Nonrigid Motion Correction in MRI Using Neural Space-Time Modelling

Yimeng Lin Jaehyeok Bae Aizada Nurdinova
Stanford University
{linym, jhbae110, nurdaiza}@stanford.edu

Abstract

Motion remains a major challenge in MRI due to long acquisition times, with nonrigid motion particularly problematic in body imaging. While several classical and deep learning-based methods have addressed motion artifacts, many rely on additional measurements, pre-training, or lack generalization across subjects and protocols. We propose a self-supervised, interpretable framework for nonrigid motion modeling and correction using implicit neural representations (INRs). Our neural space-time architecture jointly trains a scene network to represent anatomy and a motion network to predict time-resolved deformation fields. Training these networks on a per-subject basis should improve adaptability and eliminate the need for supervised datasets or external motion measurements. Incorporating a lightweight rigid motion network further stabilizes and accelerates training.

Using simulated data with combined rigid and nonrigid motion, we demonstrate accurate reconstruction across a wide range of motion amplitudes (0–50 mm). Reconstruction quality was highly sensitive to the nonrigid motion network's capacity and regularization strength, with at least 8 levels of hash encoding required to achieve SSIM > 0.90*. We also evaluated the impact of imaging parameters, finding that calibration line count had minimal effect, while increased acceleration led to a gradual decline in reconstruction performance. These results highlight the proposed framework's flexibility and data efficiency, showing strong potential for robust, motion-resolved MRI without reliance on external priors.*

1. Introduction/Related works

Magnetic Resonance Imaging (MRI) is highly susceptible to motion artifacts due to long acquisition times. This is especially problematic in high-resolution head imaging, where even small head or neck movements can blur the image, and in body imaging, where involuntary motions—such as breathing and heartbeat—introduce complex, nonrigid deformations. Although rigid-body motion models are often assumed in head imaging, they tend to break down in areas like the neck, where motion varies across space. In body imaging, motion is inherently nonrigid but can sometimes be locally approximated as rigid.

Conventional motion correction techniques typically rely on motion tracking using navigator signals or external sensors, and often assume rigid-body motion. Navigator signals may be obtained from low-resolution image navigators [1], dedicated k-space lines [2], or external sources such as optical tracking systems [3] or radiofrequency sensors like pilot tone [4]. These approaches estimate motion parameters (typically six degrees of freedom per shot) from multi-coil data using the forward model. However, these methods face several limitations: external devices require an extra calibration, solving for nonrigid motion becomes intractable with many variables, and navigator-based methods can reduce scan efficiency by taking up additional acquisition time.

To address nonrigid motion, several methods have been proposed for dynamic abdominal and cardiac imaging. One classical approach, XD-GRASP [5], separates motion phases into respiratory and cardiac bins and jointly reconstructs images across bins. This method has been widely adopted at clinical sites around the world due to its robustness and effectiveness. However, it requires manual extraction of motion states, assumes periodic motion, and limits temporal resolution due to predefined binning. As a result, it performs suboptimally in the presence of irregular or bulk motion. Recently, an end-to-end deep learning alternative was introduced, combining an auto-navigation network (RANGR) with a motion-compensated reconstruction network (Movienet). This approach automates motion state extraction and achieves higher acceleration factors in free-breathing dynamic imaging. Unlike classical methods, the deep learning pipeline requires supervised pretraining on approximately 40 clean datasets but eliminates the need for handcrafted binning and manual motion labeling.

More advanced nonrigid motion correction strategies incorporate motion directly into the image reconstruction process. Model-based approaches and motion-informed encoding (MIE) techniques integrate nonrigid deformation fields into the signal model and jointly estimate both motion and image content during reconstruction [6, 7, 8]. In parallel, several data-driven methods have emerged, including MedGAN [9], BladeNet [10], and LAPA+ (A-LIK)-Net [11], which leverage learning-based architectures to improve motion-robust imaging. Notably, LAPNet [12] is a promising cardiac imaging method that employs a learned affine parameterization for nonrigid motion correction. It demonstrates high-quality reconstruction performance but relies on pretraining with a large representative dataset and incorporates hand-crafted motion priors to guide learning.

Recently, an interpretable implicit neural representation (INR) approach was proposed for dynamic contrast-enhanced (DCE) MRI of the liver. This method models post-contrast dynamic images as a function of spatial coordinates and learnable temporal latent variables, parameterized by normally distributed variables. The learned temporal variables can be interpreted as encoding either contrast dynamics or motion, offering new insights into temporal variations in DCE-MRI. Compared to XD-GRASP-PRO [13], the approach demonstrated comparable or improved quality in both motion-resolved image reconstructions and quantitative contrast uptake curves. While promising, the method does not explicitly estimate motion fields, and its generalizability to other sampling patterns beyond radial acquisitions or to non-contrast imaging remains uncertain.

Building on these ideas, we propose an end-to-end framework for nonrigid motion estimation and correction in multi-shot MRI, leveraging space-time neural modeling [14]. Our method jointly learns two implicit neural representation (INR) functions defined over spatiotemporal coordinates. The first is a *scene network*, which maps 2D spatial coordinates (x, y) to static complex-valued image intensities. The second is a *motion network*, which maps 3D spatiotemporal coordinates (x, y, t) to a smooth, nonrigid displacement field $(\delta x, \delta y)$ that deforms the static image at each time point.

The predicted deformation fields are applied via spatial warping, followed by a non-uniform fast Fourier transform (NUFFT) and multi-coil encoding to synthesize k-space data. The scene and motion networks are jointly trained to minimize a composite loss consisting of: (1) a data fidelity term comparing the synthesized and measured k-space, and (2) a motion regularization term enforcing spatial and temporal smoothness of the deformation fields.

In contrast to prior INR-based approaches that encode temporal variation implicitly through latent variables, our method explicitly estimates motion fields with a dedicated network and regularization strategy. This formulation enables structured, interpretable modeling of motion and improves data consistency. To further enhance stability and disentangle global motion from local deformation, we introduce a lightweight *rigid motion network*, which takes acquisition time t as input and outputs in-plane translation and rotation parameters. These are used to apply an initial rigid alignment before estimating nonrigid motion.

Our framework is fully self-supervised, requiring no navigators, external sensors, or pretrained components. It directly learns both motion and image content from undersampled k-space data, enabling end-to-end optimization without auxiliary supervision. By explicitly modeling complex, potentially aperiodic nonrigid motion, the method offers a flexible and data-efficient solution for dynamic MRI reconstruction, with broad applicability across anatomical regions, contrast mechanisms, and sampling trajectories.

2. Method

2.1. Neural Space-Time Model Architecture

Our method formulates motion-corrected MRI reconstruction as a joint learning problem of a static *scene* and dynamic *motion* within a unified neural space-time framework. Specifically, we represent the static 3D image as a coordinate-based neural field F_θ, and the time-varying motion as a function M_ϕ that outputs a deformation field for each time point (or shot index). To account for both rigid and nonrigid motion components, we decompose M_ϕ into a global rigid transformation $R(t)$ and a nonrigid displacement field $D_\phi(\mathbf{x}, t)$.

Scene Network: The static scene is modeled by an implicit neural network $F_\theta(\mathbf{x})$, which maps a spatial coordinate $\mathbf{x} \in \mathbb{R}^3$ to the complex-valued image intensity at that location. We implement F_θ as a multi-layer perceptron (MLP) and apply multi-scale hashed positional encoding to $\mathbf{x}$, normalized to the range $[-1, 1]^3$, to enable high-frequency detail representation. This formulation allows F_θ to compactly represent a high-resolution volumetric MR image.

Motion Network: The motion field is defined as a function M_ϕ that produces a 3D displacement for each spatial coordinate $\mathbf{x}$ at time t. We parameterize M_ϕ as the composition of a *rigid* motion model $R(t)$ and a *nonrigid* deformation model $D_\phi(\mathbf{x}, t)$. The rigid component $R(t)$ applies a global transformation parameterized by six degrees of freedom (three translations and three rotations), which are predicted per time point. The nonrigid component $D_\phi(\mathbf{x}, t)$ is modeled by an MLP that outputs a local displacement vector $\Delta\mathbf{x}$ for each spatial coordinate and time, enabling spatially varying motion.

Coordinate Warping: The full motion at time t defines a mapping from static to displaced coordinates as:

$$M(\mathbf{x}, t) = R(t)(\mathbf{x}) + D_\phi(\mathbf{x}, t), \tag{1}$$

where $R(t)$ denotes the rigid transformation, and D_ϕ is the nonrigid displacement field.

To synthesize the image at time t, we query the scene network at the corresponding static coordinate that maps to $\hat{\mathbf{x}}$ after motion. Assuming small deformations, we apply backward warping:

$$\hat{I}_t(\hat{\mathbf{x}}) = F_\theta\left(R^{-1}(t)(\hat{\mathbf{x}}) - D_\phi(\hat{\mathbf{x}}, t)\right). \tag{2}$$

2.2. MRI Forward Model

To simulate k-space measurements from motion-corrupted images, we integrate a differentiable MRI forward model into the network training loop. For each shot t and coil c, the forward operator $\mathcal{A}$ applies coil sensitivity modulation followed by a Fourier transform. Let $S_c(\mathbf{x})$ denote the known sensitivity profile of coil c, and let $\hat{I}_t(\mathbf{x})$ be the reconstructed image at time t. The fully-sampled, continuous k-space signal for coil c is given by:

$$Y_{c,t}^{\text{full}}(\mathbf{k}) = \int_{\mathbf{x}} S_c(\mathbf{x})\, \hat{I}_t(\mathbf{x})\, e^{-j2\pi\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}. \tag{3}$$

In practice, each shot samples only a subset of k-space. We model this by applying a binary sampling mask $M_t(\mathbf{k})$ specific to shot t, resulting in:

$$Y_{c,t}(\mathbf{k}) = M_t(\mathbf{k}) \cdot Y_{c,t}^{\text{full}}(\mathbf{k}). \tag{4}$$

This forward model is implemented as a differentiable pipeline consisting of coil sensitivity weighting, a 2D FFT, and element-wise masking in k-space using precomputed undersampling patterns. The output is the predicted k-space data $\{Y_{c,t}^{(\text{pred})}\}$, which is compared against the acquired data during training to enforce data fidelity.

2.3. Loss Function and Regularization

We jointly train the scene and motion networks in a fully unsupervised manner, relying only on the discrepancy between predicted and measured k-space data. The primary data fidelity term enforces consistency with the acquired multi-coil k-space and is defined as:

$$\mathcal{L}_{\text{data}} = \sum_{t=1}^{N_t} \sum_{c=1}^{N_c} \left\| Y_{c,t}^{(\text{pred})} - Y_{c,t}^{(\text{meas})} \right\|_2^2, \tag{5}$$

where $Y_{c,t}^{(\text{pred})}$ and $Y_{c,t}^{(\text{meas})}$ denote the predicted and measured k-space for coil c at shot t.

To promote realistic and stable motion estimation, we introduce two regularization terms. The first enforces spatial smoothness and sparsity in the nonrigid deformation field:

$$\mathcal{L}_{\text{nonrigid}} = \|\nabla_x D_\phi\|_1 + \|\nabla_y D_\phi\|_1 + \|D_\phi\|_1, \tag{6}$$

where ∇_x and ∇_y denote finite differences along spatial axes.

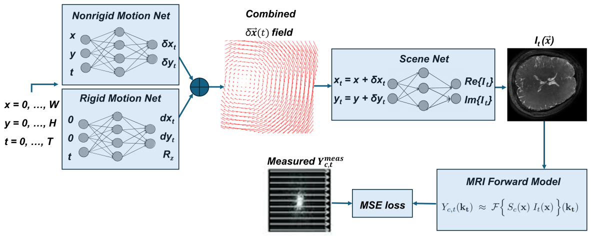

Figure 1. Schematic overview of the proposed Neural Space-Time Model (NSTM) for motion estimation and motion-corrected reconstruction. The model consists of two jointly trained networks: a motion network and a scene network. Both take spatial (and temporal) coordinates as input. The motion network outputs time-resolved motion parameters, while the scene network predicts the corresponding complex image intensities at those coordinates. The networks are optimized together using an Mean Squared Error (MSE) data consistency loss, computed by evaluating the MRI forward model on the motion-corrected images.

The second penalizes excessive global motion by constraining the predicted rigid transformations:

$$\mathcal{L}_{\text{rigid}} = \mathbb{E}_t\left[\|R(t)\|_1\right], \tag{7}$$

where $R(t)$ represents the 6-DoF rigid motion parameters at time t.

The total loss function is given by:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_{\text{rigid}}\mathcal{L}_{\text{rigid}} + \lambda_{\text{nonrigid}}\mathcal{L}_{\text{nonrigid}}, \tag{8}$$

where λ_{rigid} and $\lambda_{\text{nonrigid}}$ are tunable weights that control the contribution of the respective regularization terms.

2.4. Optimization and Training Strategy

We optimize the model parameters using the Adam optimizer with a default learning rate of 10^{-2}. Training is performed on a per-volume basis and typically converges within 1000 iterations. To improve convergence and stability, we adopt a two-phase training strategy:

1. *Rigid pretraining:* The rigid motion parameters $R(t)$ and the scene network F_θ are optimized while keeping the nonrigid motion network D_ϕ frozen.

2. *Joint optimization:* All components—including $R(t)$, F_θ, and D_ϕ—are jointly trained.

Figure 1 illustrates the overall architecture of our proposed end-to-end framework for motion-corrected MRI reconstruction.

2.5. Experiment Design

We conducted a series of experiments to isolate and evaluate the impact of key modeling decisions:

1. **Motion Model Variants:**
    - **Model A:** Full model with both rigid and nonrigid components, trained in two stages. It first learns rigid motion then adds nonrigid motion.
    - **Model B:** Nonrigid-only model using D_ϕ, without rigid motion estimation.
    - **Model C:** Rigid-only model using $R(t)$, without nonrigid deformation.
2. **Motion Magnitude:** We evaluated reconstruction performance under increasing levels of motion, varying the standard deviation of nonrigid deformation.
3. **Regularization Parameters:** We varied λ_{rigid} and $\lambda_{\text{nonrigid}}$ to assess their influence on motion field smoothness, training stability, and image reconstruction accuracy.
4. **Sampling Patterns:** We tested the effect of calibration line count by varying the number of calibration lines per shot from 0 to 32. Additionally, we varied the acceleration factor from 6× to 10× using Cartesian undersampling and evaluated reconstruction quality across different sampling patterns.
5. **Encoding Capacity:** We investigated the impact of nonrigid motion network capacity by varying the number of hash grid levels used in the motion encoder.

These experiments collectively highlight the importance of joint rigid/nonrigid motion modeling, the sensitivity of the framework to regularization and motion magnitude, and its robustness across a range of sampling patterns and network capacities.

3. Dataset and Features

High-quality, motion-free brain images were acquired on a 3T GE Signa Premier system using a 32-channel head coil, under IRB approval with informed consent. The dataset consists of central slices extracted from a single T_1-weighted acquisition using Magnetic Resonance Fingerprinting (MRF), with an isotropic resolution of 1 mm and in-plane dimensions of 220×220.

To simulate motion, we generated 100 distinct motion states by applying random in-plane rigid transformations and/or smoothly varying nonrigid deformation fields. Rigid motion was modeled by uniformly sampling rotation angles (in degrees) and translations (in pixels) from the range $[0, 5]$. Nonrigid motion fields were sampled from a zero-mean Gaussian distribution with a tunable variance parameter and subsequently smoothed using a Gaussian filter to ensure spatial continuity of the deformation.

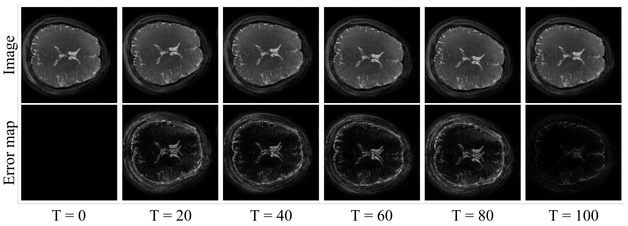

Figure 2. Example of simulated motion used in the experiment (**top**) and the corresponding error map with respect to the static ground truth (**bottom**).

An example of the simulated motion and its corresponding error map with respect to the static ground truth is shown in Figure 2. Only simulated data were used in this study, enabling controlled, quantitative evaluation of both image reconstruction quality and motion estimation accuracy.

3.1. Hash Encoding

To efficiently represent spatiotemporal input coordinates, all three neural networks in our framework—the scene network, rigid motion network, and nonrigid motion network—employ multi-resolution hash grid encodings, as introduced in the neural field literature and implemented in Tiny CUDA Neural Networks (TCNN) [15].

This encoding maps continuous input coordinates (e.g., (x, y, t)) to high-dimensional feature vectors through a hierarchy of spatial resolutions. Each level in the hierarchy consists of a learnable hash table indexed by discretized coordinates. Linear interpolation between neighboring entries ensures continuity, allowing the networks to represent both coarse structural information and fine-grained spatial detail with high memory efficiency.

- The **scene network** uses a 2D spatial encoding with `16` levels and `2` features per level. The base resolution is set to `16`, increasing by a factor of `1.35` per level, with a `log2_hashmap_size` of `16`. This configuration offers a compact yet expressive representation of static anatomical structures in the image domain.
- The **rigid motion network** employs a lightweight 1D temporal encoding with `8` levels, `2` features per level, and a `log2_hashmap_size` of `20`. The resolution increases by a factor of `1.3`, enabling smooth estimation of global motion trajectories over time.
- The **nonrigid motion network** uses the most expressive encoding, applied to 3D spatiotemporal coordinates (x, y, t). It consists of `16` levels, `2` features per level, and a `log2_hashmap_size` of `20`, with a resolution scaling factor of `1.26` per level. This con-

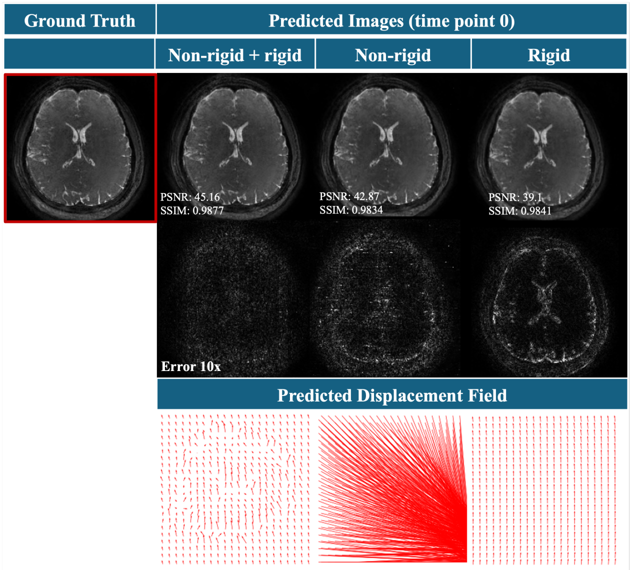

Figure 3. Predicted images and displacement fields for simulated nonrigid motion with 10 mm magnitude using different model variants: rigid+nonrigid, nonrigid-only, and rigid-only. The rigid+nonrigid model demonstrated faster convergence, improved accuracy, and greater robustness throughout training.

figuration supports dense, time-resolved modeling of locally varying deformation fields.

This hash-based encoding strategy enables the networks to learn rich, high-capacity representations directly from raw coordinates without requiring explicit spatial grids or positional encodings. It facilitates end-to-end training with rapid convergence and scalability to high-resolution data.

4. Experiments/Results/Discussion

4.1. Motion Model Variants

We evaluated the motion-correction performance of the three model variants at a fixed time point ($t = 0$). The experiments were conducted using a simulated combined 10 mm of nonrigid and 5 mm/deg of rigid motion, across $n_t = 100$ frames with 32 calibration lines per shot and an acceleration factor of $R = 2$. In the top row of Figure 3, we display the ground-truth, motion-corrupted MR image (outlined in red) alongside the predicted reconstructions produced by (1) the two-path model that learns rigid motion and adds nonrigid motion, (2) the nonrigid-only model, and (3) the rigid-only model. Under each predicted image, the numerical PSNR and SSIM values quantify reconstruction fidelity. Immediately below, a 10× amplified error map highlights residual artifacts. In the bottom row, we plot the predicted displacement field as red arrows, illustrating how each network explains the underlying motion.

A qualitative inspection of the 10× error maps reveals that the rigid+nonrigid model yields the smallest residuals. Its PSNR (45.16 dB) and SSIM (0.9877) also exceed those of the nonrigid-only model (PSNR = 42.87 dB, SSIM = 0.9834) and the rigid-only model (PSNR = 39.10 dB, SSIM = 0.9841). In the bottom row, the rigid+nonrigid model's displacement field exhibits a coherent pattern: large global translation/rotation are overlaid with smaller, spatially varying offsets from nonrigid deformations. Because the network fits the six-parameter rigid transform first, it effectively removes most of the k-space mismatch, then uses the nonrigid MLP to refine by accounting for localized warps. This makes the rigid+nonrigid approach both stable (fewer parameters to optimize initially) and accurate.

In contrast, the nonrigid-only model must solve for two motion parameters at every pixel. Although it converges at a motion field that is close to the true deformation and produces fairly high PSNR/SSIM, its displacement vectors appear noisier, reflecting the fact that the network lacks the strong inductive bias of a rigid initialization. As a result, it requires more iterations to converge and is susceptible to local minima, leading to slightly higher residuals in the error map. Finally, the rigid-only model was able to fit only a single global shift/rotation. Because this model cannot express any local variance, it leaves substantial nonrigid artifacts uncorrected—hence the larger residual in both the error map and quantitative metrics.

4.2. Motion Magnitude

We examined the performance of each model under simulated non-rigid deformation fields of increasing magnitude. Except for the amplitude of non-rigid motion, other experimental parameters are the same as in Experiment 4.1. In Figure 4, the standard deviation of the nonrigid motion was varied across 10, 20, 30, 40, 50 mm along the horizontal axis. The left plot reports the PSNR (dB) of the reconstructed image, while the right plot shows the corresponding SSIM. We compared the three model variants: rigid+nonrigid (blue), nonrigid-only (orange), and rigid-only (green).

The rigid+nonrigid model's performance curve is flat across all tested nonrigid magnitudes, demonstrating that explicitly learning a rigid transform first provides a strong inductive bias that keeps the subsequent nonrigid refinement both accurate and stable. The nonrigid-only network, although capable of matching the rigid+nonrigid model under certain conditions, exhibits severe inconsistency: its large parameter space can lead to poor local minima. The rigid-only model steadily deteriorates as nonrigid motion grows, confirming that a purely rigid assumption is insufficient whenever significant local deformations are present. These results underscore the importance of combining a lightweight rigid model with a flexible nonrigid MLP to

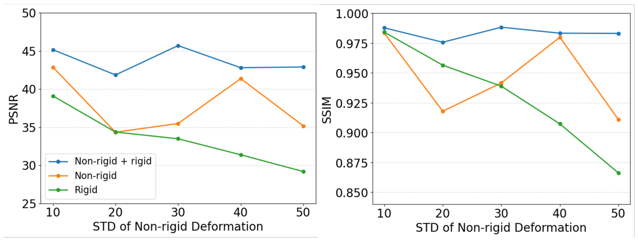

Figure 4. Performance of different models with increasing nonrigid motion magnitude. PSNR and SSIM values are plotted against the standard deviation of the synthetic nonrigid deformation fields. The rigid+nonrigid model maintains stable performance, while the nonrigid-only model shows convergence instability and the rigid-only model degrades steadily.

achieve reliable, high-fidelity motion correction across a wide range of motion amplitudes.

4.3. Regularization Parameters

We investigated the impact of regularization strength on model performance by varying the tunable parameters λ_{rigid} and $\lambda_{\text{nonrigid}}$ in Eq. 8. Figure 5 illustrates the predicted images, error maps, and displacement fields at time point 5, comparing models trained with $\lambda_{\text{nonrigid}} \in \{1\text{e}{-2}, 5\text{e}{-2}, 1\text{e}{-1}\}$. All other hyperparameters were kept at their default settings.

The experiments were conducted using a simulated combined 10 mm of nonrigid and 5 mm/deg of rigid motion, across $n_t = 20$ frames with 9 calibration lines per shot and an acceleration factor of $R = 8$. As shown in Figure 5, the reconstructed images and estimated displacement fields are highly sensitive to the choice of $\lambda_{\text{nonrigid}}$. The optimal value, $\lambda_{\text{nonrigid}} = 5\text{e}{-2}$, yielded an SSIM of 0.9529 and a PSNR of 36.92, with the estimated displacement field closely matching the designed motion.

Table 1 summarizes PSNR and SSIM results for varying $\lambda_{\text{rigid}} \in \{5\text{e}{-3}, 1\text{e}{-2}, 5\text{e}{-2}, 1\text{e}{-1}\}$. The model performance less dependent on λ_{rigid}, with the optimal value found to be $\lambda_{\text{rigid}} = 5\text{e}{-2}$.

λ_{rigid}	5e-3	1e-2	5e-2	1e-1
PSNR	31.10	36.55	**36.92**	36.15
SSIM	0.8937	0.9489	**0.9529**	0.9402

Table 1. PSNR and SSIM for the predicted motion-corrected reconstructions with different rigid motion regularization parameter λ_{rigid}.

4.4. Sampling Patterns

We evaluated the impact of sampling patterns by varying both the number of calibration lines and the effective acceleration rate. The left side of Figure 6 shows the SSIM distribution for the number of calibration lines per shot varied among $\{0, 9, 16, 32\}$. Each training was repeated 10 times. The results indicate that increasing the number of calibration lines does not consistently improve reconstruction quality. This suggests that, for the given dataset, high-frequency k-space components are in theory sufficient for both image reconstruction and motion field estimation. However, with noisier in vivo cases, having low-frequency kspace information for all shots may play a more critical role.

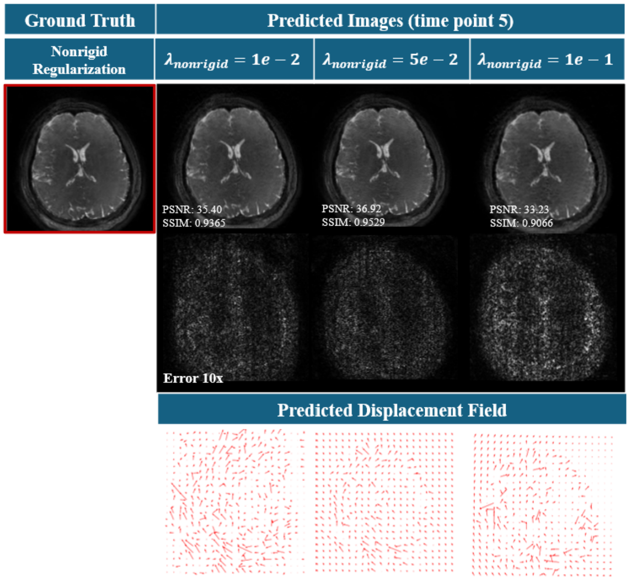

Figure 5. Reconstructed images from the rigid+nonrigid motion framework are shown for varying values of the nonrigid motion regularization parameter $\lambda_{\text{nonrigid}}$. The model is highly sensitive to this hyperparameter, likely due to the nonrigid motion network being the highest-capacity component of our NSTM framework.

The right side of Figure 6 presents the SSIM distribution for the effective acceleration rate varied among $\{5.7, 7.3, 8.7\}$. The results show that reconstruction quality improves as the acceleration rate decreases. These findings confirm that increasing the sampling density facilitates more accurate and stable reconstruction of the dynamic image sequence, although decent quality reconstruction with SSIM = 0.93 was feasible with the acceleration factor of 9.

4.5. Encoding Capacity

We evaluated the effect of nonrigid motion encoding capacity by varying the number of hash grid levels used in the nonrigid motion network. Figure 7 shows predicted images, error maps, and displacement fields at a time point 15, comparing models with $n_{\text{levels}} \in \{1, 4, 8, 16\}$. The total number of trainable encoding parameters ranged from 8K (1 level) to 15M (16 levels).

Experiments were conducted with a simulated motion

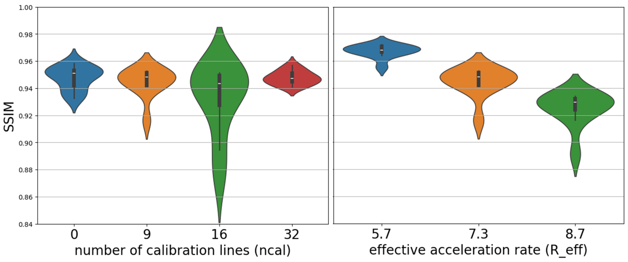

Figure 6. Distribution of SSIM values for reconstructed motion-corrected images across different k_y–t sampling patterns, varying in acceleration rate and number of calibration lines. Each experiment was repeated 10 times. (Left): SSIM as a function of the number of calibration lines per shot. Results show minimal sensitivity to this parameter. (Right): SSIM as a function of the effective acceleration rate. Higher acceleration led to a consistent decline in reconstruction quality, although decent quality with mean SSIM = 0.93 was achieved even with R_eff = 9.

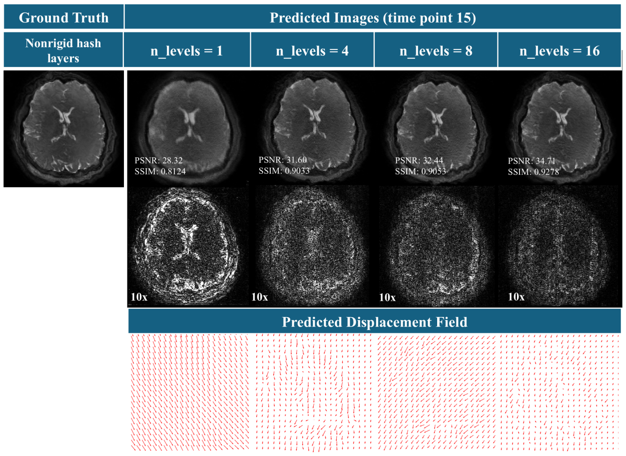

Figure 7. We evaluated the effect of encoding capacity in the nonrigid motion model by varying the number of hash encoding layers. Parameter counts ranged from 8K (1 layer) to 15M (16 layers). Experiments were conducted with combined nonrigid (50 mm) and rigid (5 mm/deg) motion, using $n_t = 20$, acceleration $R = 10$, and 32 calibration lines per time point. Models with 8 or more layers accurately fit the displacement fields and images, achieving SSIM = 0.9.

setup combining nonrigid motion of 50 mm and rigid motion of 5 mm/deg, using $n_t = 20$ frames, acceleration factor $R = 10$, and 32 calibration lines per time point. Figure 7 highlights that models with 8 or more encoding levels accurately captured both the displacement field and the image content, achieving SSIM values of 0.9 or higher. Increasing the encoding depth to 16 layers further improved PSNR to 34.74 and SSIM to 0.9278. In contrast, the model with only 1 layer produced visible residual motion artifacts and lower structural fidelity (SSIM = 0.8124).

5. Conclusion and Future Work

We proposed a self-supervised framework for motion-resolved MRI that explicitly estimates nonrigid motion using neural space-time modeling. In simulated experiments, we demonstrated that jointly training motion and scene networks on a per-case basis enables accurate image reconstruction across a wide range of motion amplitudes, without the need for pretraining or external motion measurements. Incorporating a lightweight rigid motion network further improved training stability and convergence by isolating global motion components early in the optimization.

Model performance was strongly influenced by the capacity of the nonrigid motion encoder—particularly the number of hash grid levels—with at least 8 levels required to achieve high-quality reconstructions (SSIM > 0.90). The framework showed high sensitivity to regularization strength applied to the nonrigid motion network, reflecting its role as the most expressive and flexible component. In terms of sampling strategy, we observed moderate sensitivity: the number of calibration lines per shot had minimal effect, while higher acceleration rates degraded reconstruction performance. These results demonstrate the potential of the proposed method as a flexible and data-efficient solution for robust, motion-resolved MRI reconstruction.

For future work, we plan to extend our framework to support 3D motion fields and volumetric image reconstruction. We also aim to evaluate and fine-tune the method on in vivo data as a step toward integration into clinical MRI workflows. Additionally, we will explore its application to multi-shot and multi-contrast imaging, where contrast dynamics introduce further temporal signal variation. Finally, benchmarking against state-of-the-art methods, including XD-GRASP [5] and recent INR-based motion modeling approaches [13], will be crucial for assessing reconstruction quality and generalizability.

6. Contributions & Acknowledgements

Our team collaboratively divided the workload to address the various components of this project effectively. Yimeng integrated the MRI forward model into the Neural Space-Time framework and set up the evaluation metrics. She also conducted experiments investigating different motion model variants and the impact of motion magnitude. Jaehyeok developed the simulated nonrigid dataset and conducted experiments analyzing the effects of regularization parameters, sampling strategies, and explored multi-shot acquisition cases (not presented). Aizada designed the overall architecture of the Neural Space-Time Model for joint motion estimation and motion-corrected reconstruction, acquired brain MRF images, simulated rigid motion data and contributed to the forward model implementation. She also conducted a review of related literature and led experiments evaluating the model's encoding capacity.

References

[1] Peter G Batchelor, David Atkinson, Pablo Irarrazaval, Derek LG Hill, Joseph V Hajnal, and David J Larkman. Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine*, 54(5):1273–1280, 2005.

[2] N White, C Roddey, A Shankaranarayanan, E Han, D Rettmann, J Santos, J Kuperman, and A Dale. Promo: Real-time prospective motion correction in brain mri using image-based tracking. *Magnetic Resonance in Medicine*, 63(1):91–105, 2010.

[3] Julian Maclaren, Michael Herbst, Oliver Speck, and Maxim Zaitsev. Prospective motion correction in brain imaging: A review. *Magnetic Resonance in Medicine*, 69(3):621–636, 2013.

[4] Peter Speier, Michael Buehrer, Felix A Breuer, and Sebastian Kozerke. Pilot tone navigation for continuous motion correction in mri. *Magnetic Resonance in Medicine*, 83(5):1557–1570, 2020.

[5] Li Feng, Leon Axel, Hersh Chandarana, Kai T Block, Daniel K Sodickson, and Ricardo Otazo. Xd-grasp: Golden-angle radial mri with motion-resolved compressed sensing. *Magnetic Resonance in Medicine*, 75(2):775–788, 2016.

[6] Anton Loktyushin, Hannes Nickisch, Rolf Pohmann, Bernhard Schölkopf, and Klaus Maier-Hein. Blind retrospective motion correction of mr images. *Magnetic Resonance in Medicine*, 73(5):1457–1468, 2015.

[7] Muhammad Usman, David Atkinson, Freddy Odille, Christoph Kolbitsch, Guillaume Vaillant, Tobias Schaeffter, and Claudia Prieto. Motion corrected compressed sensing for free-breathing dynamic cardiac mri. *Magnetic Resonance in Medicine*, 70(2):504–516, 2013.

[8] Andrea Bustin, Gael Cruz, Radhouene Neji, Claudia Prieto, and René M Botnar. Motion-compensated reconstruction for free-breathing 3d whole-heart coronary mri. *Magnetic Resonance in Medicine*, 81(1):102–115, 2019.

[9] Karim Armanious, Chenming Jiang, Marc Fischer, Thomas Kuestner, Tobias Hepp, Konstantin Nikolaou, Sergios Gatidis, and Bin Yang. Medgan: Medical image translation using gans. *Computerized Medical Imaging and Graphics*, 79:101684, 2020.

[10] Efrat Shimron, Alfredo De Goyeneche, Ke Wang, Alma Halgren, Ali B. Syed, Shreyas Vasanawala, and Michael Lustig. Bladenet: Rapid propeller acquisition and reconstruction for high spatio-temporal resolution abdominal mri. In *Proceedings of the 30th Annual Meeting of the ISMRM*, London, United Kingdom, 2022. International Society for Magnetic Resonance in Medicine. Abstract presented at ISMRM 2022.

[11] Aya Ghoul, Kerstin Hammernik, Andreas Lingg, Patrick Krumm, Daniel Rueckert, Sergios Gatidis, and Thomas Küstner. Highly efficient non-rigid registration in k-space with application to cardiac magnetic resonance imaging, 2024.

[12] Thomas Kuestner, Jiazhen Pan, Haikun Qi, Gastao Cruz, Christopher Gilliam, Thierry Blu, Bin Yang, Sergios Gatidis, Rene Botnar, and Claudia Prieto. Lapnet: Non-rigid registration derived in k-space for magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 40(12):3686–3697, December 2021.

[13] Jie Feng, Jingjia Chen, Yuyao Zhang, Li Feng, Dong Liang, and Hongjiang Wei. Calibration-free dce-mri with subsecond temporal resolution using interpretable implicit neural representation. In *Proceedings of the 33rd Annual Meeting of the ISMRM*, Singapore, 2025. International Society for Magnetic Resonance in Medicine.

[14] Vincent Sitzmann, Julien NP Martel, Alexander W Bergman, David B Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 7462–7473, 2020.

[15] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022.