QID^2: An Image-Conditioned Diffusion Model for *Q*-space Up-sampling of DWI Data

Zijian Chen*, Jueqi Wang* and Archana Venkataraman

Department of Electrical and Computer Engineering, Boston University
{zijianc,jueqiw,archanav}@bu.edu

Abstract. We propose an image-conditioned diffusion model to estimate high angular resolution diffusion weighted imaging (DWI) from a low angular resolution acquisition. Our model, which we call QID^2, takes as input a set of low angular resolution DWI data and uses this information to estimate the DWI data associated with a target gradient direction. We leverage a U-Net architecture with cross-attention to preserve the positional information of the reference images, further guiding the target image generation. We train and evaluate QID^2 on single-shell DWI samples curated from the Human Connectome Project (HCP) dataset. Specifically, we sub-sample the HCP gradient directions to produce low angular resolution DWI data and train QID^2 to reconstruct the missing high angular resolution samples. We compare QID^2 with two state-of-the-art GAN models. Our results demonstrate that QID^2 not only achieves higher-quality generated images, but it consistently outperforms state-of-the-art baseline methods in downstream tensor estimation across multiple metrics and in generalizing to downsampling scenario during testing. Taken together, this study highlights the potential of diffusion models, and QID^2 in particular, for q-space up-sampling, thus offering a promising toolkit for clinical and research applications.

Keywords: Diffusion Weighted Imaging · Diffusion Models · Deep Learning · Q-Space Up-sampling · Tensor Reconstruction

1 Introduction

Diffusion weighted imaging (DWI) is a non-invasive technique that capitalizes on the directional diffusivity of water to probe the tissue microstructure of the brain [4]. A typical DWI acquisition applies multiple magnetic gradients, with the field strength controlled by the b-value and the gradient directions given by the b-vectors. Mathematically, these gradients can be represented by a set of coordinates on the sphere, where the magnitude and direction of each coordinate is related to the corresponding b-value and b-vector, respectively. The domain of all such coordinates is called the q-space [28]. In general, a denser sampling of directions in the q-space, also known as the angular resolution, leads to higher quality DWI. For example, higher angular resolution acquisitions can improve

* Equal Contribution

the tensor estimation [12] and facilitates the progression from single-tensor models [1] to constrained spherical deconvolution models [23] that estimate a fiber orientation distribution function (fODF), which captures more complex fiber configurations. However, increasing the angular resolution also prolongs the acquisition time, which can be impractical in clinical settings. Not only are longer acquisitions more expensive, but they are also difficult for some patients to tolerate, which in turn increases the risk of artifacts due to subject motion [10]. Given these challenges, it is necessary to explore computational methods that can achieve high-quality DWI with a minimal number of initial scan directions.

Several studies have applied generative deep learning to DWI data. For example, the work of [29] uses a spherical U-Net to directly estimate the ODF using DWI acquired with only 60 gradient directions. More recently, generative adversarial networks (GANs) have also been used to estimate DWI volumes. Specifically, the work of [14] generates DWI for a user-specified gradient direction based on a combination of T1 and T2 images. Similarly, the authors of [22] use the Pix2Pix model introduced by [7] to synthesize DWI with 6 gradient directions from data originally captured with only 3 gradient directions [22]. Further variants of the GAN model, such as CycleGAN and DC^{2}Anet, have been applied to simulate a high b-value image from a low b-value one [15]. Beyond GANs, autoencoders have also been used to adjust the apparent b-value [8]. While these works are seminal contributions to the field, none of them consider the clinically relevant problem of up-sampling a low angular resolution DWI acquisition.

Diffusion models have emerged as powerful tool for image generation. At a high level, they work by successively adding Gaussian noise to the input and then learning to reverse this noising process [6]. Diffusion models have been employed in several medical imaging tasks, including image translation between modalities [11], super-resolution and artifact removal [26], registration [9], and segmentation [13]. We will leverage diffusion models to up-sample the DWI gradient directions, which to our knowledge, has not been explored in prior work.

In this paper, we propose an image-conditioned diffusion model, which we call QID^2, that can estimate high angular resolution DWI data from a low angular resolution acquisition[1]. One highlight is that QID^2 automatically identifies several closest available gradient directions and uses the corresponding images as prior knowledge for generating images from any target direction not included in the initial scan. This target image generation process, carried out using a U-Net based structure conditioned on this prior information, can be seen as an extrapolation based on the identified directions and images. By focusing on the most relevant data, QID^2 solicits more targeted prior information and is more computationally efficient. We train and evaluate QID^2 on DWI curated from the Human Connectome Project (HCP) dataset [24]. Our model demonstrates superior performance over GAN-based approaches, particularly when the available low angular resolution images are sparsely distributed across the sphere.

[1] Source code for our model is available online at https://github.com/jueqiw/Diffusion-Model-for-Up-sampling-Diffusion-Weighted-Imaging

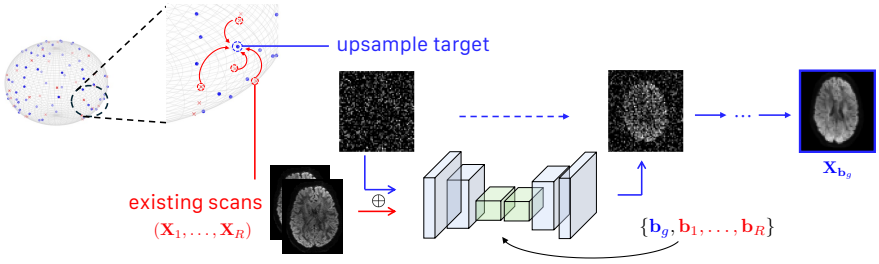

Fig. 1. QID^2 framework for up-sampling the angular resolution of DWI. The gray sphere represents the q-space. Red marks are the directions in the low angular resolution scan, and blue marks are the target gradient directions for image generation.

2 Methods

Fig. 1 provides an overview of our QID^2 framework. For any user-specified target gradient $\mathbf{b}_g$, our model will find and take as input the R closest reference b-vectors $\bar{\mathbf{b}} = (\mathbf{b}_1, \ldots, \mathbf{b}_R)$ available in the low angular resolution scan and the corresponding DWI slices $\bar{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_R)$. QID^2 will then output the estimated target image $\mathbf{X}_{\mathbf{b}_g}$. We can obtain a high angular resolution DWI by sweeping the target gradient directions across the sphere and aggregating the generated images with the original low angular resolution scan.

2.1 A Diffusion Model for Q-space Up-sampling of DWI

Inspired by recently-introduced image-conditioned Denoising Diffusion Probabilistic Models (DDPMs) [25], we design a position-aware diffusion model that leverages "neighboring" DWI data to estimate the image associated with a target gradient direction. Similar to traditional diffusion models [6], QID^2 is comprised of both a forward noising process and a reverse denoising process.

In the forward process, Gaussian noises are added successively at each time step $t \in \{0, 1, \ldots, T\}$ to the generated image. This corruption process is

$$q\left(\mathbf{X}_{\mathbf{b}_g}^{(t)} \mid \mathbf{X}_{\mathbf{b}_g}^{(t-1)}\right) = \mathcal{N}\left(\mathbf{X}_{\mathbf{b}_g}^{(t)}; \sqrt{1-\beta_t}\mathbf{X}_{\mathbf{b}_g}^{(t-1)}, \beta_t \mathbf{I}\right), \quad t \geq 1, \tag{1}$$

where $\{\beta_t\}$ are the forward process variances, and $\mathbf{X}_{\mathbf{b}_g}^{(t)}$ is the noisy image at time t. By repeatedly applying Eq. (1) to the starting image $\mathbf{X}_{\mathbf{b}_g}^{(0)}$, we have

$$q\left(\mathbf{X}_{\mathbf{b}_g}^{(t)} \mid \mathbf{X}_{\mathbf{b}_g}^{(0)}\right) = \mathcal{N}\left(\mathbf{X}_{\mathbf{b}_g}^{(t)} \sqrt{\overline{\alpha}_t}\mathbf{X}_{\mathbf{b}_g}^{(0)}, (1-\overline{\alpha}_t)\mathbf{I}\right) \tag{2}$$

where $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. Therefore, at step t, the generated image $\mathbf{X}_{\mathbf{b}_g}^{(t)}$ can be represented as a function of the initialization $\mathbf{X}_{\mathbf{b}_g}^{(0)}$:

$$\mathbf{X}_{\mathbf{b}_g}^{(t)} = \sqrt{\overline{\alpha}_t}\,\mathbf{X}_{\mathbf{b}_g}^{(0)} + \sqrt{1-\overline{\alpha}_t}\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \tag{3}$$

While the forward noising process operates solely on $\mathbf{X}_{\mathbf{b}_g}^{(0)}$, the reference DWI slices $\{\mathbf{X}_1, \ldots, \mathbf{X}_R\}$ will be used to guide the subsequent denoising process. Rather than constructing a separate network to encode the reference images, which greatly increases the number of parameters and may introduce information loss, we opt to simply concatenate these slices with the target image being generated (i.e., denoised) at each time t as $\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t)} = \mathbf{X}_{\mathbf{b}_g}^{(t)} \bigoplus_{i=1}^{R} \mathbf{X}_{\mathbf{b}_i}$.

Starting from the fully corrupted image $\bar{\mathbf{X}}_{\mathbf{b}_g}^{(T)}$, the reverse process aims to gradually recover the original image $\bar{\mathbf{X}}_{\mathbf{b}_g}^{(0)}$. We denote this process as $p_\theta(\cdot)$, where θ denotes the learnable parameters of the underlying neural network. By restricting the denoising to be Gaussian, the process $p_\theta(\cdot)$ can be written:

$$p_\theta\left(\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t-1)} \mid \bar{\mathbf{X}}_{\mathbf{b}_g}^{(t)}; \{\mathbf{b}_g, \bar{\mathbf{b}}\}\right) = \mathcal{N}\left(\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t-1)}; \mu_\theta\left(\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t)}, \{\mathbf{b}_g, \bar{\mathbf{b}}\}\right), \sigma_t^2 \mathbf{I}\right), \tag{4}$$

where the variances σ_t^2 are hyperparameters of the model. We note that the denoising process relies on the references DWI data $\{\mathbf{X}_1, \ldots, \mathbf{X}_R\}$ and the corresponding gradient directions $\bar{\mathbf{b}} = \{\mathbf{b}_1, \ldots, \mathbf{b}_R\}$, and the target direction $\mathbf{b}_g$. This combination of inputs allows QID^2 to be position-aware.

To reverse the forward noising process, we train QID^2 by minimizing the KL-divergence between $p_\theta(\cdot)$ and $q(\cdot)$ at each time step t. As shown in [6] this loss minimization is equivalent to matching the mean functions, i.e.,

$$\mathcal{L} = \mathbb{E}_{t,q}\left[\left\| \frac{1}{\sqrt{\alpha_t}}\left(\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t)} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right) - \mu_\theta\left(\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t)}, \{\mathbf{b}_g, \bar{\mathbf{b}}\}\right)\right\|^2\right]. \tag{5}$$

The mean function $\mu_\theta(\cdot)$ is generated with a U-Net architecture [17] with the cross-attention mechanism based on the concatenated gradient vectors $\mathbf{b} = \begin{bmatrix}\mathbf{b}_g\ \mathbf{b}_1 \cdots \mathbf{b}_R\end{bmatrix}$. Specifically, the encoding block is computed as follows:

$$\mathbf{H}_1 = \mathrm{FF}(\mathbf{H}_0) + \mathbf{H}_0, \quad \mathbf{H}_2 = \mathrm{Attn}(\mathbf{H}_1, \mathbf{b}) + \mathbf{H}_1,$$

where $\mathrm{FF}(\cdot)$ denotes a feed-forward network, $\mathbf{H}_0$ denotes the block input, and

$$\mathrm{Attn}(\mathbf{H}_1, \mathbf{b}) = \mathrm{Softmax}\left(\frac{(W_Q\mathbf{H}_1)(W_K\mathbf{b})^\top}{\sqrt{d_k}}\right) W_V\mathbf{b},$$

with W_Q, W_K, W_V being the learned weights and d_k being the dimension of $\mathbf{b}$. The decoding block follows a similar expression but includes skip connections from the corresponding encoding block. This design ensures that image features are effectively attended to and integrated with positional information.

Once QID^2 is trained, we can generate DWI for arbitrary gradient directions by sampling from the standard normal distribution and applying the reverse process in Eq. (4) recursively with the corresponding reference images, namely:

$$\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t-1)} = \mu_\theta\left(\bar{\mathbf{X}}_{\mathbf{b}_g}^{(t)}, \{\mathbf{b}_g, \bar{\mathbf{b}}\}\right) + \sigma_t\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \tag{6}$$

2.2 Baseline Comparison Methods

We compare QID^2 with two state-of-the-art GAN models. The first model is a conditional GAN (**cGAN**) for image generation proposed by [7]. We use the same cross-attention U-Net architecture for the generator as used in QID^2. We use a PatchGAN discriminator [7] and inject the gradient direction information $\{\mathbf{b}_g, \bar{\mathbf{b}}\}$ into the discriminator with cross-attention mechanism. We train the generator to minimize the GAN objective plus a regularization term that encourages voxel-level similarity of the generated and ground-truth images:

$$G^* = \arg\min_G \max_D \lambda_G \left[\mathbb{E}_{\mathbf{X}} \left[\log D(\mathbf{X}, \mathbf{b})\right] + \mathbb{E}_{\widetilde{\mathbf{X}}} \left[\log(1 - D(\widetilde{\mathbf{X}}, \mathbf{b}))\right]\right] + \lambda_V \mathcal{L}_1(G)$$

where $\mathbf{X} = [\mathbf{X}_{\mathbf{b}_g}, \mathbf{X}_1, \ldots, \mathbf{X}_R]$ is the concatenated real sample with $\mathbf{X}_{\mathbf{b}_g}$ drawn from the (high resolution) training data and $\widetilde{\mathbf{X}} = [G(\mathbf{X}_{1:R}, \mathbf{b}), \mathbf{X}_1, \ldots, \mathbf{X}_R]$ represents the synthesized data of generated DWI and real reference slices. Finally, λ_G and λ_V balance the adversarial and similarity L_1 losses, respectively.

The second model is the Q-space conditional GAN (**qGAN**) proposed by [14]. Unlike QID^2 and the cGAN baseline, qGAN incorporates the gradient directions and reference DWI data using a feature-wise linear modulation scheme. The qGAN discriminator is also a conditional U-Net and combines the gradient directions and reference DWI data via an inner product. Although the inputs to the original qGAN model [14] are a single structural image (e.g., B0, T1, T2) and a user-defined target gradient, we provide the same set of closest directions and corresponding images as input to ensure a fair comparison with QID^2.

Finally, as a sanity check, we compare the deep learning models to a simple interpolation scheme (**Interp**), in which we express the target gradient direction as a linear combination of the reference gradients and then use the linear coefficients to interpolate between the reference DWI slices to obtain the target.

2.3 Implementation Details

For QID^2, we use a linear noise schedule of 1000 time steps. The QID^2 U-Net employs $[128, 128, 256]$ channels across three levels with one residual block per level. We use the Adam optimizer with a learning rate of 2.5×10^{-5}, $\beta_1 = 0.5$ and $\beta_2 = 0.999$. These hyperparameters are selected based on a relevant study [16] and not fine-tuned. We use the same U-Net architecture for the cGAN generator with the same set of hyperparameters. We fix $\lambda_G = 1$ and $\lambda_V = 100$ in the loss function weights for both GAN methods. The discriminator is updated once for every two updates of the generator during training [14]. For both qGAN and cGAN, we use a learning rate of 5×10^{-5} with the Adam optimizer. We evaluate all models with both $R = 3$ and $R = 6$ reference DWI data. To avoid memory issues, we train the deep learning models to generate 2D axial slices, which we stack into 3D DWI volumes. Each 2D image has a size of $(145, 174)$. During training, we independently normalize each slice from its original intensity to a range of $[0, 1]$. Data augmentation is employed to enhance model training. Specifically, we use rotations by random angles in $[-15°, 15°]$ and random spatial

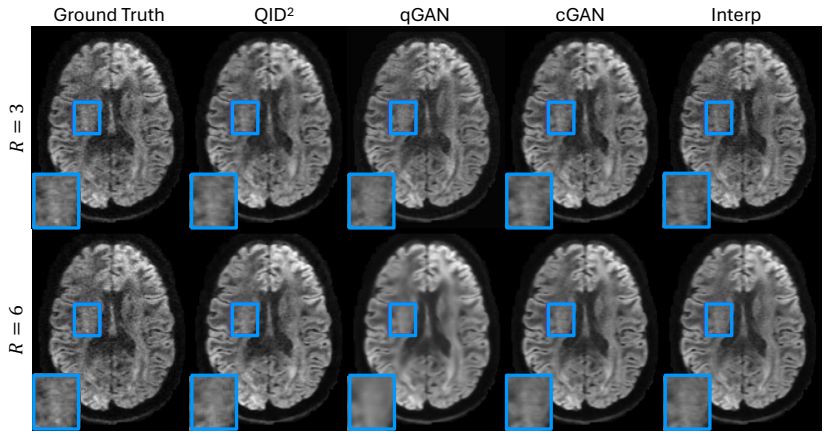

Fig. 2. Qualitative results that compare the ground-truth DWI acquisition to images generated by QID^2 and the baselines methods for $R = 3$ and $R = 6$. Zoomed-in area highlights details that are preserved by our method and do not appear in the baselines.

scaling factors in $[0.9, 1.1]$. The final output is rescaled voxel-wise back to the original intensity and masked by the subject $\mathbf{X}_{\mathbf{b}_g}$ image.

3 Experimental Results

Dataset Curation and Preprocessing: We curate a total of 720 subjects from the HCP S1200 release [24]. The remaining HCP subjects are excluded due to an inconsistent number of gradient directions at $b = 1000$ s/mm^2. The DWI is acquired on a Siemens 3T Connectome scanner at 3 shells ($b = 1000, 2000$ and 3000 s/mm^2). Each shell has exactly 90 gradient directions sampled uniformly on the sphere. The voxel size is $1.25 \times 1.25 \times 1.25$ mm^3. The data is preprocessed with distortion/motion removal and registration to the 1.25 mm structural space.

Clinical diffusion imaging typically uses lower b-values with approximately 30 gradient directions [4]. To better accommodate this situation, we focus our evaluation on the $b = 1000$ s/mm^2 shell. From here, we construct low angular resolution DWI data by subsampling the 90 gradient directions to 30 evenly spaced ones that preserve the uniformity of the sphere [2]. The data for the remaining 60 directions serve as the targets for model training and evaluation.

Each volume is broken down into 145 axial slices. The deep learning models are trained to predict the image slices for each target direction. Based on this scheme, we create 60 samples for each slice. Each sample consists of one 2D slice for the target gradient direction and R reference slices corresponding to the closest low resolution gradient directions. The distance between gradients is defined by the geodesic distance on the sphere: $d(\mathbf{b}_1, \mathbf{b}_2) = \arccos(\mathbf{b}_1 \mathbf{b}_2^\top)$ [3].

Finally, we use 576 HCP subjects for training, 72 for validation, and 30 for testing. The original DWI scans are treated as the gold standard for evaluation.

Table 1. Quantitative evaluation of the generated image quality (left) and FA estimation quality (right) of QID^2, the GAN models, and interpolation for different R. The best performance of each metric is highlighted in bold.

Methods	Image FID ↓	Image SSIM ↑	FA Error ↓	FA Map SSIM ↑
QID^2 (R=3)	14.07	0.895 ± 0.045	**0.027 ± 0.003**	**0.866 ± 0.043**
qGAN(R=3)	24.85	0.893 ± 0.046	0.037 ± 0.002	0.792 ± 0.052
cGAN(R=3)	29.93	0.913 ± 0.039	0.099 ± 0.014	0.643 ± 0.159
Interp(R=3)	**8.96**	**0.917 ± 0.038**	0.057 ± 0.026	0.750 ± 0.110
QID^2 (R=6)	**16.29**	0.900 ± 0.045	**0.027 ± 0.003**	**0.863 ± 0.042**
qGAN(R=6)	71.44	0.905 ± 0.042	0.040 ± 0.004	0.801 ± 0.045
cGAN(R=6)	36.33	0.915 ± 0.037	0.031 ± 0.004	0.851 ± 0.044
Interp(R=6)	21.46	**0.933 ± 0.044**	0.038 ± 0.006	0.815 ± 0.052

Comparing Reconstructed Image Quality: Fig. 2 presents qualitative results that compare the ground-truth DWIs to those generated by QID^2 and the baseline methods. The GAN models and Interp fail to preserve high-frequency details in the synthesized DWI data, while QID^2 succeeds in capturing the finer details more accurately, as highlighted in the zoomed-in blue boxes.

Table 1 (left) reports the Fréchet inception distance (FID) [5] and the structural similarity index measure (SSIM) [27] of the synthesized DWI data. Specifically, FID measures the realism and diversity of images by comparing the feature distributions between the generated and ground truth ones, while SSIM quantifies the similarity based on luminance, contrast, and structural information. We observe that QID^2 achieves nearly a two-fold improvement (i.e., decrease) in FID than the GAN models for both $R = 3$ and $R = 6$, which indicates that the DWI data generated by diffusion possess higher quality and greater diversity. Although the GAN models achieves a slightly higher SSIM than QID^2, the difference is not statistically significant using a two-sample (paired) t-test. Interestingly, the simple interpolation technique achieves better FID than QID^2 when $R = 3$. This is likely because the interpolation tracks the closest reference image, which is more akin to the original DWI distribution. However, the improved FID does not generalize to better tensor estimation, as seen in the next section.

Impact on Tensor Estimation: We estimate the fractional anisotropy (FA) using the standard single-tensor model [1]. Fig. 3 shows the fiber direction and FA value maps among the ground-truth, QID^2 and baseline methods for $R = 3$ and $R = 6$. Similarly to the finding in the reconstructed image, we observe that the qGAN and cGAN methods capture the general FA trends but fail to capture the high frequency features. Conversely, the diffusion-generated image by QID^2 more closely resembles the ground-truth data by capturing finer details more accurately. This shows that the visual differences in the reconstructed images in Fig. 2 are important when estimating tensors. The Interp method fails to generate realistic FA maps for $R = 3$. Empirically, we also observe quality issues with Interp for $R = 6$ even though they are less evident in the figure.

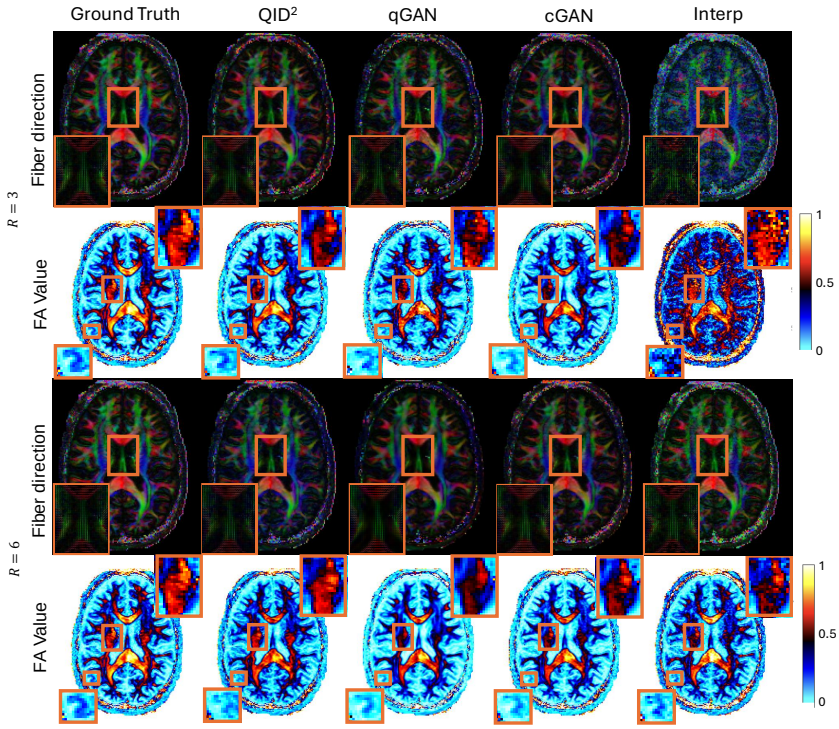

Fig. 3. Qualitative comparison between the ground truth and estimated images. **Row 1/3:** Colored fiber orientation maps with minimal visually detectable differences among the images. Zoomed-out regions show the estimated tensors in the orange box area. **Row 2/4:** FA value maps, where brighter colors indicate higher FA values. Significant differences compared to the ground truth are zoomed-out with orange boxes.

Table 1 (right) reports the mean absolute error and SSIM, as compared to the FA computed from the ground-truth high angular resolution DWI. As seen, QID^2 consistently outperforms the GAN-based model and the Interp method for both $R = 3$ and $R = 6$. Specifically for $R = 3$, the error in FA is roughly three times lower for QID^2 than for the GANs. QID^2 also achieves significantly higher SSIM values. These trends persist when the number of reference images increases to $R = 6$, i.e., even when more prior information is provided. However, the relative performance gain over the GANs shrink. Additionally, although the image-based metrics are better for the interpolation-generated (Interp) images, QID^2 outperforms this baseline by a large margin when estimating FA. Taken together, these results suggest that QID^2 is particularly effective in scenarios where the images are scarce and distributed sparsely, i.e., smaller values of R.

Fewer Available Initial Directions: As a final evaluation, we apply the models trained when assuming 30 initial gradient directions to testing data for which fewer gradient directions are available. Table 2 compares the performance

Table 2. Quantitative evaluation of the generated image quality (left) and FA estimation quality (right) of each model when only 20 and 10 initial gradient directions are available at test time. The best performance of each metric is highlighted in bold.

	Methods	Image FID ↓	Image SSIM ↑	FA Error ↓	FA Map SSIM ↑
20 initial directions	QID^2(R=3)	15.28	0.883±0.048	**0.029±0.003**	**0.851±0.048**
	qGAN(R=3)	24.75	0.886±0.04	0.056±0.003	0.624±0.067
	cGAN(R=3)	29.75	0.904±0.042	0.102±0.014	0.622±0.153
	Interp(R=3)	**9.91**	**0.934±0.042**	0.083±0.033	0.676±0.123
	QID^2(R=6)	**16.91**	0.887±0.049	**0.030±0.003**	**0.843±0.048**
	qGAN(R=6)	71.94	0.896±0.045	0.100±0.025	0.580±0.086
	cGAN(R=6)	36.31	0.906±0.041	0.073±0.011	0.591±0.096
	Interp(R=6)	19.86	**0.911±0.041**	0.050±0.011	0.757±0.074
10 initial directions	QID^2(R=3)	14.68	0.882±0.054	**0.037±0.004**	**0.798±0.069**
	qGAN(R=3)	23.74	0.881±0.054	0.060±0.004	0.591±0.078
	cGAN(R=3)	30.27	0.899±0.048	0.111±0.019	0.552±0.143
	Interp(R=3)	**10.65**	**0.914±0.051**	0.112±0.033	0.613±0.127
	QID^2(R=6)	17.48	0.883±0.056	**0.038±0.004**	**0.783±0.072**
	qGAN(R=6)	72.55	0.890±0.052	0.120±0.030	0.521±0.113
	cGAN(R=6)	36.45	**0.898±0.048**	0.083±0.014	0.507±0.120
	Interp(R=6)	**15.73**	0.880±0.057	0.091±0.023	0.643±0.110

of QID^2 and the baseline methods for 20 (top) and 10 (bottom) uniformly-distributed gradient directions at test time. Intuitively, we observe that the performance of all models worsens as the number of initial gradient directions decreases. However, QID^2 remains relatively stable from 20 to 10 directions, while baseline models experience a sharper drop. Similar to Table 1, QID^2 achieves comparable (but not the best) performance in reconstructed image quality, but it leads by a large margin with respect to tensor estimation. This result further highlights the benefits of QID^2, as even though QID^2 is trained on a dense gradient distribution, it can generalize effectively when applied to a sparser one.

4 Discussion and Future Work

Our proposed image-conditioned diffusion model, named QID^2, is designed to upsample a low angular resolution DWI acquisition to have a higher angular resolution. One key innovation is our use of the automatically identified neighboring gradient directions in the low angular resolution DWI as prior information to improve the generation of images associated with new directions. We also propose an efficient way to encode this prior information (images and corresponding gradient directions) that ensures the effective attention and integration between the two. Our real-world experiments demonstrate that QID^2 outperforms two baseline GAN models in both image quality and tensor estimation.

However, our study is not without limitations. A notable one is that the models are trained and validated on a single dataset, namely HCP. While the HCP dataset provides a large number of subjects and a state-of-the-art acquisition protocol that can be used to create both low and high angular resolution DWI for model training, it may not fully capture the types of imaging acquisitions used in a clinical setting. To address this issue, future work should train and test QID^2 on additional datasets with different scanning protocols and patient characteristics (e.g., brain lesions). An example that can be used in future work is the CDMRI Quantitative Connectivity (QuantConn) challenge dataset [20].

A second limitation of our study is the focus on a single-shell reconstruction, and correspondingly, a single tensor estimation. While the use of $b = 1000\,\mathrm{s/mm^2}$ images in this work aligns with common clinical practices [4], it only enables us to fit a basic tensor model and is insufficient for more sophisticated fODF estimation and tractography analysis. Though it is not standard in many clinical workflows due to the difficulties of acquisition, tractography is still a useful tool to optimize surgical planning and postoperative assessment for tumors and vascular malformations [4]. Our current QID^2 can be adapted to multi-shell image upsampling, but it would require training a separate model for each shell. Future work will incorporate the b-value as an auxiliary input to QID^2 to enhance efficiency and streamline the upsampling process across multiple shells.

Despite these limitations, we believe that QID^2 has promise in clinical applications. Here, clinical deployment would follow a two-step procedure. First, we can pretrain QID^2 on a large collection of publicly available datasets, from HCP to the QuantConn challenge dataset to UK Biobank. Second, individual sites can opt to fine-tune this base model using their in-house datasets in order to match the specific MRI scanning protocols and patient conditions under evaluation. On the user (clinician) end, a typical workflow using the deployed model includes performing a standard low angular DWI scan, specifying target directions (not necessarily uniform), followed by processing the data with QID^2 to generate high-angular images, and combining the outputs for downstream analysis. This approach will help reduce scan times and mitigate patient discomfort, which are both significant challenges in the clinical use of DWI [21].

5 Conclusion

We introduce an image-conditioned diffusion model (QID^2) that can generate high angular resolution DWI from low angular resolution data, effectively estimating high-quality imaging with limited initial scan directions. Our approach takes advantage of similar DWI data as prior information to predict the data for any user-specified gradient direction. The results demonstrate that diffusion-generated DWIs by QID^2 achieve superior quality and significantly outperform those generated by baseline models in downstream tensor modeling tasks. QID^2 only shows a slight performance drop when applying to a sparser initial directions distribution during testing, demonstrating its superior generalizability. Although our method currently exhibits longer training times due to the denois-

ing characteristics of DDPMs, this limitation could be mitigated by employing more efficient sampling techniques [19] and one-shot training [18] in future work.

Acknowledgements. This work was supported by the National Institutes of Health R01 HD108790 (PI Venkataraman), the National Institutes of Health R01 EB029977 (PI Caffo), the National Institutes of Health R21 CA263804 (PI Venkataraman).

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophysical journal **66**(1), 259–267 (1994)
2. Cheng, J., Shen, D., Yap, P.T., Basser, P.J.: Single-and multiple-shell uniform sampling schemes for diffusion MRI using spherical codes. IEEE transactions on medical imaging **37**(1), 185–199 (2017)
3. Chung, M.K., Chen, Z.: Embedding of functional human brain networks on a sphere. arXiv preprint arXiv:2204.03653 (2022)
4. Doshi, A., Gerke, L., Marchione, J., Bou-Haidar, P., Delman, B.: Physiologic evaluation of the brain with magnetic resonance imaging. Youmans and Winn Neurological Surgery. 7th ed. New York: Elsevier pp. 69–95 (2017)
5. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
8. Jha, R.R., Jaswal, G., Bhavsar, A., Nigam, A.: Single-shell to multi-shell dMRI transformation using spatial and volumetric multilevel hierarchical reconstruction framework. Magnetic Resonance Imaging **87**, 133–156 (2022)
9. Kim, B., Han, I., Ye, J.C.: Diffusemorph: Unsupervised deformable image registration using diffusion model. In: European conference on computer vision. pp. 347–364. Springer (2022)
10. Koh, D.M., Collins, D.J.: Diffusion-weighted MRI in the body: applications and challenges in oncology. American Journal of Roentgenology **188**(6), 1622–1635 (2007)
11. Li, Y., Shao, H.C., Liang, X., Chen, L., Li, R., Jiang, S., Wang, J., Zhang, Y.: Zero-shot medical image translation via frequency-guided diffusion models. IEEE transactions on medical imaging (2023)
12. Michailovich, O., Rathi, Y.: Fast and accurate reconstruction of HARDI data using compressed sensing. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 607–614. Springer (2010)
13. Rahman, A., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M.: Ambiguous medical image segmentation using diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11536–11546 (2023)
14. Ren, M., Kim, H., Dey, N., Gerig, G.: Q-space conditioned translation networks for directional synthesis of diffusion weighted images from multi-modal structural MRI. In: Medical Image Computing and Computer Assisted Intervention–MICCAI

2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 530–540. Springer (2021)
15. Rezaeijo, S.M., Entezari Zarch, H., Mojtahedi, H., Chegeni, N., Danyaei, A.: Feasibility study of synthetic DW-MR images with different b values compared with real DW-MR images: quantitative assessment of three models based-deep learning including CycleGAN, Pix2PiX, and DC2Anet. Applied Magnetic Resonance **53**(10), 1407–1429 (2022)
16. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
18. Shaham, T.R., Dekel, T., Michaeli, T.: Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4570–4580 (2019)
19. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
20. Strike, L.T., Blokland, G.A., Hansell, N.K., Martin, N.G., Toga, A.W., Thompson, P.M., de Zubicaray, G.I., McMahon, K.L., Wright, M.J.: "queensland twin imaging (qtim)" (2023). https://doi.org/doi:10.18112/openneuro.ds004169.v1.0.7
21. Tae, W.S., Ham, B.J., Pyun, S.B., Kang, S.H., Kim, B.J.: Current clinical applications of diffusion-tensor imaging in neurological disorders. Journal of Clinical Neurology **14**(2), 129–140 (2018)
22. Tatekawa, H., Ueda, D., Takita, H., Matsumoto, T., Walston, S.L., Mitsuyama, Y., Horiuchi, D., Matsushita, S., Oura, T., Tomita, Y., et al.: Deep learning-based diffusion tensor image generation model: a proof-of-concept study. Scientific Reports **14**(1), 2911 (2024)
23. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage **35**(4), 1459–1472 (2007)
24. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The WU-Minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)
25. Waibel, D.J., Röell, E., Rieck, B., Giryes, R., Marr, C.: A diffusion model predicts 3D shapes from 2D microscopy images. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)
26. Wang, J., Levman, J., Pinaya, W.H.L., Tudosiu, P.D., Cardoso, M.J., Marinescu, R.: Inversesr: 3d brain MRI super-resolution using a latent diffusion model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 438–447. Springer (2023)
27. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
28. Yeh, F.C., Irimia, A., de Almeida Bastos, D.C., Golby, A.J.: Tractography methods and findings in brain tumors and traumatic brain injury. Neuroimage **245**, 118651 (2021)
29. Zhao, H., Deng, C., Wang, Y., Ma, J.: Better fibre orientation estimation with single-shell diffusion MRI using spherical U-Net. In: International Conference of Pioneering Computer Scientists, Engineers and Educators. pp. 3–12. Springer (2023)