MiCo: Microstructure-Consistent Flow Matching for Diffusion MRI Angular Super-Resolution

Zhixuan Zhou **Tingting Dan** **Guorong Wu***
Departments of Computer Science and Psychiatry
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
zzhixuan@cs.unc.edu
{Tingting_Dan,grwu}@med.unc.edu

Abstract

Diffusion-weighted MRI (dMRI) requires dense sampling of q-space across many directions and shells, leading to a long scanning time that limits clinical use. Angular super-resolution (ASR) recovers a dense signal from a sparse acquisition, offering significant savings in scan time. In this context, we propose a physics-informed flow matching framework for ASR built on the medical physics principle about dMRI: under the cumulant expansion, the diffusion signal naturally splits into a known linear function of the q-vectors plus a residual carrying higher-order angular detail. We capitalize on this insight on two fronts. *First*, we initialize the flow from a closed-form dense DTI estimate, recasting ASR as residual flow matching from a physically meaningful starting point rather than uninformative noise. *Second*, we impose a microstructure-consistency loss, rooted in the physical principle that any two diffusion-weighted signals from the same voxel must be explained by a single underlying set of microstructural parameters. Together, these physics-informed designs anchor predictions wherever the acquisition is informative while leaving the learned prior to model crossing-fiber and higher-order angular detail. On the Human Connectome Project (HCP-YA) and UK Biobank, MiCo combines competitive signal fidelity with improved downstream microstructural recovery. On HCP-YA, it achieves the lowest FA and kurtosis-scalar errors among the evaluated methods and leads FOD accuracy on all three shells at 10% context, demonstrating the value of aligning reconstruction with microstructural fidelity.

1 Introduction

Diffusion-weighted MRI (dMRI) provides a unique non-invasive window into tissue microstructure and the macro-anatomy of neural pathways, and has become a standard tool for studying brain organization in both health and disease [19, 20]. Microstructural metrics derived from dMRI inform the study of neurodegenerative conditions such as Alzheimer's disease [11], stroke recovery [4], and the broader structural connectome [15, 36]. Realizing this value, however, requires sufficient angular resolution in q-space: faithful estimation of tensor models, kurtosis models, and fiber orientation distributions all depend on dense diffusion-direction sampling [40]. Acquisition time scales linearly with the number of directions, and clinical and large-cohort protocols are forced to a hard trade-off between scan duration and angular fidelity. Angular super-resolution has emerged as a principled way to address this trade-off by predicting signals along unmeasured directions from a sparse acquisition [27, 28, 31].

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Generative learners for dMRI reconstruction. Recent advances in conditional generative modeling, particularly diffusion and flow matching [16, 22, 24, 35], have opened a new paradigm for this task. Unlike classical model-based reconstruction methods (spherical harmonics [9], SHORE [32], MAP-MRI [33]), which fit a parametric basis per voxel and are brittle under sparse, noisy measurements, generative learners can leverage population-scale datasets such as HCP [41] and UK Biobank [30] to learn a strong implicit prior over the dMRI signal manifold. Recent models also accommodate flexible sampling schemes and have shown clear gains over both classical and earlier discrete deep-learning baselines on signal-level reconstruction [10, 14]. Together, generative models shed new light on angular super-resolution models that scale with data and generalize across acquisition protocols.

Signal fidelity is not microstructure fidelity. Yet a more fundamental challenge persists: signal-space fidelity does not necessarily imply microstructure fidelity. Predicted signals are passed into downstream tensor and orientation models such as DTI [2], DKI [18], and CSD [39], where small but structured errors might translate into corrupted parameter maps and fiber orientation estimates. Standard flow matching supervises all signal dimensions uniformly, failing to prevent solutions that minimize signal MSE while drifting in the downstream parameter space.

Our approach. We address both challenges with ***MiCo***, a flow matching framework grounded in a basic principle of inverse problems: *any two diffusion measurements at the same voxel should be consistent with the same underlying microstructure*. Standard flow matching, applied uniformly across all signal dimensions, treats this constraint as something to hope for, not enforce. We make it explicit. The DKI forward model is linear, which makes both halves of the resulting design closed-form. First, we replace the standard Gaussian-noise prior of flow matching with a closed-form physics-guided initialization: a diffusion tensor fit to the sparse observations gives a dense estimate on the target directions, and the network only models the residual gap to the ground-truth signal. The flow is no longer asked to recover signal content already determined by the acquisition. Second, parameter-level consistency, the requirement that predictions and ground-truth signals at the same voxel imply the same DKI parameters, becomes an exact, closed-form quadratic loss in signal space. The loss identity is exact for the stated linear DKI model; the cumulant truncation remains a physical approximation.

Contributions. We present ***MiCo***, a physics-informed flow matching framework for dMRI angular super-resolution with two design choices both capitalizing on the linearity of the DKI forward model. We propose to replace the standard Gaussian-noise prior with a closed-form DTI-based dense initialization, so that the network only models the residual gap that the forward model leaves underdetermined. We further introduce a microstructure-consistency loss that enforces parameter-level agreement between predicted and ground-truth signals, admitting an exact closed-form signal-space implementation with no per-step DKI fit. Empirically, ***MiCo*** improves FA (Fractional Anisotropy), MD (Mean Diffusivity), and MK (Mean Kurtosis) recovery relative to DISCUS. Ablations identify the signal gains from initialization and the parameter gains from consistency, while residual-stratified analysis connects the consistency benefit to DKI model fit.

2 Related Works

Classical model-based dMRI reconstruction. Continuous q-space reconstruction has historically relied on per-voxel basis fitting: spherical harmonics on a single shell [9], radial-angular bases such as SHORE [32] and MAP-MRI [33] for full 3D q-space, and compressed-sensing approaches for grid-based acquisitions [38]. These methods are interpretable and training-free but unreliable under aggressive subsampling. Furthermore, these analytic approaches cannot leverage population-level priors that modern dMRI cohorts offer.

Learning-based dMRI angular super-resolution. Early deep learning approaches such as q-Space deep learning (q-DL) [14] regressed signals or downstream microstructure metrics directly from a fixed sparse acquisition using MLPs, demonstrating that learned priors substantially outperform model-based fits under aggressive subsampling. Follow-up works scale this idea to richer architectures and outputs: CNN regressors for NODDI/GFA from subsampled q-space [13], ensemble-average-propagator estimation [42], b-vector-conditioned recurrent CNNs [27], spatio-angular convo-

lutions [28], and graph-CNN approaches treating q-space as a graph [5, 6]. DISCUS [10] achieves both input- and output-continuity via a permutation-invariant encoder–decoder that learns from arbitrary observation sets. More recently, generative models have entered this space: QID^2 [7] adapts an image-conditioned DDPM with cross-attention over neighboring b-vectors, and PGDiT [31] proposes a diffusion transformer with spherical-harmonics-guided posterior sampling at inference. Across these methods, domain knowledge is introduced primarily through architectural inductive biases or test-time correction, rather than by exploiting structural properties of the dMRI forward model during training.

Physics-informed and forward-model-guided generative models in medical physics. Beyond dMRI, a broader trend incorporates physical forward models into generative training. For example, Bastek et al. [3] formalize physics-informed diffusion via PDE-residual training losses. MRF-DiPh [29] couples a diffusion prior with the Bloch model for quantitative MRI. UPMRI [26] projects flow matching onto the range of the parallel-MRI forward operator for unsupervised k-space reconstruction. MiCo specializes range-space supervision to the DKI parameter-to-dense-q-space operator H. Whereas acquisition-operator projections enforce measured k-space consistency, this operator yields a supervised loss exactly equivalent to Fisher-metric DKI parameter discrepancy. The resulting training objective emphasizes microstructure-sensitive signal components while retaining full-space flow matching.

3 Preliminaries

dMRI acquisition. As shown in Fig.1, at each spatial voxel, a dMRI acquisition produces a paired sequence $(\mathbf{q}_i, S(\mathbf{q}_i))_{i=1}^M$ of q-vectors and scalar diffusion-weighted signals. Each q-vector decomposes into a unit direction $\hat{\mathbf{q}} \in \mathbb{S}^2$ and a b-value $b = \|\mathbf{q}\|^2$ controlling diffusion weighting. A *multi-shell* acquisition collects measurements across several distinct b-values, with directions approximately uniformly distributed on each shell. Here, we work with log-normalized signals $\ln S(\mathbf{q})/S_0$, where S_0 is the b=0 baseline.

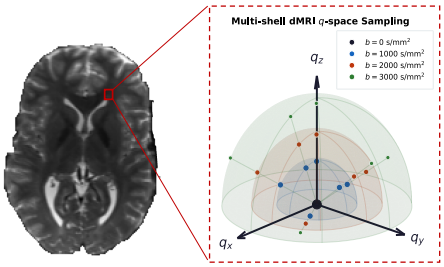

Figure 1: Multi-shell dMRI acquisition. Each voxel (left) yields a sequence of diffusion-weighted measurements at q-space sampling locations (right). Directions are distributed approximately uniformly on each shell of constant b-value (color-coded), with b=0 baseline volumes (black) defining the unweighted reference S_0. Higher b-values probe finer angular structure but require more directions for stable basis fits, motivating the angular super-resolution problem.

DTI, DKI, and the linear forward operator. Under the cumulant expansion of the diffusion propagator, the log-normalized signal is polynomial in b. Truncating at first order gives Diffusion Tensor Imaging (DTI) [2], parameterized by a symmetric 3×3 diffusion tensor with 6 independent components. Retaining the second-order term gives Diffusion Kurtosis Imaging (DKI) [18], which adds a fully symmetric rank-4 kurtosis tensor with 15 components, for $P = 22$ parameters in total. We refer to Appendix A for the full cumulant expansion and explicit parameter definitions. Crucially for our method, both models are *linear in their parameters* once the q-vectors are fixed: stacking log-signals at any set of N query directions into a vector $y \in \mathbb{R}^N$, the cumulant expansion gives

$$y \;=\; H\,x, \qquad H \in \mathbb{R}^{N\times P}, \tag{1}$$

where the design matrix H is determined entirely by the query q-vectors and b-values, and is independent of the measured signals. The parameter vector $x \in \mathbb{R}^P$ collects the linear coefficients in the cumulant expansion: 6 diffusion-tensor components, 15 fully-symmetric kurtosis-tensor terms (rescaled by the constant $\bar{D}^2$ to make H depend only on the acquisition), and a $\log S_0$ baseline absorbed by the linear regression. Within a single voxel, agreement on x is equivalent to agreement on the standard DKI microstructural parameters. The explicit basis construction of H is given in Appendix A.3.

Angular super-resolution. Let $\mathcal{C} = \{(\mathbf{q}_i, S(\mathbf{q}_i))\}_{i=1}^M$ denote a sparse *context set* observed at a voxel and $\mathcal{T} = \{(\mathbf{q}'_j, S(\mathbf{q}'_j))\}_{j=1}^N$ a dense *target set* at unobserved directions, with $M \ll N$. Stacking

log-signals yields $y_{\text{obs}} = H_{\text{obs}}\, x \in \mathbb{R}^M$ and $y_{\text{dense}} = H_{\text{dense}}\, x \in \mathbb{R}^N$. Angular super-resolution learns $p_\theta(y_{\text{dense}} \mid y_{\text{obs}})$. When H_{obs} is well-conditioned, a closed-form linear estimate is available:

$$\hat{x} = H_{\text{obs}}^{\dagger}\, y_{\text{obs}}, \qquad y_{\text{init}} = H_{\text{dense}}\, \hat{x}, \tag{2}$$

but under aggressive subsampling and measurement noise this estimate is incomplete, motivating a learned residual on top of the physics-guided prediction.

Conditional flow matching. Conditional flow matching [22] learns a velocity field $v_\theta(y, t, c)$ that transports an initial sample $y_0 \sim p_0(\cdot \mid c)$ to a target $y_1 \sim p_{\text{data}}(\cdot \mid c)$ along a path $y_t = \phi_t(y_0, y_1)$ with $\phi_0 = y_0$ and $\phi_1 = y_1$. Training regresses the network onto the path's instantaneous velocity:

$$\mathcal{L}_{\text{CFM}} \;=\; \mathbb{E}_{t,\, y_0,\, y_1}\left[\left\| v_\theta(y_t, t, c) - \tfrac{d}{dt}\phi_t(y_0, y_1)\right\|_2^2\right]. \tag{3}$$

A common choice is the linear interpolant $y_t = (1 - t)\, y_0 + t\, y_1$.

4 Method

We develop a physics-informed generative framework for dMRI angular super-resolution rooted in a basic principle of inverse problems: *any two diffusion measurements at the same voxel are consistent with the same underlying microstructure.* Standard flow matching steers predictions uniformly across signal dimensions and leaves this constraint as an implicit necessity condition (with no guarantee that it holds). We instead enforce this constraint in an explicit manner. Since the DKI forward model is linear, the required parameter-level agreement further admits an exact, closed-form projection in signal space, and can be added to flow matching as a training-time loss with no architectural changes and no per-step parameter fit.

Model interface. ***MiCo*** operates voxel-wise, using one q-conditioned Perceiver-IO model per cohort shared across all shells. Context tokens encode observed q-vectors and signals; query tokens encode target q-vectors and the current flow state. The output is the dense clean-signal prediction at those target q-vectors. No spatial neighborhood is used; Appendix C gives the architecture.

As illustrated in Fig. 2, two design choices follow. We replace the standard Gaussian prior of flow matching with a closed-form physics-guided initialization y_{init} from a diffusion-tensor fit, so the flow only models what the forward model leaves underdetermined (Section 4.1). We pair the standard flow matching loss with a parameter-consistency loss that supervises only the range of the DKI forward operator, i.e., the subspace where signal errors translate directly into microstructural-parameter errors (Section 4.2).

4.1 Physics-Guided Flow Matching with DTI Initialization

A standard flow matching pipeline transports samples from an uninformative prior, typically isotropic Gaussian noise, to the data distribution. In our setting, however, we argue this is not necessary: the sparse observation y_{obs} already contains substantial information about the target signal, and forcing the flow to recover this content from noise discards a strong inductive bias. We instead initialize the flow from a physics-guided estimate computed directly from y_{obs}, leaving the network to model only the residual gap between this estimate and the ground-truth signal.

DTI initialization. We pool context measurements across shells, retain all $b = 0$ observations, and fit a diffusion tensor with a baseline coefficient by Tikhonov-regularized least squares:

$$\hat{x}_{\text{DTI}} = (A_{\mathcal{C}}^{\top} A_{\mathcal{C}} + \Lambda_{\mathcal{C}})^{-1} A_{\mathcal{C}}^{\top}\, y_{\text{obs}}, \qquad y_{\text{init}} = A_{\mathcal{T}} \hat{x}_{\text{DTI}}. \tag{4}$$

Here $A_{\mathcal{C}} \in \mathbb{R}^{M \times 7}$ contains six tensor columns and an intercept, and $\Lambda_{\mathcal{C}}$ is diagonal. This smaller model is preferable to a 22-parameter DKI initialization under sparse sampling. The relevant sample count is joint across shells; conditioning diagnostics and the design row are given in Appendix D.3.3.

Flow matching with a physics-guided source. We instantiate the conditional flow matching framework of Section 3 with $y_0 = y_{\text{init}}$ as the source and $y_1 \sim p_{\text{data}}(\cdot \mid c)$ as the target, where c

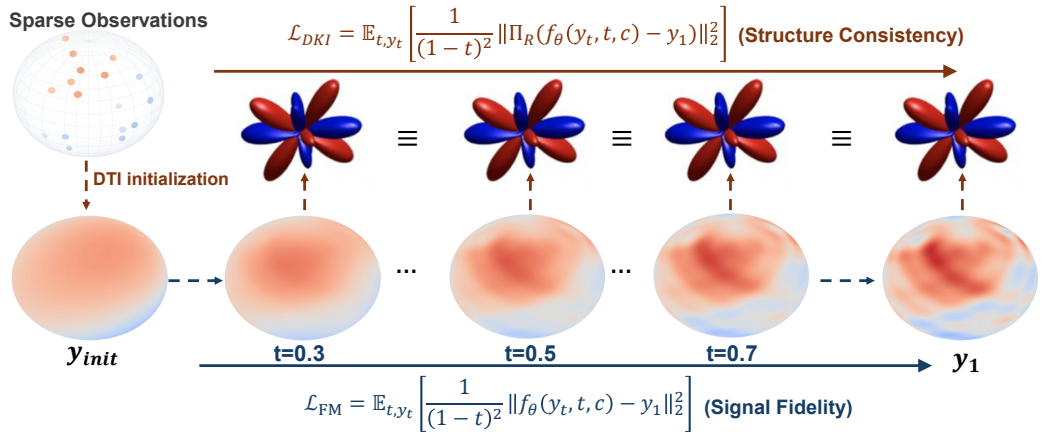

Figure 2: ***MiCo*** enforces DKI parameter consistency along the flow path. Sparse observations are first lifted to a dense initialization y_{init} via a closed-form DTI fit. Flow matching then transports y_{init} to the ground-truth signal y_1. Throughout training, ***MiCo***'s DKI loss constrains the predicted and ground-truth signals to share the same DKI parameters, equivalently, to agree on the range component $\Pi_R(y_1 - \hat{y})$. The orthogonal null component $\Pi_N(y_1 - \hat{y})$ is supervised by the standard flow matching loss alone.

encodes acquisition conditioning. Under the standard linear interpolant, the path is a deterministic trajectory anchored at y_{init}:

$$y_t \;=\; (1-t)\, y_{\text{init}} + t\, y_1, \qquad t \in [0,1]. \tag{5}$$

The associated target velocity admits two equivalent expressions on this trajectory,

$$v \;=\; \frac{dy_t}{dt} \;=\; y_1 - y_{\text{init}} \;=\; \frac{y_1 - y_t}{1-t}. \tag{6}$$

We adopt the second form throughout, since it remains well-defined for an arbitrary state y_t at time t, the regime encountered when integrating the ODE at inference. Under the x-prediction parameterization with $\hat{y}_1 = f_\theta(y_t, t, c)$, the implied velocity is obtained by replacing y_1 with $\hat{y}_1$ in Eq. 6,

$$\hat{v}_\theta \;=\; \frac{\hat{y}_1 - y_t}{1-t}. \tag{7}$$

The network thus only needs to model the gap $y_1 - y_{\text{init}}$ that the linear forward model leaves underdetermined.

Training. We supervise this prediction in velocity space via the v-loss, which under the linear interpolant is equivalent to a time-reweighted x-loss [21, 43]:

$$\mathcal{L}_{\text{FM}} \;=\; \mathbb{E}_{t\sim\mathcal{U}[0,1],\, y_1}\left[\left\|\hat{v}_\theta - v\right\|_2^2\right] \;=\; \mathbb{E}_{t,\, y_1}\left[\frac{1}{(1-t)^2}\left\|f_\theta(y_t,\, t,\, c) - y_1\right\|_2^2\right]. \tag{8}$$

The v-loss formulation places larger weight on later timesteps ($t \to 1$), where the prediction is closest to the clean signal and small errors in f_θ have the largest effect on the integrated ODE solution. The x-prediction parameterization itself is preserved because it lets the parameter-consistency loss in Section 4.2 operate on the predicted clean signal $\hat{y}_1$ directly, without an additional reparameterization step.

Inference. At inference time, the learned vector field defines the ODE $dy_t/dt = (\, f_\theta(y_t, t, c) - y_t \,) \,/\, (1-t)$, integrated from $y_0 = y_{\text{init}}$ at $t = 0$ to y_1 at $t = 1$. The dense prediction $\hat{y}$ is read off as the terminal state of the ODE.

4.2 DKI-Projected Supervision Loss

The flow matching loss of Section 4.1 supervises all signal dimensions uniformly and does not enforce that predicted and ground-truth signals at the same voxel imply the same microstructure. We now

make this constraint explicit, deriving a complementary loss that operates directly in DKI parameter space and exploits the linearity of $y = Hx$ to admit a closed-form signal-space implementation without per-step parameter fitting.

Loss kernel: parameter and signal-space forms. Let $H \in \mathbb{R}^{N\times 22}$ denote the full-column-rank DKI design matrix on the dense target directions, and for any pair of signal vectors $(\hat{y}, y) \in \mathbb{R}^N \times \mathbb{R}^N$ let $\hat{x} = (H^\top H)^{-1} H^\top \hat{y}$ and $x = (H^\top H)^{-1} H^\top y$ denote their least-squares DKI estimates. We measure parameter discrepancy in the Fisher metric of the linear estimation problem (See Appendix B.1),

$$\ell_{\mathrm{DKI}}(\hat{y}, y) \;\coloneqq\; \big\|\hat{x} - x\big\|^2_{H^\top H}, \tag{9}$$

which assigns large penalties to parameter directions tightly constrained by the acquisition geometry and small penalties to those along which H has small singular values. Substituting the definitions of $\hat{x}, x$ collapses Eq. (9) to a pure signal-space form,

$$\ell_{\mathrm{DKI}}(\hat{y}, y) \;=\; \big\|\Pi_R(\hat{y} - y)\big\|_2^2, \qquad \Pi_R \coloneqq H(H^\top H)^{-1} H^\top, \tag{10}$$

where Π_R is the orthogonal projection onto $\mathrm{range}(H)$. The full derivation is given in Appendix B.2 The parameter-space form (Eq. 9) clarifies the loss's statistical role; the signal-space form (Eq. 10) is what we implement, since it requires only one matrix-vector product through Π_R. Both are exactly equivalent.

The loss penalizes errors only in the rank-P=22 range component $\Pi_R(\hat{y} - y)$ and is by construction blind to errors in the null component $\Pi_N(\hat{y} - y) = (I - \Pi_R)(\hat{y} - y)$. Supervision of the null component therefore remains the responsibility of $\mathcal{L}_{\mathrm{FM}}$, so the combined objective reweights the range component while retaining full-space supervision.

Lifting to flow matching training. We insert this kernel into the flow matching objective by averaging over the same training distribution as $\mathcal{L}_{\mathrm{FM}}$, with a $1/(1-t)^2$ weighting:

$$\mathcal{L}_{\mathrm{DKI}} \;=\; \mathbb{E}_{t\sim\mathcal{U}[0,1],\, y_1}\left[\frac{1}{(1-t)^2}\,\ell_{\mathrm{DKI}}\big(f_\theta(y_t, t, c),\, y_1\big)\right]. \tag{11}$$

This weighting follows from a direct identity. The implied velocity under our linear interpolant satisfies $\hat{v}_\theta - v = (\hat{y}_1 - y_1)/(1-t)$, so applying the range projection to both sides yields

$$\frac{1}{(1-t)^2}\,\big\|\Pi_R(\hat{y}_1 - y_1)\big\|_2^2 \;=\; \big\|\Pi_R(\hat{v}_\theta - v)\big\|_2^2. \tag{12}$$

The factor $1/(1-t)^2$ thus emerges naturally from the velocity-space metric and concentrates the parameter-consistency penalty near $t \to 1$, where predictions are closest to the clean signal and small errors most directly affect the integrated ODE solution. $\mathcal{L}_{\mathrm{FM}}$ and $\mathcal{L}_{\mathrm{DKI}}$ are therefore both ℓ^2 losses in velocity form, differing only in the subspace on which the error is measured: the full signal space for $\mathcal{L}_{\mathrm{FM}}$, the range of H for $\mathcal{L}_{\mathrm{DKI}}$. We combine the two with a tunable weight λ, with ablations in Section 5 showing that the projected supervision improves microstructural accuracy without measurable loss in signal-level fidelity.

The matrix H depends only on the acquisition geometry and is precomputed per subject. The per-step cost of $\mathcal{L}_{\mathrm{DKI}}$ is two matrix-vector products and one 22×22 linear solve, negligible relative to the network forward pass. Per-subject masks handle differing numbers of acquired directions across subjects. We use a Perceiver-IO backbone [17]; full architectural and implementation details are in Appendix C.

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate ***MiCo*** on two large-scale dMRI cohorts. HCP-YA [41] provides high-angular-resolution acquisitions with 90 directions per shell at $b \in \{1000, 2000, 3000\}$ s/mm^2, while UK Biobank (UKB) [30] provides clinical-protocol acquisitions with 50 directions each at $b \in \{1000, 2000\}$ s/mm^2 (111 total directions per voxel vs. 288 for HCP-YA). We use 300 HCP-YA and 5000 UKB subjects with an $70\%/10\%/20\%$ subject-level train/validation/test split. Preprocessing uses QSIPrep [8]; full details in Appendix D.1.

Table 1: HCP-YA pooled voxel-median absolute errors ($\times 10^{-2}$), mean $\pm$ std over three training runs. Full seven-metric results for all methods are in Table 7; these summaries are distinct from the matched comparison in Table 5.

Context	Method	FA	MD	MK
10%	DISCUS	3.54 ± 0.36	2.99 ± 0.36	3.38 ± 0.35
10%	***MiCo***	2.96 ± 0.31	2.38 ± 0.33	2.85 ± 0.32
20%	DISCUS	2.62 ± 0.23	1.93 ± 0.21	2.27 ± 0.22
20%	***MiCo***	2.16 ± 0.21	1.60 ± 0.19	1.96 ± 0.20
30%	DISCUS	2.17 ± 0.18	1.46 ± 0.15	1.79 ± 0.17
30%	***MiCo***	1.77 ± 0.14	1.22 ± 0.13	1.53 ± 0.16

Context/target split. For each subject and shell, we sample a context set of M directions uniformly at random and use the remaining N as the target. We report sparsity regimes $|\mathcal{C}|/(M+N) \in \{10\%, 20\%, 30\%\}$ for HCP-YA; for UKB we report $\{20\%, 30\%\}$ only, since at 10% each shell yields only ~ 5 context directions, insufficient for several baselines. Splits are resampled per training step and fixed at evaluation time; b=0 volumes are always included in context.

Baselines. We compare against two families of methods. (i) *Analytical models* commonly used in clinical dMRI pipelines: DTI-WLS [2], spherical harmonics (SH-2, SH-4) [9], SHORE [32], and MAP-MRI [33]. (ii) *Deep learning models*: DISCUS [10], a continuous-domain geometric architecture specifically designed for dMRI angular super-resolution, together with two general-purpose generative bridges adapted to this task, conditional flow matching (CFM) [22] and I^{2}SB [23]. The training details can be found in Appendix D.2

Evaluation overview. We assess reconstruction quality at three complementary levels: (1) signal-level fidelity on the held-out q-space directions; (2) microstructural parameter maps derived from DTI/DKI fits; and (3) fiber-orientation distributions (FODs) recovered by constrained spherical deconvolution. Signal and scalar-map metrics are computed inside a per-subject white-matter + gray-matter mask to exclude CSF and background voxels whose statistics would otherwise dominate global averages. For learned-method signal and scalar-map summaries, we report mean $\pm$ std across 3 training runs, with voxel errors reduced to a pooled test-voxel median within each run. FOD aggregation is specified separately in Table 2. Appendix D.3.1 reports the additional matched comparisons.

5.2 Main Comparison

5.2.1 Signal-level Reconstruction

Signal NMSE measures reconstruction fidelity on held-out directions. MiCo and DISCUS have similar errors in many shell–context settings (Appendix D.4.1). In the additional matched comparison, MiCo has 4.48–5.31% higher signal NMSE alongside the downstream gains reported below.

5.2.2 Microstructural Parameter Maps

The reconstructed signals are fit with the DTI and DKI models via DIPY's[12] weighted least-squares estimator. From the DTI fit we derive fractional anisotropy (FA) and mean diffusivity (MD); from the DKI fit we additionally derive mean kurtosis (MK). Evaluation is performed voxel-wise using **Relative Error** and **Absolute Error** against the ground-truth parameter maps. Fig. 3 summarizes the resulting parameter-map quality on HCP-YA: line plots (top) report relative error of FA, MD, and MK across context fractions, while scatter plots (bottom) co-plot signal NMSE against FA MAE at each context fraction so that the two evaluation axes can be read off jointly.

Results. The complete HCP-YA numeric comparison (Table 7) reports FA, MD, AD, RD, MK, AK, and RK. MiCo has the lowest FA/MK/AK/RK errors at all three context fractions, whereas CFM or I^{2}SB has lower MD/AD/RD errors. The matched MiCo–DISCUS comparison in Table 5 gives FA/MD/MK reductions of 17.16–21.46%, despite higher signal NMSE. This separation shows why signal NMSE alone is an incomplete proxy for microstructural recovery. UKB results are in Table 9.

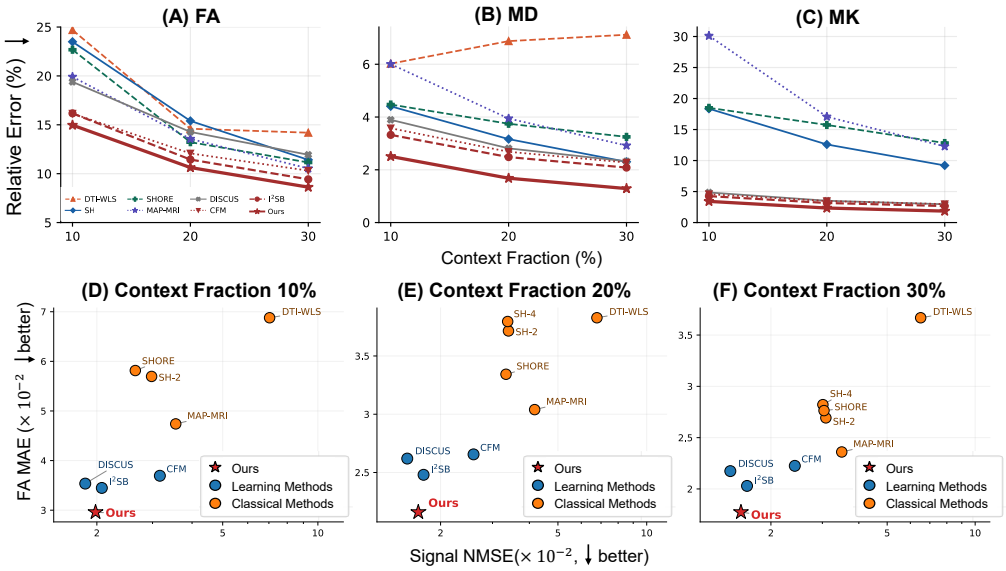

Figure 3: **Microstructural parameter recovery on HCP-YA. (A–C)** Relative error (↓) on FA, MD, MK vs. context fraction. **(D–F)** Signal NMSE vs. FA MAE per context fraction; The axes distinguish signal and downstream fidelity; lower values are better.

5.2.3 Fiber Orientation Distributions

We assess angular fidelity using CSD-derived FODs on white-matter voxels (Appendix D.4.4). Table 2 summarizes all methods and shell–context settings. MiCo leads all three shells at 10% context and $6/9$ settings overall. MAP-MRI or SHORE leads several higher-context cells; MiCo's largest gap to the leader is 1.19 ACC points. The strongest angular gains occur at the most aggressive acquisition reduction.

Table 2: HCP-YA FOD ACC ($\times 100$, higher is better), mean $\pm$ std across subjects. All methods and shell–context settings are shown. Bold marks the largest mean per shell and context. Context is the retained fraction.

	$b = 1000$			$b = 2000$			$b = 3000$		
Method	10%	20%	30%	10%	20%	30%	10%	20%	30%
DTI-WLS	56.18 ± 1.66	60.91 ± 2.31	63.35 ± 2.09	61.11 ± 2.37	67.48 ± 1.83	72.67 ± 1.40	57.71 ± 1.85	63.22 ± 2.57	67.50 ± 2.33
SH-2	51.77 ± 2.53	62.71 ± 2.26	68.64 ± 1.83	51.27 ± 14.47	62.61 ± 3.62	70.23 ± 2.56	55.23 ± 3.19	64.63 ± 3.66	71.40 ± 2.52
SH-4	N/A	64.26 ± 2.61	70.05 ± 2.08	N/A	64.91 ± 5.17	73.50 ± 2.06	N/A	68.54 ± 4.86	74.94 ± 2.39
SHORE	59.66 ± 1.96	67.36 ± 2.23	71.97 ± 1.93	63.70 ± 1.29	69.85 ± 1.17	74.95 ± 1.10	61.73 ± 2.30	68.76 ± 2.51	73.89 ± 2.49
MAP-MRI	59.78 ± 2.16	**67.61** ± 1.96	**72.28** ± 1.78	62.82 ± 1.28	70.36 ± 1.05	**75.69** ± 1.19	62.82 ± 1.42	70.62 ± 2.00	75.59 ± 1.97
CFM	52.82 ± 1.18	58.34 ± 1.13	61.42 ± 1.18	55.42 ± 1.21	61.55 ± 1.18	67.22 ± 1.22	54.81 ± 1.17	62.50 ± 1.23	67.71 ± 1.27
I^2SB	56.59 ± 1.15	61.11 ± 1.15	63.59 ± 1.30	60.32 ± 1.21	65.16 ± 1.28	70.56 ± 1.44	60.08 ± 1.19	66.41 ± 1.24	71.59 ± 1.10
DISCUS	60.53 ± 0.31	66.92 ± 0.31	71.24 ± 0.23	63.25 ± 0.09	69.81 ± 0.12	74.30 ± 0.28	63.42 ± 0.00	70.61 ± 0.39	75.35 ± 0.58
MiCo	**61.58** ± 1.35	67.04 ± 1.20	71.09 ± 1.17	**64.99** ± 1.24	**70.73** ± 1.09	74.57 ± 0.97	**64.96** ± 1.41	**71.35** ± 1.16	**75.75** ± 1.13

This ranking may reflect the methods' different objectives: MAP-MRI and SHORE use explicit radial–angular bases, while MiCo trains for dense signal recovery and DKI parameter consistency.

Table 3: **Component ablation across datasets.** Each row adds one design choice on top of the previous. We report signal NMSE and white-matter MK absolute error ($\times 10^{-2}$, $\downarrow$) at 20% context. **Bold** marks the best per column within each dataset.

Dataset	Variant	Source	λ_{DKI}	Signal NMSE $\downarrow$	MK abs. err. $\downarrow$	Δ MK vs. FM-only
HCP-YA	FM only	Gaussian	0	2.55	12.68	—
	+ DTI source	DTI	0	**1.67**	10.73	-15.4%
	+ DKI loss	DTI	0.13	1.68	**10.43**	$\mathbf{-17.7\%}$
UK Biobank	FM only	Gaussian	0	1.21	3.91	—
	+ DTI source	DTI	0	0.96	4.40	$+12.5\%$
	+ DKI loss	DTI	0.36	0.85	**3.48**	$\mathbf{-11.0\%}$

With more observations, basis fitting can recover fine angular structure; under sparse acquisition, MiCo benefits from its learned prior. Appendix D.4.4 shows the corresponding FOD reconstructions.

5.3 Ablation Study

Component contributions. We ablate the two design choices on HCP-YA and UK Biobank by successively replacing the Gaussian source with the DTI dense estimate and adding the consistency loss (Tab. 3). On HCP-YA, the DTI source delivers most of the signal-NMSE improvement (-34%) and reduces WM-MK error by 15.4%; the consistency loss tightens MK by a further 2.8% at no measurable signal-NMSE cost. On UK Biobank the DTI source alone improves signal NMSE but *degrades* WM-MK by 12.5%, and the consistency loss is required to recover and surpass the Gaussian-source baseline (-11.0% MK). The DKI-residual analysis (Appendix D.3.2) connects this effect to model fit: on UKB, the loss reduces MK error by 26.5% and 8.8% in the two lower-residual quartiles, with changes of $+9.0\%$ and $+1.3\%$ in the higher-residual quartiles. The consistency loss is most effective where the DKI approximation closely matches the signal.

Sensitivity to λ_{DKI}. Fig. 4 sweeps λ_{DKI} on HCP-YA validation and traces a clear Pareto front: moderate weights ($\lambda_{\text{DKI}} \in [0.01, 0.2]$) reduce MK error substantially at minimal signal-NMSE cost, while large weights ($\lambda_{\text{DKI}} \geq 0.6$) trade signal fidelity for diminishing microstructural gains. We select $\lambda_{\text{DKI}} = 0.13$ on HCP-YA. The same sweep on UKB (Appendix D.4.3) yields a larger optimal weight ($\lambda_{\text{DKI}} = 0.36$), with residual-stratified limitations as discussed above.

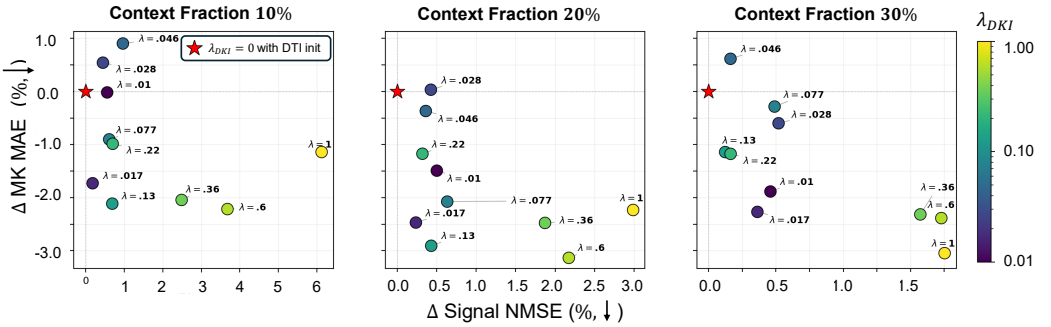

Figure 4: **Sensitivity of *MiCo* to the DKI-consistency weight λ_{DKI} on HCP-YA validation.** Each subplot shows the trade-off between signal-NMSE degradation (Δ Signal NMSE, x-axis, relative to the λ_{DKI}=0 DTI-init baseline) and microstructural improvement (Δ MK MAE, y-axis; negative = improvement) at three context fractions. Moderate λ_{DKI} achieve substantial MK improvements (1–3%) at minimal signal-NMSE cost ($< 1\%$ at ctx 20%/30%). We select λ_{DKI}=0.13 for all main HCP-YA results.

6 Conclusion

We presented **MiCo**, a flow matching framework for dMRI angular super-resolution that exploits the linearity of the DKI forward model through two closed-form components: a DTI-based source initialization that replaces the standard Gaussian prior, and a microstructure-consistency loss that supervises predictions on the rank-22 range subspace where signal errors translate directly into errors in downstream parameters. On HCP-YA and UK Biobank, MiCo combines competitive signal reconstruction with improved microstructural recovery. Its HCP-YA results lead on FA and kurtosis scalars and on FOD accuracy at 10% context across all shells. The ablations connect these gains to physics-guided initialization and DKI-range supervision. These findings establish microstructural fidelity as a useful target for angular super-resolution and show how the geometry of a forward model can align generative learning with downstream parameters.

7 Limitations

MiCo uses a DKI-based microstructural prior, whose benefit depends on how well the forward model describes the tissue signal. Our evaluation focuses on angular super-resolution in HCP-YA and UK Biobank; clinical generalization is outside the scope of this study. Future directions include incorporating spatial context, extending to more diverse populations and acquisition settings, and improving inference efficiency and fiber-orientation consistency.

Acknowledgments and Disclosure of Funding

This work was supported by the National Institutes of Health (AG091653, AG068399, AG084375) and the Foundation of Hope. This research has been conducted using the UKB resource (application number 91673), subject to a data transfer agreement.

Competing Interests

The authors declare no competing interests.

References

[1] Adam W Anderson. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(5):1194–1206, 2005.

[2] Peter J Basser, James Mattiello, and Denis LeBihan. Mr diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267, 1994.

[3] Jan-Hendrik Bastek, WaiChing Sun, and Dennis M Kochmann. Physics-informed diffusion models. *arXiv preprint arXiv:2403.14404*, 2024.

[4] Ilaria Boscolo Galazzo, Lorenza Brusini, Silvia Obertino, Mauro Zucchelli, Cristina Granziera, and Gloria Menegaz. On the viability of diffusion mri-based microstructural biomarkers in ischemic stroke. *Frontiers in neuroscience*, 12:92, 2018.

[5] Geng Chen, Yoonmi Hong, Yongqin Zhang, Jaeil Kim, Khoi Minh Huynh, Jiquan Ma, Weili Lin, Dinggang Shen, Pew-Thian Yap, and UNC/UMN Baby Connectome Project Consortium. Estimating tissue microstructure with undersampled diffusion data via graph convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 280–290. Springer, 2020.

[6] Geng Chen, Haotian Jiang, Jiannan Liu, Jiquan Ma, Hui Cui, Yong Xia, and Pew-Thian Yap. Hybrid graph transformer for tissue microstructure estimation with undersampled diffusion mri data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 113–122. Springer, 2022.

[7] Zijian Chen, Jueqi Wang, and Archana Venkataraman. Qid 2: An image-conditioned diffusion model for q-space up-sampling of dwi data. In *International Workshop on Computational Diffusion MRI*, pages 119–131. Springer, 2024.

[8] Matthew Cieslak, Philip A Cook, Xiaosong He, Fang-Cheng Yeh, Thijs Dhollander, Azeez Adebimpe, Geoffrey K Aguirre, Danielle S Bassett, Richard F Betzel, Josiane Bourque, et al. Qsiprep: an integrative platform for preprocessing and reconstructing diffusion mri data. *Nature methods*, 18(7):775–778, 2021.

[9] Maxime Descoteaux, Elaine Angelino, Shaun Fitzgibbons, and Rachid Deriche. Regularized, fast, and robust analytical q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(3):497–510, 2007.

[10] Christian Ewert, David Kügler, Rüdiger Stirnberg, Alexandra Koch, Anastasia Yendiki, and Martin Reuter. Geometric deep learning for diffusion mri signal reconstruction with continuous samplings (discus). *Imaging neuroscience*, 2:imag–2, 2024.

[11] Rutger HJ Fick, Madelaine Daianu, Marco Pizzolato, Demian Wassermann, Russell E Jacobs, Paul M Thompson, Terrence Town, and Rachid Deriche. Comparison of biomarkers in transgenic alzheimer rats using multi-shell diffusion mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 187–199. Springer, 2016.

[12] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, Ian Nimmo-Smith, and Dipy Contributors. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

[13] Eric K Gibbons, Kyler K Hodgson, Akshay S Chaudhari, Lorie G Richards, Jennifer J Majersik, Ganesh Adluru, and Edward VR DiBella. Simultaneous noddi and gfa parameter map generation from subsampled q-space imaging using deep learning. *Magnetic resonance in medicine*, 81(4): 2399–2411, 2019.

[14] Vladimir Golkov, Alexey Dosovitskiy, Jonathan I Sperl, Marion I Menzel, Michael Czisch, Philipp Sämann, Thomas Brox, and Daniel Cremers. Q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE transactions on medical imaging*, 35(5):1344–1351, 2016.

[15] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, 2008.

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[17] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv:2107.14795*, 2021.

[18] Jens H Jensen, Joseph A Helpern, Anita Ramani, Hanzhang Lu, and Kyle Kaczynski. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 53(6):1432–1440, 2005.

[19] Heidi Johansen-Berg and Timothy EJ Behrens. *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press, 2013.

[20] Derek K Jones. *Diffusion mri*. Oxford University Press, 2010.

[21] Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. *arXiv preprint arXiv:2511.13720*, 2025.

[22] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[23] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I^2sb: Image-to-image schrödinger bridge, 2023. URL https://arxiv.org/abs/2302.05872.

[24] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

[25] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[26] Xinzhe Luo, Yingzhen Li, and Chen Qin. Upmri: Unsupervised parallel mri reconstruction via projected conditional flow matching. *arXiv preprint arXiv:2512.17493*, 2025.

[27] Matthew Lyon, Paul Armitage, and Mauricio A Álvarez. Angular super-resolution in diffusion mri with a 3d recurrent convolutional autoencoder. In *International Conference on Medical Imaging with Deep Learning*, pages 834–846. PMLR, 2022.

[28] Matthew Lyon, Paul Armitage, and Mauricio A Álvarez. Spatio-angular convolutions for super-resolution in diffusion mri. *Advances in Neural Information Processing Systems*, 36: 12457–12475, 2023.

[29] Perla Mayo, Carolin M Pirkl, Alin Achim, Bjoern Menze, and Mohammad Golbabaee. Physics informed guided diffusion for accelerated multi-parametric mri reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 429–439. Springer, 2025.

[30] Karla L Miller, Fidel Alfaro-Almagro, Neal K Bangerter, David L Thomas, Essa Yacoub, Junqian Xu, Andreas J Bartsch, Saad Jbabdi, Stamatios N Sotiropoulos, Jesper LR Andersson, et al. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature neuroscience*, 19(11):1523–1536, 2016.

[31] Mu Nan, Taohui Xiao, Ruoyou Wu, Shoujun Yu, Ye Li, Hairong Zheng, and Shanshan Wang. Physics-guided diffusion transformer with spherical harmonic posterior sampling for high-fidelity angular super-resolution in diffusion mri. *arXiv preprint arXiv:2509.07020*, 2025.

[32] Evren Ozarslan, C Koay, Timothy M Shepherd, S Blackband, and Peter J Basser. Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space mri. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 17, page 1396. Citeseer, 2009.

[33] Evren Özarslan, Cheng Guan Koay, Timothy M Shepherd, Michal E Komlosh, M Okan İrfanoğlu, Carlo Pierpaoli, and Peter J Basser. Mean apparent propagator (map) mri: a novel diffusion imaging method for mapping tissue microstructure. *NeuroImage*, 78:16–32, 2013.

[34] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023.

[35] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[36] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS computational biology*, 1(4):e42, 2005.

[37] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

[38] Alexandra Tobisch, Rüdiger Stirnberg, Robbert L Harms, Thomas Schultz, Alard Roebroeck, Monique MB Breteler, and Tony Stöcker. Compressed sensing diffusion spectrum imaging for accelerated diffusion microstructure mri in long-term population imaging. *Frontiers in neuroscience*, 12:650, 2018.

[39] J-Donald Tournier, Fernando Calamante, and Alan Connelly. Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*, 35(4):1459–1472, 2007.

[40] J-Donald Tournier, Fernando Calamante, and Alan Connelly. Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR in Biomedicine*, 26(12):1775–1786, 2013.

[41] David C Van Essen, Kamil Ugurbil, Edward Auerbach, Deanna Barch, Timothy EJ Behrens, Richard Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[42] Chuyang Ye. Learning-based ensemble average propagator estimation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 593–601. Springer, 2017.

[43] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in neural information processing systems*, 36:13294–13307, 2023.

A Extended dMRI Background

A.1 Cumulant Expansion of the Diffusion Signal

Under the Gaussian phase approximation, the diffusion-weighted signal at a voxel is the characteristic function of the spin-displacement distribution. Expanding $\ln S(\mathbf{q})/S_0$ as a cumulant series in $\mathbf{q}$ and truncating at fourth order yields

$$\ln \frac{S(\mathbf{q})}{S_0} = -b\, D_{ij}\, g_i g_j \;+\; \frac{1}{6}\, b^2\, \overline{D}^{\,2}\, W_{ijkl}\, g_i g_j g_k g_l \;+\; \mathcal{O}(b^3), \tag{13}$$

where $\mathbf{q} = b^{1/2}\,\mathbf{g}$ with $\mathbf{g} \in \mathbb{S}^2$ the gradient direction, D_{ij} is the symmetric rank-2 diffusion tensor, W_{ijkl} is the fully symmetric rank-4 kurtosis tensor, and $\overline{D} = \mathrm{tr}(D)/3$ is the mean diffusivity. Einstein summation is used over indices $i, j, k, l \in \{x, y, z\}$. Truncating (13) at the b^1 term gives DTI [2]; retaining the b^2 term gives DKI [18].

A.2 Parameter Counts

The diffusion tensor D_{ij} is symmetric, so it has 6 independent components. The kurtosis tensor W_{ijkl} is fully symmetric in all four indices, yielding $\binom{4+3-1}{3} = 15$ independent components. Together with the $\ln S_0$ intercept that the linear regression must absorb, the DKI forward model has $P = 22$ parameters per voxel.

A.3 Design Matrix H

Equation (13) is linear in (D_{ij}, W_{ijkl}) once $(b, \mathbf{g})$ is fixed. Stacking the cumulant expansion at L query directions $\{(b_l, \mathbf{g}_l)\}_{l=1}^{L}$ yields the linear system $y = Hx$, where $x \in \mathbb{R}^{22}$ collects the unique parameters in the order $x = (D_{xx}, D_{yy}, D_{zz}, 2D_{xy}, 2D_{xz}, 2D_{yz}, W_{xxxx}, \ldots)^\top$. The l-th row of H is

$$H_l = \big[\, \underbrace{-b_l\, g_{l,x}^2,\; \ldots,\; -2b_l\, g_{l,y} g_{l,z}}_{\text{DTI block (6)}},\; \underbrace{\tfrac{1}{6} b_l^2\, \overline{D}^{\,2}\, g_{l,x}^4,\; \ldots}_{\text{DKI block (15)}} \,\big], \tag{14}$$

with the kurtosis monomial coefficients incorporating the multinomial multiplicities arising from the four-fold symmetry of W_{ijkl}. Following standard practice [18], $\overline{D}$ is estimated from the DTI fit and held fixed during the DKI regression. The matrix H depends only on the acquisition geometry $(b_l, \mathbf{g}_l)$ and is independent of the measured signals; for our method, H is precomputed once per subject from the b-table.

A.4 Scalar Microstructural Metrics

From the eigendecomposition $D_{ij} = \sum_{m=1}^{3} \lambda_m \, v_m^{(i)} v_m^{(j)}$ with $\lambda_1 \geq \lambda_2 \geq \lambda_3$, DTI scalars are defined as:

$$\mathrm{MD} = \tfrac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3), \tag{15}$$

$$\mathrm{FA} = \sqrt{\tfrac{3}{2}} \, \frac{\sqrt{(\lambda_1 - \mathrm{MD})^2 + (\lambda_2 - \mathrm{MD})^2 + (\lambda_3 - \mathrm{MD})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \tag{16}$$

$$\mathrm{AD} = \lambda_1, \qquad \mathrm{RD} = \tfrac{1}{2}(\lambda_2 + \lambda_3). \tag{17}$$

DKI-derived scalars are defined as directional averages of the apparent kurtosis $K(\mathbf{n}) = \overline{D}^{\,2} \, W_{ijkl} \, n_i n_j n_k n_l / (D_{ij} n_i n_j)^2$:

$$\mathrm{MK} = \tfrac{1}{4\pi} \int_{\mathbb{S}^2} K(\mathbf{n}) \, d\mathbf{n}, \tag{18}$$

$$\mathrm{AK} = K(v_1), \qquad \mathrm{RK} = \tfrac{1}{2\pi} \int_0^{2\pi} K(v_2 \cos\theta + v_3 \sin\theta) \, d\theta, \tag{19}$$

representing mean, axial, and radial kurtosis respectively.

B Mathematical Details

B.1 Fisher Information of the DKI Parameter Posterior

Under the standard linear model $y = Hx + \eta, \ \eta \sim \mathcal{N}(0, \sigma^2 I)$, the log-likelihood for the parameter x is

$$\log p(y \mid x) = -\tfrac{1}{2\sigma^2} (y - Hx)^\top (y - Hx) + \text{const.} \tag{20}$$

The Fisher information matrix of x is then

$$\mathcal{I}(x) = -\mathbb{E}\big[\nabla_x^2 \log p(y \mid x)\big] = \tfrac{1}{\sigma^2} \, H^\top H. \tag{21}$$

Up to the noise-variance factor σ^2, $H^\top H$ is therefore the Fisher information of the parameter posterior, and $\|\cdot\|_{H^\top H}$ measures parameter discrepancy in the natural geometry of the estimation problem: directions in parameter space that are tightly constrained by the data (large eigenvalues of $H^\top H$) receive proportionally larger penalties.

B.2 Derivation of the Signal-Space DKI Loss

We provide the full derivation of Eq. (10) from Eq. (9). Substituting the least-squares estimates $\hat{x} = (H^\top H)^{-1} H^\top \hat{y}$ and $x = (H^\top H)^{-1} H^\top y$:

$$\ell_{\mathrm{DKI}}(\hat{y}, y) = (\hat{x} - x)^\top H^\top H (\hat{x} - x) \tag{22}$$

$$= \big[(H^\top H)^{-1} H^\top (\hat{y} - y)\big]^\top H^\top H \big[(H^\top H)^{-1} H^\top (\hat{y} - y)\big] \tag{23}$$

$$= (\hat{y} - y)^\top H (H^\top H)^{-1} H^\top H (H^\top H)^{-1} H^\top (\hat{y} - y) \tag{24}$$

$$= (\hat{y} - y)^\top H (H^\top H)^{-1} H^\top (\hat{y} - y) \tag{25}$$

$$= (\hat{y} - y)^\top \Pi_R (\hat{y} - y) \tag{26}$$

$$= \big\|\Pi_R (\hat{y} - y)\big\|_2^2, \tag{27}$$

where the fourth equality uses $(H^\top H)^{-1} H^\top H (H^\top H)^{-1} = (H^\top H)^{-1}$, and $\Pi_R = H(H^\top H)^{-1} H^\top$ is symmetric and idempotent ($\Pi_R^2 = \Pi_R$, $\Pi_R^\top = \Pi_R$), confirming it is the orthogonal projection onto $\mathrm{range}(H)$.

C Implementation Details

Architecture. For each context measurement $(\mathbf{q}_i, S(\mathbf{q}_i))$, we form a token by concatenating D-dimensional projections of a Fourier embedding [37] of the q-vector $\mathbf{q}_i = \sqrt{b_i} \, \hat{\mathbf{q}}_i$ and of the scalar

signal. A learnable latent set $z \in \mathbb{R}^{K \times D}$ attends to these context tokens via cross-attention followed by self-attention; we stack 2 such encoder blocks. Query tokens are assembled identically from target q-vectors and the interpolant state y_t, and pass through 2 decoder blocks of cross-attention against z followed by an MLP. A final linear head outputs the scalar prediction $\hat{y}_1 \in \mathbb{R}^N$. All blocks use adaptive layer normalization [34] conditioned on the time embedding.

Hyperparameter summary. Total network configuration is listed in Table 4.

Table 4: Network architecture hyperparameters for ***MiCo***.

Hyperparameter	Value
Hidden dimension D	128
Number of latents K	64
Encoder depth (cross+self blocks)	2
Self-attention layers per encoder block	2
Decoder depth (cross blocks)	2
Number of attention heads	8
Head dimension	16
Fourier feature dimension P	128
Fourier feature bandwidth σ	4.0
Time conditioning	Adaptive Layer Norm

D Additional Experiments Details and Results

D.1 Data Processing

We preprocess raw dMRI volumes with QSIPrep [8], a standardized pipeline that performs denoising, Gibbs-ringing removal, susceptibility distortion correction, eddy-current and head-motion correction, and brain extraction. The output of this stage is a co-registered 4D volume of diffusion-weighted images together with the corresponding b-values and b-vectors.

Per-voxel normalization. On top of QSIPrep's output, we apply a three-step normalization to the DWI signals at each voxel. First, we form the normalized signal $S(\mathbf{q})/S_0$ by dividing each diffusion-weighted volume by the per-voxel mean of all b=0 volumes [10]. This removes the T2-weighted baseline and produces signals approximately in $[0, 1]$. Second, we clip outlier values that exceed physical plausibility (arising primarily from CSF and motion residuals) to the interval $[0.01, 1.2]$ before taking the logarithm. The clip floor at 0.01 prevents $-\infty$ in $\log$ and corresponds to extreme attenuation that is unreliable to estimate; the upper clip at 1.2 allows for mild noise above the unattenuated baseline without admitting physically implausible values. Third, we apply the natural logarithm and standardize globally across the training cohort:

$$\tilde{S}(\mathbf{q}) \;=\; \frac{\log\big(\mathrm{clip}(S(\mathbf{q})/S_0,\ 0.01,\ 1.2)\big) - \mu}{\sigma}, \tag{28}$$

with μ and σ estimated as the cohort-wide mean and standard deviation of the log-clipped values. The log-transform is the natural domain for DKI parameter estimation (both DTI and DKI fits operate on log-signals), and global standardization places the network's input near zero mean and unit variance, which we found stabilizes training.

Voxel-level batching. After preprocessing, each subject yields a 5D tensor of shape $(H \times W \times D, N_{\mathrm{dirs}}, 5)$, where the last dimension stores (b-value, g_x, g_y, g_z, $\tilde{S}$) for each direction. b-values are scaled by 10^{-3} before being fed to the network (i.e., b=1000 s/mm^2 enters as 1.0) to keep input magnitudes $\mathcal{O}(1)$. Spatial neighbourhood information is not used; ***MiCo*** operates on a single voxel's signal vector.

Rotation augmentation. At training time, we apply a uniformly-sampled random 3D rotation $R \in \mathrm{SO}(3)$ to all gradient directions of each voxel. This exploits the rotational equivariance of the diffusion signal under the q-vector basis: rotating both the gradient directions and the underlying tissue's principal diffusion axes leaves the signal unchanged, so applying R to $\{\hat{\mathbf{g}}_i\}$ is equivalent

to viewing the same voxel from a rotated reference frame. The augmentation enriches the angular diversity seen during training without altering signal values, and is applied independently per subject per batch.

D.2 Training Details

We train with AdamW [25] ($\beta_1 = 0.9, \beta_2 = 0.999$, weight decay 0.01), an initial learning rate of 10^{-4} with $1{,}000$-step linear warmup followed by cosine decay over $100{,}000$ total steps, gradient clipping at ℓ_2 norm 1.0, and an exponential moving average (EMA) of network weights with decay 0.9999. The EMA weights are used at evaluation. Mixed-precision (bfloat16) training is used throughout. Batch size is $2{,}048$ voxels for HCP-YA and for UKB. Models are trained on a single NVIDIA A6000 GPU for approximately 8 hours per run.

We integrate the velocity ODE $dy_t/dt = \hat{v}_\theta(y_t, t, c)$ from $y_0 = y_{\text{init}}$ at $t = 0$ to $t = 1$ using a midpoint solver with 25 steps. The predicted signal $\hat{y}$ is read off as the terminal state of the ODE and de-normalized using the inverse of the per-voxel log-standardization in Appendix D.1 for downstream evaluation.

D.3 Additional Analyses

D.3.1 Matched Comparisons

Table 5 reports the matched MiCo–DISCUS comparison. Its evaluation uses matched q-space samples, masks, and a set of $40{,}960$ voxels; its summaries differ from the pooled three-run results in Table 7. Each entry lists the error pair and the reported relative change.

Table 5: Matched HCP-YA comparison: MiCo/DISCUS error pairs followed by the reported relative change. Negative changes favor MiCo.

Metric	10% context	20% context	30% context
FA	2.97/3.64 (-18.46%)	2.17/2.65 (-18.20%)	1.77/2.20 (-19.38%)
MD	2.36/3.01 (-21.46%)	1.60/1.95 (-18.23%)	1.22/1.48 (-17.23%)
MK	2.88/3.47 (-17.16%)	1.95/2.37 (-17.67%)	1.53/1.85 (-17.22%)
AK	3.40/3.65 (-7.11%)	2.68/3.01 (-11.08%)	2.28/2.64 (-13.88%)
RK	4.77/5.87 (-18.75%)	3.47/4.40 (-21.13%)	2.91/3.67 (-20.67%)

The matched HCP-YA component comparison at 20% context reports a $+0.18\%$ signal-NMSE change and a -2.52% WM-MK change after adding the DKI loss. Table 3 gives the corresponding original three-variant ablation summaries.

D.3.2 DKI Model Validity

We measure dense ground-truth mismatch with the DKI subspace by

$$r_{\text{DKI}}(y) = \frac{\|(I - \Pi_R)y\|_2}{\|y\|_2}, \tag{29}$$

where Π_R projects onto the rank-22 DKI range. On UKB at 20% context, the median residual is 0.052 and its 95th percentile is 0.165. Stratifying test voxels into increasing residual quartiles gives MK-error changes from adding the DKI loss of -26.5%, -8.8%, $+9.0\%$, and $+1.3\%$, respectively. The reductions in the lower-residual half link the consistency benefit to the accuracy of the DKI approximation.

D.3.3 DTI Initialization Conditioning

The context pools all nonzero-b measurements and retains all $b = 0$ observations. For $g_i = (g_{ix}, g_{iy}, g_{iz})$, the seven-column DTI design row is

$$[-b_i g_{ix}^2,\ -2b_i g_{ix} g_{iy},\ -b_i g_{iy}^2,\ -2b_i g_{ix} g_{iz},\ -2b_i g_{iy} g_{iz},\ -b_i g_{iz}^2,\ -1]. \tag{30}$$

The final coefficient corresponds to the negative log-baseline under this sign convention. The Tikhonov-regularized fit is performed once across shells. At 10% context, the reported joint context counts and conditioning diagnostics are:

Cohort	Joint $N_{\mathcal{C}}$	Median $\kappa(A_{\mathcal{C}})$	95th percentile	Maximum	Fraction $> 10^3$
HCP-YA	45	3.28	4.57	9.07	0
UKB	21	7.24	16.54	130.2	0

The observed condition numbers support the stability of the joint cross-shell DTI fit in these sampled contexts.

D.3.4 Inference Cost and Fewer Solver Steps

We benchmarked in-memory HCP-YA batches of $40{,}960$ voxels on one H100 NVL GPU, including transfer, preprocessing, and terminal-signal generation, but excluding data loading, disk output, and downstream fitting.

Method	Solver steps	Throughput (voxels/s)	Peak memory (GiB)
MiCo	5	4277	3.85
MiCo	10	2127	3.85
MiCo	15	1415	3.85
MiCo	20	1058	3.85
MiCo	25	846	3.85
DISCUS	direct	35551	1.76

The tabulated throughput ratio for $25 \to 5$ steps is $4277/846 = 5.06$. DISCUS is $8.31\times$ faster than five-step MiCo and $42.02\times$ faster than 25-step MiCo in this benchmark. At 20% context, reducing from 25 to 5 steps increases NMSE/FA/MK errors by $4.3\%/1.9\%/2.3\%$ on HCP-YA and $4.4\%/0.8\%/9.9\%$ on UKB, respectively. These fixed-checkpoint results quantify the accuracy–throughput trade-off available by changing the solver step count.

D.4 Additional Results

D.4.1 Signal-level Reconstruction on HCP-YA

Table 6: Signal prediction NMSE ($\times 10^{-2}$) across different shells and sparsity levels. Results are reported as mean $\pm$ standard deviation (std); lower values indicate better reconstruction performance. Sparsity levels ($10\%, 20\%, 30\%$) represent the percentage of sampled signal retained. **Bold** and underlined values denote the best and second-best results in each column, respectively.

Method	b=1000 s/mm^2			b=2000 s/mm^2			b=3000 s/mm^2		
	10%	20%	30%	10%	20%	30%	10%	20%	30%
DTI-WLS	4.27 ± 0.10	5.46 ± 0.13	6.30 ± 0.19	3.77 ± 0.18	3.49 ± 0.13	3.55 ± 0.19	12.02 ± 0.16	10.36 ± 0.23	8.70 ± 0.17
SH-2	1.45 ± 0.33	1.72 ± 0.17	1.70 ± 0.22	2.84 ± 0.36	3.38 ± 0.07	3.04 ± 0.23	3.89 ± 0.21	4.19 ± 0.34	3.78 ± 0.36
SH-4	N/A	1.82 ± 0.26	1.79 ± 0.22	N/A	3.40 ± 0.20	3.00 ± 0.22	N/A	4.09 ± 0.19	3.65 ± 0.20
SHORE	1.37 ± 0.47	1.65 ± 0.28	1.62 ± 0.40	2.48 ± 0.15	3.14 ± 0.42	2.94 ± 0.12	3.76 ± 0.74	4.66 ± 0.84	4.04 ± 0.82
MAP-MRI	1.74 ± 0.52	2.16 ± 0.37	1.94 ± 0.37	3.49 ± 0.37	4.13 ± 0.50	3.46 ± 0.96	4.71 ± 0.80	5.08 ± 0.44	4.19 ± 0.78
CFM	1.64 ± 0.13	1.35 ± 0.18	1.28 ± 0.18	3.15 ± 0.36	2.60 ± 0.36	2.44 ± 0.41	4.33 ± 0.36	3.63 ± 0.26	3.39 ± 0.37
I^{2}SB	1.10 ± 0.15	0.88 ± 0.15	0.85 ± 0.13	2.10 ± 0.21	1.84 ± 0.30	1.69 ± 0.43	**2.78** ± 0.28	2.44 ± 0.22	2.32 ± 0.36
DISCUS	1.00 ± 0.18	**0.77** ± 0.13	**0.72** ± 0.08	2.11 ± 0.15	1.73 ± 0.23	**1.45** ± 0.21	2.98 ± 0.38	2.48 ± 0.24	2.31 ± 0.27
MiCo	**0.96** ± 0.15	0.82 ± 0.13	0.77 ± 0.12	**1.98** ± 0.20	**1.68** ± 0.18	1.58 ± 0.17	2.80 ± 0.19	**2.41** ± 0.17	**2.27** ± 0.16

D.4.2 Full HCP-YA Downstream Results

Table 7: HCP-YA pooled voxel-median absolute errors ($\times 10^{-2}$, lower is better). The $\pm$ entries are reported dispersion; learned-method summaries use three training runs. These pooled summaries are distinct from the separately matched effects in Table 5.

Method	FA	MD	AD	RD	MK	AK	RK
10% context							
DTI-WLS	6.88 ± 0.71	3.14 ± 0.45	5.97 ± 0.76	3.56 ± 0.52	74.89 ± 8.47	70.75 ± 8.36	73.73 ± 9.12
SH-2	5.70 ± 0.58	2.26 ± 0.31	5.24 ± 0.63	3.06 ± 0.42	9.46 ± 1.02	11.98 ± 1.40	16.03 ± 1.86
SH-4	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SHORE	5.81 ± 0.67	2.14 ± 0.30	5.10 ± 0.62	3.16 ± 0.47	9.71 ± 1.10	12.13 ± 1.44	18.17 ± 2.20
MAP-MRI	4.74 ± 0.52	3.18 ± 0.43	6.10 ± 0.73	2.49 ± 0.36	14.55 ± 1.64	21.14 ± 2.51	20.32 ± 2.47
CFM	3.69 ± 0.38	1.32 ± 0.15	3.53 ± 0.40	1.96 ± 0.25	8.34 ± 0.83	8.88 ± 1.00	12.47 ± 1.33
I^2SB	3.45 ± 0.41	1.33 ± 0.19	3.36 ± 0.43	1.78 ± 0.26	9.75 ± 1.10	10.98 ± 1.42	12.76 ± 1.58
DISCUS	3.54 ± 0.36	2.99 ± 0.36	5.30 ± 0.61	3.52 ± 0.47	3.38 ± 0.35	3.59 ± 0.43	5.53 ± 0.64
MiCo	2.96 ± 0.31	2.38 ± 0.33	4.76 ± 0.53	2.66 ± 0.35	2.85 ± 0.32	3.40 ± 0.42	4.74 ± 0.55
20% context							
DTI-WLS	3.83 ± 0.37	3.64 ± 0.44	6.39 ± 0.64	2.50 ± 0.31	82.16 ± 8.63	78.88 ± 8.32	78.92 ± 8.06
SH-2	3.71 ± 0.33	1.52 ± 0.17	3.43 ± 0.35	1.98 ± 0.25	6.27 ± 0.58	9.54 ± 0.98	12.33 ± 1.18
SH-4	3.80 ± 0.34	1.47 ± 0.16	3.45 ± 0.32	1.98 ± 0.23	6.25 ± 0.56	8.34 ± 0.78	12.08 ± 1.15
SHORE	3.34 ± 0.31	1.70 ± 0.19	3.28 ± 0.32	2.28 ± 0.28	8.09 ± 0.85	8.79 ± 0.90	13.56 ± 1.51
MAP-MRI	3.04 ± 0.27	1.89 ± 0.20	4.40 ± 0.43	1.55 ± 0.19	8.01 ± 0.71	13.05 ± 1.23	13.14 ± 1.39
CFM	2.66 ± 0.21	1.00 ± 0.12	2.64 ± 0.25	1.41 ± 0.15	5.74 ± 0.49	6.91 ± 0.61	9.37 ± 0.85
I^2SB	2.48 ± 0.24	0.98 ± 0.12	2.59 ± 0.28	1.24 ± 0.18	5.54 ± 0.61	7.19 ± 0.74	8.51 ± 0.95
DISCUS	2.62 ± 0.23	1.93 ± 0.21	3.81 ± 0.39	2.37 ± 0.27	2.27 ± 0.22	2.98 ± 0.28	4.13 ± 0.40
MiCo	2.16 ± 0.21	1.60 ± 0.19	3.38 ± 0.37	1.83 ± 0.23	1.96 ± 0.20	2.69 ± 0.28	3.49 ± 0.36
30% context							
DTI-WLS	3.67 ± 0.31	3.78 ± 0.36	6.00 ± 0.59	2.79 ± 0.31	75.19 ± 6.44	67.26 ± 6.54	78.99 ± 7.29
SH-2	2.69 ± 0.21	0.97 ± 0.10	2.56 ± 0.23	1.40 ± 0.16	4.41 ± 0.38	8.27 ± 0.71	9.53 ± 0.82
SH-4	2.82 ± 0.23	0.93 ± 0.09	2.58 ± 0.24	1.42 ± 0.15	4.47 ± 0.37	7.00 ± 0.60	9.43 ± 0.83
SHORE	2.76 ± 0.22	1.42 ± 0.14	2.72 ± 0.27	1.89 ± 0.21	6.58 ± 0.59	7.07 ± 0.65	11.14 ± 1.02
MAP-MRI	2.36 ± 0.19	1.28 ± 0.13	3.00 ± 0.26	1.26 ± 0.14	5.71 ± 0.48	9.08 ± 0.81	9.95 ± 0.94
CFM	2.23 ± 0.17	0.82 ± 0.09	2.21 ± 0.19	1.17 ± 0.11	4.62 ± 0.34	5.88 ± 0.47	8.07 ± 0.61
I^2SB	2.03 ± 0.18	0.78 ± 0.09	2.09 ± 0.21	1.02 ± 0.12	4.42 ± 0.43	5.89 ± 0.54	7.27 ± 0.70
DISCUS	2.17 ± 0.18	1.46 ± 0.15	3.04 ± 0.27	1.91 ± 0.19	1.79 ± 0.17	2.61 ± 0.23	3.46 ± 0.29
MiCo	1.77 ± 0.14	1.22 ± 0.13	2.74 ± 0.26	1.47 ± 0.16	1.53 ± 0.16	2.25 ± 0.21	2.87 ± 0.27

D.4.3 Results on UK Biobank

Signal level evaluation on UK Biobank. Table 8 reports signal-level NMSE on UK Biobank, where the classical low-order spherical-harmonic basis (SH-2) becomes substantially more competitive with learned methods than on HCP-YA. This is consistent with UKB's reduced angular complexity: with only two shells at $b \in \{1000, 2000\}$ s/mm^2. We omit the 10% sparsity column because UKB's N=111 total directions yield only ~ 10 context directions at that level, which is below the identifiability threshold for several baselines.

Downstream metrics evaluation on UK Biobank. The signal-level competitiveness of SH-2 at b=2000 raises the question of whether this parity is preserved at the level of microstructural parameters, or whether ***MiCo***'s training-time parameter-consistency loss recovers a downstream advantage that voxel-level NMSE alone does not surface. Table 9 addresses this by reporting the median absolute error on DTI- and DKI-derived parameters under the same evaluation protocol.

Table 8: Signal prediction NMSE ($\times 10^{-2}$) on UK Biobank across shells and sparsity levels. Results reported as median over voxels (mean $\pm$ std across runs). Sparsity ($20\%, 30\%$) is the percentage of sampled signal retained as context. **Bold** and underlined denote best and second-best per column.

Method	b=1000 s/mm^2		b=2000 s/mm^2	
	20%	30%	20%	30%
DTI-WLS	2.73 $\pm$ 0.28	1.98 $\pm$ 0.24	3.70 $\pm$ 0.53	2.51 $\pm$ 0.42
SH-2	1.00 $\pm$ 0.11	0.85 $\pm$ 0.13	**1.74** $\pm$ 0.23	**1.27** $\pm$ 0.16
SH-4	N/A	N/A	N/A	N/A
SHORE	1.11 $\pm$ 0.17	0.82 $\pm$ 0.16	1.82 $\pm$ 0.37	1.31 $\pm$ 0.37
MAP-MRI	1.50 $\pm$ 0.29	0.92 $\pm$ 0.23	3.69 $\pm$ 0.30	2.09 $\pm$ 0.33
CFM	1.20 $\pm$ 0.20	0.88 $\pm$ 0.18	2.24 $\pm$ 0.27	2.02 $\pm$ 0.24
I^{2}SB	0.95 $\pm$ 0.19	0.83 $\pm$ 0.12	1.90 $\pm$ 0.18	1.73 $\pm$ 0.14
DISCUS	0.88 $\pm$ 0.17	0.80 $\pm$ 0.13	1.95 $\pm$ 0.21	1.75 $\pm$ 0.19
MiCo	**0.85** $\pm$ 0.14	**0.79** $\pm$ 0.14	1.82 $\pm$ 0.22	1.65 $\pm$ 0.20

Table 9: Downstream microstructure parameter recovery on UK Biobank: median absolute error (↓) on DTI-derived parameters (FA, MD, AD, RD) and DKI-derived parameters (MK, AK, RK), at 20% and 30% context fractions. All values are reported in units of $\times 10^{-2}$. Lower is better. **Bold** and underlined denote best and second-best per column.

Context Fraction 20% (values $\times 10^{-2}$)							
Method	FA ↓	MD ↓	AD ↓	RD ↓	MK ↓	AK ↓	RK ↓
DTI-WLS	9.34	22.79	19.90	24.33	53.13	43.54	58.41
SH-2	6.08	4.65	9.20	5.95	5.09	9.72	9.87
SH-4	6.09	4.65	9.17	5.94	5.08	9.73	9.85
SHORE	7.16	5.59	7.37	8.27	7.87	8.25	14.04
MAP-MRI	4.77	7.70	9.52	8.14	7.85	10.95	9.27
CFM	3.18	3.62	5.99	3.90	3.91	6.13	5.82
I^{2}SB	3.14	3.87	**5.58**	4.30	4.40	6.13	6.42
DISCUS	3.41	3.75	5.78	4.27	3.91	6.21	5.90
MiCo	**2.98**	**3.43**	5.61	**3.83**	**3.48**	**5.52**	**5.45**
Context Fraction 30% (values $\times 10^{-2}$)							
Method	FA ↓	MD ↓	AD ↓	RD ↓	MK ↓	AK ↓	RK ↓
DTI-WLS	7.04	20.50	18.13	21.52	45.50	35.14	52.75
SH-2	5.88	3.67	8.45	5.01	3.93	8.58	8.88
SH-4	5.80	3.68	8.38	4.97	3.82	8.69	8.70
SHORE	6.17	3.82	6.78	6.27	5.41	7.14	11.20
MAP-MRI	3.92	5.44	7.37	6.08	5.93	8.93	7.60
CFM	2.71	**2.77**	5.01	3.19	3.20	5.58	**5.09**
I^{2}SB	2.67	3.15	4.77	3.57	3.67	5.56	5.69
DISCUS	2.75	2.95	4.75	3.34	3.21	5.58	**5.09**
MiCo	**2.61**	2.79	**4.74**	**3.04**	**3.17**	**5.32**	5.15

Sensitivity Analysis on UK Biobank We repeat the λ_{DKI} sensitivity sweep on UKB validation (Table 10) and select $\lambda_{\mathrm{DKI}} = 0.36$. At 20% context, MK error falls from the DTI-source baseline's 4.40 to 3.48×10^{-2} (20.9%); relative to the Gaussian-source baseline's 3.91, the reduction is 11.0%. Appendix D.3.2 relates the benefit to DKI residuals.

Table 10: Joint ablation on UK Biobank, context fraction 20%: source-init prior (Gaussian vs. DTI-fitted) and DKI-consistency loss weight λ_{DKI} (log-uniform sweep over $[0.01, 1.0]$, 10 settings). Median absolute error ($\times 10^{-2}$, $\downarrow$) on DTI- and DKI-derived scalar maps. The DKI-loss strength has the largest effect on the DKI-specific scalars (MK, RK). **Bold** marks the best per column; underline second-best.

λ_{DKI}	DTI-derived				DKI-derived		
	FA	MD	AD	RD	MK	AK	RK
Gaussian source							
0	3.16	3.52	5.96	3.89	3.91	6.15	**5.86**
DTI-fitted source							
0	3.15	3.77	5.60	4.31	4.40	6.13	6.42
0.010	3.08	4.19	5.60	4.81	4.88	5.97	7.17
0.017	3.19	3.74	5.66	4.30	4.43	**5.92**	6.85
0.028	**2.96**	4.13	5.84	4.53	4.83	6.18	6.57
0.046	3.03	**3.42**	5.53	3.84	3.81	6.01	5.97
0.077	3.04	3.63	5.79	3.97	4.56	6.29	6.62
0.129	2.98	4.18	5.67	4.67	5.32	6.27	7.27
0.215	3.30	3.69	**5.46**	4.38	4.27	5.98	6.33
0.359	2.98	3.43	5.61	3.83	**3.48**	5.92	5.95
0.600	2.99	3.92	5.74	4.32	4.79	6.16	6.80
1.000	3.00	3.43	5.61	**3.79**	3.86	6.09	5.89

D.4.4 FOD Reconstruction Quality on HCP-YA

We assess angular fidelity directly via fiber orientation distribution (FOD) reconstruction. For each method's predicted dense signal, we fit single-shell constrained spherical deconvolution [39] (CSD) at $\ell_{\max} = 8$ using the white-matter response function estimated from the test subject, and report the angular correlation coefficient (ACC) between predicted and ground-truth FODs over white-matter voxels. Higher ACC indicates better orientational agreement (see Anderson 2005 [1] for the definition). Table 2 reports the result across three shells and three sparsity levels.

Results analysis. First, ***MiCo*** achieves the best ACC at the most aggressive sparsity (10%) on every shell, with margins of 1–2 ACC points over the strongest baseline — the regime in which the angular prior the network learns is most consequential. Second, at higher sparsity (20%, 30%), classical radial-angular bases (MAP-MRI, SHORE) become competitive on the lower shells, where the increased angular sampling provides enough constraint that explicit basis fitting recovers most of the signal structure. ***MiCo*** is within 1.19 ACC points of the leader in every column, and attains the best ACC on 6 of 9 (shell, sparsity) cells overall.

The advantage in the aggressive-sparsity regime is most evident in the qualitative behavior of the FODs. Fig. 5 compares whole-slice and crossing-region FODs at 10% context against DISCUS, the strongest learned baseline at this sparsity. ***MiCo***'s FODs preserve both the dominant fiber direction and the secondary crossing lobes characteristic of regions like centrum semiovale. DISCUS, by contrast, produces FODs whose primary lobes are correctly oriented in many voxels but whose secondary structure is degraded, consistent with a model that captures dominant angular structure but loses fidelity in the higher-order content that distinguishes crossing fibers from single-fiber voxels. The FOD visualization complements the ACC results by showing the recovered crossing-fiber structure.

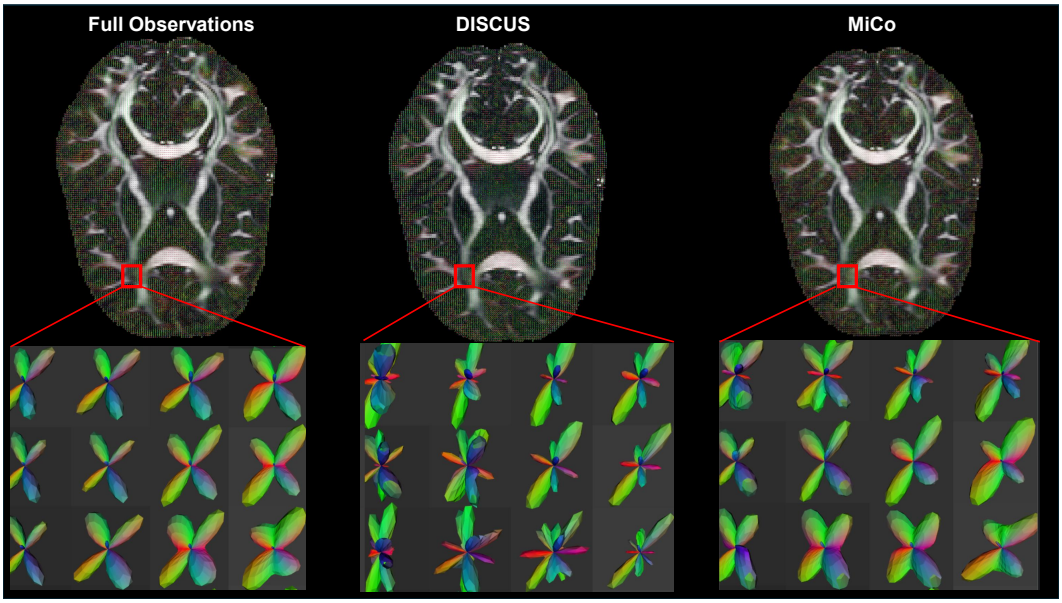

Figure 5: **Qualitative FOD reconstruction on HCP-YA at 10% context fraction**, $b = 2000\,\mathrm{s/mm^2}$. Top row: whole-slice FOD glyphs computed from constrained spherical deconvolution applied to the full 90-direction acquisition (left), DISCUS prediction (middle), and ***MiCo*** prediction (right). Bottom row: zoom into a crossing-fiber region in the centrum semiovale (red box, top row). ***MiCo*** recovers the lobe structure of the reference FODs faithfully, including the secondary lobes characteristic of crossing fibers, while DISCUS exhibits spurious lobe orientations and reduced angular sharpness in the same voxels. Glyphs are colored by direction.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract identifies two contributions, the DTI-based source initialization and the microstructure-consistency loss, both of which are derived in the methods and evaluated in Section 5.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Section 7.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The two main analytical results are stated in Section 3 and Section 4.2. Proofs are given in Appendix B

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide the network architecture, hyper-parameters, preprocessing pipeline , training protocol (optimizer, schedule, EMA, batch size, hardware), and inference settings.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Code will not be released. HCP-YA and UK Biobank are available under their respective data-use agreements and cannot be redistributed by us. Architecture, training, and evaluation details are described in the paper and appendices.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: Training and testing details can be found in the main text and in appendix

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Learned-method signal and scalar summaries report dispersion across three training runs; the FOD table specifies its aggregation separately. Appendix D.3.1 reports the matched numerical comparisons.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix reports training on a single NVIDIA A6000 GPU for approximately 8 hours per run, with batch size 2,048 voxels and bfloat16 mixed precision.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research uses only publicly released, IRB-approved imaging cohorts (HCP-YA, UK Biobank) accessed under their respective data-use agreements, and conforms to the NeurIPS Code of Ethics in all respects.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

Justification: Reducing dMRI scan time at fixed angular fidelity has direct positive impact on patient comfort and clinical accessibility. Reconstruction errors could affect downstream scientific or clinical interpretation.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper releases neither pre-trained models with high misuse potential nor scraped data. The trained model produces diffusion-MRI signal predictions on standard public datasets and poses no dual-use risk

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All datasets (HCP-YA, UK Biobank) and software (QSIPrep, DIPY, PyTorch) are properly cited. HCP-YA and UK Biobank were accessed under their respective Data Use Agreements; QSIPrep is released under the BSD-3-Clause license and DIPY under the BSD-3-Clause license, and we comply with their terms of use.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: No new datasets are released.

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This work uses pre-existing anonymized public neuroimaging datasets (HCP-YA and UK Biobank) and does not involve crowdsourcing or new data collection from human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: We use only existing public datasets (HCP-YA and UK Biobank) whose original studies obtained IRB approval and informed participant consent at the time of acquisition. No new human-subject research is conducted in this work.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were not used as part of the core methodology. They were used only for writing assistance, which does not require declaration under the NeurIPS policy.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.