network. From the results, we can see that the ME and ErrSD values of both azimuth and elevation angles for WIFPCA are lower than that values for PCATMIP and the other two image fusion networks. But the proposed network gave the best results by significantly reducing estimation errors of fiber angles especially the elevation angles.

Fig. 8 shows the elevation and azimuth angle maps obtained by running the proposed network and state-of-the art ones directly on cardiac DW images of an *in vivo* human heart. It is observed that the original elevation angles (computed from the original *in vivo* DW images) in the LV vary irregularly from epicardium to

endocardium, and that the original azimuth angles in the LV do not exhibit circularly periodic variation pattern. After signal loss compensation using the PCATMIP, WIFPCA, DIF-net, U2Fusion-net and the proposed method, the elevation angles in the LV showed a roughly linear variation from epiocardium to endocardium, and their corresponding circularly periodic variation of the azimuth angles were also recovered. In particular, regular elevation and azimuth angle variations for the WIFPCA and the proposed method were observed more clearly compared with the other methods. But angles for the proposed method showed a relatively lower noise level compared with that for the WIFPCA. Table V summarizes the mean and standard deviation (Mean±SD) of FA and MD values obtained after using the proposed method and state-of-the art ones on the cardiac DTI datasets of all the six *in vivo* human hearts. We observe that the mean and variations of both FA and MD for WIFPCA were lower than those for PCATMIP, which is consistent with the results in [19]. In contrast, the proposed method yielded the lowest mean and variations of FA (0.42 ± 0.08) and MD ($0.65\pm0.09\times10^{-3}$ mm^2/s) compared with all the other methods.

IV. DISCUSSION

We proposed an efficient GAN-based method to compensate for motion-induced signal loss in *in vivo* cardiac DTI. It improved the angular resolution of acquired motion-induced DW images by training a GAN network based on three major mechanisms: 1) utilizing the adversarial learning strategy to generate realistic higher angular-resolution DW images, 2) incorporating L2 norm to guarantee the similarity between generated and ground-truth pixels, 3) using a regularized loss to minimize angular inconsistency in DW image sets generated for the same angular resolution.

There are mainly two major challenges in implementing the proposed network. One is that cardiac DTI data required for training the present network to perform super angular resolution are quite difficult to obtain in practice. These data include the DW images associated with the mapping from 12 to 24 diffusion gradient directions (used as the network's input and ground-truth images, respectively) and the DW images associated with different sets of of diffusion gradient directions (used for applying the trained network to different imaging machines or schemes). The pre-processing strategies proposed in Section III-B.1 make it possible to generate enough DW images associated with these desired directions from existing *ex vivo* DW images acquired in 192 directions. The other challenge is lacking of *in vivo* cardiac DTI data for enabling the network to achieve signal loss compensation in *in vivo* DTI. The method presented in III-B.3 allows us to cope with this problem by adding realistic motion-induced signal losses to *ex vivo* DW images to imitate *in vivo* DW images corrupted by cardiac motion. A similar method for imitating *in vivo* cardiac DTI data was proposed in [37], where motion-induced signal losses were estimated from DENSE acquisitions [44]. In contrast, our imitation method is capable of estimating signal losses directly from a given *in vivo* DTI dataset and does not rely on any specific imaging sequence, which is more flexible and easier to implement.

During the training process, the data fed to the proposed network contain 36 DW images including 12 input images and 24 ground-truth images to learn the mapping from 12 to 24 directions (as mentioned in Section III-C). In fact, since we specified that the directions of input are a subset of output, an alternative way to perform super angular resolution is to input 12 images associated with original 12 directions to the network and let it output 12 images associated with 12 new directions. In this study, since the DW images of all 24 directions were served as network's output, the correlation between the DW images associated with the original 12 directions and new 12 directions were both enhanced after optimizing the network with adversarial learning and angular consistency. This is also an important reason contributing to the impressive performance of the proposed method.

From the results in Fig. 7, it can be observed that the angular maps are noisy. This is because we directly derived these angle maps from the DW images and without using additional pre-processings such as interpolation or denoising. Even so, we also found some main features of elevation and azimuth angles, which present in normal cardiac fibers, e.g., regular variations from epicardium to endocardium of LV.

In Table III, an interesting phenomenon is that the estimation errors of ME and ErrSD for the four infant hearts (from H1 to H4) are generally smaller than the errors from the two adult hearts (H5 and H6). An intuitive understanding is that this difference reflects the fact that the network compensates differently for signal loss depending on whether the myocardial structures are infant or adults. However, this phenomenon might also be caused by the differences in imaging parameters used to acquire datasets, such as angular resolution, slice thickness and number, etc, or due to the use of different DTI machines. Further study would be needed to investigate the impact of these factors on the performance of the proposed method.

The present network is dedicated to super angular resolution of motion-induced DW images acquired in 12 directions. As such it cannot take directly the data associated with other angular resolutions as input. Future work is required to modify the network so that it can be applied to DW images of new angular resolutions. Besides, the data used for training the proposed network were only obtained by adding motion-induced signal loss to *ex vivo* DW images. It would be interesting in our next step to train the proposed network using *in vivo* cardiac DTI datasets. In addition, the results of Table II also show that the network performance can be improved by constraining the diffusion tensor estimated from the generated images. From there on, another interesting attempt might be to introduce new DTI-derived parameters (e.g., FA or fiber orientation) as an additional regularized term in loss function, or redesign new sub-networks based on these parameters.

V. CONCLUSIONS

We proposed a GAN-based super angular-resolution method for compensating for motion-induced signal loss in *in vivo* DTI of the human heart. The constructed network used an objective function based on adversarial learning, image similarity and angular consistency. Several pre-processing strategies of imitating lower and higher angular resolutions in *in vivo* DTI data were designed to train and test the network. The experimental results showed that the proposed method presents the ability to compensate for motion-induced signal loss in terms of accurately estimating diffusion tensors and DTI-derived parameters such as FA and MD. Moreover, the results also demonstrated that the use of angular consistency improved the network in producing DW images from which fiber orientations were more accurately derived. The proposed method then provides an interesting method for the compensation of motions in *in vivo* cardiac DTI.

REFERENCES

[1] P. B. Kingsley, "Introduction to diffusion tensor imaging mathematics: Part i. tensors, rotations, and eigenvectors," *Concepts in Magnetic Resonance Part A*, vol. 28, no. 2, pp. 101–122, 2006.

[2] ——, "Introduction to diffusion tensor imaging mathematics: Part ii. anisotropy, diffusion-weighting factors, and gradient encoding schemes," *Concepts in Magnetic Resonance Part A*, vol. 28, no. 2, pp. 123–154, 2006.

[3] ——, "Introduction to diffusion tensor imaging mathematics: Part iii. tensor calculation, noise, simulations, and optimization," *Concepts in Magnetic Resonance Part A*, vol. 28, no. 2, pp. 155–179, 2006.

[4] M.-T. Wu, W.-Y. I. Tseng, M.-Y. M. Su, C.-P. Liu, K.-R. Chiou, V. J. Wedeen, T. G. Reese, and C.-F. Yang, "Diffusion tensor magnetic resonance imaging mapping the fiber architecture remodeling in human myocardium after infarction: correlation with viability and wall motion," *Circulation*, vol. 114, no. 10, pp. 1036–1045, 2006.

[5] W.-Y. I. Tseng, J. Dou, T. G. Reese, and V. J. Wedeen, "Imaging myocardial fiber disarray and intramural strain hypokinesis in hypertrophic cardiomyopathy with mri," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 23, no. 1, pp. 1–8, 2006.

[6] R. R. Edelman, J. Gaa, V. J. Wedeen, E. Loh, J. M. Hare, P. Prasad, and W. Li, "In vivo measurement of water diffusion in the human heart," *Magnetic resonance in medicine*, vol. 32, no. 3, pp. 423–428, 1994.

[7] M.-T. Wu, M.-Y. M. Su, Y.-L. Huang, K.-R. Chiou, P. Yang, H.-B. Pan, T. G. Reese, V. J. Wedeen, and W.-Y. I. Tseng, "Sequential changes of myocardial microstructure in patients postmyocardial infarction by diffusion-tensor cardiac mr: correlation with left ventricular structure and function," *Circulation: Cardiovascular Imaging*, vol. 2, no. 1, pp. 32–40, 2009.

[8] J. Dou, T. G. Reese, W.-Y. I. Tseng, and V. J. Wedeen, "Cardiac diffusion mri without motion effects," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 48, no. 1, pp. 105–114, 2002.

[9] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri," *Journal of magnetic resonance*, vol. 213, no. 2, pp. 560–570, 2011.

[10] E. X. Wu, Y. Wu, J. M. Nicholls, J. Wang, S. Liao, S. Zhu, C.-P. Lau, and H.-F. Tse, "Mr diffusion tensor imaging study of postinfarct myocardium structural remodeling in a porcine model," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 4, pp. 687–695, 2007.

[11] D. E. Sosnovik, C. Mekkaoui, S. Huang, H. H. Chen, G. Dai, C. T. Stoeck, S. Ngoy, J. Guan, R. Wang, W. J. Kostis, *et al.*, "Microstructural impact of ischemia and bone marrow–derived cell therapy revealed with diffusion tensor magnetic resonance imaging tractography of the heart in vivo," *Circulation*, vol. 129, no. 17, pp. 1731–1741, 2014.

[12] U. Gamper, P. Boesiger, and S. Kozerke, "Diffusion imaging of the in vivo heart using spin echoes–considerations on bulk motion sensitivity," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 57, no. 2, pp. 331–337, 2007.

[13] T. G. Reese, R. M. Weisskoff, R. N. Smith, B. R. Rosen, R. E. Dinsmore, and V. J. Wedeen, "Imaging myocardial fiber architecture in vivo with magnetic resonance," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 786–791, 1995.

[14] W.-Y. I. Tseng, T. G. Reese, R. M. Weisskoff, and V. J. Wedeen, "Cardiac diffusion tensor mri in vivo without strain correction," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 2, pp. 393–403, 1999.

[15] S. Nielles-Vallespin, C. Mekkaoui, P. Gatehouse, T. G. Reese, J. Keegan, P. F. Ferreira, S. Collins, P. Speier, T. Feiweier, R. De Silva, *et al.*, "In vivo diffusion tensor mri of the human heart: reproducibility of breath-hold and navigator-based approaches," *Magnetic resonance in medicine*, vol. 70, no. 2, pp. 454–465, 2013.

[16] C. T. Stoeck, C. Von Deuster, M. Genet, D. Atkinson, and S. Kozerke, "Second-order motion-compensated spin echo diffusion tensor imaging of the human heart," *Magnetic resonance in medicine*, vol. 75, no. 4, pp. 1669–1676, 2016.

[17] C. Nguyen, Z. Fan, B. Sharif, Y. He, R. Dharmakumar, D. S. Berman, and D. Li, "In vivo three-dimensional high resolution cardiac diffusion-weighted mri: a motion compensated diffusion-prepared balanced steady-state free precession approach," *Magnetic resonance in medicine*, vol. 72, no. 5, pp. 1257–1267, 2014.

[18] V. Pai, S. Rapacchi, P. Kellman, P. Croisille, and H. Wen, "Pcatmip: enhancing signal intensity in diffusion-weighted magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 65, no. 6, pp. 1611–1619, 2011.

[19] H. Wei, M. Viallon, B. M. Delattre, K. Moulin, F. Yang, P. Croisille, and Y. Zhu, "Free-breathing diffusion tensor imaging and tractography of the human heart in healthy volunteers using wavelet-based image fusion," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 306–316, 2014.

[20] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[21] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[22] Z. Deng, L. Wang, Q. Wu, Q. Chen, Y. Cao, L. Wang, X. Cheng, J. Zhang, and Y. Zhu, "Investigation of in vivo human cardiac diffusion tensor imaging using unsupervised dense encoder-fusion-decoder network," *IEEE Access*, vol. 8, pp. 220 140–220 151, 2020.

[23] J.-D. Tournier, F. Calamante, and A. Connelly, "Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging," *NMR in Biomedicine*, vol. 26, no. 12, pp. 1775–1786, 2013.

[24] D. K. Jones, "The effect of gradient sampling schemes on measures derived from diffusion tensor mri: a monte carlo study," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 51, no. 4, pp. 807–815, 2004.

[25] L. Zhan, N. Jahanshad, D. B. Ennis, Y. Jin, M. A. Bernstein, B. J. Borowski, C. R. Jack Jr, A. W. Toga, A. D. Leow, and P. M. Thompson, "Angular versus spatial resolution trade-offs for diffusion imaging under time constraints," *Human brain mapping*, vol. 34, no. 10, pp. 2688–2706, 2013.

[26] Y. He, L. Wang, F. Yang, Y. Xia, P. Clarysse, and Y. Zhu, "Systematic study of joint influence of angular resolution and noise in cardiac diffusion tensor imaging," in *International Conference on Functional Imaging and Modeling of the Heart*. Springer, 2021, pp. 200–210.

[27] L.-A. McGill, T. F. Ismail, S. Nielles-Vallespin, P. Ferreira, A. D. Scott, M. Roughton, P. J. Kilner, S. Y. Ho, K. P. McCarthy, P. D. Gatehouse, *et al.*, "Reproducibility of in-vivo diffusion tensor cardio-vascular magnetic resonance in hypertrophic cardiomyopathy," *Journal of Cardiovascular Magnetic Resonance*, vol. 14, no. 1, pp. 1–15, 2012.

[28] A. D. Scott, P. F. Ferreira, S. Nielles-Vallespin, P. Gatehouse, L.-A. McGill, P. Kilner, D. J. Pennell, and D. N. Firmin, "Optimal diffusion weighting for in vivo cardiac diffusion tensor imaging," *Magnetic resonance in medicine*, vol. 74, no. 2, pp. 420–430, 2015.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[31] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.

[32] J. Cheng, D. Shen, P.-T. Yap, and P. J. Basser, "Single-and multiple-shell uniform sampling schemes for diffusion mri using spherical codes," *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 185–199, 2017.

[33] T. Zhou, Y. Jae Lee, S. X. Yu, and A. A. Efros, "Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1191–1200.

[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[36] F. Yang, Y.-M. Zhu, G. Michalowicz, P.-S. Jouk, L. Fanton, M. Viallon, P. Clarysse, P. Croisille, and Y. Usson, "Quantitative comparison of human myocardial fiber orientations derived from dti and polarized light imaging," *Physics in Medicine & Biology*, vol. 63, no. 21, p. 215003, 2018.

[37] H. Wei, M. Viallon, B. M. Delattre, L. Wang, V. M. Pai, H. Wen, H. Xue, C. Guetter, P. Croisille, and Y. Zhu, "Assessment of cardiac motion effects on the fiber architecture of the human heart in vivo," *IEEE transactions on medical imaging*, vol. 32, no. 10, pp. 1928–1938, 2013.

[38] R. V. Carlson, K. M. Boyd, and D. J. Webb, "The revision of the declaration of helsinki: past, present and future," *British journal of clinical pharmacology*, vol. 57, no. 6, pp. 695–713, 2004.

[39] E. Caruyer, C. Lenglet, G. Sapiro, and R. Deriche, "Design of multishell sampling schemes with uniform coverage in diffusion mri," *Magnetic resonance in medicine*, vol. 69, no. 6, pp. 1534–1540, 2013.

[40] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient," *The journal of chemical physics*, vol. 42, no. 1, pp. 288–292, 1965.

[41] F. Yang, Y.-M. Zhu, S. Rapacchi, J. Luo, M. Robini, and P. Croisille, "Interpolation of vector fields from human cardiac dt-mri," *Physics in Medicine & Biology*, vol. 56, no. 5, p. 1415, 2011.

[42] H. Jung, Y. Kim, H. Jang, N. Ha, and K. Sohn, "Unsupervised deep image fusion with structure tensor representations," *IEEE Transactions on Image Processing*, vol. 29, pp. 3845–3858, 2020.

[43] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2fusion: A unified unsupervised image fusion network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[44] A. H. Aletras, S. Ding, R. S. Balaban, and H. Wen, "Dense: displacement encoding with stimulated echoes in cardiac functional mri," *Journal of magnetic resonance (San Diego, Calif.: 1997)*, vol. 137, no. 1, p. 247, 1999.