

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Accelerated motion correction with deep generative diffusion models

Brett Levac¹  | Sidharth Kumar¹ | Ajil Jalal²  | Jonathan I. Tamir¹ 

¹Chandra Family Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas USA

²Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, California USA

Correspondence

Brett Levac, Chandra Family Department of Electrical and Computer Engineering, University of Texas at Austin, TX 78712, USA.

Email: blevac@utexas.edu

Funding information

ARO, Grant/Award Numbers: W911NF2110117, 051242-002; NSF, Grant/Award Number: IFML 2019844; Oracle Research Fellowship; NIH, Grant/Award Number: U24EB029240; Aspect Imaging; Google research scholars program; NSF, Grant/Award Number: CCF2239687(NSFCAREER)

Abstract

Purpose: The aim of this work is to develop a method to solve the ill-posed inverse problem of accelerated image reconstruction while correcting forward model imperfections in the context of subject motion during MRI examinations.

Methods: The proposed solution uses a Bayesian framework based on deep generative diffusion models to jointly estimate a motion-free image and rigid motion estimates from subsampled and motion-corrupt two-dimensional (2D) k-space data.

Results: We demonstrate the ability to reconstruct motion-free images from accelerated two-dimensional (2D) Cartesian and non-Cartesian scans without any external reference signal. We show that our method improves over existing correction techniques on both simulated and prospectively accelerated data.

Conclusion: We propose a flexible framework for retrospective motion correction of accelerated MRI based on deep generative diffusion models, with potential application to other forward model corruptions.

KEYWORDS

deep generative diffusion models, deep learning, motion correction, MRI reconstruction

1 | INTRODUCTION

MRI is a highly effective medical imaging modality which owes much of its utility to having superior soft tissue contrast without any ionizing radiation. Unfortunately, MRI is notoriously slow when compared to other imaging methods. This limitation can lead to increased operating costs

and decreased image quality due to a variety of factors. A common way to reduce scan time is to acquire less data and thus subsample k-space. This process, however, makes the task of image reconstruction an ill-posed inverse problem. To better handle this task, many techniques have been developed such as parallel imaging,^{3–5} handcrafted image regularization,^{6–8} dictionary learning,⁹ subspace constraints,¹⁰ and more recently deep learning.^{11–17}

Although highly subsampling k-space reduces the likelihood of motion occurring during the scan, MRI is still susceptible to subject motion due to physical con-

Portions of this work were presented at IEEE International Symposium on Biomedical Imaging (ISBI) 2023¹ and ISMRM Workshop on Data Sampling and Image Reconstruction 2023.²

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

straints during a given scan such as the repetition time (TR) needed between excitations. The resulting artifacts can often render the image nondiagnostic, and may ultimately require the corrupted scan to be reacquired.¹⁸ The severity of motion artifacts is related to a variety of factors including acquisition parameters (sampling trajectory, echo train ordering, signal preparation) and the degree of motion. See Figure 1 for examples of how these artifacts manifest. These artifacts have tangible costs in clinical settings, especially when scanning pediatric patients where motion artifacts are extremely common.¹⁹ Many approaches to address motion corruption have been proposed. These methods can be separated into two broad categories: prospective and retrospective.

Prospective methods are categorized as those which can be used at scan time to modify the acquisition in response to subject motion. To measure motion during the scan, a variety of approaches have been proposed that either leverage additional pulse sequence actions like motion navigators or external measurement devices such as respiratory bellows, radiofrequency tones, nuclear magnetic resonance probes, and optical tracking.^{20–24} These measurements can be used to correct motion by binning data into different motion states to later inform the reconstruction process, or even discarding corrupted measurements and guiding re-acquisition of corrupted data. Due to the overhead created by additional measurement equipment and re-acquisition of data, these methods may still increase operating costs and even scan time, as well as require modifications to the sequence or addition of peripheral hardware.

Retrospective methods assume no control of the imaging procedure and correct for motion artifacts after measurement data have been collected. This means techniques that require sequence modification are not possible. They also typically assume no access to direct measurements of the true motion states which occurred during the exam. Retrospective techniques are more widely compatible with existing clinical protocols but also face a more difficult task than prospective methods.

In light of recent advancements in the area of deep learning, perhaps the most straightforward approach to retrospectively correct motion is to directly map motion corrupt images to clean images. One line of such approaches successfully trained a conditional generative adversarial network (cGAN) to translate motion corrupt images to clean images.^{25,26} This technique falls into the class of end-to-end deep learning methods. However, as previously stated, artifact appearance is heavily dependent on the chosen forward operator, for example, the sampling trajectory (Figure 1). Due to this, a network's performance at test time is highly dependent on how motion is synthesized at training time to create training pairs, and therefore

end-to-end performance will degrade if changes to the acquisition protocol occur such as sampling trajectory and echo train ordering.²⁷

Motion can be described as an unknown perturbation to the assumed forward model that gives rise to artifacts at reconstruction time. This has led previous works that aim to jointly solve an optimization problem for the target image and the unknown motion that occurred at scan time.^{28–32} These methods have primarily been applied to the low acceleration regime. To build upon joint optimization, supervised learning has been used as one step in a larger iterative algorithm that jointly solves for the image and the motion parameters.³³ Although this method shows notable improvements over prior methods, it is still likely susceptible to distribution shifts in the forward operator (changes to acquisition and sampling parameters), as to train the end-to-end network component it is necessary to preselect the manifestation of the motion artifacts to learn the proper inversion. It also still relies on a linear reconstruction backbone for solving the accelerated reconstruction task which is not as powerful as recent deep learning-based reconstruction techniques, and therefore performance will degrade at higher acceleration factors.

Therefore there is an unmet need to solve accelerated image reconstruction in the presence of unknown and arbitrary motion without requiring retraining for every possible configuration of scan parameters. In pursuit of this goal, in this work we propose a retrospective motion correction technique that builds off of recent advances in deep generative models.^{34,35} In particular, we formulate the reconstruction under the lens of posterior sampling.^{14,16,17} We extend the framework to joint posterior sampling over the image and motion parameters. To demonstrate feasibility of the framework, we restrict our focus to two-dimensional (2D) rigid motion, though extensions to three-dimensional (3D) rigid and nonrigid motion are possible. Our goal is to develop a method that is (1) effective at correcting in-plane, rigid motion from subsampled data while (2) being agnostic to choices in the forward model which can greatly affect the manifestation of the motion artifacts observed. We are also motivated by the recent success of foundation models in other areas of deep learning including vision and language, where a single pretrained model can be used for a variety of tasks.

2 | THEORY

2.1 | Accelerated multicoil MRI reconstruction

The goal of accelerated image reconstruction in MRI is to recover an image $\mathbf{x} \in \mathbb{C}^N$ from subsampled Fourier



FIGURE 1 Retrospective simulation of the impact of motion on reconstruction. (Left) Motion-free image, (Center) simulated motion-corrupt image resulting from Cartesian fast spin-echo, (Right) simulated motion-corrupt image resulting from PROPELLER fast spin-echo. All acquisitions were fully sampled and the same motion states were used for each repetition time (TR). No motion correction was applied.

measurements (k-space) $\mathbf{y} \in \mathbb{C}^M$. We can denote the measurement (forward) process in MRI as

$$\mathbf{y}^{(l)} = N_K S_l \mathbf{x} + \boldsymbol{\eta}_l, \quad (1)$$

where $\mathbf{y}^{(l)}$ is k-space of the l th coil, $N_K \in \mathbb{C}^{M \times N}$ denotes the (possibly nonuniform) Fourier transform operator (2D or 3D) evaluated at coordinates $K \in \mathbb{R}^{d \times M}$, $d = 2, 3$ for 2D and 3D imaging, respectively, $S_l \in \mathbb{C}^{N \times N}$ is the l th coil sensitivity map, and $\boldsymbol{\eta}_l \sim \mathcal{N}(0, \sigma^2 I)$ is additive noise. We can consolidate the forward operator for all coils into one operator $A = N_K S$.

Viewed from the perspective of regularized inverse problems, reconstruction can then be formulated as solving the optimization problem

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda R(\mathbf{x}), \quad (2)$$

where $R(\mathbf{x})$ can be a handcrafted image regularization term such as L1-wavelet sparsity,⁷ or low-rank structure.^{6,8} Reconstruction can also be solved with deep networks by learning a mapping (f_θ) from measurement to image space using training data^{11,15}

$$\mathbf{x}^* = f_\theta(\mathbf{y}). \quad (3)$$

More recently, there has been a push to use deep generative models to learn useful statistical priors for regularization.^{14,16,17,36,37} In these techniques, reconstruction takes on a Bayesian formulation where the goal is to solve the inverse problem with a variety of estimators such as maximum a posteriori estimation, minimum mean square error estimation, or posterior sampling

which have perceptual quality benefits over many other formulations.³⁸

2.2 | Generative diffusion models for inverse problems

Recent work in the deep generative modeling space has been focused on diffusion processes.^{34,35,39,40} For the remainder of this section we will adopt the notation introduced in Reference 39. Generative diffusion models can be understood through viewing two complementary stochastic differential equations (SDE). The first SDE is called the forward process. In the forward process, noise is gradually added to the data distribution of interest:

$$d\mathbf{x} = \frac{s'(t)}{s(t)} \mathbf{x} dt + s(t) \sqrt{2\sigma'(t)\sigma(t)} d\boldsymbol{\omega}. \quad (4)$$

Here $s(t)$ and $\sigma(t)$ are commonly called the signal scaling and noise scaling schedules, respectively. Additionally, $\frac{s'(t)}{s(t)}$ is often referred to as the drift coefficient, while $s(t)\sqrt{2\sigma'(t)\sigma(t)}$ is commonly called the diffusion coefficient, and $\boldsymbol{\omega}$ defines a Brownian motion process. This process can be reversed via a complementary SDE or ordinary differential equation (ODE).^{39–41} We will focus on the reverse ODE which is given by:

$$d\mathbf{x} = \left(\frac{s'(t)}{s(t)} \mathbf{x} - s^2(t) \sigma'(t) \sigma(t) \nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}; \sigma(t)) \right) dt, \quad (5)$$

where $\tilde{\mathbf{x}} = \frac{\mathbf{x}}{s(t)}$. When run from time $t = T$ to $t = 0$ the procedure results in sampling from the original clean data

distribution $p_{\text{data}}(\mathbf{x})$. Here we note that $s(t)$ and $\sigma(t)$ are analytically defined in the forward equation (Equation 4), so the only portion of the reverse ODE which needs to be learned is the score ($\nabla_{\mathbf{x}} \log p(\mathbf{x}; \sigma(t))$) at each time t . The score can be approximated by training a neural network ($D_{\theta}(\mathbf{x}, \sigma(t))$) via denoising score matching.⁴² For clarity in future equations, we note here that $D_{\theta}(\mathbf{x}; \sigma(t))$ is not a direct approximation to the score function but rather is trained to predict the denoised signal at each noise level leading to the following relation with the score function at each time point during the reverse process:

$$\nabla_{\mathbf{x}} \log p(\mathbf{x}; \sigma(t)) = \frac{D(\mathbf{x}; \sigma(t)) - \mathbf{x}}{\sigma^2(t)}. \quad (6)$$

With access to an approximation of the true score function, the reverse ODE can be solved using ODE solvers like Euler's method (first order). To solve inverse problems, we can instead use the following reverse ODE:

$$d\mathbf{x} = \left(\frac{s'(t)}{s(t)} \mathbf{x} - s^2(t) \sigma'(t) \sigma(t) [\nabla_{\mathbf{x}} \log p(\mathbf{y}|\tilde{\mathbf{x}}; \sigma(t)) + \nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}; \sigma(t))] \right) dt \quad (7)$$

Following this ODE, we will be sampling from the posterior distribution $p(\mathbf{x}|\mathbf{y})$. The key issue with this approach is that we only analytically know the form of the likelihood at time $t = 0$ (e.g., $p(\mathbf{y}|\mathbf{x}; \sigma(0)) = \mathcal{N}(A\mathbf{x}, \sigma^2 I)$). Prior works like Diffusion Posterior Sampling⁴³ have approximated the likelihood at intermediate times steps with

$$p(\mathbf{y}|\mathbf{x}(t); \sigma(t)) \approx p(\mathbf{y}|\hat{\mathbf{x}}; \sigma(0)), \quad (8)$$

where $\hat{\mathbf{x}} = E[\mathbf{x}(0)|\mathbf{x}(t)]$ is an estimate of the denoised image at time $t = 0$ and is given by Tweedie's formula⁴⁴ to be

$$E[\mathbf{x}(0)|\mathbf{x}(t)] = \frac{1}{s(t)} (\mathbf{x}(t) + s^2(t) \sigma^2(t) \nabla_{\mathbf{x}} \log p(\mathbf{x}(t); \sigma(t))). \quad (9)$$

$$= \frac{1}{s(t)} (\mathbf{x}(t) + s^2(t) (D_{\theta}(\mathbf{x}(t); \sigma(t)) - \mathbf{x}(t))). \quad (10)$$

This leads to the approximate posterior sampling (PS) inference procedure for solving inverse problems as shown in Algorithm 1.

2.3 | Measurement formation in the presence of motion

We consider motion which is rigid, in-plane, and occurs between readout lines. The assumption that motion does not occur during the readout period is not too restrictive as the readout duration is typically much shorter than the time between readouts. This assumption means that issues such as spin-history effects are not considered. Prior works

Algorithm 1. Diffusion posterior sampling⁴³

```

procedure PS( $D_{\theta}(\mathbf{x}; \sigma(t)), \sigma(t), s(t), t_{i \in 0, \dots, N}, \mathbf{y}$ )
  samplex( $t_0$ )  $\sim \mathcal{N}(\mathbf{0}, \sigma^2(t_0) s^2(t_0) \mathbf{I})$ 
  for  $i \in \{0, \dots, N - 1\}$  do
     $\tilde{\mathbf{x}} = \frac{\mathbf{x}(t_i)}{s(t_i)}$ 
     $\nu_i = \left( \frac{\sigma'(t_i)}{\sigma(t_i)} + \frac{s'(t_i)}{s(t_i)} \right)$ 
     $\zeta_i = \frac{\sigma'(t_i) s(t_i)}{\sigma(t_i)}$ 
     $\mathbf{d}_P = \nu_i \mathbf{x}(t_i) - \zeta_i D_{\theta}(\tilde{\mathbf{x}}; \sigma(t_i))$ 
     $\hat{\mathbf{x}} = \frac{1}{s(t_i)} \left( \mathbf{x}(t_i) + s(t_i)^2 \sigma(t_i)^2 \frac{D_{\theta}(\tilde{\mathbf{x}}; \sigma(t_i)) - \mathbf{x}(t_i)}{\sigma(t_i)^2} \right)$ 
     $\mathbf{d}_L = \nabla_{\mathbf{x}} \|A\hat{\mathbf{x}}(\mathbf{x}(t_i)) - \mathbf{y}\|_2^2$ 
     $\mathbf{x}(t_{i+1}) \leftarrow \mathbf{x}(t_i) + (t_{i+1} - t_i) \mathbf{d}_P + \gamma_{t_i} \mathbf{d}_L$ 
  end for
end procedure

```

have, however, investigated the effects of spin-history for retrospective motion correction.⁴⁵ Under these assumptions we can characterize the effects of rigid body motion (rotation and translation) on k-space measurements using simple Fourier theory. In particular, rotation in image space leads to the same rotation in k-space, while translation in image space causes linear phase shifts in k-space. Both of these effects can be captured in a modified forward operator:

$$\mathbf{y}_i = P_{\phi_i} N_{R_{\theta_i} K_i} S \mathbf{x} + \eta, \quad \eta \sim \mathcal{N}(\mathbf{0}, \sigma^2 I), \quad (11)$$

where $\mathbf{x} \in \mathbb{C}^N$ is the motion-free image, $S \in \mathbb{C}^{N_c \times N \times N}$ contains the N_c sensitivity maps, R_{θ_i} is a rotation matrix for the i th motion state, P_{ϕ_i} is a diagonal matrix implementing a linear phase shift describing the horizontal and vertical translations during the i th motion state, K_i are the coordinates for the intended k-space trajectory during the i th motion state, and $N_{R_{\theta_i} K_i}$ is the Nonuniform Fast Fourier Transform of $S\mathbf{x}$ at the coordinates $R_{\theta_i} K_i$. We note here that although $\mathbf{y}_i$ and $\mathbf{x}$ are linearly related, ϕ_i , θ_i and $\mathbf{y}_i$ are not.

For ease of notation we combine all motion states for the rotation angles, translation distances and intended sampling trajectories into the variables θ , ϕ , and K , respectively:

$$\mathbf{y} = P_{\phi} N_{R_{\theta} K} S \mathbf{x} + \eta, \quad \eta \sim \mathcal{N}(\mathbf{0}, \sigma^2 I). \quad (12)$$

To further simplify this expression we combine all unknown motion parameters (θ , ϕ) into a single variable κ and get the expression

$$\mathbf{y} = A_{\kappa} \mathbf{x} + \eta, \quad \eta \sim \mathcal{N}(\mathbf{0}, \sigma^2 I), \quad (13)$$

where A_{κ} includes all linear operators in Equation (12). We note here that the motion operators (R_{θ} , P_{ϕ}) are the same for all coils. We also assume the coil sensitivity maps

are the same for every motion state, which is not strictly correct as coil maps are a function of the object's position.⁴⁶

For some of the experimentation in this paper we not only assume constant motion states during a single readout but also fixed motion states for each TR. This is not required for our method but it fits with the observation that time between TRs is much longer, in general, than time between readouts within a TR. This is, for example, the case in many fast spin-echo imaging sequences. We wish to note that although we explicitly consider 2D rigid body motion in our forward model formulation, 3D and nonrigid motion can also be modeled as modifications to the forward model. However, nonrigid motion requires parameterization of a deformation field which increases problem complexity. We also note that other parameterized corruptions in the forward process could be modeled in a similar way, for example, to model magnetic field inhomogeneity.

2.4 | Accelerated motion correction with generative diffusion models

As stated above, prior works have shown promising results when using deep generative diffusion models to solve ill-posed inverse problems like subsampled image reconstruction.^{14,16,17} In most prior work, however, the forward model is assumed to be fixed and known throughout the reconstruction procedure. In our work, we assume that our forward operator A_κ belongs to a restricted class of operators with unknown parameters κ which must be learned during inference. Another way of viewing this problem is as posterior sampling from the joint distribution $p(\mathbf{x}, \kappa | \mathbf{y})$. Under this joint posterior, we arrive at the following reverse time ODEs:

$$\begin{aligned} d\mathbf{x} &= \left(\frac{s'(t)}{s(t)} \mathbf{x} - s^2(t) \sigma'(t) \sigma(t) [\nabla_{\mathbf{x}} \log p(\mathbf{y} | \tilde{\mathbf{x}}, \kappa; \sigma(t)) \right. \\ &\quad \left. + \nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}, \kappa; \sigma(t))] \right) dt, \\ d\kappa &= \left(\frac{s'(t)}{s(t)} \kappa - s^2(t) \sigma'(t) \sigma(t) [\nabla_{\kappa} \log p(\mathbf{y} | \tilde{\mathbf{x}}, \kappa; \sigma(t)) \right. \\ &\quad \left. + \nabla_{\kappa} \log p(\tilde{\mathbf{x}}, \kappa; \sigma(t))] \right) dt. \end{aligned} \quad (14)$$

If we assume independence between $\mathbf{x}$ and κ we arrive at the following reverse ODEs:

$$\begin{aligned} d\mathbf{x} &= \left(\frac{s'(t)}{s(t)} \mathbf{x} - s^2(t) \sigma'(t) \sigma(t) [\nabla_{\mathbf{x}} \log p(\mathbf{y} | \tilde{\mathbf{x}}, \kappa; \sigma(t)) \right. \\ &\quad \left. + \nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}; \sigma(t))] \right) dt, \\ d\kappa &= \left(\frac{s'(t)}{s(t)} \kappa - s^2(t) \sigma'(t) \sigma(t) [\nabla_{\kappa} \log p(\mathbf{y} | \tilde{\mathbf{x}}, \kappa; \sigma(t)) \right. \\ &\quad \left. + \nabla_{\kappa} \log p(\kappa; \sigma(t))] \right) dt. \end{aligned} \quad (15)$$

Algorithm 2. General Motion Informed Posterior Sampling

```

procedure GEN-MI-PS( $D_\theta(\mathbf{x}; \sigma(t)), \sigma(t), s(t), t_{i \in 0, \dots, N}, \mathbf{y}$ )
  sample  $\mathbf{x}(t_0) \sim \mathcal{N}(\mathbf{0}, \sigma^2(t_0) s^2(t_0) \mathbf{I})$ 
  sample  $\kappa(t_0) \sim \mathcal{N}(\mathbf{0}, \sigma^2(t_0) s^2(t_0) \mathbf{I})$ 
  for  $i \in \{0, \dots, N-1\}$  do
     $\tilde{\mathbf{x}} = \frac{\mathbf{x}(t_i)}{s(t_i)}$ 
     $\nu_i = \left( \frac{\sigma'(t_i)}{\sigma(t_i)} + \frac{s'(t_i)}{s(t_i)} \right)$ 
     $\zeta_i = \frac{\sigma'(t_i) s(t_i)}{\sigma(t_i)}$ 
     $\mathbf{d}_{P_x} = \nu_i \mathbf{x}(t_i) - \zeta_i D_\theta(\tilde{\mathbf{x}}; \sigma(t_i))$ 
     $\hat{\mathbf{x}} = \frac{1}{s(t_i)} \left( \mathbf{x}(t_i) + s(t_i)^2 \sigma(t_i)^2 \frac{D_\theta(\tilde{\mathbf{x}}; \sigma(t_i)) - \mathbf{x}(t_i)}{\sigma(t_i)^2} \right)$ 
     $\mathbf{d}_{L_x} = \nabla_{\mathbf{x}} \|A_\kappa \hat{\mathbf{x}}(\mathbf{x}(t_i)) - \mathbf{y}\|_2^2$ 
     $\mathbf{x}(t_{i+1}) \leftarrow \mathbf{x}(t_i) + (t_{i+1} - t_i) \mathbf{d}_{P_x} + \gamma_{t_i} \mathbf{d}_{L_x}$ 
     $\mathbf{d}_{P_\kappa} = \nabla_{\kappa} \log p(\kappa; \sigma(t))$ 
     $\mathbf{d}_{L_\kappa} = \nabla_{\kappa} \|A_\kappa \hat{\mathbf{x}}(\mathbf{x}(t_i)) - \mathbf{y}\|_2^2$ 
     $\kappa(t_{i+1}) \leftarrow \kappa(t_i) + (t_{i+1} - t_i) \mathbf{d}_{P_\kappa} + \xi_{t_i} \mathbf{d}_{L_\kappa}$ 
  end for
end procedure

```

From here we note that $p(\mathbf{y} | \mathbf{x}, \kappa; \sigma(0)) \sim \mathcal{N}(A_\kappa \mathbf{x}, \sigma^2 I)$ and we arrive at the final algorithm in Algorithm 2, which we call General motion-informed posterior sampling (MI-PS). The formulation here is intentionally general to allow for arbitrary drift coefficients.

Under the assumption of independence between the image and the motion states, our method has the benefit that the deep generative model can be pretrained for the standard task of posterior sampling reconstruction using motion-free data. Incorporating motion only requires a change to the inference algorithm and does not require retraining the generative model, even if the acquisition parameters change. This is in stark contrast to end-to-end methods which must be trained with specific scan parameters in mind at inference time.²⁷ We find that in practice the independence assumption is not too restrictive.

3 | METHODS

To evaluate our proposed approach, we perform both retrospective simulated motion experiments as well as prospectively accelerated motion-corrupt scans. In the retrospective simulation, we evaluate the impact of changing sampling pattern, echo train length (ETL), and grouping of motion states. In the prospective scanning, we evaluate our approach on a fast spin-echo acquisition without making any modifications to the pulse sequence.

3.1 | Evaluation methods

We investigate a variety of methods that span both non-learned and learned retrospective motion correction. For simulation studies where motion is generated retrospectively, we use upper- and lower-bound comparisons which use either no motion estimate, or exact motion states for reconstruction. These provide worst-case and best-case performance for various methods. In particular, we display the results for seven different methods:

1. **NAMER/TAMER Lower Bound (NT LB):** We perform a linear reconstruction of the motion-corrupt data using the Conjugate Gradient algorithm (CG-SENSE)⁴⁷ without access to the true motion states and we perform no motion estimation. We use this as a lower bound for the performance of methods like NAMER and TAMER^{30,33} which are based on (1) estimating motion parameters and (2) performing a linear reconstruction such as CG-SENSE.
2. **NAMER/TAMER Upper Bound (NT UB):** We perform a CG-SENSE reconstruction of motion-corrupt data with access to the true motion states. We use this as an upper bound for the performance of methods like NAMER and TAMER^{30,33} as in the best case, if NAMER and TAMER perfectly estimate the motion states then the reconstruction will match NT UB.
3. **UNet:** A UNet⁴⁸ model was trained on pairs of motion corrupt and motion free images. During training, the UNet uses as input the CG-SENSE reconstruction of motion corrupted T2 brain images using simulated Cartesian $R = 3$, ETL = 16 motion corrupt data. The output of the network is then compared to the fully sampled and motion-free ground truth image using an $L2$ loss function as in previous work.³³ The model was trained on 14 000 image pairs from fastMRI⁴⁹ and was trained for 10 epochs.
4. **Conditional GAN (cGAN):** A cGAN-based correction method similar to References 25 and 26 was trained on pairs of corrupted and clean images. Like the UNet, the input is the CG-SENSE reconstruction assuming no motion, and the data were simulated Cartesian $R = 3$, ETL = 16. We use UNet⁴⁸ and ResNet-18⁵⁰ models for the generator and the discriminator networks, respectively. To promote better visual quality in the generated images, the training loss includes $L2$, adversarial, and perceptual components through an Imagenet pre-trained VGG model.⁵¹ We observed during training that including the perceptual loss helps in improving the visual quality of final motion-free images. During training the generator is given motion corrupted images, from a given sampling pattern and acceleration level, as inputs and trained to generate images as close as

possible to the clean images. For training the cGAN network we followed the training pipeline from Reference 52, where the generator and discriminator are trained alternatively. The best learning rate was determined heuristically and found to be 0.0001. Both input and output images are normalized with the max value before passing through the network.

5. **Motion informed posterior sampling lower bound (MI-PS LB):** Diffusion-based posterior sampling reconstruction of motion-corrupt data with no motion correction (i.e., assumes zeros for estimated motion parameters). The inference procedure for MI-PS LB is shown in Algorithm 1. This serves as a lower bound on posterior sampling reconstruction as it reflects the result of not accounting for motion in the forward model.
6. **Motion informed posterior sampling (MI-PS):** Diffusion-based posterior sampling reconstruction of both image and motion states using our proposed approach shown in Algorithm 2.
7. **Motion informed posterior sampling upper bound (MI-PS UB):** Diffusion based posterior sampling reconstruction of motion-corrupt data with access to the true motion states throughout the entire inference procedure. The inference procedure for MI-PS UB is found in Algorithm 1. This serves as an upper bound of our proposed approach when motion states are perfectly estimated.

3.2 | Experiments

Our experiments are subdivided into retrospective simulated motion and prospective in vivo scanning.

3.2.1 | Simulated motion

To test the robustness of our method at a variety of acceleration levels and sampling patterns, we simulated motion on T2 brain images from the fastMRI dataset⁴⁹ for two different sampling patterns at three different accelerations and two different ETLs. Specifically, we use Cartesian and PROPELLER⁵³ based sampling patterns each at ETLs of 8 and 16 for accelerations of $R = 3, 4, 5$. All sampling patterns used a readout of 384 points. Example trajectories for each sampling pattern and ETL are shown at $R = 4$ in Figure 2. For each TR we simulate a single independent motion state triplet (rotation, x-translation, y-translation). This means that, for example, the case of Cartesian (or PROPELLER) sampling at $R = 4$ with an ETL = 8 resulted in 12 TRs and thus 12 motion states to estimate along with the corrected image. The motion states for each TR

were sampled independently from a uniform distribution. Following previous work²⁵ translation was sampled from $\mathcal{U}(-2, 2)$ pixels and rotation from $\mathcal{U}(-2, 2)$ degrees. See Figure 2 for an example of simulated motion states for a given (R, ETL) pair.

Prior to simulating motion corruption over the raw k-space data we first resized the fully sampled k-space to be 384×384 and performed inverse Fourier transform to get fully sampled multicoil images. Next we calculated sensitivity maps using ESPIRiT.⁵⁴ We then applied motion to k-space measurements by drawing random motion states and passing the coil images through the motion-corrupt forward operator. Finally, we added noise to the sampled k-space data. We note here that some parts of the preprocessing here constitute an inverse crime.⁵⁵ However, all competing methods used the same data so our method should not have gained an unfair advantage in this respect.

3.2.2 | Grouped versus nongrouped echo trains

To investigate the impact of assuming constant motion throughout entire echo trains we also applied our method without assuming shared motion states for all phase encodes in one TR. For example, if a Cartesian acquisition with ETL = 16 and TR = 8 was collected and we used a grouped (G) motion state prior, for lines in the same TR, we would only estimate eight motion states (one for each TR). If on the other hand we do not assume any specific grouping (NG) of phase encode lines we would instead estimate $ETL \times TR = 128$ lines. We conducted this experiment for all sampling patterns and accelerations listed above.

3.2.3 | Prospective motion-corrupt scanning

We acquired T2 brain scans from three healthy volunteers with institutional review board approval and informed consent. The data were collected on a Siemens Vida 3 Tesla MRI scanner with a 16-channel head coil at our institution, and we emphasize that the scanner hardware and imaging protocol differed from the fastMRI training data. We first collected scans at $R = 3, 4, 5$ for each participant where we asked the participants to stay still during the scan; we used this as motion-free baseline. Next, we collected scans at $R = 3, 4, 5$ where we asked the participants to rotate and translate their head (approximately) in-plane during the scan. Scan parameters were: ETL = 16, slice thickness = 4 mm, FOV = 220 mm $\times$ 220 mm, resolution = 0.57 mm $\times$ 0.57 mm. We applied the NT LB, UNet, cGAN, MI-PS LB, and MI-PS methods to motion-corrupt scans while using NT UB and MI-PS UB on the motion-free scans. For MI-PS, instead of estimating one motion state for each TR

we estimated separate motion states for each phase encode (MI-PS NG).

3.3 | Implementation details

3.3.1 | Training

All model training used 14 000 T2 brain images from the fastMRI brain dataset.⁴⁹ The deep generative diffusion network used a UNet style architecture with two input/output channels for real and imaginary components of the images. The network contained about 65 million trainable parameters. Training followed the same procedure as described in Reference 39. Specifically, we selected the following parameters to define the drift and diffusion coefficients in the forward process: $s(t) = 1$, $\sigma(t) = t$ with $\sigma_{\max} = 5$ and $\sigma_{\min} = 0.002$. The UNet and cGAN methods were trained by simulating motion corrupt-images from a Cartesian trajectory over the full training set with $R = 3$, ETL = 16. We used the Adam optimizer with a learning rate of 0.0003 and 10 epochs. Training for all models was done on a mix of A100, A40, and RTX3090 GPUs.

3.3.2 | Inference

For both PS and MI-PS methods, inference on simulated motion was performed by running the reverse ODE with $N = 300$ inference time steps and a time step schedule of $t_i = \left(\sigma_{\max}^{1/\rho} + \frac{i}{N-1} \left(\sigma_{\min}^{1/\rho} - \sigma_{\max}^{1/\rho} \right) \right)^\rho$ where $\rho = 7$. Similar to Reference 43 we selected the likelihood weighting $\gamma_{t_i} = \frac{1}{\|A_k \hat{x} - y\|_2}$. As motion was simulated, we did not wish to unfairly assume a prior over motion states. Therefore we set $\mathbf{d}_{p_k} = 0$. Finally we used a fixed step size (ξ) for updating motion estimates. For Cartesian and PROPELLER sampling patterns, $\xi = 1$ and $\xi = 0.3$ were used, respectively, when we used grouped echo train assumptions during inference. When we did not assume grouped echo trains $\xi = 10$ and $\xi = 1$ were used for Cartesian and PROPELLER, respectively. We found it best to initialize the motion estimates to zero ($\kappa = 0$). These details lead to the final update procedure shown in Algorithm 3. We recorded inference time for a single slice for the posterior sampling methods.

3.4 | Image quality evaluation

We quantitatively evaluate the retrospective results using normalized root mean squared error (NRMSE) and structural similarity index measure (SSIM) on a test set of 500

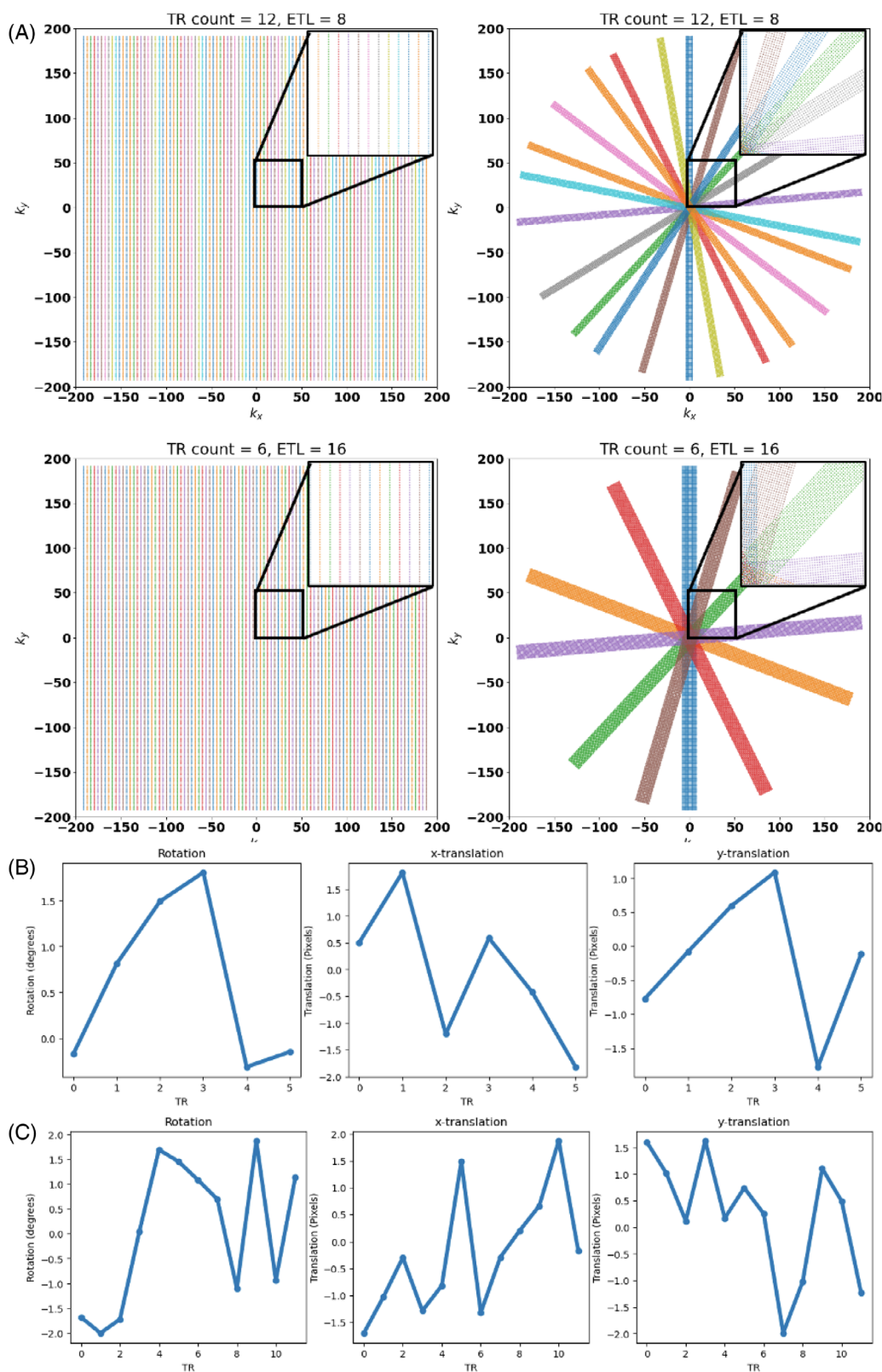


FIGURE 2 (A) For each plot shown the differing colors denote separate repetition times (TRs). For each TR we simulate a single separate motion state triplet (rotation, x-translation, y-translation). (Top Left) $R = 4$ Cartesian sampling trajectory with ETL = 8, number of TRs = 12. (Bottom Left) $R = 4$ Cartesian sampling trajectory with ETL = 16, number of TRs = 6. (Top Right) $R = 4$ PROPELLER sampling trajectory with ETL = 8, number of TRs = 12. (Bottom Right) $R = 4$ PROPELLER sampling trajectory with ETL = 16, number of TRs = 6. Example motion states. (B) motion trajectory for $R = 4$, ETL = 16 (TR count = 6) sampling pattern. (C) Motion trajectory for $R = 4$, ETL = 8 (TR count = 12) sampling pattern.

Algorithm 3. Motion informed posterior sampling (MI-PS)

```

procedure MI-PS( $D_\theta(\mathbf{x}; \sigma(t)), \sigma(t), s(t), t_{i \in 0, \dots, N}, \mathbf{y}$ )
  samplex( $t_0$ )  $\sim \mathcal{N}(\mathbf{0}, \sigma^2(t_0)s^2(t_0)\mathbf{I})$ 
   $\kappa(t_0) = 0$ 
  for  $i \in \{0, \dots, N-1\}$  do
     $\hat{\mathbf{x}} = D_\theta(\mathbf{x}; \sigma(t_i))$ 
     $\mathbf{d}_{P_x} = \frac{\mathbf{x}(t_i) - \hat{\mathbf{x}}}{t_i}$ 
     $\mathbf{d}_{L_x} = \nabla_{\mathbf{x}} \|A_\kappa \hat{\mathbf{x}}(\mathbf{x}(t_i)) - \mathbf{y}\|_2^2$ 
     $\mathbf{d}_{L_\kappa} = \nabla_{\kappa} \|A_\kappa \hat{\mathbf{x}}(\mathbf{x}(t_i)) - \mathbf{y}\|_2^2$ 
     $\mathbf{x}(t_{i+1}) \leftarrow \mathbf{x}(t_i) + (t_{i+1} - t_i)\mathbf{d}_{P_x} + \gamma_{t_i}\mathbf{d}_L$ 
     $\kappa(t_{i+1}) \leftarrow \kappa(t_i) + \xi_{t_i}\mathbf{d}_{L_\kappa}$ 
  end for
end procedure

```

T2 brain images from fastMRI (five slices from 100 different volumes) with simulated motion. We also report peak signal-to-noise ratio in Appendix S1. As there is an ambiguity between two data-consistent reconstructions with a fixed motion offset, we first align the reconstruction to the motion-free ground-truth before evaluating NRMSE, SSIM, and peak signal-to-noise ratio. For the prospective scanning, we qualitatively evaluate the image quality as we do not have a fully sampled motion-free reference image.

4 | RESULTS

We present results comparing retrospective motion correction in both simulation and prospective in vivo scanning. First, we show that MI-PS can reconstruct faithful motion-free images without a prior motion estimate in simulated motion corruption for various sampling trajectories and without requiring retraining. Next, we show in simulations that grouping the motion states based on echo train in fast spin-echo provides appreciable benefits. Finally, MI-PS removes motion artifacts from prospectively accelerated in vivo scans where the subjects were instructed to move during the scan. Inference time for a single slice using our proposed MI-PS technique took approximately 90 s, compared to 80 s for PS.

4.1 | Retrospective simulation

Quantitative NRMSE and SSIM metrics are shown for each simulated motion case in Table 1 with peak signal-to-noise ratio shown in Table S1. Example reconstructions for simulated motion using Cartesian and PROPELLER trajectories are presented in Figures 3 and 4, respectively.

Clearly, reconstruction without accounting for motion (NT LB and MI-PS LB) leads to large error, necessitating the use of motion correction. The NT UB reconstruction, which has access to ground-truth motion parameters, performs poorly due to residual aliasing artifacts at higher accelerations. While the UNet and cGAN methods are able to remove aliasing, the resulting images are heavily smoothed and many high frequency features are lost even for in distribution test images that exactly match the training setting (Cartesian, ETL = 16, $R = 3$). Performance dips even more for these deep learning methods when applied to PROPELLER sampling, likely because of the differences in both aliasing and motion artifact manifestation at test time. Finally, MI-PS is able to handle the acceleration as well as the motion through joint posterior sampling, and its performance is only marginally worse than MI-PS UB which has access to ground-truth motion states. The inference progression across different steps of the MI-PS solver is shown in Figure S1.

Unsurprisingly, using a PROPELLER-based acquisition leads to lower error for NT UB and NT LB, likely due to the presentation of incoherent artifacts from subsampling and from motion. Interestingly, PROPELLER reconstructions are somewhat worse quantitatively for posterior sampling compared to Cartesian acquisition. This may be due to the heavier subsampling of high frequency k-space.

4.1.1 | Effect of grouping motion states

Table 2 shows NRMSE and SSIM for MI-PS when motion states are either grouped by echo train or separately estimated for each phase encode (See Table S2 for peak signal-to-noise ratio results). In the case of Cartesian sampling, there is a clear benefit to grouping motion states by echo train, likely due to the difficulty in estimating motion from high-frequency phase encodes where signal-to-noise ratio is low. In the case of PROPELLER, grouping the motion states is not as crucial, likely because all readouts have points close to the center of k-space which has higher signal-to-noise ratio. Nonetheless, grouping motion states still helps.

Next we show example reconstructions comparing the impact of grouping motion states. Figures 5A and 6A show example reconstructions for Cartesian and PROPELLER acquisitions, respectively, for grouped versus ungrouped motion states. Residual motion artifacts are clearly visible when motion states are not grouped. Additionally, we show the motion estimates for each scenario in Figures 5B and 6B. For Cartesian sampling, the motion estimates are poor except near the center of the echo train (corresponding to the center of k-space). While PROPELLER also

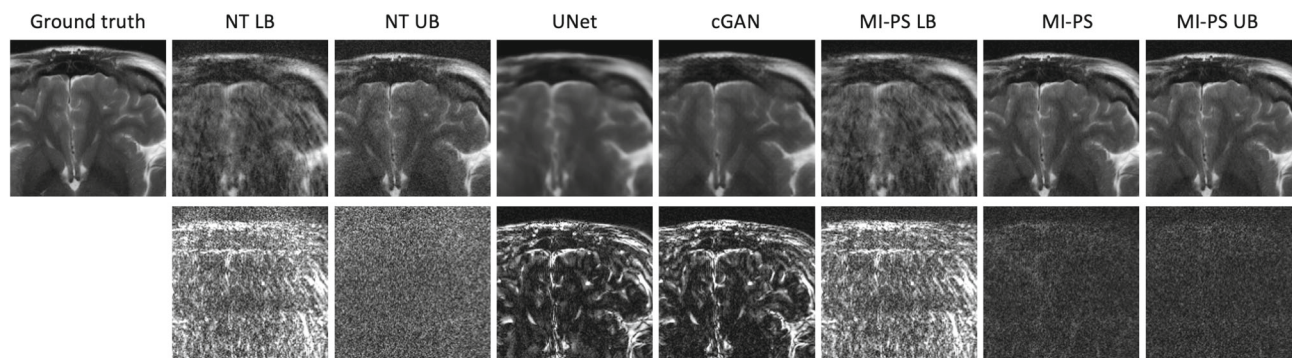


FIGURE 3 Zoom up of example reconstruction of simulated motion for a Cartesian sampling pattern with $R = 3$, $ETL = 16$. From left to right: motion-free fully sampled ground-truth, NAMER/TAMER lower bound, NAMER/TAMER upper bound, UNet, conditional generative adversarial network, motion informed posterior sampling lower bound, motion informed posterior sampling, motion informed posterior sampling upper bound. The bottom row includes difference images which are scaled $5\times$ compared to the top row.

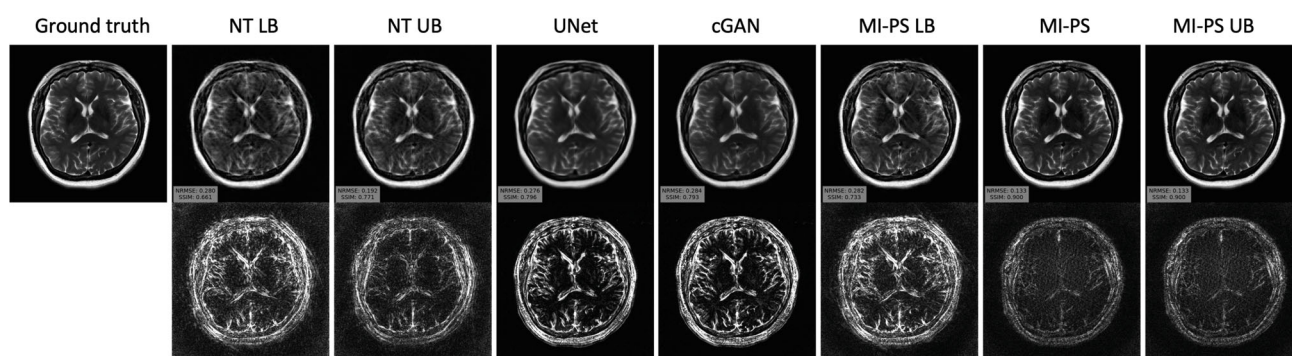


FIGURE 4 Example reconstruction of simulated motion for a PROPELLER sampling pattern with $R = 5$, $ETL = 8$. From left to right: motion-free fully sampled ground-truth, NAMER/TAMER lower bound, NAMER/TAMER upper bound, UNet, conditional generative adversarial network, motion informed posterior sampling lower bound, motion informed posterior sampling, motion informed posterior sampling upper bound. The bottom row includes difference images which are scaled $5\times$ compared to the top row.

over/undershoots motion estimates when motion states are ungrouped, the impact on image quality is small.

4.2 | Prospective in vivo scanning

Finally, we apply each of the reconstruction techniques to the prospectively accelerated scans. For brevity, we show results at $R = 3$ and $R = 5$ on the second subject in Figure 7. We show results for all acceleration factors and for each subject in Figures S2–S4. Even when the subjects stayed still, a CG-SENSE reconstruction (which represents NT UB) was not able to remove all residual aliasing artifacts. A motion-free posterior sampling reconstruction (which represents MI-PS UB), in contrast, performs well at $R = 5$, which is in line with the existing literature.¹⁶ However, when motion does occur during the scan, posterior sampling alone (MI-PS LB) is not able to remove these motion artifacts and thus qualitatively matches NT LB at lower accelerations. At higher accelerations, just as in the

simulation results, MI-PS provides much higher quality images than even NT UB due to the acceleration level and it retains much more of the high frequency detail when compared to the other deep learning techniques (UNet, cGAN). Note that here the motion states were not grouped for MI-PS due to technical reasons and thus results could further improve.

5 | DISCUSSION

Deep generative diffusion models have recently emerged as an extremely powerful advancement for solving inverse problems due to their ability to decouple the measurement model from the prior, which can be well-modeled with deep neural networks. In particular, this means that in contrast to end-to-end deep learning-based inverse problem solvers, deep diffusion priors can be used in a more modular fashion for a variety of imaging problems which vary based on their measurement forward operator

TABLE 1 Quantitative results for simulated motion experiments.

R	NT LB		NT UB		UNET		cGAN		MI-PS LB		MI-PS		MI-PS UB	
	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑
Cartesian, ETL = 8														
3	0.352	0.430	0.212	0.577	0.242	0.817	0.229	0.846	0.289	0.559	0.074	0.897	0.071	0.894
4	0.408	0.347	0.298	0.450	0.266	0.802	0.254	0.830	0.290	0.576	0.085	0.882	0.082	0.878
5	0.458	0.296	0.360	0.432	0.290	0.782	0.279	0.808	0.294	0.579	0.100	0.866	0.095	0.866
Cartesian, ETL = 8														
3	0.354	0.422	0.211	0.580	0.236	0.819	0.228	0.843	0.293	0.552	0.073	0.892	0.071	0.894
4	0.412	0.338	0.298	0.492	0.258	0.805	0.250	0.828	0.294	0.568	0.087	0.872	0.082	0.878
5	0.455	0.290	0.341	0.440	0.278	0.786	0.273	0.807	0.297	0.571	0.096	0.859	0.087	0.872
PROPELLER, ETL = 16														
3	0.284	0.720	0.125	0.865	0.283	0.799	0.289	0.810	0.292	0.771	0.101	0.887	0.099	0.885
4	0.291	0.713	0.158	0.835	0.296	0.789	0.302	0.801	0.299	0.772	0.116	0.887	0.112	0.885
5	0.308	0.689	0.224	0.774	0.320	0.768	0.327	0.780	0.304	0.767	0.162	0.868	0.157	0.870
PROPELLER, ETL = 8														
3	0.286	0.706	0.122	0.874	0.272	0.803	0.280	0.814	0.298	0.706	0.112	0.861	0.110	0.862
4	0.296	0.693	0.148	0.842	0.288	0.793	0.294	0.805	0.308	0.717	0.117	0.873	0.114	0.875
5	0.302	0.684	0.180	0.804	0.299	0.785	0.305	0.797	0.311	0.730	0.127	0.878	0.123	0.880

Abbreviations: cGAN, conditional generative adversarial network; ETL, echo train length; MI-PS, motion-informed posterior sampling; NRMSE, normalized root mean squared error; SSIM, structural similarity index measure.

(e.g., changes to acceleration factor, sampling trajectory, ETL, etc.). In many inverse problem settings, however, the true forward operator used to collect measurements may be unknown as is the case in motion-corrupt MRI scans. This motivates the use of deep generative diffusion models when correcting motion artifacts as it decouples the learning of the image prior from a specific motion artifact pattern that arises due to forward model specifics (i.e., sampling pattern, echo train ordering). In essence, motion estimation comes for free if the generative model was pre-trained for the task of motion-free image reconstruction.

Several prior methods have approached the problem of motion correction by incorporating the parameterization of motion in the forward operator and jointly optimizing over both the image and motion variables.^{29,30,33} In a similar fashion, our proposed approach solves this problem by treating the (unknown) parameterization of the motion as a complementary random variable which must be jointly sampled alongside the clean MR image of interest from motion corrupted measurements (i.e., $\mathbf{x}, \kappa \sim p(\mathbf{x}, \kappa | \mathbf{y})$). This is similar to other recently proposed techniques for solving blind inverse problems.^{56,57} The treatment of the motion correction problem enables us to decouple the training of the prior from the motion correction task which allows our method to be transferable

between differing sampling trajectories. This is extremely important as the manifestation of motion artifacts in the final image is highly dependent on the sampling trajectory used to collect measurements.²⁷ It is this property of motion artifacts that makes it difficult to generalize end-to-end methods to arbitrary motion corruption.

In our simulated experiments and in vivo scans, we see that our method outperforms both prior deep learning methods^{25,26} and the best-case reconstructions for previously investigated joint optimization techniques.^{30,33} This can be owed to the powerful image prior provided by the generative diffusion model which discourages motion estimates that give rise to motion-corrupt images at reconstruction time. Importantly, this property alone is not enough to mitigate motion, as evidenced by the result of diffusion posterior sampling without considering motion (represented by MI-PS LB). In Figures S5 and S6, we also show that L1-Wavelet regularization is similarly insufficient at $R = 5$, even if the subject stays still.

Although we did not use a trained prior to regularize the motion variables, we did use the prior of enforcing all motion states in the same TR to be the same. We showed on prospectively acquired data that we can drop this assumption and still obtain good reconstruction results. However, it is clear from Figure 5 that this may not

TABLE 2 Echo train grouping results.

R	MI-PS LB		MI-PS (NG)		MI-PS (G)		MI-PS UB	
	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑	NRMSE↓	SSIM↑
Cartesian, ETL = 16								
3	0.289	0.559	0.162	0.785	0.074	0.897	0.071	0.894
4	0.290	0.576	0.202	0.743	0.085	0.882	0.082	0.878
5	0.294	0.579	0.225	0.715	0.100	0.866	0.095	0.866
Cartesian, ETL = 8								
3	0.293	0.552	0.156	0.788	0.073	0.892	0.071	0.894
4	0.294	0.568	0.192	0.748	0.087	0.872	0.082	0.878
5	0.297	0.571	0.209	0.726	0.096	0.859	0.087	0.872
PROPELLER, ETL = 16								
3	0.292	0.771	0.129	0.872	0.101	0.887	0.099	0.885
4	0.299	0.772	0.150	0.868	0.116	0.887	0.112	0.885
5	0.304	0.767	0.194	0.850	0.162	0.868	0.157	0.870
PROPELLER, ETL = 8								
3	0.298	0.706	0.132	0.853	0.112	0.861	0.110	0.862
4	0.308	0.717	0.141	0.862	0.117	0.873	0.114	0.875
5	0.311	0.730	0.152	0.866	0.127	0.878	0.123	0.880

Abbreviations: ETL, echo train length; MI-PS, motion-informed posterior sampling; MI-PS LB, motion informed posterior sampling lower bound; MI-PS UB, motion informed posterior sampling upper bound; NRMSE, normalized root mean squared error; SSIM, structural similarity index measure.

always hold for heavier motion corruption. Specifically, in simulated motion corrupt Cartesian data we saw that a more informative prior, for example, by explicitly grouping the phase encodes in a single ETL, improves reconstruction quality due to more accurate estimation of motion parameters at high frequency k-space locations, likely due to the low signal-to-noise ratio at high-frequency k-space regions.

Recently, adaptive end-to-end methods have been proposed under the lens of hybrid networks. In these settings, motion parameters are estimated and used to choose the weights of the neural network that was trained for those weights.⁵⁸ Similarly, unrolled methods that explicitly solve for motion using principled optimization have also been proposed.^{59,60} These methods may provide a solution to out-of-distribution error, though they still must be retrained for different sampling trajectories.

We assumed a flat prior on the motion states, which is quite naive. For example, we do not account for the likely causal transition between one motion state and the next. A stronger prior on the motion, for example, through a Gaussian process or other Markov chains, could be a flexible way to model such time-dependencies. However, this may not be suitable for rapid and sporadic motion. For these reasons, we chose to keep the prior simple. The effects of

the design choice should be explored further to determine the potential benefits of richer motion priors.

Our prospective scan results sheds light on the power of our framework when the training set does not match the test set—both in protocol settings as well as scanner hardware. In essence, it is not necessary to retrain the model for every specific MRI protocol. However, we note that we only scanned a single slice, whereas nearly all MRI protocols will collect multislice data. While in principle this could be handled by our framework, we did not apply it to multislice data. Our framework was also formulated for 2D rigid motion which occurs only between readouts. Therefore, in its current state, our method is not able to model out-of-plane motion, nonrigid motion, or complicated spin-history effects. Our framework could be easily extended to 3D rigid motion estimation by adding additional rotation and translation parameters. Nonrigid motion could also be handled through proper parameterization of the nonrigid space, for example, thorough spline interpolation or optical flow. Deeper consideration will need to be made to account for through plane motion and general spin-history effects as these factors are not modeled in our forward operator parameterization and thus would lead to a measurement inconsistent likelihood function at inference time. We also note that although

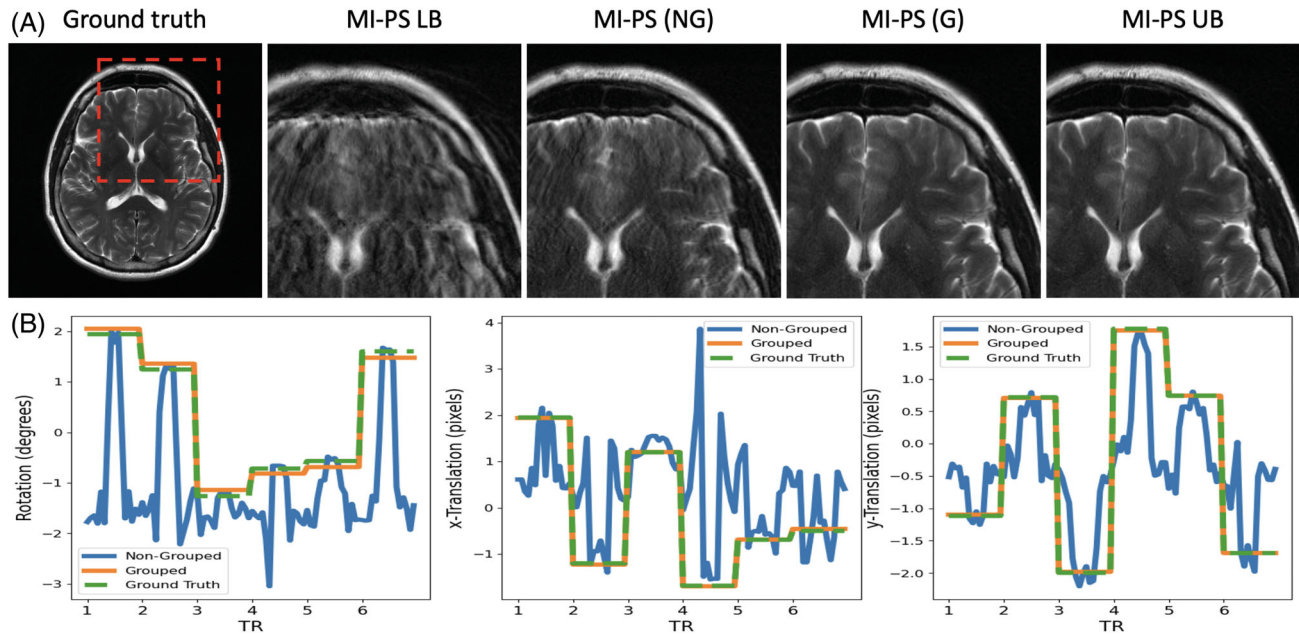


FIGURE 5 (A) Reconstruction comparison between using our method with (G) and without (NG) grouped echo trains for motion estimation at $R = 4$, ETL = 16 Cartesian sampling. (B) Estimated motion comparison between using our method with (G) and without grouped (NG) echo trains.

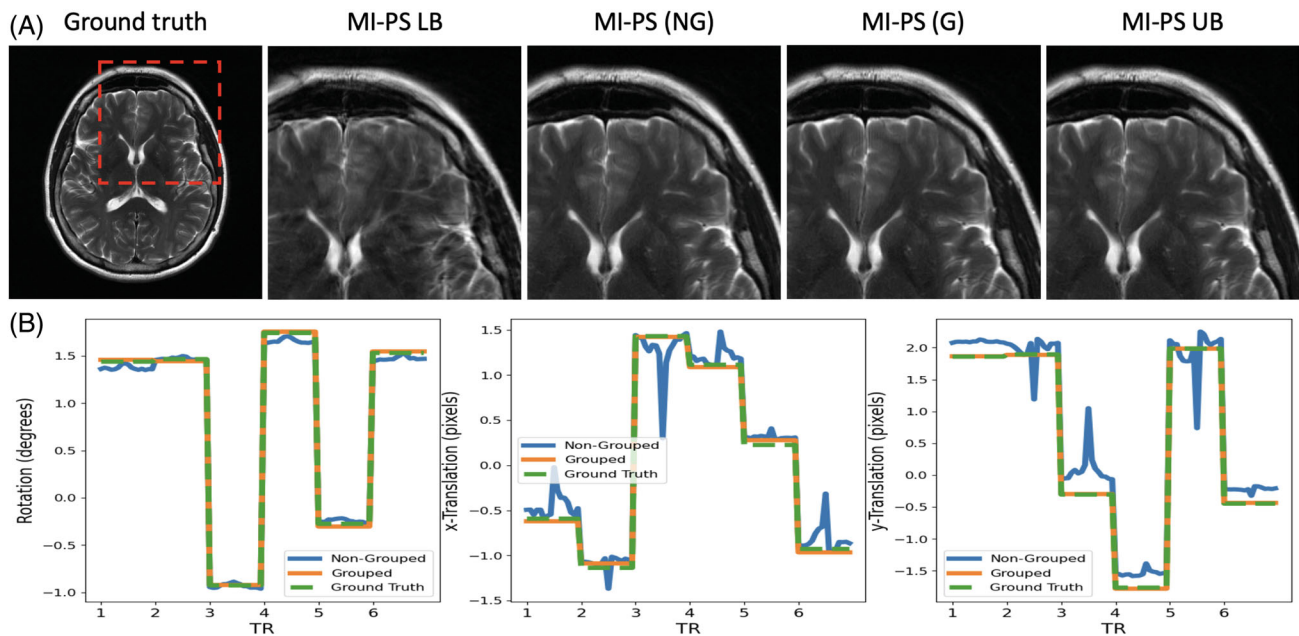


FIGURE 6 (A) Reconstruction comparison between using our method with (G) and without (NG) grouped echo trains for motion estimation at $R = 4$, ETL = 16 PROPELLER sampling. (b) Estimated motion comparison between using our method with (G) and without grouped (NG) echo trains.

our method can be applied to a variety of shifts in the forward operator without retraining (e.g., ETL, sampling trajectory), to apply our method to different image contrasts or anatomies the generative model would need to be retrained for the new target image contrast of interest. This limitation is common to all deep learning approaches, and

could be partially mitigated by increasing the diversity in training data.

As we used the fastMRI dataset for simulation, information about echo train ordering was unavailable. Therefore, a limitation of our simulation results is that the effect of signal decay was not modeled.²⁷ It is possible that the

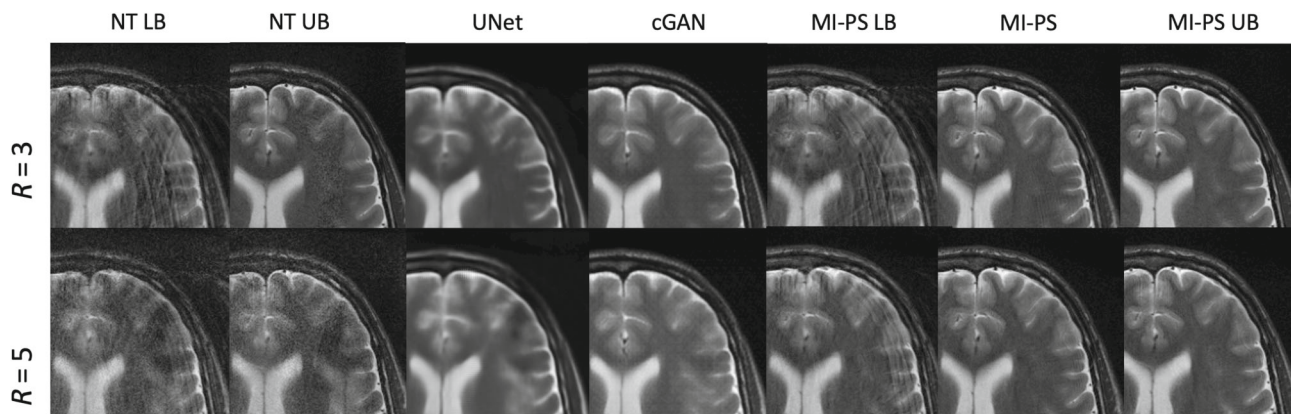


FIGURE 7 Zoomed reconstruction results for prospectively accelerated scan of subject 2 at $R = 3, 5$.

change in contrast due to varying ETLs could lead to a mismatch with our pretrained prior, though this would also affect end-to-end methods similarly. In practice, we found that prospective scanning was relatively unaffected by these mismatches.

Another limitation of our approach is reconstruction time. We found that a single 2D slice required about 90 s of compute on a GPU for reconstruction. The reason for this is twofold: first, the inference procedure requires 300 passes through the deep generative model and also includes an auto-differentiation step as part of diffusion posterior sampling.⁴³ Secondly, the formulation of motion corruption in k -space necessitates the use of a nonuniform Fast Fourier Transform operator, even for the case of Cartesian sampling.

Posterior sampling is a promising approach to image reconstruction from subsampled measurements compared to other point estimators. One reason is for the ability to quantify uncertainty in the result. Similar to conventional posterior sampling, joint posterior sampling over image and motion could also be used to learn uncertainty in the motion parameters. In this work we did not explicitly explore uncertainty in the motion, though it could be interesting for future work.

6 | CONCLUSION

We proposed a method to correct motion artifacts from accelerated MRI by parameterizing motion as inconsistencies in the forward operator which can be jointly estimated as random variables alongside a clean reconstructed MR image. To solve the joint estimation problem we leveraged advancements in deep generative diffusion models to perform motion-informed posterior sampling. We displayed our proposed technique's ability to correct 2D rigid body motion on both simulated and prospectively corrupted scan data.

ACKNOWLEDGMENTS

This work was supported by NSF IFML 2019844, NIH U24EB029240, NSF CCF-2239687 (CAREER), ARO W911NF2110117, 051242-002, the Biomedical Imaging Center (RRID:SCR_021898) at UT Austin, and by research support from Aspect Imaging, Google Research Scholar Program, and Oracle for Research Fellowship.

CONFLICTS OF INTEREST STATEMENT

The authors do not have any conflicts of interest.

DATA AVAILABILITY STATEMENT

Source code to reproduce the results in this work is available at https://github.com/utcsilab/motion_score_mri.

ORCID

Brett Levac  <https://orcid.org/0000-0002-1182-3438>

Ajil Jalal  <https://orcid.org/0009-0006-9244-8575>

Jonathan I. Tamir  <https://orcid.org/0000-0001-9113-9566>

REFERENCES

- Levac B, Jalal A, Tamir JI. Accelerated motion correction for mri using score-based generative models. Paper presented at: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). 2023a:1-5. doi:10.1109/ISBI53787.2023.10230457
- Levac B, Jalal A, Tamir JI. Motion robust reconstruction with score-based generative models. Paper presented at: ISMRM Workshop on Data Sampling and Image Reconstruction. vol. 1. 2023b.
- Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (grappa). *Magn Reson Med*. 2002;47:1202-1210. doi:10.1002/mrm.10171
- Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. Sense: sensitivity encoding for fast mri. *Magn Reson Med*. 1999;42:952-962. doi:10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S
- Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (smash): fast imaging with

- radiofrequency coil arrays. *Magn Reson Med*. 1997;38:591-603. doi:10.1002/mrm.1910380414
6. Haldar JP. Low-rank modeling of local k -space neighborhoods (loraks) for constrained mri. *IEEE Trans Med Imaging*. 2014;33:668-681. doi:10.1109/TMI.2013.2293974
7. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid mr imaging. *Magn Reson Med*. 2007;58:1182-1195.
8. Trzasko JD. Exploiting local low-rank structure in higher-dimensional mri applications. In: Ville DVD, Goyal VK, Papadakis M, eds. *Wavelets and Sparsity XV*. Vol 8858. International Society for Optics and Photonics. SPIE; 2013:885821. doi:10.1117/12.2027059
9. Ravishankar S, Bresler Y. Mr image reconstruction from highly undersampled k -space data by dictionary learning. *IEEE Trans Med Imaging*. 2011;30:1028-1041. doi:10.1109/TMI.2010.2090538
10. Tamir JI, Uecker M, Chen W, et al. T2 shuffling: sharp, multi-contrast, volumetric fast spin-echo imaging. *Magn Reson Med*. 2017;77:180-195. doi:10.1002/mrm.26102
11. Aggarwal HK, Mani MP, Jacob M. Modl: model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2019;38:394-405. doi:10.1109/TMI.2018.2865356
12. Arvinte M, Vishwanath S, Tewfik AH, Tamir JI. Deep j-sense: accelerated mri reconstruction via unrolled alternating optimization. *arXiv:2103.02087*. 2021. doi:10.48550/ARXIV.2103.02087
13. Bora A, Jalal A, Price E, Dimakis AG. Compressed sensing using generative models. In: Precup D, Teh YW, eds. *Proceedings of the 34th International Conference on Machine Learning*. Vol 70. *Proceedings of Machine Learning Research*. PMLR; 2017:537-546.
14. Chung H, Ye JC. Score-based diffusion models for accelerated mri. *Med Image Anal*. 2022;80:102479. doi:10.1016/j.media.2022.102479
15. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated mri data. *Magn Reson Med*. 2018;79:3055-3071. doi:10.1002/mrm.26977
16. Jalal A, Arvinte M, Daras G, Price E, Dimakis AG, Tamir JI. Robust compressed sensing MRI with deep generative priors. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Wortman VJ, eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc; 2021:14938-14954.
17. Luo G, Heide M, Uecker M. MRI reconstruction via data driven markov chain with joint uncertainty estimation. *arXiv preprint arXiv:2202.01479*. 2022.
18. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: a complex problem with many partial solutions. *J Magn Reson Imaging*. 2015;42:887-901.
19. Slipsager JM, Glimberg SL, Søgaard J, et al. Quantifying the financial savings of motion correction in brain mri: a model-based estimate of the costs arising from patient head motion and potential savings from implementation of motion correction. *J Magn Reson Imaging*. 2020;52:731-738.
20. Cheng JY, Zhang T, Ruangwattanapaisarn N, et al. Free-breathing pediatric mri with nonrigid motion correction and acceleration. *J Magn Reson Imaging*. 2015;42:407-420.
21. De Zanche N, Barmet C, Nordmeyer-Massner JA, Pruessmann KP. Nmr probes for measuring magnetic fields and field dynamics in mr systems. *Magn Reson Med*. 2008;60:176-186.
22. Ludwig J, Speier P, Seifert F, Schaeffter T, Kolbitsch C. Pilot tone-based motion correction for prospective respiratory compensated cardiac cine mri. *Magn Reson Med*. 2021;85:2403-2416.
23. Maclaren J, Herbst M, Speck O, Zaitsev M. Prospective motion correction in brain imaging: a review. *Magn Reson Med*. 2013;69:621-636.
24. Nael K, Pawha PS, Fleysher L, et al. Prospective motion correction for brain mri using an external tracking system. *J Neuroimaging*. 2021;31:57-61.
25. Johnson PM, Drangova M. Conditional generative adversarial network for 3d rigid-body motion correction in mri. *Magn Reson Med*. 2019;82:901-910.
26. Usman M, Latif S, Asim M, Lee B-D, Qadir J. Retrospective motion correction in multishot MRI using generative adversarial network. *Sci Rep*. 2020;10:4786.
27. Levac B, Kumar S, Kardoni S, Tamir JI. Fse compensated motion correction for mri using data driven methods. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Springer Nature Switzerland; 2022:707-716.
28. Cordero-Grande L, Ferrazzi G, Teixeira RPAG, O'Muircheartaigh J, Price AN, Hajnal JV. Motion-corrected mri with disorder: distributed and incoherent sample orders for reconstruction deblurring using encoding redundancy. *Magn Reson Med*. 2020;84:713-726. doi:10.1002/mrm.28157
29. Cordero-Grande L, Teixeira RPAG, Hughes EJ, Hutter J, Price AN, Hajnal JV. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Trans Comput Imaging*. 2016;2:266-280.
30. Haskell MW, Cauley SF, Wald LL. Targeted motion estimation and reduction (TAMER): data consistency based motion mitigation for mri using a reduced model joint optimization. *IEEE Trans Med Imaging*. 2018;37:1253-1265.
31. Lingala SG, DiBella E, Jacob M. Deformation corrected compressed sensing (dc-cs): a novel framework for accelerated dynamic mri. *IEEE Trans Med Imaging*. 2015;34:72-85. doi:10.1109/TMI.2014.2343953
32. Rizzuti G, Sbrizzi A, van Leeuwen T. Joint retrospective motion correction and reconstruction for brain mri with a reference contrast. *IEEE Trans Comput Imaging*. 2022;8:490-504. doi:10.1109/TCI.2022.3183383
33. Haskell MW, Cauley SF, Bilgic B, et al. Network accelerated motion estimation and reduction (NAMER): convolutional neural network guided retrospective motion correction using a separable motion model. *Magn Reson Med*. 2019;82:1452-1461.
34. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. 2020. doi:10.48550/ARXIV.2006.11239
35. Song Y, Ermon S. Generative modeling by estimating gradients of the data distribution. Paper presented at: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). 2019; Vancouver, Canada:11918-11930.
36. Luo G, Zhao N, Jiang W, Hui ES, Cao P. Mri reconstruction using deep bayesian estimation. *Magn Reson Med*. 2020;84:2246-2261. doi:10.1002/mrm.28274
37. Tezcan KC, Baumgartner CF, Luechinger R, Pruessmann KP, Konukoglu E. Mr image reconstruction using deep density priors. *IEEE Trans Med Imaging*. 2019;38:1633-1642. doi:10.1109/TMI.2018.2887072
38. Blau Y, Michaeli T. The perception-distortion tradeoff. Paper presented at: Proceedings of the IEEE Conference on Computer

- Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA, 2018, pp. 6228-6237. doi: 10.1109/CVPR.2018.00652
39. Karras T, Aittala M, Aila T, Laine S. Elucidating the design space of diffusion-based generative models. *arXiv:2206.00364*. 2022.
 40. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. *arXiv:2011.13456*. 2021.
 41. Anderson BDO. Reverse-time diffusion equation models. *Stoch Process Appl*. 1982;12:313-326.
 42. Vincent P. A connection between score matching and denoising autoencoders. *Neural Comput*. 2011;23:1661-1674. doi:10.1162/NECO-a-00142
 43. Chung H, Kim J, Mccann MT, Klasky ML, Ye JC. Diffusion posterior sampling for general noisy inverse problems. *arXiv:2209.14687*. 2023.
 44. Efron B. Tweedie's formula and selection bias. *J Am Stat Assoc*. 2011;106:1602-1614. doi:10.1198/jasa.2011.tm11181
 45. Nurdinova A, Ruschke S, Karampinos DC. On the limits of model-based reconstruction for retrospective motion correction in the presence of spin history effects. Paper presented at: Annual Meeting of ISMRM 2022. 2022. Abstract 0862.
 46. Kellman P, Epstein FH, Mcveigh ER. Adaptive sensitivity encoding incorporating temporal filtering(TSENSE). *Magn Reson Med*. 2001;45:846-852.
 47. Pruessmann KP, Weiger M, Börner P, Boesiger P. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn Reson Med*. 2001;46:638-651. doi:10.1002/mrm.1241
 48. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer; 2015:234-241.
 49. Knoll F, Zbontar J, Sriram A, et al. fastMRI: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiol Artif Intell*. 2020;2.
 50. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
 51. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*. 2015.
 52. Kumar S, Saber H, Charron O, Freeman L, Tamir JI. Correcting synthetic MRI contrast-weighted images using deep learning. *Magn Reson Imaging*. 2024;106:43-54.
 53. Pipe JG. Motion correction with propeller mri: application to head motion and free-breathing cardiac imaging. *Magn Reson Med*. 1999;42:963-969. doi:10.1002/(SICI)1522-2594(199911)42:5<963::AID-MRM17i3.0.CO;2-L
 54. Uecker M, Lai P, Murphy MJ, et al. Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magn Reson Med*. 2014;71:990-1001. doi:10.1002/mrm.24751
 55. Shimron E, Tamir JI, Wang K, Lustig M. Implicit data crimes: machine learning bias arising from misuse of public data. *Proc Natl Acad Sci*. 2022;119:e2117203119. doi:10.1073/pnas.2117203119
 56. Chung H, Kim J, Kim S, Ye JC. Parallel diffusion models of operator and image for blind inverse problems. Paper presented at: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023:6059-6069.
 57. Murata N, Saito K, Lai CH, et al. Gibbsddrm: a partially collapsed gibbs sampler for solving blind inverse problems with denoising diffusion restoration. *arXiv preprint arXiv:2301.12686*. 2023.
 58. Singh NM, Dey N, Hoffmann M, et al. Data consistent deep rigid mri motion correction. Paper presented at: Medical Imaging with Deep Learning. 2023. <https://proceedings.mlr.press/v227/singh24a.html>
 59. Pan J, Rueckert D, Küstner T, Hammernik K. Learning-based and unrolled motion-compensated reconstruction for cardiac mr cine imaging. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2022:686-696.
 60. Xu X, Gan W, Kothapalli SV, Yablonskiy DA, Kamilov US. Correct: a deep unfolding framework for motion-corrected quantitative r2* mapping. *arXiv:2210.06330*. 2022.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Table S1. PSNR results for simulated motion experiments.

Table S2. Echo Train Grouping Results (PSNR).

Figure S1. MI-PS reconstruction across different iterations at inference time. The image is initialized to all-noise and the motion states are initialized to all-zero. Over the iterations, the motion states begin to track the true motion (up to a global shift), and the image progressively converges.

Figure S2. Reconstruction results for prospectively accelerated scan of subject 1 at R = 3, 4, 5.

Figure S3. Reconstruction results for prospectively accelerated scan of subject 2 at R = 3, 4, 5.

Figure S4. Reconstruction results for prospectively accelerated scan of subject 3 at R = 3, 4, 5.

Figure S5. Reconstruction of in vivo subject at R = 5 using L1-Wavelet regularization at various regularization strengths when the subject stayed still. Wavelet regularization is not sufficient to remove aliasing artifacts at this acceleration factor.

Figure S6. Reconstruction of in vivo subject at R = 5 using L1-Wavelet regularization at various regularization strengths when the subject was instructed to move during the scan. Wavelet regularization is not sufficient to remove aliasing and motion artifacts at this acceleration factor.

How to cite this article: Levac B, Kumar S, Jalal A, Tamir JI. Accelerated motion correction with deep generative diffusion models. *Magn Reson Med*. 2024;1-16. doi: 10.1002/mrm.30082