Data Consistent Deep Rigid MRI Motion Correction

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

Motion artifacts are a pervasive problem in MRI, leading to misdiagnosis or mischaracterization in population-level imaging studies. Current retrospective rigid intra-slice motion correction techniques jointly optimize estimates of the image and the motion parameters. In this paper, we use a deep network to reduce the joint image-motion parameter search to a search over rigid motion parameters alone. Our network produces a reconstruction as a function of two inputs: corrupted k-space data and motion parameters. We train the network using simulated, motion-corrupted k-space data generated from known motion parameters. At test-time, we estimate unknown motion parameters by minimizing a data consistency loss between the motion parameters, the network-based image reconstruction given those parameters, and the acquired measurements. Intra-slice motion correction experiments on simulated and realistic 2D fast spin echo brain MRI achieve high reconstruction fidelity while retaining the benefits of explicit data consistency-based optimization.

1. Introduction

Subject motion frequently corrupts brain magnetic resonance imaging (MRI) and obfuscates anatomical interpretation (Andre et al., 2015). *Prospective* motion correction strategies adapt the acquisition in real-time to adjust for measured rigid-body motion (Tisdall et al., 2012; Frost et al., 2019). Unfortunately, prospective strategies require altering clinical workflows, prolonging scans, or interfering with standard acquisition parameters. *Retrospective* strategies correct motion algorithmically after acquisition, with or without additional motion measurements (Batchelor et al., 2005; Bammer et al., 2007; Polak et al., 2022). Retrospective motion correction without additional motion information is particularly appealing, because it does not require external hardware or clinical pulse sequence modifications and enables retroactive correction of large, previously collected k-space datasets. This work develops a deep learning method for retrospective rigid motion correction. We focus on *intra-slice* motion correction in multi-shot acquisitions, which acquire multiple k-space segments per slice and are a workhorse of clinical screening exams (Mehan et al., 2014).

Retrospective intra-slice motion correction is often formulated as joint optimization of rigid motion parameters and the underlying image (Haskell et al., 2019; Cordero-Grande et al., 2016). The result is a highly non-convex, ill-posed optimization problem with undesirable local minima. The challenging nature of joint estimation has inspired several deep learning alternatives that train standard supervised convolutional neural networks (CNNs) to reconstruct motion-free images directly from motion-corrupted inputs (Pawar et al., 2018; Duffy et al., 2021; Levac et al., 2022b; Singh et al., 2022). Alternatively, GANs can be used to provide priors on the reconstructed images (Küstner et al., 2019; Johnson and Drangova, 2019). However, none of these works incorporate the structure of the forward MR imaging

© 2023 Anonymous.

model or offer the flexibility for the reconstruction strategy to depend on the motion parameters. Further, while these approaches provide visually appealing reconstructions, they do not enforce *data consistency* between the estimated output and the acquired measurements. As a result, they are susceptible to producing hallucinations, i.e., visually plausible features not appearing in the true image and inconsistent with the acquired measurements.

Our work is most closely related to two methods that combine neural networks with enforced data consistency. Haskell et al. (2019) use an iterative procedure where a CNN produces an initial motion-corrected image, which is used to initialize an optimization over the motion parameters and motion-free image estimates. Levac et al. (2022a) use a score-based generative model that estimates the log probability of a reconstruction. At test-time, motion parameters and the underlying image are optimized to find a solution which is both data consistent and has a high prior probability under the generative model. Both of these methods incorporate optimizations of both the underlying image and motion parameters. Our proposed strategy reduces this joint optimization to an optimization over just the motion parameters.

Our key insight is that the joint optimization can be simplified using a deep neural network to learn a mapping from proposed motion parameters and corrupted k-space to a motion-corrected reconstruction. Then, inference simplifies to a search at test-time solely over the motion parameters, each of which maps to a reconstructed image. This strategy eliminates the search over images while retaining the benefits of iterative, optimization-based motion parameter estimation. In particular, the discrepancy between acquired measurements and the estimated image and motion parameters can be automatically monitored during the optimization to reject failure cases where the reconstruction is of poor quality.

We demonstrate our intra-slice motion correction method on 2D fast spin echo (FSE) brain MRI. Our method produces reconstructions that are consistent with the acquired k-space measurements in the presence of inter-shot motion or provides an indicator when they are not. The method is trained on simulated pairs of corrupted and corrected examples and generalizes to a proof-of-concept acquired k-space example.

2. Methods

There are two stages to our motion correction procedure: (1) training a network that maps motion parameters to reconstructions and (2) performing a test-time optimization that solves for the motion parameters for an unseen k-space example, which then maps to the final reconstruction via the trained network (Fig. 1). We first describe the MRI forward model and then detail its use in neural network training and motion parameter estimation.

Forward model. MRI acquires Fourier space measurements. The imaging process is described by the forward model:

$$y_i = A_i x + \epsilon, \tag{1}$$

where x is the underlying 2D image we wish to recover, y_i are the acquired MRI measurements from the ith coil element, and ϵ is Rician noise. In the ideal setting where no motion is present, the forward imaging operator A_i for the ith coil is

$$A_i = UFC_i, \tag{2}$$

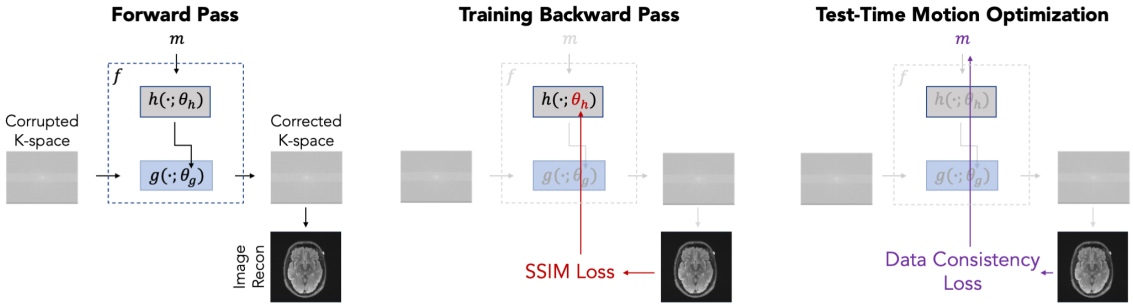

Fig. 1: Information flow through our hypernetwork. During a forward pass, true or estimated motion parameters m serve as input to hypernetwork $h(\cdot;\theta_h)$ which generates the weights θ_g of a reconstruction subnetwork $g(\cdot;\theta_g)$. Reconstruction subnetwork $g(\cdot;\theta_g)$ takes corrupted k-space data as input and produces a reconstruction. The hypernetwork weights θ_h are the only network parameters directly updated during training. At test-time, we freeze θ_h and use the data consistency loss to optimize the motion parameters m.

where C_i denotes a diagonal matrix which encodes the coil-sensitivity profile of the ith coil, F is the 2D Fourier transform, and U encodes the undersampling pattern for the acquisition, such that acquired k-space points have an entry of 1 at the appropriate location on the diagonal, while all other points have a 0.

In the presence of motion, the forward operator A_i is a function of unknown motion parameters m. We consider in-plane rigid-body motion and assume a quasi-static motion model where the 2D object moves between the acquisition of different shots but remains stationary during the acquisition of each individual shot. Under this model,

$$A_i(m) = \sum_s U_s F C_i M_s(m), \tag{3}$$

where U_s encodes the undersampling pattern for shot s and $M_s(m)$ is a motion matrix encoding in-plane rigid translation and rotation during shot s. This forward model differs from Eq. 2 in two ways: (1) it varies across acquired images, which are characterized by different motion parameters, and (2) it is in general unknown because m is unknown.

Training. We instantiate the reconstruction network $f(y, m; \theta)$ with a hypernetwork (Ha et al., 2016) that uses fully-connected layers h operating on motion parameters m to generate weights θ_g of a reconstruction subnetwork g:

$$f(y, m; \theta) = g(y; \theta_g(m)) \text{ where } \theta_g(m) = h(m; \theta_h). \tag{4}$$

Unlike standard supervised deep learning methods which do not incorporate information about the motion parameters into the structure of the neural network (Pawar et al., 2018; Duffy et al., 2021; Levac et al., 2022b; Singh et al., 2022; Küstner et al., 2019; Johnson and Drangova, 2019), our method's parameterization is motion-dependent. The hypernetwork architecture enables flexible prediction of a reconstruction subnetwork g specific to the motion parameters, just as the forward model A is a function of the motion parameters.

Importantly, the weights θ_g of the reconstruction subnetwork are never trained directly. The only trainable parameters in f are θ_h, the weights of the fully connected layers h that

produce θ_g. We use the forward model in Eqns. 1 and 3 to simulate data to train our network. The parameters θ_h are trained using simulated, motion corrupted data:

$$\theta_h^* = \arg\min_{\theta} \mathop{\mathbb{E}}_{x,m} L(f(A(m)x + \epsilon, m; \theta), x), \tag{5}$$

where L is the structural similarity index measure (SSIM) (Wang et al., 2004) loss function, though our method is general and could be used with other losses.

Test-time optimization. Applying Bayes' rule to Eq. 1 and assuming the image estimate x and estimated motion parameters m to be independent yields

$$\begin{aligned} \log p(x, m|y) &= \log p(y|x, m) + \log p(x) + \log p(m) + c \\ &= -||A(m)x - y||^2 + \log p(x) + \log p(m) + c, \end{aligned} \tag{6}$$

where c is a constant term. Maximizing the posterior probability $p(x, m|y)$ involves minimizing the *data consistency* term $||A(m)x - y||^2$ and maximizing priors over the underlying image and motion parameters. The data consistency term is an analytical expression describing how well the estimated image and motion parameters agree with the acquired measurements y under the forward imaging model A, while the priors are often expressed as hand-crafted regularizers (Rudin et al., 1992) or explicitly or implicitly modeled by a neural network (Goodfellow et al., 2020; Kingma and Welling, 2013; Song et al., 2020).

At test-time, we freeze the weights of the reconstruction network and optimize the motion parameters with a data consistency loss for each corrupted measurement y as defined by the first term on the right hand side in Eq. 6:

$$\hat{m} = \arg\min_{m} ||A(m)f(y, m; \theta_h^*) - y||^2. \tag{7}$$

The final image reconstruction is then simply $\hat{x}(y, \hat{m}) = f(y, \hat{m}; \theta_h^*)$. Monitoring the loss in Eq. 7 provides information about how well the motion correction procedure is performing. When this loss is high relative to the total signal energy of the acquired k-space data even after optimization, we discard the reconstruction and/or decide to reacquire the image.

3. Experiments

Data. We demonstrate our approach on 2D T2 FLAIR FSE brain MRI k-space data (3T GE Signa Premier, 48-channel head coil, 6-shot acquisition, TR=10s, TE=118ms, TI=2.6s, FOV=260×260mm^2, slice thickness=5mm, slice spacing=1mm) under an approved IRB protocol. The shots are designed such that the second shot contains the central k-space line and the highest signal energy. The first, third, and fourth shots also contain significant central k-space data and relatively high signal energy, while the fifth and sixth shots largely contain data from the periphery of k-space and have much lower signal energy. We split the dataset into 553/197/100 training/validation/test 2D slices from 31/11/9 subjects respectively, with no subject overlap. We treat the acquired data as motion-artifact-free ground truth and simulate motion artifacts during training and testing. We also generate a realistic test example by mixing acquired k-space measurements $y_{i,1}$ and $y_{i,2}$ from two scans of the same subject in different head positions: $y_i = U_{pre}y_{i,1} + U_{post}y_{i,2}$.

Motion simulation. We simulate motion-corrupted data from a collection of acquired motion-free k-space data. We apply ARC reconstruction (Brau et al., 2008) to data followed by the inverse Fourier transform and root-sum-of-squares coil combination, yielding motion-free image x. We estimate coil sensitivity profiles S_i using ESPIRiT (Uecker et al., 2014; Iyer et al., 2020) and extend the profiles to the image edge via B-spline interpolation. We synthesize motion-corrupted measurements according to Eqns. 1 and 3. We define motion matrix M by selecting a random shot affected by motion and sampling 2D translation parameters $(\Delta_h, \Delta_v) \sim \mathcal{U}(-10\text{mm}^2, 10\text{mm}^2)$ and rotation parameter $\theta \sim \mathcal{U}(-10°, 10°)$. The motion simulation is diagrammed in detail in Appendix A, Fig. 5.

Network architectures. Our hypernetwork takes in motion parameters and, via fully-connected layers, outputs the reconstruction subnetwork weights. The reconstruction subnetwork input is 88-channel real and imaginary components of the ARC parallel imaging reconstruction (Brau et al., 2008) (Orchestra SDK - GE Healthcare) of k-space data from 44 receive coils (4 neck channels were discarded). This input is processed via 6 convolution layers in both image and frequency space (Anonymized), where each intermediate 3×3 convolution outputs 32 channel features. The network has 2.7 million trainable parameters in total. The reconstruction subnetwork outputs 88 k-space channels, which we reconstruct via the inverse Fourier transform and root-sum-of-squares coil combination.

The input to the fully-connected layers are the motion parameters comprising 18 scalars representing x- and y- translations and a rotation for each of 6 shots of k-space. The motion parameters represent the object pose in each of these 6 k-space shots. The motion parameters are processed via 6 fully connected layers with 256 units each.

Implementation details. We normalize input/output k-space by the maximum intensity in the corrupted image. Networks are trained to reconstruct the image (x or $M(m)x$) corresponding to the central k-space line. All models are trained with the SSIM loss function using the Adam optimizer (learning rate 10^{-3}) and batch size 6. The test-time optimization on simulated data uses 4 trials of gradient descent with a cyclical learning rate decaying exponentially from 10^{-6} to 10^{-7}, while the test-time optimization on the realistic test example uses a learning rate of 10^{-9}. We tune hyperparameters and perform model selection on the validation set and hold out the test set for final evaluation. Each network is trained for 250,000 iterations. Optimizing the motion parameters for `HN` takes ∼5 minutes per 2D slice to run the 4 trials sequentially and is trivially parallelizable. After reconstruction, we automatically reject poor reconstructions indicated by a data consistency loss greater than 5% of the total signal energy of the corrupted k-space measurements. We use SSIM to evaluate similarity between reconstructed and ground truth images and exclude a 10 pixel border to prevent reconstruction edge-effects from affecting our results.

Baselines. We analyze two versions of our method: `HN`, which includes all results of our test-time optimization, and `HN-R`, which rejects examples where the data consistency error is high relative to the signal energy. We compare our strategy to four baseline methods.

`ARC` (Brau et al., 2008) is a classical autocalibrating parallel imaging reconstruction that interpolates undersampled k-space regions based on a fully-sampled central calibration region. This method performs no motion correction and is commonly used clinically.

`Conv` (Anonymized) uses a CNN comprised of only $g(\cdot; \theta_g)$ with θ_g trained directly. The network has no dependence on motion parameters and requires no test-time optimization.

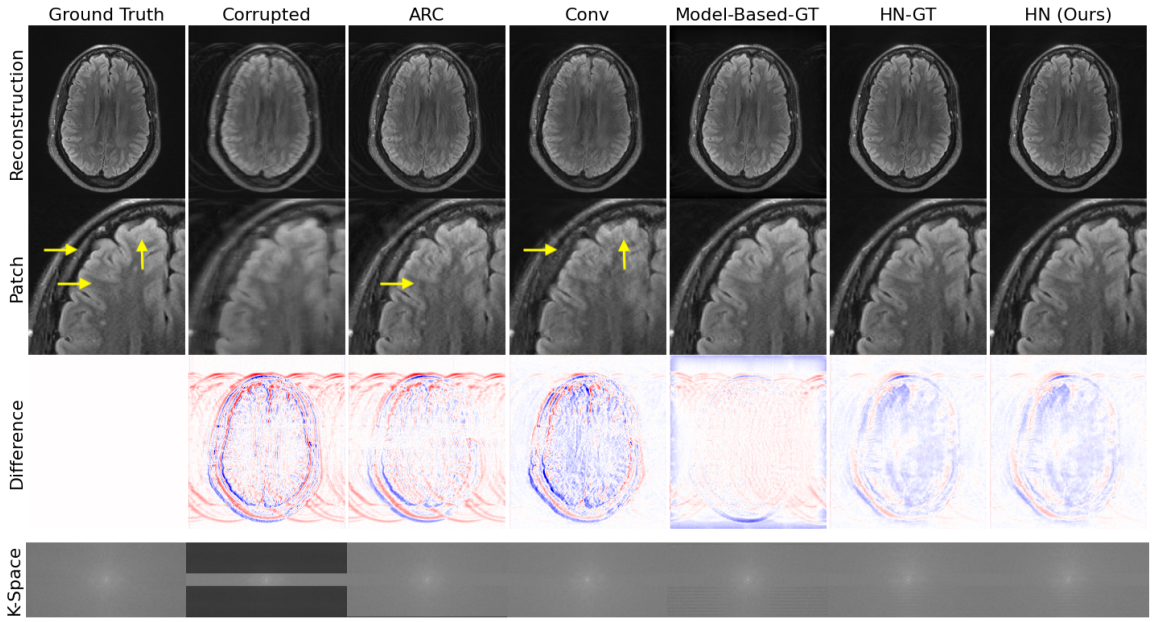

Fig. 2: Simulated test example. Our method, `HN`, outperforms motion-naive baselines (`ARC` and `Conv`) and performs similarly to motion-aware methods (`Model-Based-GT` and `HN-GT`) without access to any motion parameters.

The network architecture is identical to that of $g(\cdot;\theta_g)$, but convolution size and depth are varied such that the number of trainable parameters matches our method, where the only trainable parameters are the weights of the fully connected hypernetwork. This baseline characterizes a state-of-the-art deep learning alternative to `ARC` and serves as an ablation to isolate the difference between a standard feedforward network and our hypernetwork.

`Model-Based-GT` (Gallichan et al., 2016) assumes known motion parameters and applies analytical corrections. Translations and rotations are corrected via k-space phase shifts and rotations and the result is reconstructed via an inverse Non-Uniform Fast Fourier Transform (NUFFT) (Fessler and Sutton, 2003; Muckley et al., 2020). This method is an upper bound on joint estimation performance because it assumes known motion parameters.

`HN-GT` computes the output of our hypernetwork when ground truth motion parameters are provided as input. This characterizes a motion-informed deep learning method and, similar to `Model-Based-GT`, represents the best-case scenario for reconstruction.

4. Results

Fig. 2 shows reconstructions on a simulated example. Our method, `HN`, produces sharper, more accurate reconstructions than `ARC` and `Conv`, which are 'motion-naive', i.e., do not incorporate motion information. It also provides reconstructions of similar quality to 'motion-aware' methods that have access to known motion parameters (`Model-Based-GT` and `HN-GT`).

Fig. 3a shows that our method outperforms motion-naive methods across the test set and is on par with the motion-aware methods. Fig. 3b shows that our method yields better reconstructions than motion-naive baselines for the vast majority of subjects. Fig. 3c

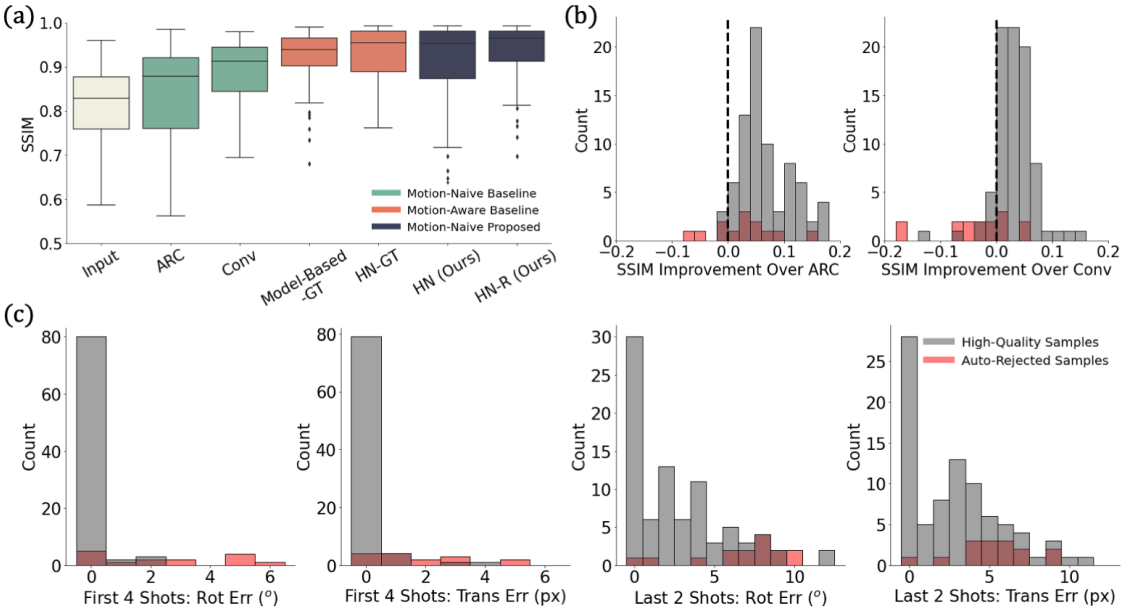

Fig. 3: Quantitative simulated results. (a) Reconstruction quality of all methods. (b) SSIM improvement over motion-naive methods. Our method improves on the baseline for bars right of the dashed line. (c) Motion parameter estimate errors from our method in the four high-energy shots and two low-energy shots. Our method outperforms motion-naive baselines and performs comparably to motion-aware ones. The automated rejection strategy enabled by our method effectively identifies examples where the optimization fails.

demonstrates that our method recovers accurate motion parameters for the first four high-energy shots, with most inaccurate cases being automatically rejected.

Our method does not reliably recover motion parameters for the last two shots. These low-energy shots play a relatively small role in the data consistency optimization but also have low impact on the resulting reconstruction. Of fifteen automatically rejected cases, motion occurred between the second and third shots in nine examples. This splits the first four high-energy shots between two positions and causes hard-to-correct artifacts.

Finally, Fig. 4 visually illustrates the performance of motion-naive methods on the realistic test example that combines acquired k-space data from the same subject in two different positions. `HN` provides the best reconstruction among methods that do not require access to the true motion parameters, demonstrating that it successfully reconstructs motion-corrupted real k-space signals despite using only simulated data during training.

5. Discussion and Conclusions

Limitations. While we demonstrate that our approach generalizes to acquired k-space data, our evaluation focuses on simulated data. We do not model through-plane motion, a known source of significant artifact. Less well-studied effects including intra-shot motion, spin history, and signal decay during the FSE echo train may also affect our method.

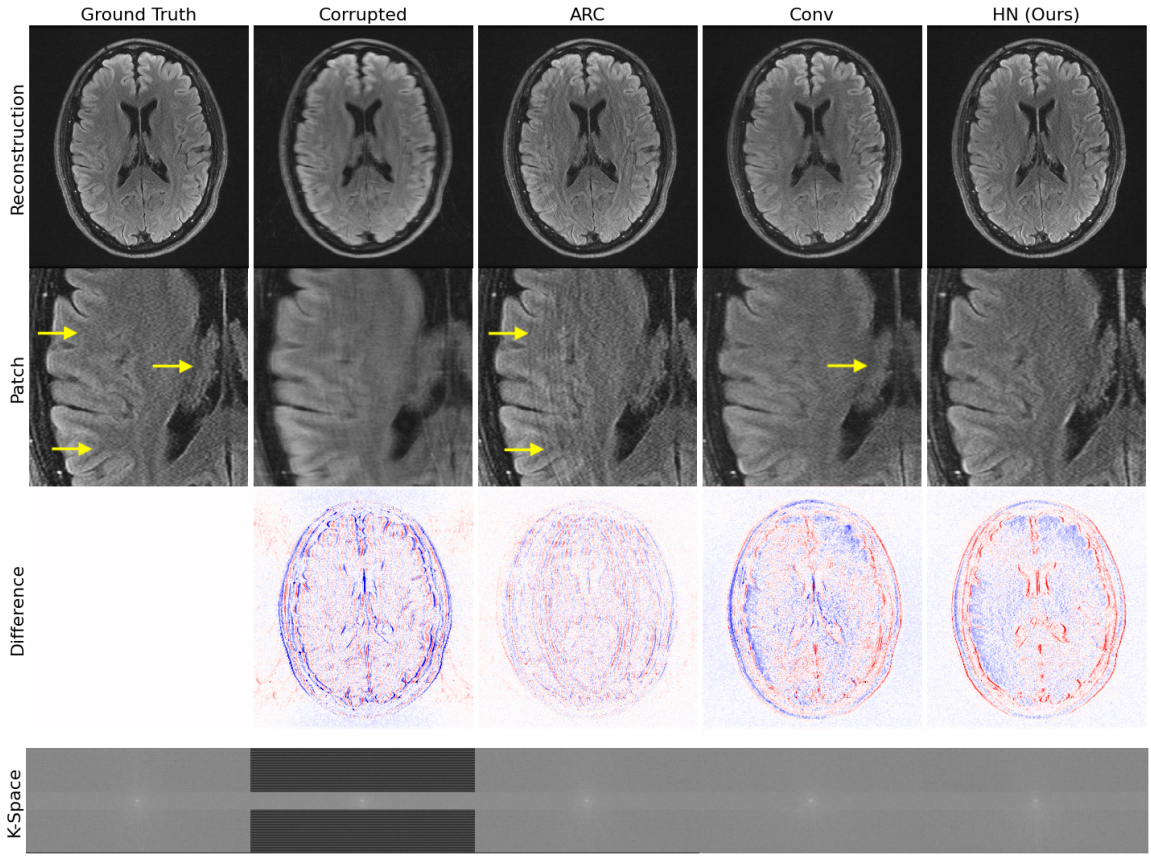

Fig. 4: Realistic test example. For this example, we do not have access to the true motion parameters, and our method (`HN`) outperforms the baselines which do not require this information. `HN` removes the artifact in `ARC` and is sharper than `Conv` (see arrows). Despite being trained on simulated data, our method generalizes to this test example based on acquired k-space data and produces a high-quality reconstruction.

Future work. Future work will thus investigate, model, and correct these additional motion effects while retaining data-consistent reconstructions in larger scale, clinical k-space datasets. We also aim to extend our method to MRI sequences with distinct k-space acquisition patterns and contrasts. Beyond MRI, our strategy for learning physically consistent reconstructions given a partially-unknown forward model could be applied to other semi-blind inversion problems, e.g. in seismic imaging or computational photography.

Conclusions. This work develops a general deep rigid motion correction approach for multi-shot MRI that we evaluate on 2D FSE T2 FLAIR data. The proposed framework first learns a mapping between corrupted k-space data, true motion parameters, and high-quality reconstructions. At test-time, only the motion parameter estimates are optimized, yielding data-consistent reconstructions without the complexity of a joint search over images and motion parameters. The approach includes a strategy for automatically rejecting samples where the optimization fails. Our method reliably produces high-quality reconstructions in simulation and generalizes to a proof-of-concept realistic acquired k-space example.

References

Jalal B Andre, Brian W Bresnahan, Mahmud Mossa-Basha, Michael N Hoff, C Patrick Smith, Yoshimi Anzai, and Wendy A Cohen. Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical MR examinations. *Journal of the American College of Radiology*, 12(7):689–695, 2015.

Anonymized.

Roland Bammer, Murat Aksoy, and Chunlei Liu. Augmented generalized sense reconstruction to correct for rigid body motion. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 57(1):90–102, 2007.

PG Batchelor, D Atkinson, P Irarrazaval, DLG Hill, J Hajnal, and D Larkman. Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(5):1273–1280, 2005.

Anja CS Brau, Philip J Beatty, Stefan Skare, and Roland Bammer. Comparison of reconstruction accuracy and efficiency among autocalibrating data-driven parallel imaging methods. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 59(2):382–395, 2008.

Lucilio Cordero-Grande, Rui Pedro AG Teixeira, Emer J Hughes, Jana Hutter, Anthony N Price, and Joseph V Hajnal. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Transactions on Computational Imaging*, 2(3):266–280, 2016.

Ben A Duffy, Lu Zhao, Farshid Sepehrband, Joyce Min, Danny JJ Wang, Yonggang Shi, Arthur W Toga, Hosung Kim, Alzheimer's Disease Neuroimaging Initiative, et al. Retrospective motion artifact correction of structural MRI images using deep learning improves the quality of cortical surface reconstructions. *Neuroimage*, 230:117756, 2021.

Jeffrey A Fessler and Bradley P Sutton. Nonuniform fast fourier transforms using min-max interpolation. *IEEE transactions on signal processing*, 51(2):560–574, 2003.

Robert Frost, Paul Wighton, F Işık Karahanoğlu, Richard L Robertson, P Ellen Grant, Bruce Fischl, M Dylan Tisdall, and André van der Kouwe. Markerless high-frequency prospective motion correction for neuroanatomical MRI. *Magnetic resonance in medicine*, 82(1):126–144, 2019.

Daniel Gallichan, José P Marques, and Rolf Gruetter. Retrospective correction of involuntary microscopic head movement using highly accelerated fat image navigators (3D fatnavs) at 7T. *Magnetic resonance in medicine*, 75(3):1030–1039, 2016.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

Melissa W Haskell, Stephen F Cauley, Berkin Bilgic, Julian Hossbach, Daniel N Splitthoff, Josef Pfeuffer, Kawin Setsompop, and Lawrence L Wald. Network accelerated motion estimation and reduction (NAMER): convolutional neural network guided retrospective motion correction using a separable motion model. *Magnetic resonance in medicine*, 82 (4):1452–1461, 2019.

Siddharth Iyer, Frank Ong, Kawin Setsompop, Mariya Doneva, and Michael Lustig. SURE-based automatic parameter selection for ESPIRiT calibration. *Magnetic Resonance in Medicine*, 84(6):3423–3437, 2020.

Patricia M Johnson and Maria Drangova. Conditional generative adversarial network for 3d rigid-body motion correction in MRI. *Magnetic Resonance in Medicine*, 82(3):901–910, 2019.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Thomas Küstner, Karim Armanious, Jiahuan Yang, Bin Yang, Fritz Schick, and Sergios Gatidis. Retrospective correction of motion-affected MR images using deep learning frameworks. *Magnetic resonance in medicine*, 82(4):1527–1540, 2019.

Brett Levac, Ajil Jalal, and Jonathan I Tamir. Accelerated motion correction for MRI using score-based generative models. *arXiv preprint arXiv:2211.00199*, 2022a.

Brett Levac, Sidharth Kumar, Sofia Kardonik, and Jonathan I Tamir. Fse compensated motion correction for MRI using data driven methods. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 707–716. Springer, 2022b.

William A Mehan, R Gilberto Gonzalez, Bradley R Buchbinder, John W Chen, William A Copen, Rajiv Gupta, Joshua A Hirsch, George J Hunter, Scott Hunter, Jason M Johnson, et al. Optimal brain MRI protocol for new neurological complaint. *PloS one*, 9(10): e110803, 2014.

Matthew Muckley, Ruben Stern, Tullie Murrell, and Florian Knoll. TorchKbNufft: A high-level, hardware-agnostic non-uniform fast Fourier transform. In *ISMRM Workshop on Data Sampling & Image Reconstruction*, 2020. Source code available at https://github.com/mmuckley/torchkbnufft.

Kamlesh Pawar, Zhaolin Chen, N Jon Shah, and Gary F Egan. Motion correction in MRI using deep convolutional neural network. In *Proceedings of the ISMRM Scientific Meeting & Exhibition, Paris*, volume 1174, 2018.

Daniel Polak, Daniel Nicolas Splitthoff, Bryan Clifford, Wei-Ching Lo, Susie Y Huang, John Conklin, Lawrence L Wald, Kawin Setsompop, and Stephen Cauley. Scout accelerated motion estimation and reduction (SAMER). *Magnetic Resonance in Medicine*, 87(1): 163–178, 2022.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Nalini M Singh, Juan Eugenio Iglesias, Elfar Adalsteinsson, Adrian V Dalca, Polina Golland, et al. Joint frequency and image space learning for MRI reconstruction and analysis. *The Journal of Machine Learning in Biomedical Imaging*, 2022.

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

M Dylan Tisdall, Aaron T Hess, Martin Reuter, Ernesta M Meintjes, Bruce Fischl, and André JW van der Kouwe. Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magnetic resonance in medicine*, 68(2): 389–399, 2012.

Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic resonance in medicine*, 71(3):990–1001, 2014.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Appendix A. Motion Simulation

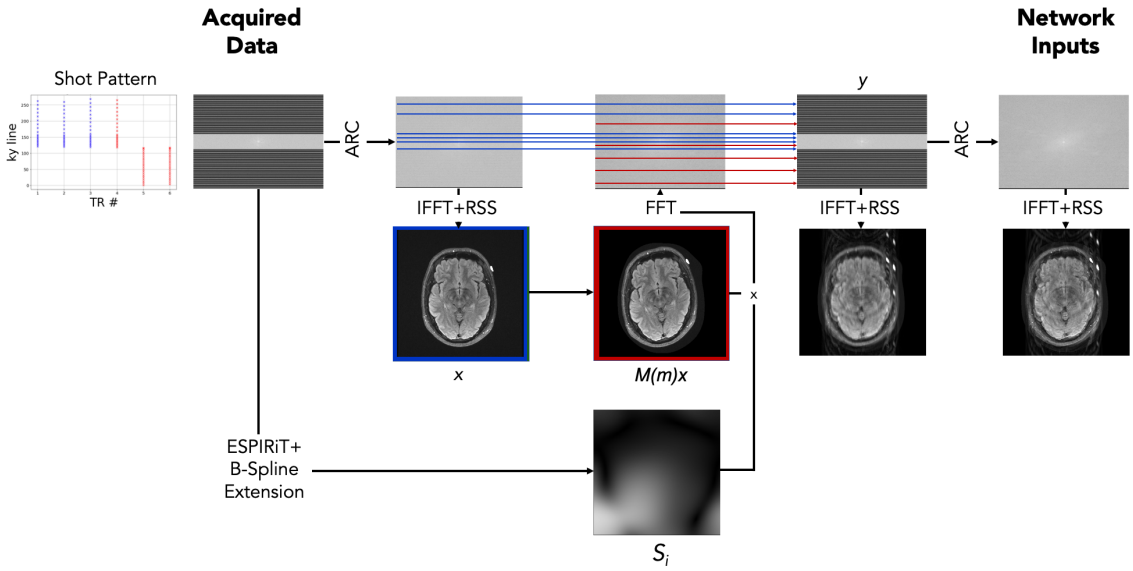

Fig. 5: Motion simulation schematic. We apply ARC reconstruction (Brau et al., 2008) to motion-free data followed by the inverse Fourier transform and root-sum-of-squares coil combination, yielding the initial motion-free image x. We also estimate coil sensitivity profiles S_i from the acquired data using ESPIRiT (Uecker et al., 2014) with learned parameter estimation (Iyer et al., 2020) and extend the profiles to the image edge via B-spline interpolation. Next, we simulate an image under a sampled random motion $M(m)$ by applying rotations and translations to the image and use the sensitivity maps and a Fourier transform to simulate k-space data corresponding to the moved position. Based on the shot pattern for the acquisition, we combine k-space data from the appropriate lines corresponding to the position pre- (blue) and post-motion (red) to form simulated, motion-corrupted k-space measurements y. This simulates the k-space data that would have been acquired had the subject moved from the blue position to the red position over the course of the acquisition. The input to our networks is the ARC reconstruction of this simulated y. In practice, we sample two versions of $M(m)x$ and treat one of the two as x, to avoid discrepancies between simulated and acquired data when mixing the k-space.