RESEARCH ARTICLE

Unified Motion Correction and Reconstruction for Accelerated Zero Echo Time MRI

Yan Chen[1] | Curtis A. Corum[2] | James H. Holmes[3,4] | Mathews Jacob[1]

[1]Electrical and Computer Engineering, University of Virginia, Virginia, U.S.
[2]Champaign Imaging LLC, St. Paul, Minnesota, U.S.
[3]Department of Radiology, University of Iowa, Iowa, U.S.
[4]Biomedical Engineering and Hawk-IDDRC, University of Iowa, Iowa, U.S.

Correspondence
Mathews Jacob
Email: wzr8sn@virginia.edu

Funding Information
The work is supported by R01-AG067078 and R01-EB019961. The data acquisition was supported by R43MH122028 and was acquired on an MRI instrument funded by S10OD025025. The University of Iowa receives research support from GE Healthcare. University of Virginia receives support from Siemens Healthineers and Canon Medical Systems.

arXiv:2611.05169v1 [eess.IV] 3 Nov 2026

Purpose: To develop an accelerated Zero Echo Time (ZTE) MRI reconstruction framework that jointly estimates the 3D image and rigid-motion parameters, while compensating for missing k-space center data and angular undersampling artifacts within a regularized maximum a posteriori (MAP) formulation.

Methods: We formulate the joint recovery as a MAP estimation, termed Energy-based Motion Correction (EnMoCo), which is solved by alternating between k-space rigid-motion updates (trajectory rotation and phase modulation) and conjugate-gradient data-consistency updates. Regularization is provided by a pre-trained Multi-Scale Energy-based model (MuSE), implemented in a data-efficient 2.5D form by combining slice-wise energies from two orthogonal views to reduce 3D training demands.

Results: We demonstrate the effectiveness of the proposed method on in vivo 3D ZTE data and simulated data to replicate more challenging hardware scenarios, including dead time gaps spanning 1.5–5 samples, rotations up to $\pm 17°$, translations up to ± 10 mm, and a range of scan times (153 to 611 seconds). In simulations, the proposed method improves both PSNR and SSIM relative to CG-SENSE and Autofocusing (e.g., $\sim$10–15 dB higher PSNR for gap $\geq$3.5), and reduces rotational motion error in accelerated experiments. Experimental reconstructions show reduced motion blur and inconsistencies in the tissue contrast compared with the same baselines across larger dead time gaps of 3–4.5 samples.

Conclusion: A unified MAP formulation with k-space motion modeling and an explicit energy-based prior can improve reconstruction quality for motion-corrupted, accelerated ZTE relative to standard reconstruction and motion-correction baselines evaluated in this work.

KEYWORDS:
ZTE, motion correction, dead time gap, energy model

WORD COUNT: 5000

Received 26 April 2016; | Revised 6 June 2016; | Accepted 6 June 2016

DOI: xxx/xxxx

1 | INTRODUCTION

Zero echo-time (ZTE) MRI is actively explored[1,2,3,4,5,6] for its near-silent and time-efficient 3D isotropic image acquisition. Silent ZTE relies on radial acquisitions, where gradient directions are incremented between repetitions instead of being ramped up and down repeatedly[7]. Slow gradient increments reduce scan noise compared to regular radial acquisitions, enhancing patient comfort, especially for pediatric subjects and individuals with autism spectrum disorder (ASD), Down syndrome, or dementia. Despite its advantages, ZTE introduces several intrinsic reconstruction challenges. Because the readout gradient is active during RF excitation and no prewinder lobes are used, the acquisition inherently misses k-space samples near the center[5]. Although reducing the readout bandwidth can decrease the dead time gap, it may also introduce chemical-shift artifacts and soft-tissue intensity nonuniformity. These trade-offs motivate high-bandwidth ZTE protocols that rely on algorithmic strategies to recover missing data[8]. Although silent ZTE improves comfort and reduces the likelihood of motion during scanning, the prolonged acquisition can make it challenging for subjects to remain still throughout the scan, particularly for individuals with ASD. Hence, the simultaneous presence of missing central k-space samples and subject motion further complicates accurate ZTE MRI reconstruction.

Traditionally, missing k-space samples are filled using additional acquisitions designed to sample the central k-space region (e.g., WASPI[9], PETRA[10], and HYFI[11]). A key limitation of such hybrid strategies is that data from the center of k-space may differ from data at the outer regions collected using the ZTE acquisition, due to either motion or magnetization preparation state. To avoid separate central acquisitions, several purely computational infilling methods have been proposed[12,13,14,15,16]. These approaches exploit finite image support[17,13,14] and/or parallel imaging[18] to estimate the missing samples directly from the acquired data. However, these approaches are proposed for motion-free settings. Applying these methods to motion-corrupted ZTE data poses additional challenges due to inconsistency across spokes. Moreover, some infilling strategies, such as Low Rank[15], are computationally expensive when the dead time gap is large, making them difficult to integrate with iterative motion estimation and correction.

A common approach for motion estimation is to recover low-resolution navigator volumes from the central k-space signal[19,20,21,22], followed by the estimation of rigid motion parameters from the navigator volumes[20] by image registration in the image domain. In addition to being computationally expensive, approaches using image domain navigators have a trade-off: smaller bins improve the temporal resolution of motion estimates, while larger bins are often needed to produce high-SNR navigators for accurate registration and reliable motion parameters. An alternative approach is Autofocusing (AF)[23,24,25], where rigid motion parameters are iteratively adjusted to optimize an image sharpness metric (e.g., entropy) or data consistency that exploits parallel imaging redundancy[26,27,28,29,30,31]. These approaches are not directly applicable for accelerated ZTE, as dead time gaps and acceleration enlarge the null-space of the encoding operator, and because motion and undersampling errors are strongly coupled, making motion-compensated recovery ill-posed. For ZTE motion correction, MERLIN[4] adopts registration between navigator images using coil sensitivity maps provided by WASPI, which resolves the missing samples issue using an additional low-resolution scan and is constrained by the SNR of navigator images.

Regularized maximum a posteriori (MAP) reconstruction provides a natural framework for ZTE, where multiple degradations can be handled jointly, while a strong image prior can make the problem well-posed. While learning-based priors have recently improved MRI reconstruction, many of the widely used formulations are difficult to deploy in the 3D ZTE setting. For example, end-to-end unrolled architectures[32,33,34] can be memory-demanding for 3D radial applications, because reconstruction cannot be decomposed into independent 2D slices. Plug-and-play reconstruction with pre-trained denoisers[35,36] alleviates some of these practical constraints, but convergence guarantees often require conservative assumptions on the denoiser (e.g., Lipschitz continuity or nonexpansiveness), which may reduce reconstruction quality[37,38]. Although self-supervised approaches do not require training data, their reconstruction process is often time-consuming[39,40]. Custom-trained deep learning models for motion compensation[41,42] have also been developed using paired motion-free and motion-corrupted images or k-space data[43,28,44]. Recently, diffusion-based approaches have also been proposed for 2D image recovery and motion correction[45,46], using diffusion posterior sampling (DPS) together with motion compensation applied to intermediate noisy latents. Translating these benefits to ZTE is challenging, primarily due to limited training data and the substantial computational burden of diffusion inference in 3D.

We develop a regularized reconstruction framework for 3D ZTE acquisitions affected by (i) a missing k-space center, (ii) subject motion, and (iii) undersampling from accelerated scans. We cast image recovery as a

regularized MAP estimation problem, solving for the image jointly with nuisance variables, including rigid-motion parameters and unmeasured k-space samples (both the central gap and undersampled locations), within a unified optimization scheme. For regularization, we use the Multi-Scale Energy-based model (MuSE)[38], which provides an explicit energy (negative log-prior) parameterized by a compact CNN trained via multi-scale denoising score matching. The MuSE prior used in this work is algorithmically unrelated to MUSE (Multiplexed Sensitivity-Encoding[47]) and AMUSE[48], which are SENSE-based reconstruction methods for multishot diffusion MRI. The explicit prior distinguishes MuSE from plug-and-play approaches based on implicit denoisers[35,36,41] and allows the use of optimization algorithms with established convergence properties under standard assumptions[37]. We compute a MAP estimate using a fast majorization-minimization strategy as in 38. From a practical standpoint, MuSE-based MAP inference is typically less computationally demanding than diffusion posterior sampling methods, which can make it more tractable for 3D settings. In addition, the compact CNN prior can be trained with relatively limited data, which is helpful in our constrained-data regime. To further reduce training requirements, we adopt a 2.5D construction in which the 3D prior is approximated as a sum of slice-wise energies computed from two orthogonal views. For motion, we model rigid-body effects directly in k-space through trajectory rotations and the corresponding phase modulations, enabling motion compensation within the data-consistency step rather than relying on reconstructions from binned navigator images. This approach is intended to reduce dependence on binning choices and the associated trade-off between temporal resolution and navigator fidelity. The resulting algorithm alternates between motion-parameter estimation, data-consistency updates, and MuSE-based regularization. We term the proposed method EnMoCo (**En**ergy-based **Mo**tion **Co**rrection).

2 | METHODS

We propose a regularized MAP reconstruction framework for 3D ZTE that jointly estimates the image, rigid-motion parameters, and unmeasured k-space samples. The method uses an explicit MuSE energy prior[38] in a data-efficient 2.5D form, alternating majorization-minimization image updates with k-space motion compensation.

2.1 | Forward Model with Rigid Motion

We denote the multichannel MRI measurements of an image $\boldsymbol{x}$ at a time τ and the k-space location $\mathbf{k}_\tau$ as

$$\boldsymbol{y}_\tau = \mathcal{F}_{\mathbf{k}_\tau}\ \boldsymbol{x} + \boldsymbol{n}_\tau, \tag{1}$$

where $\mathcal{F}_{\mathbf{k}}$ denotes the three-dimensional Fourier transform with coil sensitivity weighting. Here, $\boldsymbol{n}_\tau$ and $\boldsymbol{y}_\tau$ represent the noise and measurements in k-space, respectively.

We assume the brain to be a rigid body and model the motion at time τ by $\boldsymbol{\theta}_\tau$ and $\boldsymbol{t}_\tau$, which are the three rotation and three translation parameters. The measured k-space data at a specific time point τ thus correspond to the Fourier transform of the rotated and translated version $\boldsymbol{x}'_\tau$, where

$$\boldsymbol{x}'_\tau = \mathbf{R}_{\boldsymbol{\theta}_\tau}\boldsymbol{x} - \boldsymbol{t}_\tau. \tag{2}$$

We concatenate the rotation parameters $\boldsymbol{\theta} \in \mathbb{R}^{T\times 3}$ and translation parameters $\boldsymbol{t} \in \mathbb{R}^{T\times 3}$ along the feature dimension to obtain the joint motion parameter matrix $\boldsymbol{\xi} \in \mathbb{R}^{T\times 6}$ across all T time points, as shown below:

$$\boldsymbol{\xi} = \begin{bmatrix}\boldsymbol{\theta}\ \boldsymbol{t}\end{bmatrix} \in \mathbb{R}^{T\times 6}. \tag{3}$$

Using the properties of the Fourier transform, the operation $\mathcal{F}_{\mathbf{k}_\tau}\boldsymbol{x}'_\tau$ can be expressed as:

$$\mathcal{F}_{\mathbf{k}_\tau}\ \boldsymbol{x}'_\tau = \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}_\tau}\boldsymbol{x}. \tag{4}$$

This reformulation allows us to model the k-space signal of the rotated image $\boldsymbol{x}'_\tau$ as the modified forward model $\boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}_\tau}$ applied to the static image $\boldsymbol{x}$. In particular, a rotation $\mathbf{R}_\theta$ in the image domain corresponds to the rotation by the same angle in k-space. Similarly, a phase modulation in k-space accounts for a translation $\boldsymbol{t}$ in the image domain. The above properties allow us to account for motion exclusively in k-space, without having to generate intermediate images.

2.2 | Joint Recovery of Motion & Image

We propose to jointly estimate the image $\boldsymbol{x}$ and the 3D motion parameters $\boldsymbol{\xi}$ by minimizing the cost function $\mathcal{L}(\boldsymbol{x};\boldsymbol{\xi})$:

$$\mathcal{L}(\boldsymbol{x};\boldsymbol{\xi}) \ = \ \tfrac{1}{2\sigma^2}\left\|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}\right\|_2^2 \ + \ \mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}), \tag{5}$$

where $\mathcal{R}_{\boldsymbol{w}} : \mathbb{C}^{M_x\times M_y\times M_z} \to \mathbb{R}$ is an image domain regularizer, represented by the 3D MuSE[38], and σ^2 is the variance of the measurement noise. To train with limited data, we adopt a 2.5D strategy, where the 3D prior is the sum of the MuSE energies of the 2D slices. The motion parameters $\boldsymbol{\xi}$ are estimated simultaneously during the reconstruction of the image volume $\boldsymbol{x}$.

2D MuSE regularizer

Let $\boldsymbol{\nu} \in \mathbb{C}^{M_x \times M_y}$ denote a 2D slice in the 3D volume. MuSE models the prior distribution of $\boldsymbol{\nu}$ as

$$p_{\boldsymbol{w}}(\boldsymbol{\nu}) \;=\; \frac{1}{Z_{\boldsymbol{w}}} \exp\big(-r_{\boldsymbol{w}}(\boldsymbol{\nu})\big), \tag{6}$$

where $Z_{\boldsymbol{w}}$ is the partition function (normalization constant), $r_{\boldsymbol{w}}(\boldsymbol{\nu})$ denotes the energy of a 2D image $\boldsymbol{\nu}$, and $\boldsymbol{w}$ denotes the parameters of the energy model $r_{\boldsymbol{w}}(\cdot) : \mathbb{C}^{M_x \times M_y} \to \mathbb{R}$. The energy is parameterized as[38]

$$r_{\boldsymbol{w}}(\boldsymbol{\nu}) \;=\; \frac{1}{2\varsigma_0^2} \big\|\boldsymbol{\nu} - \mathcal{D}_{\boldsymbol{w}}(\boldsymbol{\nu})\big\|_2^2, \tag{7}$$

where $\mathcal{D}_{\boldsymbol{w}}(\cdot) : \mathbb{C}^{M_x \times M_y} \to \mathbb{C}^{M_x \times M_y}$ is a compact DRUNet-style CNN, and $\varsigma_0 > 0$ is a fixed small constant.

MuSE learns $\boldsymbol{w}$ using multi-scale denoising score matching. In particular, for noise levels ς drawn from a prescribed schedule and Gaussian perturbations $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, the parameters are obtained by minimizing

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \mathbb{E}_{\varsigma}\big(\mathbb{E}_{\boldsymbol{\nu}}\mathbb{E}_{\boldsymbol{z}}\big[\|\nabla_{\boldsymbol{\nu}} \mathbf{r}_{\boldsymbol{w}}(\boldsymbol{\nu} + \varsigma\boldsymbol{z}) - \varsigma\boldsymbol{z}\|_2^2\big]\big), \tag{8}$$

which encourages $\nabla_{\boldsymbol{\nu}} r_{\boldsymbol{w}}$ to match the Gaussian corruption process across multiple noise scales. Training across multiple ς values has been reported to improve robustness during optimization[38]. Consequently, the learned CNN $\mathcal{D}_{\mathbf{w}}(\cdot)$ robustly removes noise across multiple acceleration rates and varying degrees of blur.

2.5D regularization using the 2D model

Let $\boldsymbol{x} \in \mathbb{C}^{M_x \times M_y \times M_z}$ denote the 3D volume. We first define an energy by summing MuSE slice energies along the z direction:

$$\mathcal{R}^z_{\boldsymbol{w}}(\boldsymbol{x}) \;=\; \sum_{s_z=1}^{M_z} \mathbf{r}_{\boldsymbol{w}}\big(P_{s_z}\boldsymbol{x}\big), \tag{9}$$

where P_{s_z} is a linear slice selection operator to extract the x–y slice at index $z = s_z$. $\mathcal{R}^z_{\boldsymbol{w}}$ denotes the energy of a 3D volume as a summation of 2D energies along z direction, $\mathcal{R}^z_{\boldsymbol{w}}(\cdot) : \mathbb{C}^{M_x \times M_y \times M_z} \to \mathbb{R}$. This construction allows training to leverage all 2D slices extracted from each 3D volume, which can reduce the demand for fully independent 3D training examples.

To mitigate view-dependent artifacts (i.e., artifacts in planes orthogonal to the training orientation), we form a 2.5D prior by averaging energies computed from two complementary views and obtain $\mathcal{R}_{\boldsymbol{w}}(\cdot) : \mathbb{C}^{M_x \times M_y \times M_z} \to \mathbb{R}$:

$$\mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}) \;=\; \frac{1}{2}\Big(\mathcal{R}^z_{\boldsymbol{w}}(\boldsymbol{x}) + \mathcal{R}^y_{\boldsymbol{w}}(\boldsymbol{x})\Big), \tag{10}$$

where $\mathcal{R}^y_{\boldsymbol{w}}(\boldsymbol{x}) = \sum_{s_y=1}^{M_y} r_{\boldsymbol{w}}(P_{s_y}\boldsymbol{x})$ sums energies of x–z slices extracted at $y = s_y$.

2.3 | Maximum-a-posteriori (MAP) optimization

Given the prior $\mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x})$, we utilize the Majorization-minimization (MM) optimization scheme to solve the cost function in Eq. (5).

$$\begin{aligned}\{\boldsymbol{x}^{n+1}, \boldsymbol{\xi}^{n+1}\} = \arg\min_{\boldsymbol{x},\boldsymbol{\xi}} \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}\|_2^2 + \mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}^n) \\ + \nabla_{\boldsymbol{x}}\mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}^n)(\boldsymbol{x} - \boldsymbol{x}^n) + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}^n\|_2^2,\end{aligned} \tag{11}$$

where L is the Lipschitz constant of the score network, and n is the index of the iteration. MM reformulates the optimization into an alternating minimization scheme, which alternates between motion update using steepest descent:

$$\boldsymbol{\xi}^{n+1} \;=\; \boldsymbol{\xi}^n - \alpha\nabla_{\boldsymbol{\xi}} \tfrac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^n}\boldsymbol{x}^n\|_2^2, \tag{12}$$

and image update using the Conjugate Gradient (CG) algorithm:

$$\boldsymbol{x}^{n+1} \;=\; \left(\frac{\boldsymbol{\mathcal{A}}^H_{\boldsymbol{\xi}^{n+1}}\boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^{n+1}}}{\sigma^2} + L\boldsymbol{I}\right)^{-1} \left(\frac{1}{\sigma^2}\boldsymbol{\mathcal{A}}^H_{\boldsymbol{\xi}^{n+1}}\boldsymbol{y} + L\hat{\mathbf{x}}^n\right), \tag{13}$$

where $\hat{\mathbf{x}}^n$ is the denoised version of $\boldsymbol{x}^n$:

$$\hat{\mathbf{x}}^n = \left(\boldsymbol{x}^n - \frac{1}{L}\nabla_{\boldsymbol{x}}\mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}^n)\right), \tag{14}$$

obtained by subtracting the MuSE score. The algorithm implicitly recovers missing samples in the center k-space.

We note from (12) that the motion parameters $\boldsymbol{\xi}$ are estimated only using the measured k-space samples and the forward model that depends on the motion parameters; it does not depend on the missing central k-space samples, and hence the algorithm may not benefit from clever infilling strategies. The alternation between (13) and (14) results in the infilling of the missing k-space regions and suppresses noise artifacts.

Multi-resolution MoCo

Prior work suggests that rigid motion can often be estimated reliably from low-resolution images[29]. Because the central (low-frequency) k-space region is relatively densely sampled, low-resolution reconstructions derived from these samples tend to be less affected by aliasing and high-frequency inconsistencies than full-resolution images. We use a coarse-to-fine (multi-resolution) motion-estimation strategy. Specifically, we introduce a k-space mask $\mathbf{M}_s$ that initially retains only a central low-resolution region, which gradually expands to include higher spatial frequencies as the optimization progresses. The motion update in (12) is modified to

$$\boldsymbol{\xi}^{n+1} \;=\; \boldsymbol{\xi}^n - \alpha\nabla_{\boldsymbol{\xi}} \tfrac{1}{2\sigma^2} \big\|\mathbf{M}_s\big(\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}^n\big)\big\|_2^2. \tag{15}$$

We start with a k-space mask that corresponds to a resolution of 16 mm and progressively (16 mm, 4 mm, and 1 mm) enlarge it until it covers the complete sampled region. Importantly, masking is applied only during the motion-parameter update; the image update still uses the full data-consistency step in (13). Fig. 1 shows the full optimization pipeline.

Algorithm 1 EnMoCo

Inputs: measurements $\boldsymbol{y}$, forward operator $\boldsymbol{\mathcal{A}}$, pre-trained $\mathcal{R}_{\boldsymbol{w}}$, other hyper-parameters
Outputs: image $\boldsymbol{x}$, motion parameters $\boldsymbol{\xi}$

Initialize:
$\boldsymbol{\xi} \leftarrow \mathbf{0}$
$\boldsymbol{x} \leftarrow \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}\|_2^2 \; + \; \lambda\|\boldsymbol{x}\|_2^2$

for scale $\in \{16, 4, 1\}$ mm **do**
  get mask $\boldsymbol{M}_s$ of *scale*
  for $n = 1$ to N **do**
    $\boldsymbol{\xi}^n \leftarrow \arg\min_{\boldsymbol{\xi}} \frac{1}{2\sigma^2} \|\boldsymbol{M}_s(\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}^{n-1})\|_2^2$
    repeat
      $\hat{\mathbf{x}}^n \leftarrow \left(\boldsymbol{x}^{(n-1)} - \frac{1}{L}\nabla_{\boldsymbol{x}}\mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}^{(n-1)})\right)$
      $\boldsymbol{x}^n \leftarrow \left(\frac{\boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^n}^H \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^n}}{\sigma^2} + L\boldsymbol{I}\right)^{-1} \left(\frac{1}{\sigma^2}\boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^n}^H \boldsymbol{y} + L\hat{\mathbf{x}}^n\right)$
    until convergence
  end for
end for
return $\boldsymbol{x}$, $\boldsymbol{\xi}$

2.4 | Experiments

2.4.1 | ZTE Acquisition

We used a magnetization prepared Zero Echo-time (MP-ZTE) sequence to acquire the T_1-weighted images; the sequence is illustrated in Fig. S2 of the supplementary material. Imaging was performed on a 3T clinical MRI (Signa Premier, GE Healthcare, Waukesha, WI) with a 48-channel receive array head coil (GE Healthcare). Acquisition parameters for the ZTE-based inversion-recovery scan were 25.6 cm isotropic field of view, matrix size=256, readout bandwidth=±31.25 kHz/FOV, and flip angle=5 degrees. T_1-weighted image contrast was achieved using intermittent magnetization preparation pulses at the start of each segmented readout block, each consisting of 192 angularly optimized ZTE free-induction decay radial lines using HEALPix radial view ordering[49]. The delay between the inversion pulse and the first excitation was 1058.9 ms, followed by 192 readouts of TR=2.32 ms, and an 800 ms delay to allow the longitudinal magnetization to recover before the next inversion. Spokes from all RF excitations were equally weighted in the reconstruction.

Each excitation radiofrequency pulse was followed by the readout block, with the minimum possible delay. Accounting for hardware switching delays (RF ring-down, coil un-blanking, and signal stabilization) alongside half the duration of the excitation pulse, these constraints combined to a missing k-space radius of ∼2.5 readout samples at a readout bandwidth of ±31.25 kHz/-FOV. Data were acquired from seven normal volunteers who provided informed consent under this IRB-approved study. We acquired one motion-free and one motion-corrupted scan with controlled head motion from each subject. For one of the subjects, two additional motion-corrupted scans were acquired, resulting in four motion levels for that subject. In total, 16 datasets were used for testing. We used two acquisition settings: 406 s scans with 170 IR segments and 611 s scans with 256 IR segments. Four datasets from two subjects used the 406 s acquisition, while the remaining datasets used the 611 s acquisition. To evaluate the impact of acceleration, retrospective undersampling was performed on the 611 s acquisitions by retaining partial blocks, e.g., 128 and 64 blocks corresponding to acquisition times of 306 and 153 seconds, respectively. Additionally, we acquired eight motion-free datasets from eight subjects for training.

2.4.2 | Pre-processing

In this work, we aimed to recover motion parameters ξ of dimension $T \times 6$, where T denotes the number of inversion blocks. We neglected the motion within each inversion block in this study. We used the Stoch–Olejniczak approach[14,13] to infill 2.5 samples, from which the coil sensitivity maps were subsequently estimated using JSENSE. The reference image was reconstructed using TV regularization with a weight of 0.002. We simulated gap sizes of 1.5–5 in our experiments, which cover the large range reported in the literature[15].

2.4.3 | Simulated 3D ZTE Data

A simulated ZTE dataset derived from MPnRAGE[50] was used for quantitative validation. We assumed a multichannel non-uniform Fourier transform model mimicking the ZTE acquisition, where Gaussian noise was added to simulate data with four different signal-to-noise ratios (SNR): S1=24.4 dB; S2=19.8 dB; S3=14.0 dB; S4=10.1 dB. Center k-space samples were retrospectively removed to create a range of discrete dead time gaps. Three scan durations were used, where we retained

256, 128, and 64 inversion blocks with acquisition times of 611, 306, and 153 seconds, respectively. The motion parameters used for simulation are shown in Fig. 4 (A).

The proposed method assumed the motion parameters to be the same for all spokes within an inversion block. To assess intra-block motion from continuous movement, an additional simulation was performed. We generated the motion-corrupted data by assuming that the motion parameters were linear across spokes at two different motion levels as shown in Fig. 6 . The proposed block-wise motion estimation was applied to both profiles to evaluate image quality degradation.

2.4.4 | Network Architectures and Implementation Details

We used a 2D DRUNet model $\mathcal{D}_{\boldsymbol{w}}$ with 64, 128, 256, and 512 channels for four levels, respectively. The model was trained with normalized slices corrupted by noise, with standard deviations ς uniformly chosen from 0 to 0.1. The training dataset included eight motion-free ZTE datasets, acquired and reconstructed using Total Variation (TV), from 611 s scans. We retrieved 2D slices from the x–y and x–z planes, discarding those with low SNR, resulting in a total of 2480 valid slices. Each complex 2D slice was treated as a two-channel real image. $\boldsymbol{w}$ was trained for 80 epochs using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 32. Training dataset augmentation, including rotation and brightness adjustments, was applied to account for motion and contrast shifts.

We set the Lipschitz constant L to 2, σ^2 to 0.1, and the MM exit tolerance to 10^{-3}. The maximum number of CG iterations was set to 200. An ablation study evaluating hyperparameter tuning is included in Fig. S7–S8 of the Supporting Information. Three spatial resolutions were adopted for multi-resolution motion estimation: 16 mm, 4 mm, and 1 mm. 50 iterations of motion estimation were adopted in each alternating loop with a learning rate of 0.01. All hyperparameters were fixed across all in vivo experiments. All experiments were conducted on a single NVIDIA A100 GPU with 80 GB of memory.

2.4.5 | Methods used for Comparisons

In this section, we describe the baseline methods evaluated for comparison. The proposed approach was compared against ZTE infilling methods both with and without motion correction. CG-SENSE and Low Rank represented state-of-the-art (SOTA) infilling methods without motion correction. For motion-corrected baselines, motion estimation was incorporated into CG-SENSE using either ground-truth motion parameters (CG-SENSE Upper Bound) or motion estimation methods (AF and MERLIN). Additionally, a classical regularization term, Total Variation (TV-MoCo), was integrated alongside motion correction.

CG-SENSE

CG-SENSE recovers k-space holes using coil sensitivity maps and quadratic regularization.

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}\boldsymbol{x}\|_2^2 \; + \; \lambda\|\boldsymbol{x}\|_2^2. \tag{16}$$

Here, we did not account for motion (i.e., $\boldsymbol{\xi} = 0$) during reconstruction.

CG-SENSE Upper Bound

CG-SENSE Upper Bound (CG-UP) performed a CG-SENSE reconstruction with ground-truth motion parameters on simulated data. It serves as the upper bound for the CG-SENSE when the motion parameters are known. The objective function is

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}\|_2^2 \; + \; \lambda\|\boldsymbol{x}\|_2^2, \tag{17}$$

where $\boldsymbol{\xi}$ represents the ground truth motion parameters.

Autofocusing

Autofocusing (AF) is a motion-correction method that alternates between image updates and motion searches[27,26]. We used Tikhonov regularization over the image in this work; the algorithm alternates between the following data consistency updates for the image and motion parameters:

$$\begin{aligned} \boldsymbol{x}^{n+1} &= \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^n}\boldsymbol{x}\|_2^2 \; + \; \lambda\|\boldsymbol{x}\|_2^2 \\ \boldsymbol{\xi}^{n+1} &= \boldsymbol{\xi}^n - \alpha\nabla_{\boldsymbol{\xi}}\|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}^{n+1}\|_2^2. \end{aligned} \tag{18}$$

MERLIN

MERLIN[4] employs image-domain registration to align multiple navigator images to a single template, where a single WASPI acquisition at the start generates coil sensitivity maps. Navigator images from consecutive k-space segments are registered pairwise to the first navigator image, yielding rotation and translation parameters. The corrected k-space data are combined and reconstructed into a high-resolution image, with WASPI data explicitly filling the dead time gap. Inspired by MERLIN's registration-based motion correction framework, we adopted a similar approach: low-resolution navigator images recovered from each segment using SENSE were used to recover motion parameters, followed by motion-compensated reconstruction. Due to the lack of WASPI acquisition in our data, we infilled the central k-space regions implicitly using CG-SENSE. We included this

baseline to compare the effectiveness of image-domain registration against the k-space-based motion correction for ZTE neuroimaging.

TV-MoCo

We follow an alternating minimization approach similar to Aligned SENSE[51], where we alternate between Total Variation (TV) regularization and motion-compensated reconstruction; we term this approach TV-MoCo. TV-MoCo alternates between the following steps:

$$\begin{aligned} \boldsymbol{x}^{n+1} &= \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}^n}\boldsymbol{x}\|_2^2 \; + \; \lambda\|\boldsymbol{G}\boldsymbol{x}\|_1 \\ \boldsymbol{\xi}^{n+1} &= \boldsymbol{\xi}^n - \alpha\nabla_{\boldsymbol{\xi}}\|\boldsymbol{y} - \boldsymbol{\mathcal{A}}_{\boldsymbol{\xi}}\boldsymbol{x}^{n+1}\|_2^2. \end{aligned} \tag{19}$$

where $\boldsymbol{G}$ is the gradient operator.

Low Rank Infilling

The Low Rank iterative infilling method reformulates the finite support constraint in the image domain to the locally low rank property in k-space[15]. It also takes advantage of the coil sensitivity map information, similar to the CG-SENSE method. The optimization is posed as follows:

$$\min_{\boldsymbol{x}} \|\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda\|\mathbf{HNS}\boldsymbol{x}\|_*. \tag{20}$$

where $\boldsymbol{\mathcal{A}}$ denotes the forward operator excluding the missing samples trajectory, $\mathbf{N}$ denotes the NuFFT operator including the missing samples trajectory, $\mathbf{S}$ refers to the coil sensitivity map, and $\mathbf{H}$ represents the matrix lifting operator to form a structured Hankel matrix from 3D Cartesian data. The high computational complexity of this method makes its extension to motion correction challenging.

2.4.6 | Evaluation Metrics

We relied on both qualitative and quantitative metrics for evaluation. In the simulated 3D ZTE settings where ground truth images were available, we evaluated image reconstruction quality using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). We used ℓ_1 distance measures for motion accuracy,

$$\ell_1(\hat{\boldsymbol{\xi}}) = \frac{1}{T}\sum_{t=1}^{T}|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}|, \tag{21}$$

where T represents the number of k-space segments. The estimated motion parameters from all data were registered to a common reference image to eliminate linear shift errors.

3 | RESULTS

A preview of in vivo reconstructions on 611 s and 153 s data is shown in Fig. 2 . CG-SENSE reconstructions without motion correction suffer from extensive motion artifacts, while EnMoCo is able to significantly reduce these artifacts. In the 153 s data, CG-SENSE exhibits a contrast shift resulting from missing k-space center and has motion artifacts. EnMoCo is able to reduce the motion artifacts and contrast shift. We perform rigorous experiments on simulated data and in vivo data in the following sections to validate the proposed method.

3.1 | Validation using numerical simulations

We first evaluate the proposed algorithm on simulated ZTE data to understand the trade-offs and to perform quantitative comparisons with other methods.

Impact of dead time gap

Fig. 3 illustrates the impact of dead time gap under intermediate motion (M2: ±15°, ±8 mm) and an SNR of 19.8 dB (S2 from Fig. 4). CG-SENSE partially compensates for missing data, restoring image contrast for dead time gaps, but it exhibits motion-induced blurring. MERLIN (CG-SENSE + registration-based motion estimation) reduces motion distortions using image-domain registration, while AF (CG-SENSE + k-space motion estimation) further improves motion estimation accuracy. TV-MoCo effectively removes aliasing artifacts; however, contrast shifts caused by missing central samples are amplified at larger gaps (gap $\geq$ 3.5). EnMoCo consistently yields reduced motion and aliasing artifacts, particularly for larger gaps ($\geq$ 4). EnMoCo effectively suppresses motion and aliasing artifacts by combining motion estimation with an energy-based denoiser. The reduction in errors is also evident in the difference map. We attribute the gains to the improved regularization provided by MuSE. The results of the gaps below 3 and other methods, including Low Rank and CG-UP, are included in Fig. S3 and Table S1 of the Supporting Information. Table S3 summarizes the ℓ_1 motion estimation errors for AF and EnMoCo. Fig. S5 presents an ablation study evaluating the relative contributions of energy-based regularization and motion correction.

Runtime analysis

The execution times across varying gap sizes are summarized in Table 2 . Conventional methods (CG-SENSE, AF, and MERLIN) rely on standard conjugate-gradient iterations, resulting in execution times that remain largely invariant to gap size. By contrast, TV-MoCo and Low Rank involve sharp increases in runtime as the gap increases. For example, the complexity of the Low Rank method[15] increases rapidly from 3 min at a gap of 1.5 to 920 min at gaps $\geq$ 4, while TV-MoCo reaches up to

247 min at a gap of 5 due to degraded convergence under severe central k-space loss. We note that the runtime of EnMoCo is higher for small gaps (28 min); its runtime scales far more gracefully as the gap size expands (reaching only 40 min at a gap of 5).

Impact of motion severity and SNR

Fig. 4 summarizes results for a constant gap size of 4 samples, four SNR levels (S1–S4: 24.4–10.1 dB), and four rigid-motion conditions with increasing motion severity (M1–M4: ±2° to ±17°).

The reconstructed images remain sharp across motion levels at all SNR levels, indicating that motion correction is less sensitive to noise (Fig. 4 (B)). At a fixed SNR, increasing motion severity results in relatively slow rises in residual errors in the difference maps shown in Fig. 4 (C). A sharp increase in errors would be expected due to effective non-uniform k-space sampling from spoke rotations caused by motion, leading to error amplification and alias artifacts. This degradation is minimized by the MuSE regularizer, which controls noise amplification and artifacts. From Fig. 3 , we note that at a gap size of 4, an SNR of S2: 19.8 dB, and motion level M2, AF exhibits significantly more noise amplification, indicating greater sensitivity of SENSE to noise and/or imperfect conditioning. Table S5 and S6 report the ℓ_1 error of motion estimation and PSNR/SSIM across multiple SNR and motion levels. An ablation study investigating the robustness of the 2.5D prior to rotation is presented in Fig. S9 of the Supporting Information. Rotation about the x–axis exhibits a slight degradation compared to y– and z–axis rotations at larger angles, as x–axis rotation causes out-of-plane motion across both the x–y and x–z training planes, whereas y– and z–axis rotations maintain in-plane transformations for one of the trained planes.

Impact of acceleration

Fig. 5 evaluates performance under accelerated 3D ZTE acquisitions by retrospectively reducing the number of inversion blocks to 128 and 64, corresponding to scan durations of 306 s and 153 s, respectively.

At the 153 s scan duration, AF exhibits pronounced contrast deviations and aliasing even at a gap size of 3.5, which worsen at a gap of 4. While TV-MoCo provides reasonable artifact suppression at 306 s, its performance degrades substantially on the 153 s dataset: the reduced data redundancy destabilizes its motion parameter estimation, leaving visible residual blurring and localized structural errors. In contrast, EnMoCo maintains high-fidelity reconstructions across gap sizes up to 4 across both acceleration factors. Only mild contrast degradation is observed at the extreme gap size of 4.5 under the 153 s setting, highlighting the stabilizing role of the learned prior in joint motion and image estimation under severe data undersampling. PSNR and SSIM are summarized in Table S2 of the Supporting Information. Table S4 summarizes the ℓ_1 motion estimation errors.

Impact of intra-block motion

Under moderate, realistic continuous head motion (Motion 1 in Fig. 6), residual blurring and artifact levels are negligible relative to reconstructions without intra-block corruption, demonstrating that the block-wise constant approximation holds remarkably well in typical practical scenarios. However, under severe continuous motion (Motion 2), residual motion artifacts become visible in difference maps, reflecting localized breakdowns of the piecewise constant assumption during high-velocity movement. Fig. S4 in the Supporting Information shows a comparison between simulation and in vivo data with more motion patterns. The results confirm that performance degradation caused by continuous intra-block motion remains minimal for typical slow-to-moderate motion.

3.2 | In vivo Analysis

Comparison with SOTA methods

Fig. 7 compares reconstructions on in vivo ZTE data across gap sizes 3 to 5 (±15°, ±10 mm). CG-SENSE exhibits severe noise and motion blur as the gap size increases. AF mitigates motion artifacts up to gap 4, but suffers from blurring and contrast loss at larger gaps. TV-MoCo improves artifact suppression at small gaps, but degraded motion estimates at gaps ≥ 4.5 introduce residual blur. In contrast, EnMoCo maintains stable image quality across nearly all gap sizes, effectively suppressing artifacts and preserving contrast. Comparisons of accelerated in vivo data and multiple motion levels are included in Fig. S10–S11 of the Supporting Information. Across various acceleration factors, motion severity levels, and dead time gaps, the proposed method robustly suppresses the majority of motion artifacts while preserving fine anatomical structures. Fig. S6 in the Supporting Information demonstrates the advantage of multi-scale over single-scale motion correction, showing improved sharpness on in vivo data.

Qualitative evaluation on multiple subjects

Fig. 8 compares: CG-SENSE reconstruction of the motion-corrupted scan (no motion correction), EnMoCo of the motion-corrupted scan, and EnMoCo of the motion-free scan from seven subjects with a gap size of 3.

The extent of motion varies across subjects, as reflected in the uncorrected CG-SENSE reconstructions.

Across subjects, EnMoCo reduces visible motion-related artifacts in the motion-corrupted scans and does not introduce obvious degradation in the motion-free scans. In cases with more pronounced motion (e.g., Subjects 6 and 7), residual blurring remains after correction, although the reconstructed images show improved sharpness and reduced artifact burden relative to the uncorrected reconstructions. Fig. S12 in the Supporting Information shows additional axial and sagittal reconstruction views.

4 | DISCUSSION

This work presents a unified reconstruction framework for 3D ZTE that jointly addresses dead-time–induced missing central k-space samples, rigid subject motion, and undersampling within a regularized MAP formulation. Regularization is provided by a pre-trained Multi-Scale Energy-based model (MuSE), used as an explicit energy and optimized using a majorization-minimization strategy, which is well-suited to the ill-conditioning introduced by missing-center data and acceleration. To reduce 3D training demands, we use a data-efficient 2.5D construction that aggregates slice-wise energies from two orthogonal views, aiming to mitigate view-dependent artifacts while remaining practical in limited-data settings.

Several methodological and practical considerations merit discussion for future work. First, hyperparameter selection required careful balancing across varying gap sizes and motion levels; while fixed parameters performed robustly under moderate conditions, adaptive tuning could further optimize extreme cases. Second, evaluating in vivo reconstructions is inherently constrained by the absence of a gold-standard reference. Moreover, hyperparameters tuned on synthetic data may not be optimal for in vivo datasets. This performance gap between simulated and experimental data primarily stems from lower signal-to-noise ratio, imperfect coil sensitivity maps, and continuous off-resonance effects.

Our implementation assumes rigid-body motion and neglects intra-block motion during one inversion block. This assumption may break down under severe, continuous, high-velocity movement. Consequently, improving the temporal resolution of motion correction remains an important direction for future investigation. Directly comparing our approach with the full MERLIN[4] pipeline using WASPI infilling is left for future work, as this requires a dedicated acquisition protocol not retrospectively available in our current dataset. The 2.5D MuSE prior substantially reduces training data and memory requirements compared with a full 3D prior, but it is trained only with in-plane rotation augmentation. Augmenting the training data with slices extracted from 3D volumes rotated out-of-plane could improve the robustness of the 2.5D prior under extreme motion.

Finally, this work does not account for individual spoke-dependent T_1-weighting variations during the acquisition of multiple FIDs that result in small effective TI changes over the course of each inversion recovery. Because the T_1 recovery curve is smooth and central k-space dominates overall image contrast, the effective T_1 weighting represents a blended average across all 192 readouts rather than a single discrete TI. A more principled approach is to explicitly resolve the TI variation along the readout train, reconstructing images at multiple inversion times as in MPnRAGE[50] and performing motion correction on these TI-resolved images, as demonstrated for cardiac T_1 mapping[52]. Such T_1-resolved motion correction would avoid contrast mismatches between the static template image and spokes acquired across a wide inversion time range.

Although evaluated on ZTE, the framework is not ZTE-specific. The k-space motion model and the MuSE prior apply directly to conventional 3D radial trajectories, which correspond to the special case of no dead time gap. Because the k-space center is measured directly, we expect the image update to be better conditioned and to converge faster. Validation on non-ZTE 3D radial data remains future work.

5 | CONCLUSION

We introduced a unified MAP-based reconstruction method for accelerated 3D ZTE that jointly estimates the image and rigid-motion parameters while implicitly accounting for unmeasured k-space samples arising from dead time and undersampling. The approach alternates k-space motion updates with conjugate-gradient data-consistency steps and MuSE-based regularization using an explicit energy prior in a data-efficient 2.5D form. Results on simulated and experimental datasets demonstrate improved robustness to motion and missing-center effects relative to the tested baselines, with remaining failure modes primarily at the largest dead time gaps and highest motion/acceleration settings. These findings suggest that combining explicit k-space motion modeling with an energy-based learned prior is a practical direction for improving motion robustness and image quality in accelerated ZTE acquisitions.

Acknowledgments

The work is supported by R01-AG067078 and R01-EB019961. The data acquisition was supported by R43MH122028 and was acquired on an MRI instrument funded by S10OD025025. The University of Iowa receives research support from GE Healthcare. University of Virginia receives support from Siemens Healthineers and Canon Medical Systems.

ORCID

Yan Chen 0009-0003-6886-1886
Curtis A. Corum 0000-0002-4994-6206
James H. Holmes 0000-0002-0936-3700
Mathews Jacob 0000-0001-6196-3933

REFERENCES

1. Aydıngöz Üstün, Yıldız Adalet Elçin, Ergen F. Bilge. Zero Echo Time Musculoskeletal MRI: Technique, Optimization, Applications, and Pitfalls. *RadioGraphics.* 2022;42(5):1398–1414.
2. Ljungberg Emil, Damestani Nikou L., Wood Tobias C., et al. Silent Zero TE MR Neuroimaging: Current State-of-the-Art and Future Directions. *Progress in nuclear magnetic resonance spectroscopy.* 2021;123:73–93.
3. Weiger Markus, Pruessmann Klaas P., Hennel Franciszek. MRI with Zero Echo Time: Hard versus Sweep Pulse Excitation. *Magnetic Resonance in Medicine.* 2011;66(2):379–389.
4. Ljungberg Emil, Wood Tobias C., Solana Ana Beatriz, Williams Steven C. R., Barker Gareth J., Wiesinger Florian. Motion Corrected Silent ZTE Neuroimaging. *Magnetic Resonance in Medicine.* 2022;88(1):195–210.
5. Wiesinger Florian, Ho Mai-Lan. Zero-TE MRI: principles and applications in the head and neck. *The British Journal of Radiology.* 2022;95(1136):20220059.
6. Ljungberg Emil, Wood Tobias, Solana Ana Beatriz, et al. Silent T1 mapping using the variable flip angle method with B1 correction. *Magnetic Resonance in Medicine.* 2020;84(2):813–824.
7. Wiesinger Florian, Sacolick Laura I, Menini Anne, et al. Zero TE MR bone imaging in the head. *Magnetic Resonance in Medicine.* 2016;75(1):107–114.
8. Engström Mathias, McKinnon Graeme, Cozzini Cristina, Wiesinger Florian. In-phase zero TE musculoskeletal imaging. *Magnetic Resonance in Medicine.* 2020;83(1):195–202.
9. Wu Yaotang, Dai Guangping, Ackerman Jerome L, et al. Water- and fat-suppressed proton projection MRI (WASPI) of rat femur bone. *Magnetic Resonance in Medicine.* 2007;57(3):554–567.
10. Grodzki David M, Jakob Peter M, Heismann Bjoern. Ultrashort echo time imaging using pointwise encoding time reduction with radial acquisition (PETRA). *Magnetic Resonance in Medicine.* 2012;67(2):510–518.
11. Froidevaux Romain, Weiger Markus, Rösler Manuela B., Brunner David O., Pruessmann Klaas P.. HYFI: Hybrid Filling of the Dead-Time Gap for Faster Zero Echo Time Imaging. *NMR in Biomedicine.* 2021;34(6):e4493.
12. Froidevaux Romain, Weiger Markus, Brunner David O., Dietrich Benjamin E., Wilm Bertram J., Pruessmann Klaas P.. Filling the Dead-Time Gap in Zero Echo Time MRI: Principles Compared. *Magnetic Resonance in Medicine.* 2018;79(4):2036–2045.
13. Corum Curtis A, Kruger Stanley J, Jacob Mathews, Magnotta Vincent A, Holmes James H. Robust and computationally efficient missing point and phase estimation for multi-channel acquisition with ZTE sequences. *Joint Annual Meeting ISMRM: ISMRT 31st Annual Meeting.* 2022;30:4068.
14. Stoch Grzegorz, Olejniczak Zbigniew. Missing first points and phase artifact mutually entangled in FT NMR data—noniterative solution. *Journal of Magnetic Resonance.* 2005;173(1):140–152.
15. Huo Zimu, de Arcos José, Wiesinger Florian, Kaggie Joshua D., Graves Martin J.. Low-Rank Iterative Infilling for Zero Echo-Time (ZTE) Imaging. *Magnetic Resonance in Medicine.* 2025;93(3):1149–1162.
16. Wood Tobias C, Ljungberg Emil, Chiew Mark. ZTE infilling from auto-calibration neighbourhood elements. *Proc Intl Soc Mag Reson Med.* 2021;29:1169.
17. Kuethe Dean O, Caprihan Arvind, Lowe Irving J, Madio David P, Gach H Michael. Transforming NMR data despite missing points. *Journal of Magnetic Resonance.* 1999;139(1):18–25.
18. Wood Tobias C, Brackenier Yannick. Implicit ZTE deadtime gap filling with iterative reconstruction. *ESMRMB 2021 Online 38th Annual Scientic Meeting.* 2021;:7–9.
19. Welch Edward Brian, Rossman Phillip J, Felmlee Joel P, Manduca Armando. Self-navigated motion correction using moments of spatial projections in radial MRI. *Magnetic Resonance in Medicine.* 2004;52(2):337–345.
20. Kecskemeti Steven, Samsonov Alexey, Velikina Julia, et al. Robust Motion Correction Strategy for Structural MRI in Unsedated Children Demonstrated with Three-dimensional Radial MPnRAGE. *Radiology.* 2018;289(2):509–516.
21. Wallace Tess E, Piccini Davide, Kober Tobias, Warfield Simon K, Afacan Onur. Rapid motion estimation and correction using self-encoded FID navigators in 3D radial MRI. *Magnetic Resonance in Medicine.* 2024;91(3):1057–1066.
22. Vaillant Ghislain, Prieto Claudia, Kolbitsch Christoph, Penney Graeme, Schaeffter Tobias. Retrospective rigid motion correction in k-space for segmented radial MRI. *IEEE transactions on medical imaging.* 2013;33(1):1–10.
23. Atkinson David, Hill Derek L.G., Stoyle Peter N.R., et al. Automatic Compensation of Motion Artifacts in MRI. *Magnetic Resonance in Medicine.* 1999;41(1):163–170.
24. Loktyushin Alexander, Nickisch Hannes, Pohmann Rolf, Schölkopf Bernhard. Blind Retrospective Motion Correction of MR Images. *Magnetic Resonance in Medicine.* 2013;70(6):1608–1618.

25. Kuzmina Ekaterina, Razumov Artem, Rogov Oleg Y, Adalsteinsson Elfar, White Jacob, Dylov Dmitry V. Autofocusing+: Noise-resilient motion correction in magnetic resonance imaging. *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2022;:365–375.
26. Haskell Melissa W, Cauley Stephen F, Wald Lawrence L. TArgeted Motion Estimation and Reduction (TAMER): data consistency based motion mitigation for MRI using a reduced model joint optimization. *IEEE transactions on medical imaging.* 2018;37(5):1253–1265.
27. Cordero-Grande Lucilio, Ferrazzi Giulio, Teixeira Rui Pedro A. G., O'Muircheartaigh Jonathan, Price Anthony N., Hajnal Joseph V.. Motion-corrected MRI with DISORDER: Distributed and Incoherent Sample Orders for Reconstruction Deblurring Using Encoding Redundancy. *Magnetic Resonance in Medicine.* 2020;84(2):713–726.
28. Al-Haj Hemidi Ziad, Weihsbach Christian, Heinrich Mattias P.. IM-MoCo: Self-supervised MRI Motion Correction Using Motion-Guided Implicit Neural Representations. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024.* 2024;:382–392.
29. Wu Qing, Du Chenhe, Tian Xuanyu, Yu Jingyi, Zhang Yuyao, Wei Hongjiang. Moner: motion correction in undersampled radial MRI with unsupervised neural representation. *International Conference on Learning Representations.* 2025;:70680–70701.
30. Chen Yan, Holmes James H, Corum Curtis, Magnotta Vincent, Jacob Mathews. Deep factor model: a novel approach for motion compensated multi-dimensional MRI. *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI).* 2023;:1–4.
31. Liu Weijia, Tang Runyu, Yao Ye, et al. SAMoCo: A Learning-Initialized and Physics-Refined Framework for Sampling-Agnostic Motion Correction in MRI. *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2026;488–498.
32. Hammernik Kerstin, Klatzer Teresa, Kobler Erich, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine.* 2018;79(6):3055-3071.
33. Aggarwal Hemant K, Mani Merry P, Jacob Mathews. MoDL: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging.* 2018;38(2):394–405.
34. Monga Vishal, Li Yuelong, Eldar Yonina C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine.* 2021;38(2):18–44.
35. Venkatakrishnan Singanallur V, Bouman Charles A, Wohlberg Brendt. Plug-and-play priors for model based reconstruction. *2013 IEEE global conference on signal and information processing.* 2013;:945–948.
36. Ahmad Rizwan, Bouman Charles A, Buzzard Gregery T, et al. Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE signal processing magazine.* 2020;37(1):105–116.
37. Hurault Samuel, Leclaire Arthur, Papadakis Nicolas. Gradient step denoiser for convergent plug-and-play. *International Conference on Learning Representations.* 2022;.
38. Chand Jyothi Rikhab, Jacob Mathews. Multi-scale energy (muse) framework for inverse problems in imaging. *IEEE transactions on computational imaging.* 2024;10:1250–1265.
39. Lena Beatrice, Najac Chloé, Börnert Peter, Webb Andrew, Dong Yiming. Time-Conditioned Zero-Shot Self-Supervised Reconstruction for Accelerated 3D Ultra-Low-Field MRI. *Magnetic Resonance in Medicine.* 2026;96(3):1303–1312.
40. Chen Yan, Kecskemeti Steve R, Holmes James H, et al. Accelerating 3D radial MPnRAGE using a self-supervised deep factor model. *Magnetic Resonance in Medicine.* 2025;94(3):1191–1201.
41. Haskell Melissa W., Cauley Stephen F., Bilgic Berkin, et al. Network Accelerated Motion Estimation and Reduction (NAMER): Convolutional Neural Network Guided Retrospective Motion Correction Using a Separable Motion Model. *Magnetic Resonance in Medicine.* 2019;82(4):1452–1461.
42. Johnson Patricia M., Drangova Maria. Conditional Generative Adversarial Network for 3D Rigid-body Motion Correction in MRI. *Magnetic Resonance in Medicine.* 2019;82(3):901–910.
43. Singh Nalini M, Dey Neel, Hoffmann Malte, et al. Data consistent deep rigid MRI motion correction. *Medical imaging with deep learning.* 2024;:368–381.
44. Dabrowski Oscar, Falcone Jean-Luc, Klauser Antoine, et al. SISMIK for brain MRI: Deep-learning-based motion estimation and model-based motion correction in k-space. *IEEE Transactions on Medical Imaging.* 2024;44(1):396–408.
45. Levac Brett, Kumar Sidharth, Jalal Ajil, Tamir Jonathan I. Accelerated motion correction with deep generative diffusion models. *Magnetic Resonance in Medicine.* 2024;92(2):853–868.
46. Chen Lixuan, Tian Xuanyu, Wu Jiangjie, et al. Joint coil sensitivity and motion correction in parallel MRI with a self-calibrating score-based diffusion model. *Medical Image Analysis.* 2025;102:103502.
47. Chen Nan-kuei, Guidon Arnaud, Chang Hing-Chiu, Song Allen W. A robust multi-shot scan strategy for high-resolution diffusion weighted MRI enabled by multiplexed sensitivity-encoding (MUSE). *NeuroImage.* 2013;72:41–47.
48. Guhaniyogi Shayan, Chu Mei-Lan, Chang Hing-Chiu, Song Allen W, Chen Nan-kuei. Motion immune diffusion imaging using augmented MUSE for high-resolution multi-shot EPI. *Magnetic Resonance in Medicine.* 2016;75(2):639–652.
49. Corum Curtis A, Kruger Stanley J, Magnotta Vincent A. HEALPix view-order for 3D radial self-navigated motion-corrected ZTE MRI. *arXiv preprint arXiv:1910.10276.* 2019;.
50. Kecskemeti Steven, Samsonov Alexey, Hurley Samuel A, Dean Douglas C, Field Aaron, Alexander Andrew L. MPnRAGE: A technique to simultaneously acquire hundreds of differently contrasted MPRAGE images with

applications to quantitative T1 mapping. *Magnetic Resonance in Medicine.* 2016;75(3):1040–1053.

51. Cordero-Grande Lucilio, Teixeira Rui Pedro AG, Hughes Emer J, Hutter Jana, Price Anthony N, Hajnal Joseph V. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Transactions on Computational Imaging.* 2016;2(3):266–280.
52. Becker Kirsten M, Blaszczyk Edyta, Funk Stephanie, et al. Fast myocardial T1 mapping using cardiac motion correction. *Magnetic Resonance in Medicine.* 2020;83(2):438–451.

SUPPORTING INFORMATION

The following supporting information is available as part of the online article:

Table S1. PSNR (dB)/SSIM of simulated 611 s ZTE data with varied gaps. All reconstructions are registered to the reference image. TV-MoCo and EnMoCo outperform other methods across gap sizes with spatial regularization and k-space motion correction. EnMoCo recovers images with the highest PSNR and SSIM for gaps > 3, demonstrating the superior performance of a learned prior at larger gap sizes.

Table S2. PSNR (dB)/SSIM of simulated 306 s and 153 s ZTE data with gaps 3 – 4.5. EnMoCo maintains high-fidelity reconstructions across gap sizes up to 4 across both acceleration factors. Note that the motion of the 306 s acquisition is heavier than the 153 s data due to the variation of motion across blocks, as reflected by the PSNR difference between CG-UP for both acquisition times.

Table S3. ℓ_1 Motion error (mean $\pm$ std) of simulated 611 s ZTE data. Rotation (degrees) and Translation (pixels). EnMoCo achieves lower ℓ_1 error for rotation parameters, which we attribute to multi-resolution search. The ℓ_1 error of translation from both AF and EnMoCo is higher than that of rotation, as shown in Fig. S1. Small rotational perturbations are easier to detect and estimate accurately, as they induce pronounced motion artifacts in 3D MRI.

Table S4. ℓ_1 motion error (mean $\pm$ std) of simulated 306 s and 153 s ZTE data. Rotation in degrees and translation in pixels. EnMoCo achieves substantially lower rotation errors than AF for both accelerated acquisitions, while translation errors are comparable.

Table S5. ℓ_1 error of motion estimation with varying SNR and motion levels: Four motion levels, M1–M4, are simulated with increasing motion amplitude and number of movements across inversion blocks, as shown in Fig. S1. Upper: Rotation (degrees), Lower: Translation (pixels). The ℓ_1 error degrades gracefully as the SNR decreases.

Table S6. PSNR (dB)/SSIM for simulated ZTE data with varying SNR and motion levels. Four SNR levels are simulated, ranging from 24.4 dB to 10.1 dB. Four motion levels, M1–M4, are simulated with increasing motion amplitude and number of movements across inversion blocks. Simulated motion parameters are shown in Fig. S1. At a given SNR level, reconstructed image quality is primarily governed by the severity of motion. Notably, a single set of hyperparameters was utilized across all SNR and motion levels; consequently, the

reported metrics reflect fixed parameter performance rather than individually optimized global solutions.

Figure S1. Motion estimation from simulated ZTE with varying SNR and motion corruption. Small rotational perturbations lead to pronounced motion artifacts in 3D MRI, making them highly detectable and easier to estimate accurately. In contrast, translations often produce subtler effects, resulting in less precise estimation.

Figure S2. Details of ZTE sequence: (a) Magnetization prepared Zero Echo-time (MP-ZTE) sequence. We use a non-selective inversion pulse, followed by a delay to obtain the desired T_1 weighting. (b) We use 192 5° excitations within each individual inversion segment. Gradients increment between RF pulses. (c) HEALPix segment gradient values and view endpoints for 4×192 view segments (768 views). (d) Diagram of the outer point from each k-space segment demonstrating relatively uniform angular coverage of k-space.

Figure S3. Comparisons from 611 s simulated data across gaps 1.5-5. Low Rank effectively reduces the artifacts and contrast shift but exhibits blurring artifacts caused by motion. While Low Rank and TV-MoCo suppress noise more effectively than CG-SENSE, their computational runtimes scale sharply with increasing gap size. By contrast, EnMoCo consistently yields reduced motion and aliasing artifacts while maintaining favorable runtime scaling as the gap size expands (Table 2 in the manuscript).

Figure S4. Comparison of simulated and in vivo experiments under realistic motion and noise conditions. (A) Impact of motion pattern and intra-block motion: Reconstruction results across increasing motion severities (light to heavy) for (a) in vivo acquisitions, (b) synthetic data simulated with piecewise-constant per-block motion, and (c) synthetic data simulated with continuous intra-block motion. (d) Corresponding rotation and translation trajectories extracted from in vivo datasets. Continuous motion simulation (c) reproduces the motion-dependent blurring observed in vivo, whereas piecewise-constant simulation (b) underestimates degradation under heavy, rapid motion. (B) Impact of signal-to-noise ratio (SNR): Synthetic reconstructions evaluated across decreasing SNR levels under piecewise-constant and continuous motion models. Decreasing SNR combined with continuous motion modeling closes the performance gap between synthetic simulations and in vivo observations.

Figure S5. Ablation study evaluating regularization strategy and motion correction components. Reconstruction results, PSNR/SSIM metrics, and absolute difference maps relative to the motion-free reference (Ref):(A) vs. (B) Regularization Impact (MuSE vs. Tikhonov): The proposed MuSE prior significantly reduces noise and gap-infilling artifacts compared to standard Tikhonov-regularized CG-SENSE. (A) vs. (C) Regularization Impact (MuSE vs. Total Variation): The MuSE prior yields superior artifact suppression and detail preservation over TV regularization. (A) vs. (D) Joint Reconstruction vs. Image-Only: Reconstructing without the motion correction step (EnMoCo w/o MoCo) results in a 7.1 dB PSNR drop (25.5 dB), confirming that image estimation and motion correction are tightly coupled and must be solved jointly.

Figure S6. Ablation study comparing multi-scale and single-scale MoCo on the 611 s in vivo data. Multi-scale MoCo incorporates spatial resolutions of 16 mm, 4 mm, and 1 mm. Single-scale MoCo applies optimization at 1 mm. Compared to single-scale MoCo, which exhibits residual blurring, multi-scale MoCo achieves sharper reconstruction of subtle brain structures and superior preservation of fine details.

Figure S7. Ablation study of Lipschitz constants and outer loop steps. A smaller Lipschitz constant provides faster convergence but suffers from instability, e.g., Lipschitz=0.8. We adopt a Lipschitz constant equal to 2 to balance convergence stability and speed. We evaluate the global loss and the accuracy of motion estimation across 10 outer iterations. Both the global loss and motion estimation converge within 10 outer iterations.

Figure S8. Ablation study of σ^2. $1/2\sigma^2$ determines the weights of data consistency and prior. Lower σ^2 indicates higher importance of data consistency. We observe that the reconstructed image from $\sigma^2 = 0.01$ is corrupted by aliasing artifacts due to an insufficient prior term. $\sigma^2 = 1$ is over-regularized by higher weights of the prior term and is not consistent with measurements, as shown by the lower PSNR. We choose $\sigma^2 = 0.1$, which achieves a better trade-off between data consistency and prior.

Figure S9. Robustness evaluation of the 2.5D prior across rotation angles (0°–30°) along the x-, y-, and z-axes. EnMoCo demonstrates stable reconstruction performance (measured in PSNR) for rotation angles between 0° and 20°, covering typical clinical motion regimes and the in vivo data tested in this study. Rotation about the x-axis exhibits a slight degradation compared to y- and z-axis rotations at larger angles, as x-axis rotation causes out-of-plane motion across both the x-y and x-z training planes, whereas y- and z-axis rotations maintain in-plane transformations for one of the trained planes.

Figure S10. Evaluation of reconstructions under four levels of motion corruption. CG-SENSE reconstruction without motion correction exhibits increasing blurriness and structural degradation as motion severity increases. The proposed method suppresses the majority of motion artifacts across different motion levels and dead time

gaps. Notably, we observe that the image quality is more sensitive to the dead time gap than to the motion level. For a gap of 5, severe motion leads to noticeable contrast shifts and structural loss. When the gap is 3, the motion-induced degradation is substantially reduced.

Figure S11. Results of accelerated in vivo ZTE acquisition. We retain 128 and 64 inversion blocks out of 256, corresponding to a scan time of 306 and 153 seconds, respectively. The dead time gap ranges from 3 to 4.5. Aliasing artifacts dominate the image quality at higher acceleration, as visible in the AF result shown above. The performance of the proposed method is robust, with improved anatomical structures for different acceleration factors.

Figure S12. Evaluation on seven subjects showing coronal (A), sagittal (B), and axial (C) reformats of the 3D volumes. Two acquisitions were obtained from each subject, with and without intentional motion. EnMoCo improves the sharpness of the image reconstructed from corrupted data. EnMoCo successfully reconstructs high-quality ZTE images across a wide range of motion levels, from minimal, e.g., Subject 1, to extreme motion, e.g., Subject 7. By contrast, CG-SENSE is corrupted by motion artifacts for scans with heavy motion, e.g., Subjects 4 – 7. For motion-free acquisitions, EnMoCo still performs motion estimation but does not degrade the reconstruction quality.

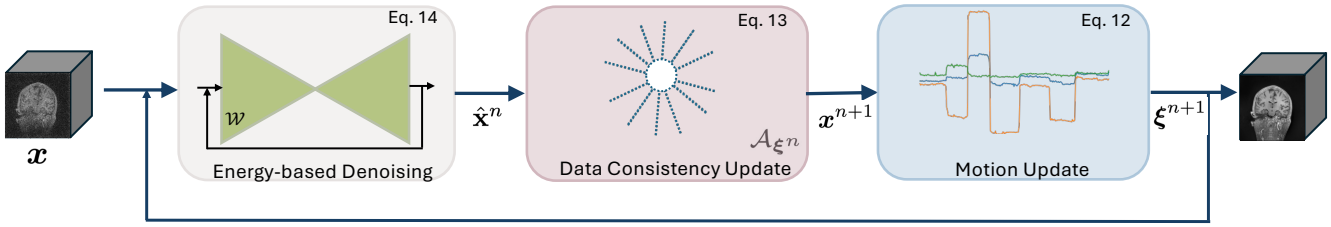


FIGURE 1 Overview of joint reconstruction and motion correction. We initialize the image volume $\boldsymbol{x}$ using CG-SENSE with no motion correction, while the motion parameters $\boldsymbol{\xi}$ are initialized as zeros. The optimization alternates between energy-based denoising specified by Eq. (14), data consistency update using CG specified by Eq. (13), and motion update specified by Eq. (12). The iterative algorithm recovers the motion-compensated and k-space center-infilled volume.

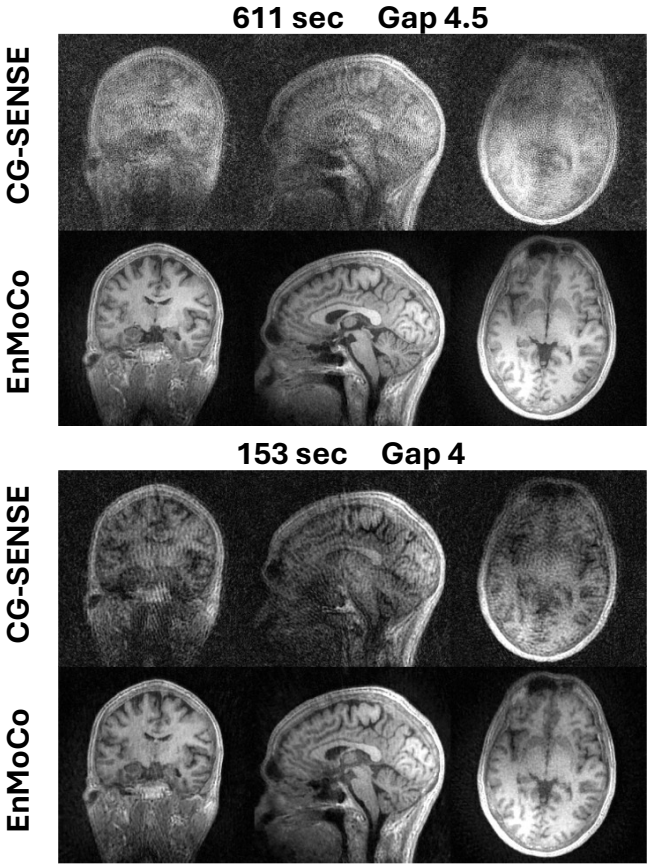


FIGURE 2 Preview of CG-SENSE and EnMoCo reconstructions from 611 s and 153 s motion-corrupted in vivo acquisitions. CG-SENSE reconstruction exhibits blurriness caused by motion in 611 s data and contrast shift due to acceleration in 153 s data. EnMoCo resolves motion artifacts using motion correction and reduces contrast shift using MuSE regularization.

TABLE 1 PSNR (dB)/SSIM of simulated 611 s ZTE data with varied gaps. All reconstructions are registered to the reference image. Results of Low Rank and CG-UP are included in the Supporting Information Table S1.

Methods\Gaps	3	3.5	4	4.5	5
Gridding	16.9/0.460	15.9/0.434	15.6/0.422	15.4/0.411	14.8/0.401
CG-SENSE	20.9/0.292	20.5/0.281	19.0/0.259	17.2/0.244	15.3/0.145
AF	22.6/0.415	21.9/0.402	19.7/0.372	17.8/0.346	15.4/0.200
MERLIN	21.5/0.345	20.8/0.332	18.9/0.305	17.2/0.287	16.6/0.264
TV-MoCo	32.8/**0.841**	30.4/0.826	21.4/0.743	20.6/0.728	19.8/0.693
EnMoCo	**33.6**/0.839	**32.8/0.841**	**32.6/0.829**	**32.4/0.824**	**28.4/0.805**

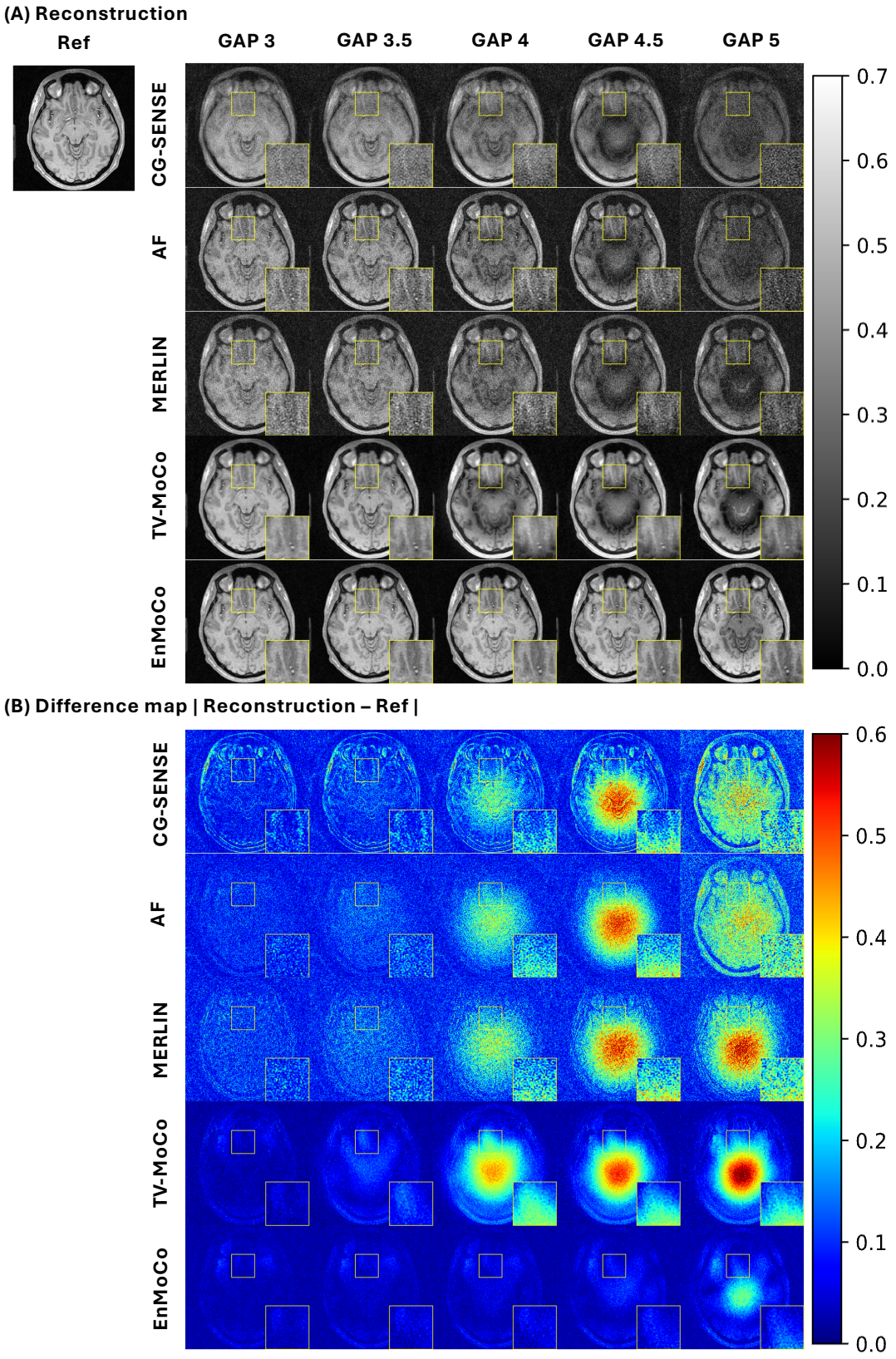


FIGURE 3 Impact of the dead time gap on motion-compensated recovery in simulated 611 s ZTE experiments. Reconstructions are shown for dead time gaps of 3–5 samples using CG-SENSE, AF, MERLIN, TV-MoCo, and EnMoCo. CG-SENSE suffers from both aliasing artifacts and motion distortions. MERLIN reduces motion distortions using image-domain registration, while AF further reduces motion distortions. TV-MoCo effectively removes aliasing artifacts; however, contrast shifts caused by missing central samples are amplified at larger gaps (gap ≥ 3.5). EnMoCo consistently yields reduced motion and aliasing artifacts, particularly for larger gaps (≥ 4). The PSNR and SSIM are reported in Table 1 .

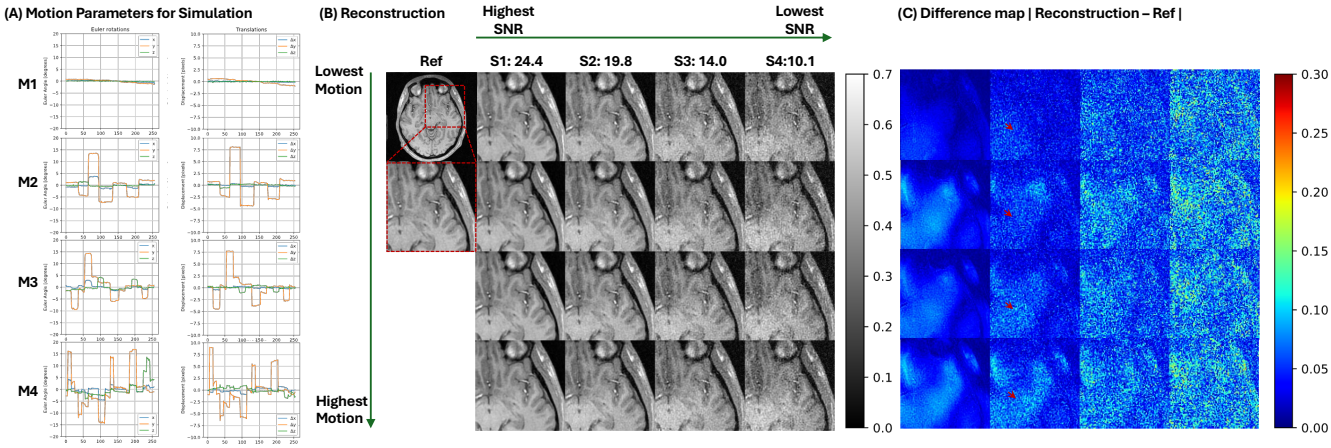


FIGURE 4 Impact of motion severity and SNR in reconstruction quality on simulated data: Data were simulated at four SNR levels (S1–S4: 24.4–10.1 dB) and four rigid-motion conditions (M1–M4). (A) shows the prescribed block-wise rigid-motion parameters (three rotations and three translations per block) used to generate the corrupted measurements; motion magnitude and/or the number of motion events increase from M1 (mild) to M4 (severe). Reconstructions are shown in (B), and the difference maps in (C). Across SNR levels, the reconstructions are observed to remain visually sharp under mild-to-moderate motion, indicating that the motion updates are relatively robust to SNR. At a fixed SNR, the difference maps show a moderate increase in residual errors and noise variance with increasing motion severity. We note that motion alters the effective location of the k-space spokes, resulting in some regions of the k-space sphere being undersampled; this translates to error amplification and aliasing artifacts, as indicated by red arrows in (C). We observe that MuSE regularization keeps the increase in noise variance moderate. By contrast, Fig. 3 shows that AF exhibits significantly higher noise amplification even at M2.

TABLE 2 Runtime (in minutes) for each method across different gap sizes from a 611 s acquisition. Conventional methods (CG-SENSE, AF, and MERLIN) rely on standard conjugate-gradient iterations, resulting in execution times that remain largely invariant to gap size. TV-MoCo and Low Rank involve sharp increases in runtime as the gap increases. The runtime of EnMoCo scales far more gracefully as the gap size expands.

Method	Gap Size						
	1.5	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
CG-SENSE	**1.4**	**1.5**	**1.6**	**1.8**	**2.0**	**2.5**	**2.6**
Low Rank	3	33	101	340	920	920	920
AF	10	11	11	11	11	12	12
MERLIN	11	11	11	11	11	12	12
TV-MoCo	6	8	38	181	183	239	247
EnMoCo	28	32	32	35	37	40	40

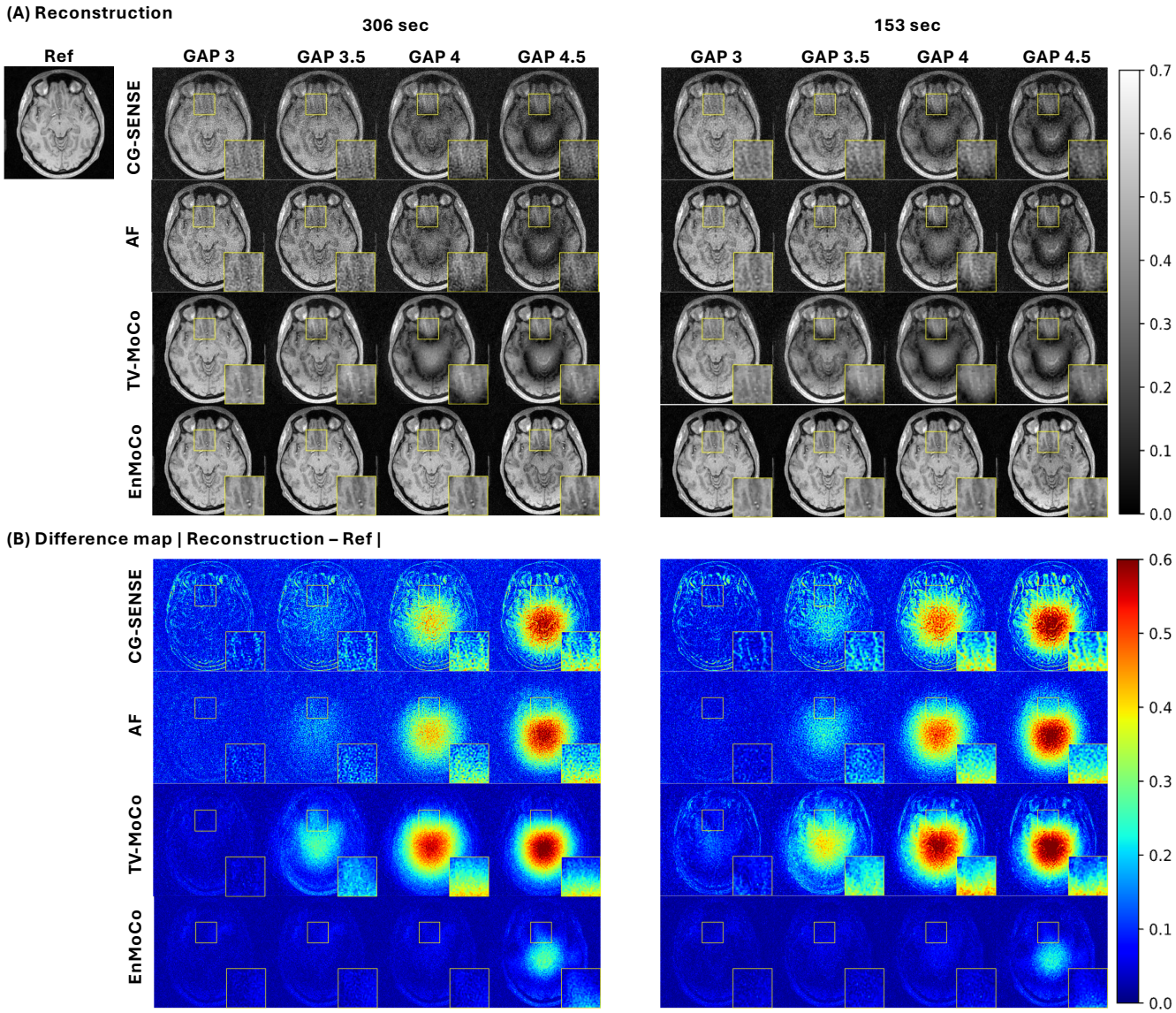


FIGURE 5 Impact of scan acceleration and dead time gap size on simulated image quality and motion estimation. Reconstructions are shown for undersampled acquisitions retaining 128 blocks (306 s) and 64 blocks (153 s) out of 256 full inversion blocks across gap sizes ranging from 3.0 to 4.5. CG-SENSE and AF exhibit severe aliasing artifacts and contrast distortion due to missing central k-space data, particularly for gaps > 3.5. TV-MoCo exhibits motion distortion at the 153 s scan duration. In contrast, EnMoCo consistently suppresses motion artifacts and recovers missing k-space samples up to gap size 4.0, demonstrating superior robustness under high acceleration.

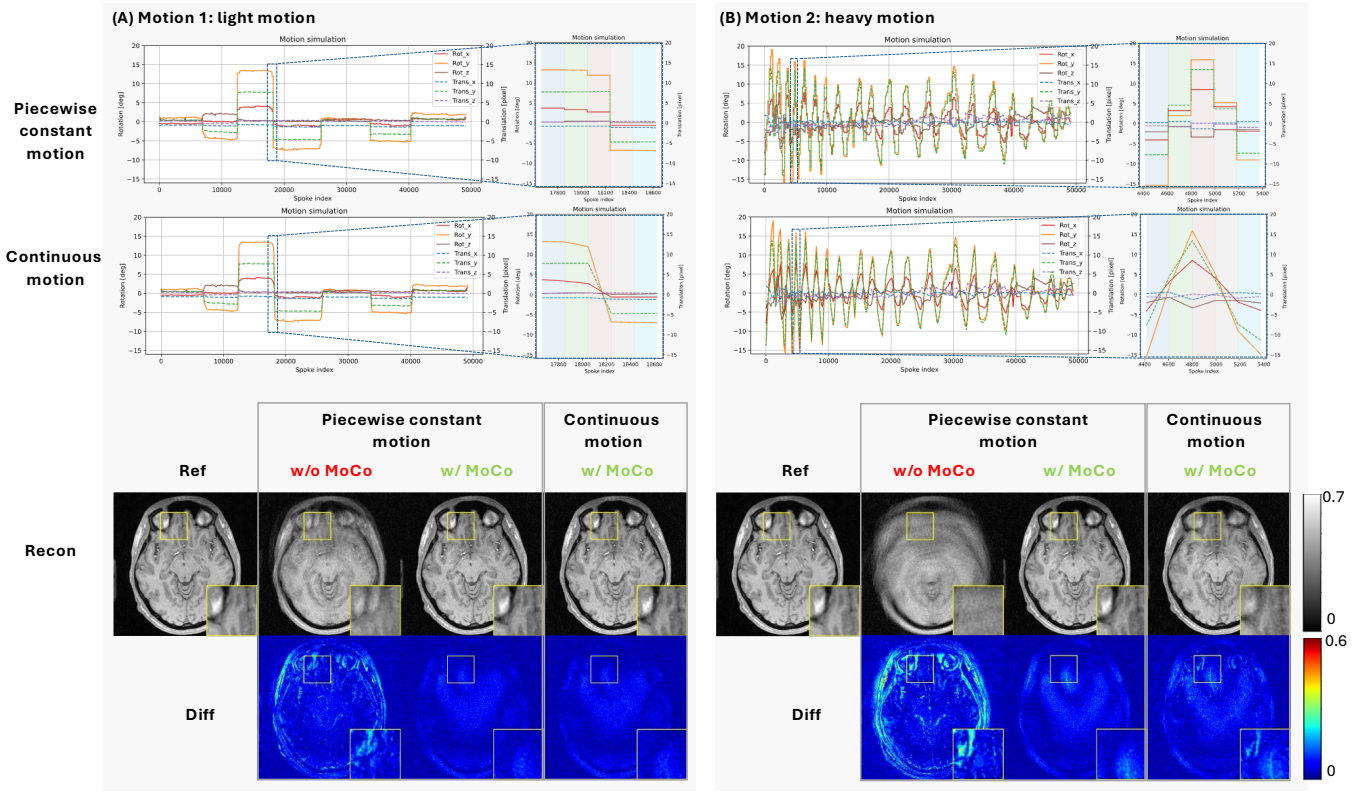


FIGURE 6 Evaluation of continuous intra-block motion modeling. (A) Motion 1 with light motion and (B) Motion 2 with heavy motion, showing corresponding motion parameters per spoke, reconstructions from EnMoCo with (w/) and without (w/o) applying motion correction (MoCo), and difference maps. Parameter trajectories from five selected blocks are zoomed in and color-coded. Reconstructions from Motion 1 with continuous motion corruption are comparable to piecewise constant motion corruption, demonstrating that the per-block approximation holds well under moderate, realistic continuous head motion. For Motion 2, residual motion artifacts become visible with continuous motion corruption, reflecting localized breakdowns of the block-wise constant assumption during high-velocity movement.

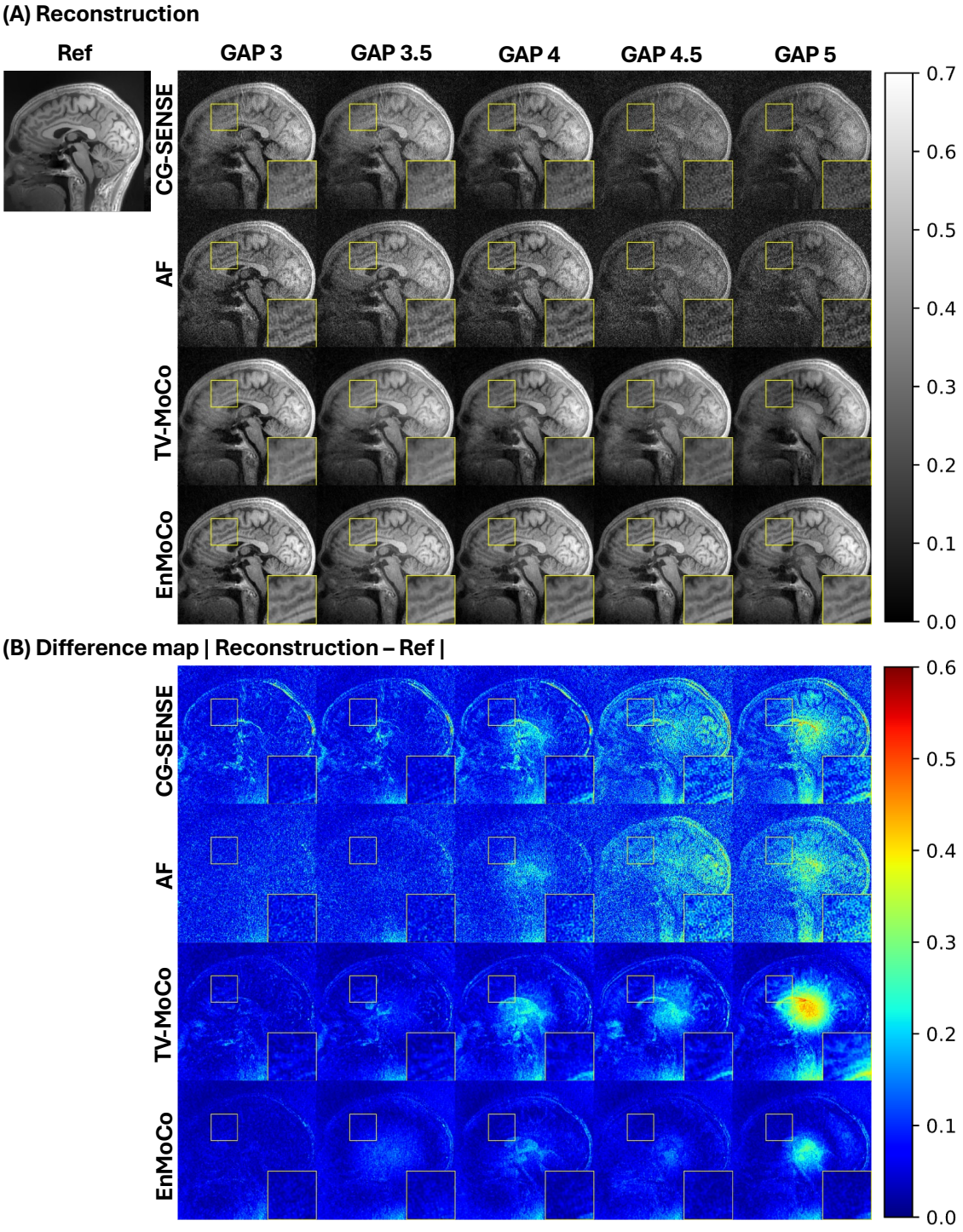


FIGURE 7 Evaluation of missing samples infilling and motion correction on in vivo 611 s ZTE data. The motion level is comparable to M2 shown in Fig. 4 . We observe that CG-SENSE preserves global image contrast, but exhibits pronounced structural artifacts resulting from uncorrected motion. AF mitigates motion artifacts for gaps up to 4, but suffers from blurriness and contrast inconsistencies as the number of missing samples increases. TV-MoCo performs well on gaps of 3 and 3.5, but exhibits contrast shift for higher gap sizes. The proposed method further improves image quality, reducing both structural artifacts and contrast shift. However, the image quality is less effective than in the simulated experiments, which we attribute to the imperfections in the reference image, inaccuracies in the coil sensitivity maps, intra-block motion, and the inherently lower SNR of real data.

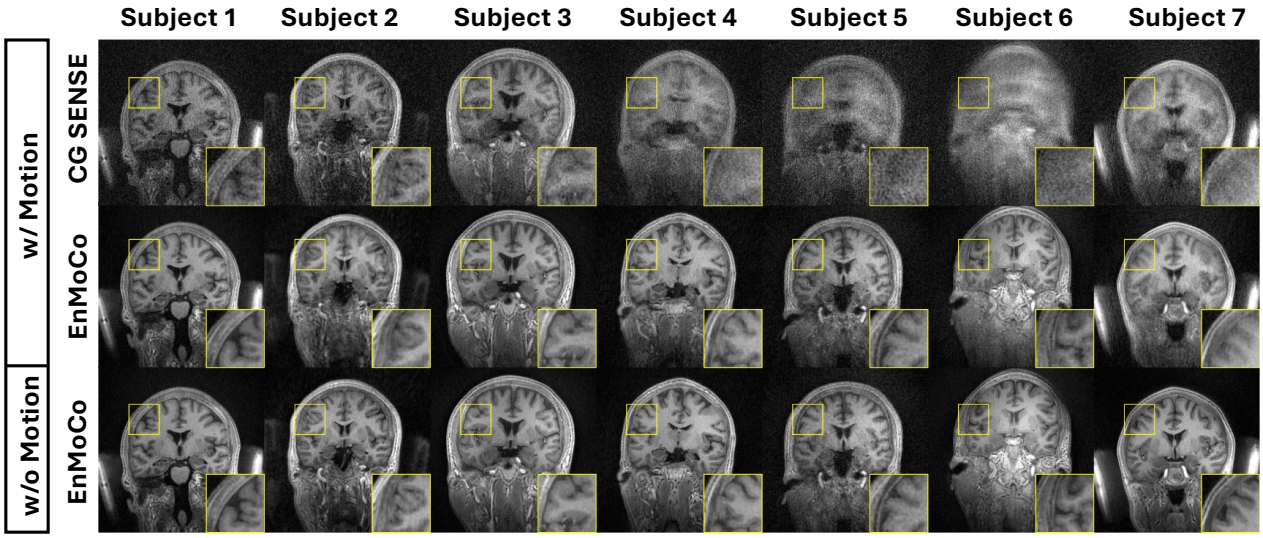


FIGURE 8 Evaluation on seven subjects. EnMoCo improves the sharpness of the image reconstructed from corrupted data. EnMoCo successfully reconstructs high-quality ZTE images across a wide range of motion levels, from minimal, e.g., Subject 1, to extreme motion, e.g., Subject 7. By contrast, CG-SENSE is corrupted by motion artifacts for scans with heavy motion, e.g., Subjects 4 – 7. For motion-free acquisitions, EnMoCo still performs motion estimation but does not degrade the reconstruction quality.