



Published in final edited form as:

*Magn Reson Med.* 2020 August ; 84(2): 800–812. doi:10.1002/mrm.28177.

## Reconstruction of undersampled 3D non-Cartesian image-based navigators for coronary MRA using an unrolled deep learning model

Mario O. Malavé<sup>1</sup>, Corey A. Baron<sup>2</sup>, Srivathsan P. Koundinyan<sup>1</sup>, Christopher M. Sandino<sup>1</sup>, Frank Ong<sup>1</sup>, Joseph Y. Cheng<sup>1,3</sup>, Dwight G. Nishimura<sup>1</sup>

<sup>1</sup>Magnetic Resonance Systems Research Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA

<sup>2</sup>Department of Medical Biophysics, Western University, London, ON, Canada

<sup>3</sup>Department of Radiology, Stanford University, Stanford, CA

### Abstract

**Purpose:** To rapidly reconstruct undersampled 3D non-Cartesian image-based navigators (iNAVs) using an unrolled deep learning (DL) model, enabling nonrigid motion correction in coronary magnetic resonance angiography (CMRA).

**Methods:** An end-to-end unrolled network is trained to reconstruct beat-to-beat 3D iNAVs acquired during a CMRA sequence. The unrolled model incorporates a nonuniform FFT operator in TensorFlow to perform the data-consistency operation, and the regularization term is learned by a convolutional neural network (CNN) based on the proximal gradient descent algorithm. The training set includes 6,000 3D iNAVs acquired from 7 different subjects and 11 scans using a variable-density (VD) cones trajectory. For testing, 3D iNAVs from 4 additional subjects are reconstructed using the unrolled model. To validate reconstruction accuracy, global and localized motion estimates from DL model-based 3D iNAVs are compared with those extracted from 3D iNAVs reconstructed with  $I_1$ -ESPIRiT. Then, the high-resolution coronary MRA images motion corrected with autofocusing using the  $I_1$ -ESPIRiT and DL model-based 3D iNAVs are assessed for differences.

**Results:** 3D iNAVs reconstructed using the DL model-based approach and conventional  $I_1$ -ESPIRiT generate similar global and localized motion estimates and provide equivalent coronary image quality. Reconstruction with the unrolled network completes in a fraction of the time compared to CPU and GPU implementations of  $I_1$ -ESPIRiT (20× and 3× speed increases, respectively).

**Conclusions:** We have developed a deep neural network architecture to reconstruct undersampled 3D non-Cartesian VD cones iNAVs. Our approach decreases reconstruction time for

---

**Correspondence:** Mario O. Malavé, Packard Electrical Engineering, Room 355, 350 Serra Mall, Stanford, CA 94305-9510. momalave@gmail.com.

#### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

3D iNAVs, while preserving the accuracy of nonrigid motion information offered by them for correction.

## Keywords

3D cones trajectory; convolutional neural networks; coronary MRA; non-Cartesian

## 1 | INTRODUCTION

We have previously developed an approach for free-breathing whole-heart coronary magnetic resonance angiography (CMRA)<sup>1,2</sup> using an alternating-repetition time (ATR) balanced steady-state free precession (bSSFP) sequence.<sup>3</sup> The high-resolution imaging data is collected with a non-Cartesian 3D cones trajectory. For translational and nonrigid respiratory motion tracking, beat-to-beat 3D image-based navigators (iNAVs) of the heart are acquired<sup>2,4</sup> using an accelerated variable-density (VD) 3D cones sampling technique.<sup>1,5</sup> The reconstruction of 3D iNAVs using compressed sensing (i.e.,  $l_1$ -ESPIRiT<sup>1,6</sup>) is a time-consuming process, as each scan involves the collection of several hundred 3D iNAVs.

Deep learning (DL) has the potential to reduce reconstruction times for undersampled MRI data. Convolutional neural networks (CNNs) have recently become a powerful tool for image reconstruction. CNNs are popular due to their ease of use, accuracy, and fast inference time. CNN architectures generally operate in the image domain and are trained to minimize a specific loss function with respect to a “ground truth” image. One of the issues with a pure CNN architecture is the lack of incorporating physics specific to the application, leading to a “black-box” DL approach. Such an approach requires a very large number of training datasets, and can lead to issues with image quality as well as convergence during training. Through the use of a DL model-based architecture, more sophisticated techniques that incorporate CNNs with previous iterative reconstruction methods can provide improved accuracy while reducing the demand for training data and training time.<sup>7</sup>

Among the several DL model-based approaches that have been proposed,<sup>8,9</sup> an unrolled network architecture<sup>10–14</sup> has emerged as a promising technique. Here, images are reconstructed by unfolding the proximal gradient descent (PGD) algorithm<sup>15</sup> and learning the regularization functions and coefficients. Prior studies leveraging unrolled networks have been limited to contexts involving Cartesian acquisitions. For reconstruction of non-Cartesian datasets, only image-to-image CNNs have been investigated.<sup>16</sup> In this work, we modify the unrolled model architecture to accommodate non-Cartesian 3D k-space datasets by incorporating a nonuniform Fast Fourier Transform (NUFFT) operator. This would enable the rapid reconstruction of the undersampled 3D iNAV datasets acquired in our CMRA sequence.

## 2 | METHODS

### 2.1 | Imaging data and 3D iNAV acquisition

Beat-to-beat undersampled 3D iNAVs are acquired as part of the CMRA sequence shown in Supporting Information Figure S1A. Specifically, free-breathing high-resolution CMRA data

( $28 \times 28 \times 14 \text{ cm}^3$  FOV, 1.2 mm isotropic spatial resolution, and 500–600 total heartbeat scan time) are collected with a 3D cones trajectory using ATR-bSSFP.<sup>1,2,17</sup> The 3D iNAVs are acquired in the same volumetric region and after the segmented full-resolution acquisition (i.e., 500–600 separate 3D iNAVs are acquired and reconstructed) by continuing the ATR-bSSFP sequence to maintain similar image contrast.<sup>1</sup> The 3D cardiac datasets were acquired on a 1.5T GE Signa system with an 8-channel cardiac coil using VD trajectories consisting of 32 cone readouts with 4.4 mm isotropic spatial resolution, yielding an acceleration factor of 9 due to undersampling<sup>18</sup> two different trajectories that were based on either sequential or phyllotaxis 3D iNAV designs.<sup>19</sup> Details for both trajectories are shown in Supporting Information Figure S1B,C.

To correct for rigid and nonrigid respiratory motion, an autofocusing technique is applied to the high-resolution data. This motion compensation method requires accurate localized motion estimates, which are derived from the 3D iNAVs following their reconstruction with a computationally expensive iterative optimization approach (i.e.,  $I_1$ -ESPIRiT). The solution to this optimization problem, however, can be derived in an accelerated fashion using an unrolled DL model.

## 2.2 | 3D iNAV reconstruction

**2.2.1 | Iterative algorithm and unrolled network overview**—The unrolled network is based on PGD, which solves the following inverse problem with the image  $x$ , k-space data  $y$ , encoding operator  $A$ , and regularization term  $R(x)$ :

$$\underset{x}{\text{minimize}} \frac{1}{2} \|Ax - y\|_2^2 + \lambda R(x) \quad (1)$$

The solution, which is found using proximal gradient descent, iterates between the data-consistency and proximal operator steps:

$$x^{k+1} = P_{\lambda R}(x^k - \alpha A^T(Ax^k - y)) \quad (2)$$

The proximal operator of the regularization function is  $P_{\lambda R}$ , defined as:

$$P_{\lambda R}(v) = \underset{u}{\text{arg min}} \left( R(u) + \frac{1}{2\lambda} \|u - v\|_2^2 \right) \quad (3)$$

When using non-Cartesian data, the acquisition model  $A$  incorporates the SENSE reconstruction<sup>20</sup> operator  $\mathcal{S}$  (coil sensitivity maps computed using ESPIRiT<sup>6</sup>), and the NUFFT operator,  $F_{\text{NUFFT}}$ . The regularization term,  $\lambda R(x)$ , is implicitly learned by replacing the proximal operator  $P_{\lambda R}$  with a CNN to obtain the next iteration  $x^{k+1}$ . For  $I_1$ -ESPIRiT, the regularization function is the  $I_1$ -norm of the wavelet transform applied to the image,  $x$ . In this case, PGD simplifies to the iterative soft-shrinkage algorithm (ISTA).<sup>21</sup>

The data-consistency step is important because it allows the model to incorporate the measured k-space data in each iteration. There are two ways of implementing data-consistency: hard-projection, and soft-projection. To apply hard-projection, the trajectory undersampling is performed using subsampling (i.e., not collecting certain data points from

the fully sampled k-space trajectory). After the acquisition, the uncollected data points are zero-filled. For each iteration of  $I_1$ -ESPIRiT or the unrolled model, the Fourier transform of the 3D image is taken (after previous regularization, i.e., wavelet or CNN), and the data in the acquired k-space trajectory locations are replaced with the original measured data while allowing for the zero-filled locations to update (Figure 1A).

When using compressed sensing, VD sampling has been shown to work well.<sup>22</sup> Many different k-space sampling techniques have been developed using VD Cartesian subsampling.<sup>23–26</sup> For example, in Cheng et al,<sup>26</sup> a Cartesian subsampling design called variable-density sampling and radial view ordering (VDRad) is used which approximates VD spirals on the Cartesian grid. The current 3D iNAVs use a true VD non-Cartesian design<sup>1</sup> instead of the subsampled Cartesian approach. The VD non-Cartesian cones design is not a subsampled version of a fully sampled trajectory; therefore, there are not any uncollected k-space data points to zero-fill and utilize during the regularization and hard-projection steps. Thus, soft-projection is used for data-consistency. In summary, for the Cartesian approach, hard-projection or soft-projection can be applied, but the VD non-Cartesian design is limited to soft-projection. To further illustrate the trajectory differences, a subsampled Cartesian trajectory, and a VD non-Cartesian trajectory are shown in Figure 1B.

One key difference with the prior Cartesian model<sup>11</sup> and proposed non-Cartesian model is the replacement of the data-consistency step using the FFT with a gradient descent (GD) update step (i.e., soft-projection) using the NUFFT, which maintains consistency with the measured non-Cartesian k-space data. The step size for the GD update step,  $\alpha$ , is initialized to 2 and left as a learnable parameter for the model to allow for improved training flexibility. Also, soft-projection for data-consistency can potentially give improved results when the k-space measurements are noisy, since a GD approach is used.<sup>27</sup> The non-Cartesian unrolled model is summarized below.

Unrolled Network Problem:

$$J_{ObjectiveFunction} = \frac{1}{2} \|Ax - y\|_2^2 + \lambda R(x)$$

$$J_{DataConsistency} = \frac{1}{2} \|Ax - y\|_2^2$$

$A = F_{NUFFT}S$  (acquisition model = NUFFT operator, SENSE operator)

$CNN = \lambda R(x)$  (wavelet regularization is replaced with a CNN)

$y$  = input raw k-space data

$x_{out}$  = output of the network

$\alpha$  = initialize to 2 (learnable parameter)

$k$  = index for each iteration ( $N$  = number of iterations)

$x^k$  = output of the data-consistency step

$x_{cnn}^k$  = output of the CNN

1. Initialize  $x_{cnn}^0 = A^T y$
2. Data-consistency (soft projection/gradient update):  

$$x^{k+1} = x_{cnn}^k - \alpha \nabla (J_{DataConsistency}) = x_{cnn}^k - \alpha A^T (A x_{cnn}^k - y)$$
3. CNN (regularization step):  $x_{cnn}^{k+1} = CNN(x^{k+1})$  (Equation 2 with  $P_{\lambda R} = CNN$ )
4. Repeat Steps 2–3 for  $(N-1)$  iterations
5.  $x_{out} = x_{cnn}^N$

**2.2.2 | Neural network architecture**—The unrolled model architecture uses 4 gradient steps ( $N=4$  iterations) consisting of 2 ( $M=2$ ) residual network (ResNet)<sup>28</sup> blocks/step. The hyperparameters were initially chosen to match<sup>13</sup> and empirically tuned to ensure convergence given the memory constraints for the current application. The input into the network is the undersampled 3D complex k-space data, k-space coordinates (to generate the NUFFT operator), and the respective coil sensitivity maps for each channel (for the SENSE operator). The ground truth used for training is the reconstructed image when using  $I_1$ -ESPIRiT. Each gradient step begins with data-consistency which uses the forward and transpose acquisition model  $A$  and  $A^T$  to apply soft projection where  $x^0$  is initialized as  $A^T y$ . The complex image is then separated into 2 channels consisting of the real and imaginary components. Next, the network uses  $M$  ResNet blocks comprised of two 3D convolutional layers with a kernel size of  $3 \times 3 \times 3$  and filter depth of 64. Also, each convolutional layer is preceded by a rectified linear unit preactivation layer as recommended in He et al.<sup>29</sup> An additional layer is added to the end of each unrolled step which outputs 2 channels for the real and imaginary parts of the k-space data and uses a linear activation to preserve the sign of the data. The final layer is also added to a skip connection from the input of the first ResNet block to accelerate training convergence. For previous Cartesian approaches,<sup>13,14</sup> circular convolutions were used to handle the “wrap-around” coherent aliasing artifacts. However, while the FFT causes periodic boundary conditions in the image domain, here the NUFFT operation does not because of image cropping and zero-padding that it uses, and the noise-like aliasing properties of undersampled 3D cones.<sup>4</sup> Accordingly, zero-padded convolutions (i.e., noncircular convolutions) were applied in each convolutional layer. The gradient step block is then repeated 3 (i.e.,  $N-1$ ) more times for a total of  $N=4$  iterations. Also, the network is trained using the complex  $I_1$  loss and a batch size of 1 (stochastic gradient descent). While the  $I_2$  loss has been used to successfully train other reconstruction networks (e.g.,<sup>7</sup>), using the  $I_1$  loss has been shown to improve convergence and produce sharper images.<sup>30,31</sup> A graphical representation of the unrolled architecture is shown in Figure 2A,B. In Figure 2B, the data-consistency for the prior<sup>11</sup> and proposed models are shown using a hard-projection and soft-projection respectively.

## 2.3 | Training data

The training set includes a total of seven subjects. Four of the seven subjects were scanned with both the sequential-based and phyllotaxis-based 3D iNAVs acquired each heartbeat which gives a total of 11 scans. Of the 11 scans, 5 were acquired with the trajectories rotated by the golden angle between each heartbeat (Supporting Information Figure S1D) to vary the aliasing artifacts, which serves as a form of data augmentation to improve the performance and ability of the model to generalize. Each scan collected data for 500–600 heartbeats, thus yielding approximately 6,000 total iNAV datasets used for training. The ground truth images were reconstructed with  $I_1$ -ESPIRiT using the Berkeley Advanced Reconstruction Toolbox (BART) toolbox.<sup>32</sup> The  $I_1$ -ESPIRiT reconstruction parameters were empirically determined using 50 iterations, step size of  $1e-6$ , and wavelet regularization ( $\lambda = 0.05$ ). To alleviate the aliasing artifacts from objects outside of the FOV during training, the ground truth datasets were reconstructed at  $(2 \times FOV_x, 1.25 \times FOV_y, 1.5 \times FOV_z)$ ; that is,  $(128_x, 80_y, 48_z)$  with a native matrix size of  $(64_x, 64_y, 32_z)$ . Additionally, training was performed without data-consistency using one iteration for a “black-box” approach (i.e.,  $N=1$  iteration and  $M=8$  ResNet blocks/step with the same kernel size of  $3 \times 3 \times 3$  and filter depth of 64) to examine the differences compared to the unrolled model (i.e., data-consistency and CNN blocks). The reconstructions were run on two different Linux systems with 2.20 GHz Xeon E5–2650 v4 CPU, 512 GB RAM with 48 total cores, and a 3.70 GHz Intel i7-8700K CPU, 64 GB RAM with 12 total cores. Also, the reconstructions were performed on two different GPUs using an NVIDIA Titan XP with 12 GBs of GDDR5X memory, and an NVIDIA Titan RTX GPU with 24 GBs of ultra-fast GDDR6 memory.

**2.3.1 | Computation**—When using the NUFFT operator, there is an increase in computation compared to the standard FFT. The NUFFT operator requires additional steps involving density compensation, convolution with a gridding kernel (Kaiser-Bessel),<sup>33</sup> sampling on the Cartesian grid, oversampled FFTs, and an apodization correction. This can introduce challenges when reconstructing undersampled datasets which require larger matrix sizes to alleviate aliasing artifact. When using the NUFFT operator with the unrolled model, the increased computation increases the memory requirements for training, thus limiting the matrix size.

An efficient NUFFT operator is implemented in Python using native TensorFlow functions so that the overall operator can be back-propagated. To reduce memory overhead, a pruned FFT implementation is used to compute the oversampled FFT step.<sup>34</sup> In particular, pruned FFTs avoid extra memory allocation by exploiting structures of the radix-2 Cooley-Tukey FFT.<sup>35</sup> Note that pruned FFTs are also used in BART so that our computation time comparisons are fair. The gridding and re-gridding steps are implemented using the gather and scatter functions provided by Tensorflow. At its essence, our 3D NUFFT implementation entails multiplications and additions, which are operations that are compatible with standard back-propagation procedures in TensorFlow.

The overall model is an end-to-end unrolled network where the NUFFT operator used during the data-consistency step is incorporated into the TensorFlow graph using python. In this way, the auto-differentiation framework in TensorFlow is leveraged with the NUFFT

operator (i.e., the parameters are trained as part of the entire network). To satisfy memory requirements, the proposed model architecture was trained on an NVIDIA Titan RTX GPU with 24 GBs of ultra-fast GDDR6 memory.

## 2.4 | Motion correction with 3D iNAVs

To correct for respiratory motion in the high-resolution data, 3D global and localized motion estimates are calculated<sup>1,5,36</sup> using both the  $I_1$ -ESPIRiT and model-reconstructed 3D iNAVs. A technique similar to the previous state-of-the-art method in Luo et al<sup>5</sup> was applied for motion-estimate calculation when using the 3D iNAVs:

1. The motion estimates are generated by selecting a region of interest (ROI) mask that covers the heart in the axial, sagittal, and coronal planes. Then, a reference 3D iNAV time frame (heartbeat) is determined using mutual information<sup>37,38</sup> and the similarity matrix approach.
2. Global translational motion estimates are calculated by minimizing the mean-squared difference cost function between the current 3D iNAV frame with the previously determined reference 3D iNAV.
3. To take advantage of the localized spatial information from the 3D iNAVs, residual 3D displacement fields are also calculated, using the MATLAB Imaging Processing Toolbox (The Mathworks, Natick, Massachusetts), relative to the reference 3D iNAV (after aligning the 3D iNAVs using the global motion estimates) and used to determine five unique spatial regions (or bins) of localized residual motion. The bins are obtained with k-means clustering (minimizing the  $l_2$ -norm distance metric), and using the displacement field estimates (in the selected ROI) as the features.
4. The mean of all the features within each calculated bin (plus the global motion estimate) is then used as the residual motion estimate for the bin.

We then apply a linear phase modulation term generated using the motion estimates for each heartbeat in k-space to generate a bank of 6 3D motion-compensated reconstructions from one global motion estimate, and 5 residual localized motion estimates (all 5 applied on top of the global estimates). For the final nonrigid autofocused image, the reconstruction is performed on a pixel-by-pixel basis by choosing the pixel from the bank of 3D motion-compensated reconstructions that minimizes the gradient entropy value at each pixel.<sup>1,36</sup>

## 2.5 | Inference and testing

Four additional subjects were scanned with the previously mentioned navigator designs (2 sequential-based, 1 rotated sequential-based, and 1 rotated phyllotaxis-based) to test the generalization of the unrolled model. The motion information of the 3D iNAVs when using  $I_1$ -ESPIRiT and the unrolled model is then assessed by examining the motion estimates, autofocusing outcomes, and right coronary artery (RCA) and left coronary artery (LCA) images. The motion estimate similarity between the  $I_1$ -ESPIRiT and DL model-based 3D iNAVs is determined by calculating the correlation coefficients of the left/right (L/R), anterior/posterior (A/P), and superior/inferior (S/I) motion estimates for the global and five spatial bins. Furthermore, the autofocusing outcomes were analyzed by computing the

“autofocusing histograms” for each volunteer. The autofocusing histograms show the occurrence of each selected bin that minimized the gradient entropy for each pixel in the final high-resolution image. Also, oblique reformatted maximum intensity projection (MIP) images of the RCA and LCA are shown with cross-sectional views of the vessels before and after motion correction when using the  $l_1$ -ESPIRiT and DL model-based 3D iNAV's for autofocusing. To assess the vessel sharpness of the motion-corrected images, quantitative measurements of coronary vessel sharpness were obtained using the image edge profile acutance (IEPA) metric<sup>39</sup> similar to the previous method in Malavé et al.<sup>17</sup> The IEPA measurements were performed on the RCA and LCA (left main coronary artery (LMCA) and the left anterior descending (LAD) artery) along 50 mm coronary segments made at 1 mm intervals, and with 10 evenly spaced profile lines drawn perpendicular to the lumen axis. The IEPA values range from 0 to 1 where higher values correspond to sharper edges.

### 3 | RESULTS

#### 3.1 | 3D iNAV reconstructions

Subjects 3–4 (test dataset) 3D iNAV inputs (gridded images), outputs (CNN), outputs (data-consistency and CNN), and ground truths (after  $l_1$ -ESPIRiT) are shown in Figure 3 with their respective axial, sagittal, and coronal slices. Subjects 1–2 are shown in Supporting Information Figure S2. For the CNN output, a denoising/smoothing effect is observed, but introduces artifacts for all subjects (Figure 3 and Supporting Information Figure S2) which are not present in the unrolled model (data-consistency and CNN) and  $l_1$ -ESPIRiT outputs. Also, for subjects 3–4 (Figure 3), the denoising for the CNN did not perform as well and blurred parts of the image compared to the unrolled model (e.g., in the coronal slices near the apex of the heart). A few differences are highlighted using white arrows in Figure 3 and Supporting Information Figure S2. Perhaps, with more subjects for training, the “black-box” approach can become more competitive with our proposed approach. However, the need for large amounts of training data may be a disadvantage for many applications. Additionally, the  $l_1$  loss for each subject is shown in Table 1. For each subject, the unrolled model (data-consistency and CNN) outperformed the “black-box” CNN approach. For the unrolled model, an example output of each iteration during training is shown in Supporting Information Figure S3. For the initial iterations, certain structures may not be visible or may appear blurred due to the undersampling which introduces aliasing and interpolation errors when compared to later iterations. For each iteration, image depiction is improved by denoising the image and enhancing structure which mimics multiple iterations for  $l_1$ -ESPIRiT. When employing the trained architecture, the undersampled cardiac images (compared to the outcomes from gridding) retained structural features as a result of the denoising/smoothing operation. More specifically, the aliasing artifacts arising from undersampling a cones trajectory were effectively reduced after evaluation by the network. The training was run for 40 epochs which took a total of approximately 16.7 days (15.3 days for the CNN without data-consistency). The average inference time per 3D iNAV for the proposed architecture is approximately 0.5 seconds (4.33–5.20 total minutes) on GPU (Titan RTX), while  $l_1$ -ESPIRiT (using BART) requires approximately 10 seconds (87.42–104.90 total minutes) on CPU (Xeon E5–2650 v4, and Intel i7-8700K) and 1.5 seconds (12.92–15.50 total minutes) on GPU (Titan RTX). Average reconstruction times for both the



unrolled model and CNN are shown in Table 2. All four subject datasets have a total of 500–600 3D iNAV.

### 3.2 | Motion estimates

Due to the improved performance of the unrolled network architecture compared to the CNN without data-consistency, motion estimates were acquired using the 3D iNAVs reconstructed with the unrolled model. The global motion estimates for the first 100 heartbeats of the four different subjects are shown in Supporting Information Figure S4 for all three directions (A/P, L/R, S/I). The scatter plots for  $I_1$ -ESPIRiT vs. Model and the correlation coefficients (R) are shown in Figure 4. Also, the measurements are normalized by subtracting the mean S/I, A/P, and L/R displacements similar to<sup>36</sup> when comparing the estimates. The additional correlation coefficients (including the global and five spatial bins) are shown in Supporting Information Table S1. The R values for all four subjects show a strong positive correlation of the global motion estimates between the  $I_1$ -ESPIRiT and DL model-based 3D iNAVs, thus indicating similar motion estimates in all directions (Supporting Information Figure S4).

### 3.3 | Autofocusing histograms

The autofocusing histograms for all 4 subjects are shown in Supporting Information Figure S5. For subjects 2 and 3 (Supporting Information Figure S5B,C, respectively), the global bin is the most selected by the autofocusing algorithm which demonstrates that there was less residual motion beyond the rigid-body translational motion. For subjects 1 and 4 (Supporting Information Figure S5A,D), bins 4 and 5 are the most selected, respectively, which shows that there was additional residual motion which minimized the gradient entropy metric. Subjects 1 and 2 histograms lack noticeable difference, and subjects 3 and 4 show minor differences in bin 2. Additionally, example residual motion estimate scatter plots and correlation coefficients (R) for all four subjects are shown in Supporting Information Figure S6. The first 100 heartbeats corresponding to residual motion estimates (A/P, L/R, S/I) are also shown in Supporting Information Figure S7. Small differences (submillimeter) can be seen in the residual motion estimates between the  $I_1$ -ESPIRiT and DL model-based 3D iNAV estimates. These minor differences are further investigated in the high-resolution images to verify the effects on the coronary image quality.

### 3.4 | Motion-corrected images

The 3D autofocused images using both 3D iNAV reconstruction schemes are shown for all 4 subject scans. The total time for autofocusing reconstruction takes approximately 8 minutes (2 minutes to calculate the global translations, 5 minutes to calculate the displacement fields and perform k-means clustering, and 1 minutes to calculate gradient entropy minimization of each pixel). In Figure 5, the right coronary artery (RCA) is shown before and after motion correction with cross-sectional views demonstrating the improvements when using the  $I_1$ -ESPIRiT and DL model-based 3D iNAVs. The RCA images for all four subjects show nearly identical vessel sharpness. Also, in Figure 6, the left coronary artery (LCA) is shown with the corresponding cross-sectional views. Similar to the RCA, the LCA sharpness increased after motion correction and maintained similar and comparable improvements when using  $I_1$ -ESPIRiT and DL model-based 3D iNAVs for motion correction. Additionally,

the average RCA and LCA vessel sharpness values (IEPA), shown in Supporting Information Table S2, correspond to the similar vessel sharpness seen for the motion-corrected images in Figures 5 and 6.

## 4 | DISCUSSION

We have shown that the proposed non-Cartesian unrolled network architecture generates similar 3D iNAV reconstruction results in a fraction of the time, with a 20× and 3× speed increase for CPU and GPU, respectively, and leads to high correlations of the derived global and localized motion estimates compared to  $l_1$ -ESPIRiT 3D iNAV reconstructions. The most computationally expensive part of the training involves the NUFFT operation. The NUFFT is used to apply the data-consistency (soft projection) step which can require more calculations for the gradients during training, thus increasing training time and the GPU memory requirements. Similar issues arise during the  $l_1$ -ESPIRiT iterative reconstruction. To account for this, BART uses the Toeplitz method<sup>40</sup> for the compressed sensing reconstruction. However, the Toeplitz method requires a matrix size that is twice as large as the nominal matrix (i.e., twice the FOV in all three dimensions) to encompass the support of the object in the image domain. This helps to alleviate aliasing errors that can arise from objects outside of the FOV when using the Toeplitz method. Due to the computation benefits, further investigation may be warranted by using a Toeplitz-based NUFFT for the data-consistency (soft projection) steps even though the 2× oversampling constraints are not required when using a normal NUFFT operator. Benefiting from using the Toeplitz-based NUFFT operator would depend on the amount of undersampling, trajectory type, and native matrix size.<sup>41</sup>

When using the proposed non-Cartesian unrolled model, there are improvements and limitations to address. For this implementation, the complex data was separated into 2 channels without any noticeable training issues, but architectures that can process the full complex data channel may prove beneficial for better generalization and potentially lead to faster training since fewer filter weights are used. In Virtue et al,<sup>42</sup> this was done by using complex activation layers that attenuate the magnitude based on the input phase, essentially acting as a “complex rectified linear unit” activation function. Also, currently an  $l_1$  loss is used, but other approaches such as using generative adversarial networks (GANs)<sup>43</sup> have the ability to learn better loss functions that take into account diagnostic image quality.<sup>44–46</sup> This would have the potential for improving the “perceived” image quality for clinicians which an  $l_1$  or  $l_2$  loss may not fully quantify. Furthermore, when collecting the 3D iNAV datasets, the fully sampled data is not obtained due to the finite acquisition window after collecting the segmented high-resolution data within one heartbeat; thus, the ground truth was obtained through the compressed sensing reconstruction of VD cones. Even though the ground truth is biased towards the compressed sensing reconstruction, we have shown that the unrolled network was able to successfully reconstruct the 3D iNAVs with improved image quality compared to gridding and allow for similar motion estimates compared to using compressed sensing.

Additionally, we have primarily trained the unrolled model with 3D VD cones cardiac datasets with 4.4 mm resolutions. Fortunately, due to flexibility of the model architecture,

the training dataset for the current application is not limited to these specific undersampled 3D cardiac iNAV. Thus, to help further generalization of the model, datasets acquired using different trajectories, resolutions, and anatomies can be added to the current training dataset for retraining (fine-tuning).<sup>47</sup> Also, when training with higher resolutions, the aliasing artifacts may become more severe and require larger matrix sizes to avoid aliasing from objects outside of the FOV during the data-consistency step in the model. These modifications can potentially require prolonged training times, as opposed to classical compressed sensing methods. Therefore, the GPU implementation of  $l_1$ -ESPIRiT may have an advantage over DL model-based methods when training time is limited. Nevertheless, the proposed approach has successfully shown the reconstruction of undersampled low-resolution 3D iNAVs that were used for motion estimation. Further analysis of different resolutions and anatomies may be warranted to test the limitations of the unrolled model.

In the current work, autofocusing<sup>1,5,36</sup> and binning with localized motion correction were investigated when using the proposed DL model-based 3D iNAV reconstruction approach. Further techniques for motion correction involve using a deep-learning framework<sup>48–50</sup> to obtain the motion estimates or motion-corrected images directly from the k-space data can potentially replace autofocusing but would likely require a larger training set. This could allow for an end-to-end model using the proposed unrolled PGD architecture for 3D iNAV reconstruction and an additional model which takes in the 3D iNAVs with the motion-corrupted high-resolution k-space data and outputs the motion-corrected images. This framework has the potential for a substantial reduction in reconstruction time. Even if the end-to-end model produces less optimal motion-corrected images compared to previous techniques, the model can be used as a tool for quickly validating scan quality. Then, if further improvements are required, reconstruction can be applied using the standard more time-consuming  $l_1$ -ESPIRiT and autofocusing techniques.

Future work includes implementing the non-Cartesian unrolled model for higher resolution data. Current limitations include the 24 GBs of memory on the NVIDIA RTX GPU which limits the potential matrix size used for training. To solve this problem, filter depth, number of iterations, and ResNet blocks can be decreased but may reduce the ability for proper generalization of the model. As previously mentioned, a Toeplitz-based NUFFT can be used but would lead to a  $2\times$  FOV oversampling requirement. Also, multi-GPU training can be implemented to allow for larger matrix sizes but would increase training time due to introduced overhead. The technique that currently has the most potential and feasibility involves the bandpass approach<sup>13</sup> that segments k-space into multiple overlapping patches. Each patch is then reconstructed in parallel using different unrolled DL models that are trained to reconstruct specific regions of k-space. To generate the final image, all individual segments are stitched together in k-space. Training a non-Cartesian high-resolution unrolled model has many challenges and constraints, but there are many options for solving the memory and computation problems.

## 5 | CONCLUSION

A deep neural network architecture was developed for the reconstruction of undersampled VD cones 3D iNAVs acquired during a CMRA sequence. The unrolled network architecture

was designed to solve the PGD reconstruction problem and for the reconstruction of undersampled non-Cartesian datasets. It was shown that the reconstruction of 3D iNAVs using the DL model-based reconstruction compared to using  $l_1$ -ESPIRiT can be performed in a fraction of the time (1/20<sup>th</sup> on CPU and 1/3<sup>rd</sup> on GPU) while generating similar motion estimates and, after motion correction of the high-resolution data, equivalent RCA and LCA image quality.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding information

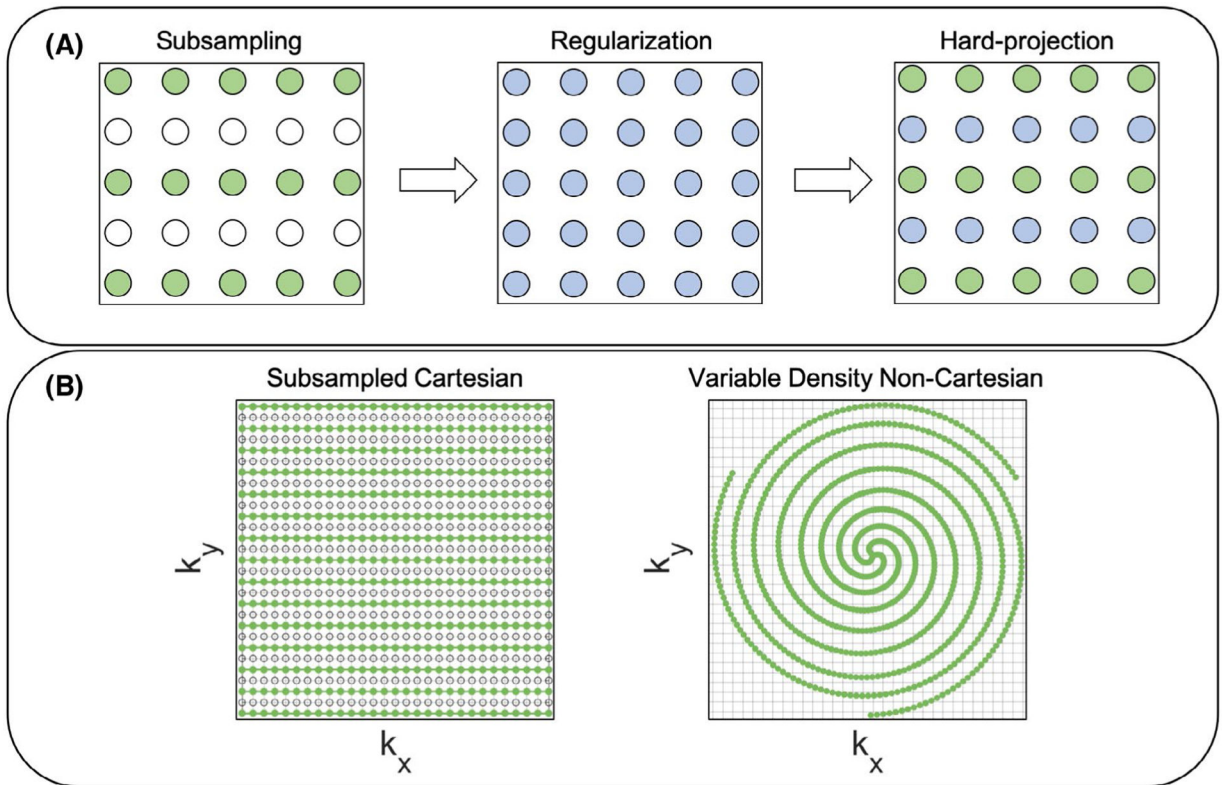
NSF Graduate Research Fellowship Program, Grant/Award Number: NIH R01 HL127039 and T32HL007846; GE Healthcare

## REFERENCES

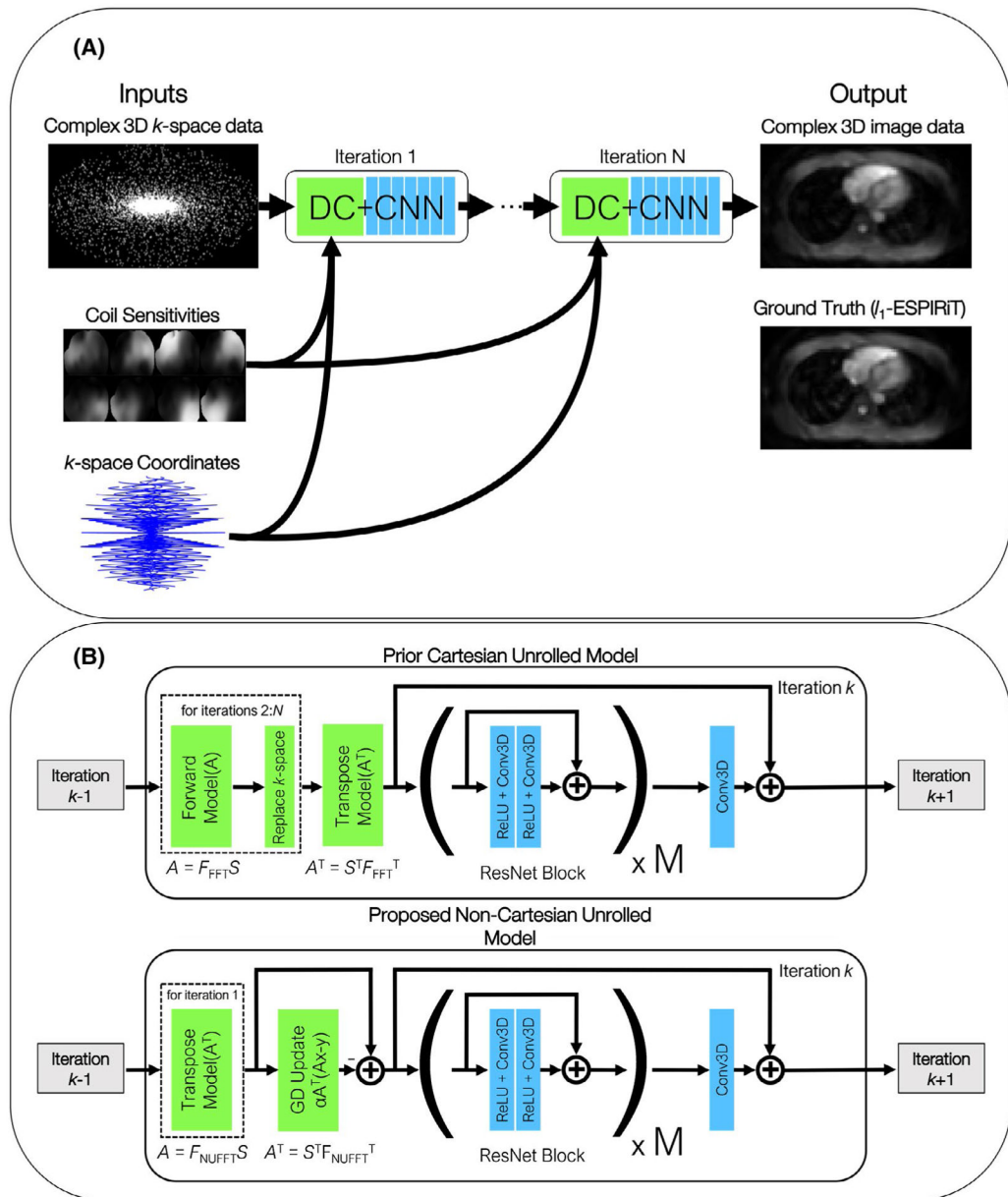
1. Addy NO, Ingle RR, Luo J, et al. 3D image-based navigators for coronary mr angiography. *Magn Reson Med*. 2017;77: 1874–1883. [PubMed: 27174590]
2. Wu HH, Gurney PT, Hu BS, Nishimura DG, McConnell MV. Free-breathing multiphase whole-heart coronary mr angiography using image-based navigators and three-dimensional cones imaging. *Magn Reson Med*. 2013;69:1083–1093. [PubMed: 22648856]
3. Leupold J, Hennig J, Scheffler K. Alternating repetition time balanced steady state free precession. *Magn Reson Med*. 2006;55:557–565. [PubMed: 16447171]
4. Gurney PT, Hargreaves BA, Nishimura DG. Design and analysis of a practical 3D cones trajectory. *Magn Reson Med*. 2006; 55:575–582. [PubMed: 16450366]
5. Luo J, Addy NO, Ingle RR, Baron CA, Cheng JY, Hu BS, Nishimura DG. Nonrigid motion correction with 3D image-based navigators for coronary MR angiography. *Magn Reson Med*. 2017;77:1884–1893. [PubMed: 27174673]
6. Uecker M, Lai P, Murphy MJ, et al. Espirit—an eigenvalue approach to autocalibrating parallel MRI: where sense meets grappa. *Magn Reson Med*. 2014;71:990–1001. [PubMed: 23649942]
7. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055–3071. [PubMed: 29115689]
8. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2018;38:280–290. [PubMed: 30080145]
9. Biswas S, Aggarwal HK, Jacob M. Dynamic MRI using model-based deep learning and STORM priors: MoDL-STORM. *Magn Reson Med*. 2019;82:485–494. [PubMed: 30860286]
10. Diamond S, Sitzmann V, Heide F, Wetzstein G. Unrolled optimization with deep priors. *arXiv preprint 2017; 1705.08041*.
11. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2017;37:491–503. [PubMed: 29035212]
12. Aggarwal HK, Mani MP, Jacob M. Modl: Model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2018;38:394–405. [PubMed: 30106719]
13. Cheng JY, Chen F, Alley MT, Pauly JM, Vasanawala SS. Highly scalable image reconstruction using deep neural networks with bandpass filtering. *arXiv preprint 2018;1805.03300*.
14. Cheng JY, Chen F, Sandino C, Mardani M, Pauly JM, Vasanawala SS. Compressed sensing: from research to clinical practice with data-driven learning. *arXiv preprint 2019;1903.07824*.
15. Parikh N, Boyd S. Proximal algorithms. *Foundations Trends® Optim*. 2014;1:127–239.

16. Hauptmann A, Arridge S, Lucka F, Muthurangu V, Steeden JA. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning-proof of concept in congenital heart disease. *Magn Reson Med*. 2019;81:1143–1156. [PubMed: 30194880]
17. Malavé MO, Baron CA, Addy NO, et al. Whole-heart coronary MR angiography using a 3D cones phyllotaxis trajectory. *Magn Reson Med*. 2019;81:1092–1103. [PubMed: 30370941]
18. Addy NO, Ingle RR, Wu HH, Hu BS, Nishimura DG. High-resolution variable-density 3D cones coronary MRA. *Magn Reson Med*. 2015;74:614–621. [PubMed: 26172829]
19. Malavé MO, Koundinyan SP, Sandino CM, Cheng JY, Nishimura DG. Improved design and reconstruction of 3D image based navigators for coronary mr angiography. In *Proceedings of the International Society for Magnetic Resonance in Medicine, Paris, France, 2018*;26:0920.
20. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. Sense: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42:952–962. [PubMed: 10542355]
21. Daubechies I, Defrise M, DeMol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun Pure Appl Math*. 2004;57:1413–1457.
22. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182–1195. [PubMed: 17969013]
23. Vasanawala S, Murphy M, Alley MT, et al. Practical parallel imaging compressed sensing MRI: summary of two years of experience in accelerating body MRI of pediatric patients. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, 2011*;2011:1039–1043.
24. Liu J, Saloner D. Accelerated MRI with circular cartesian undersampling (circus): a variable density cartesian sampling strategy for compressed sensing and parallel imaging. *Quant Imaging Med Surgery*. 2014;4:57–67.
25. Prieto C, Doneva M, Usman M, et al. Highly efficient respiratory motion compensated free-breathing coronary MRA using golden-step cartesian acquisition. *J Magn Reson Imaging*. 2015;41:738–746. [PubMed: 24573992]
26. Cheng JY, Zhang T, Ruangwattanapaisarn N, et al. Free-breathing pediatric MRI with nonrigid motion correction and acceleration. *J Magn Reson Imaging*. 2015;42:407–420. [PubMed: 25329325]
27. Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math*. 2006;59:1207–1223.
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2016*;770–778.
29. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. *European Conference on Computer Vision*. 2016;630–645.
30. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging*. 2016;3:47–57.
31. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging*. 2018;38:167–179. [PubMed: 30040634]
32. Uecker M, Ong F, Tamir JI, et al. Berkeley advanced reconstruction toolbox. In *Proceedings of the International Society for Magnetic Resonance in Medicine, Toronto, Ontario, Canada, 2015*;23:2486.
33. Beatty PJ, Nishimura DG, Pauly JM. Rapid gridding reconstruction with a minimal oversampling ratio. *IEEE Trans Med Imaging*. 2005;24:799–808. [PubMed: 15959939]
34. Ong F, Uecker M, Jiang W, Lustig M. Fast non-cartesian reconstruction with pruned fast fourier transform. In *Proceedings of the International Society for Magnetic Resonance in Medicine, Toronto, Ontario, Canada, 2015*;23:3639.
35. Norton A, Silberger AJ. Parallelization and performance analysis of the Cooley-Tukey FFT algorithm for shared-memory architectures. *IEEE Trans Comput*. 1987;36:581–591.
36. Ingle RR, Wu HH, Addy NO, et al. Nonrigid autofocus motion correction for coronary mr angiography with a 3D cones trajectory. *Magn Reson Med*. 2014;72:347–361. [PubMed: 24006292]
37. Studholme C, Hill DL, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*. 1999;32:71–86.

38. Pluim JP, Maintz JA, Viergever MA. Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imaging*. 2003;22:986–1004. [PubMed: 12906253]
39. Biasioli L, Lindsay AC, Choudhury RP, Robson MD. Loss of fine structure and edge sharpness in fast-spin-echo carotid wall imaging: measurements and comparison with multiple-spin-echo in normal and atherosclerotic subjects. *J Magn Reson Imaging*. 2011;33:1136–1143. [PubMed: 21509872]
40. Baron CA, Dwork N, Pauly JM, Nishimura DG. Rapid compressed sensing reconstruction of 3D non-cartesian MRI. *Magn Reson Med*. 2018;79:2685–2692. [PubMed: 28940748]
41. Ou T, Ong F, Uecker M, Waller L, Lustig M. Nufft: fast auto-tuned gpu-based library. In *Proceedings of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii, USA, 2017*;25:3807.
42. Virtue P, Stella XY, Lustig M. Better than real: complex-valued neural nets for MRI fingerprinting. *IEEE International Conference on Image Processing (ICIP)*; 2017:3953–3957.
43. Goodfellow I, PougetAbadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Infor Process Syst*. 2014:2672–2680.
44. Kulkarni K, Lohit S, Turaga P, Kerviche R, Ashok A. Reconnet: non-iterative reconstruction of images from compressively sensed measurements. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2016*;449–458.
45. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial networks for compressed sensing automates MRI. *arXiv preprint 2017*; 1706.00051.
46. Quan TM, NguyenDuc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging*. 2018;37:1488–1497. [PubMed: 29870376]
47. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*. 2019;81:116–128. [PubMed: 29774597]
48. Lv J, Yang M, Zhang J, Wang X. Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study. *British J Radiol*. 2018;91:20170788.
49. Pawar K, Chen Z, Shah NJ, Egan GF. Moconet: motion correction in 3D MPRAGE images using a convolutional neural network approach. *arXiv preprint 2018*; 1807.10831.
50. Lossau T, Nickisch H, Wissel T, et al. Motion estimation and correction in cardiac ct angiography images using convolutional neural networks. *Comput Med Imaging Graphics*. 2019;76:101640.

**FIGURE 1.**

A, A subsampled trajectory is shown where the green circles and white circles represent measured and zero-filled data in  $k$ -space respectively. After regularization (blue circles) using wavelets (i.e., soft thresholding in the wavelet domain) or a CNN (i.e., applying convolutions in the image domain), the Fourier transform of the 3D image is taken and the original  $k$ -space trajectory locations are replaced with the measured data by performing a hard-projection. B, The Cartesian trajectory (left) uses subsampling while the non-Cartesian trajectory (right) uses a variable-density (VD) design to achieve undersampling

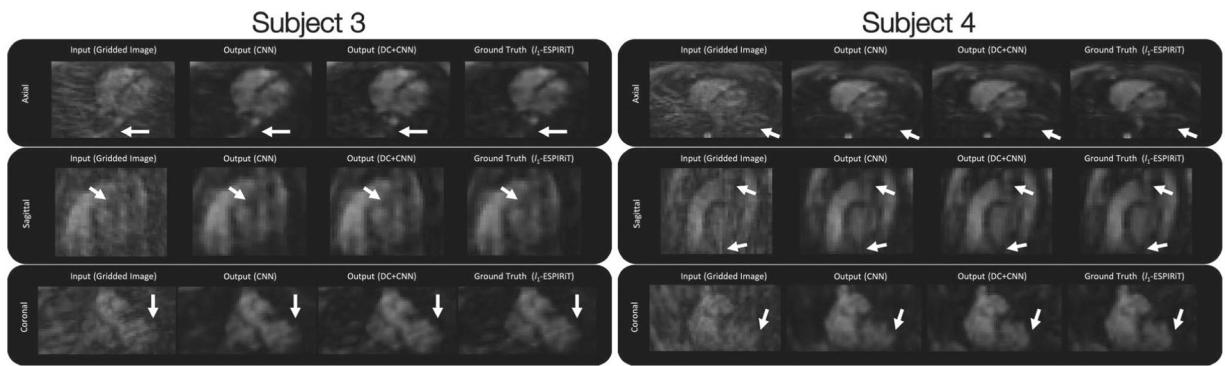


**FIGURE 2.**

A, The input into the network is the 3D  $k$ -space data,  $k$ -space coordinates (to generate the NUFFT operator,  $F_{\text{NUFFT}}$ ), and the coil sensitivity maps for each channel (for the SENSE reconstruction operator  $S$ ). The ground truth is the  $l_1$ -ESPIRiT reconstruction of the input  $k$ -space data. Each iteration consists of a data-consistency (DC) and CNN block. Initially before entering the CNN block, the transpose model ( $A^T$ ) is used to transform the data to image space. B, The architecture uses  $N=4$  iterations (gradient steps) consisting of  $M=2$  ResNet blocks/step. One key difference between the prior Cartesian model<sup>11</sup> (top) and the proposed model (bottom) is the replacement of the data-consistency step using an FFT with a gradient-descent (GD) update step (i.e., soft-projection) using the NUFFT, which maintains consistency with the measured non-Cartesian  $k$ -space data. The first iteration

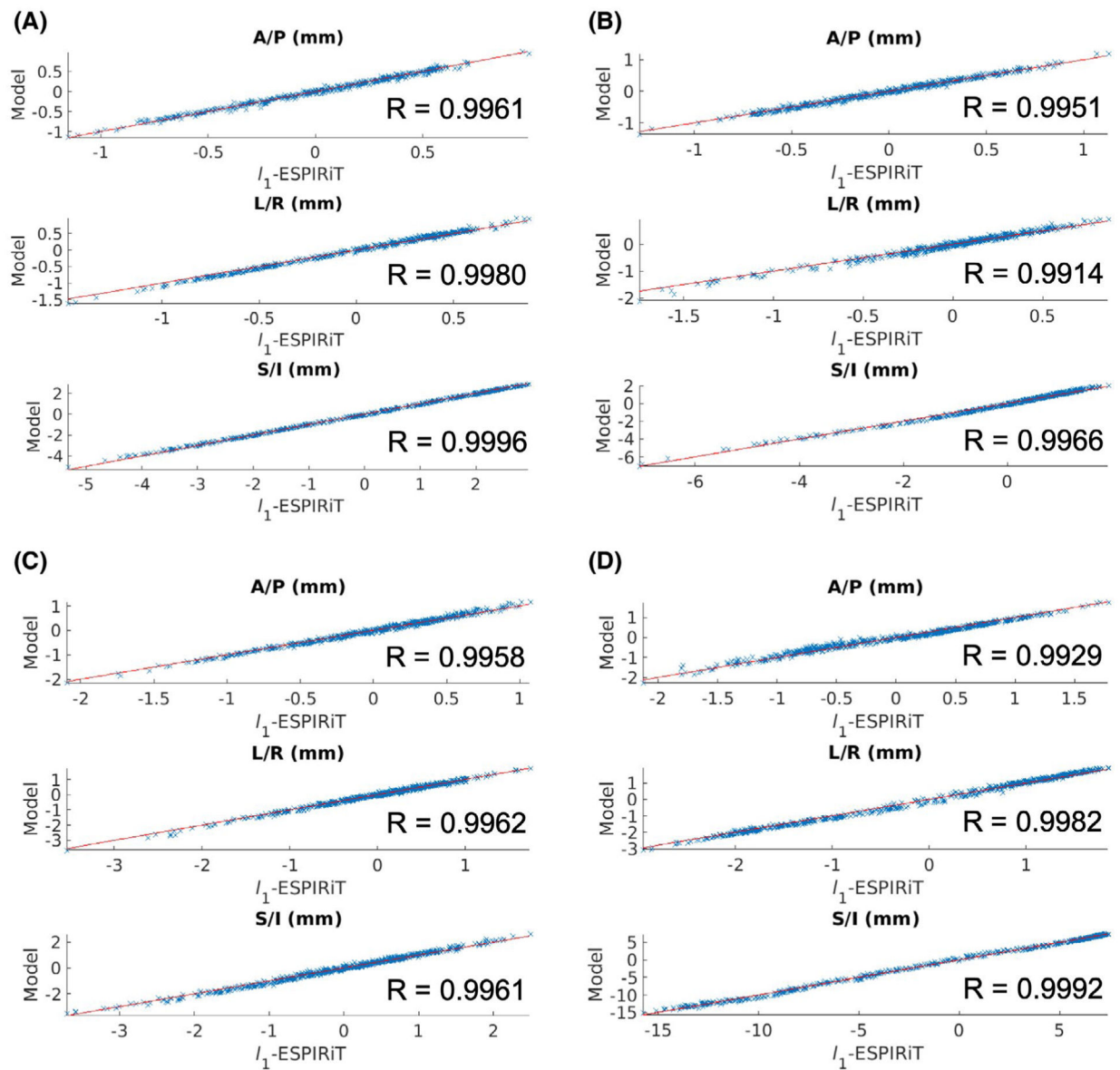


applies the transpose model ( $A^T$ ), followed by the CNN. The remaining iterations (i.e., 2 to  $N$ ) for hard-projection (top) and soft-projection (bottom) use the architecture shown in the dotted box, and gradient update step, respectively, followed by the CNN



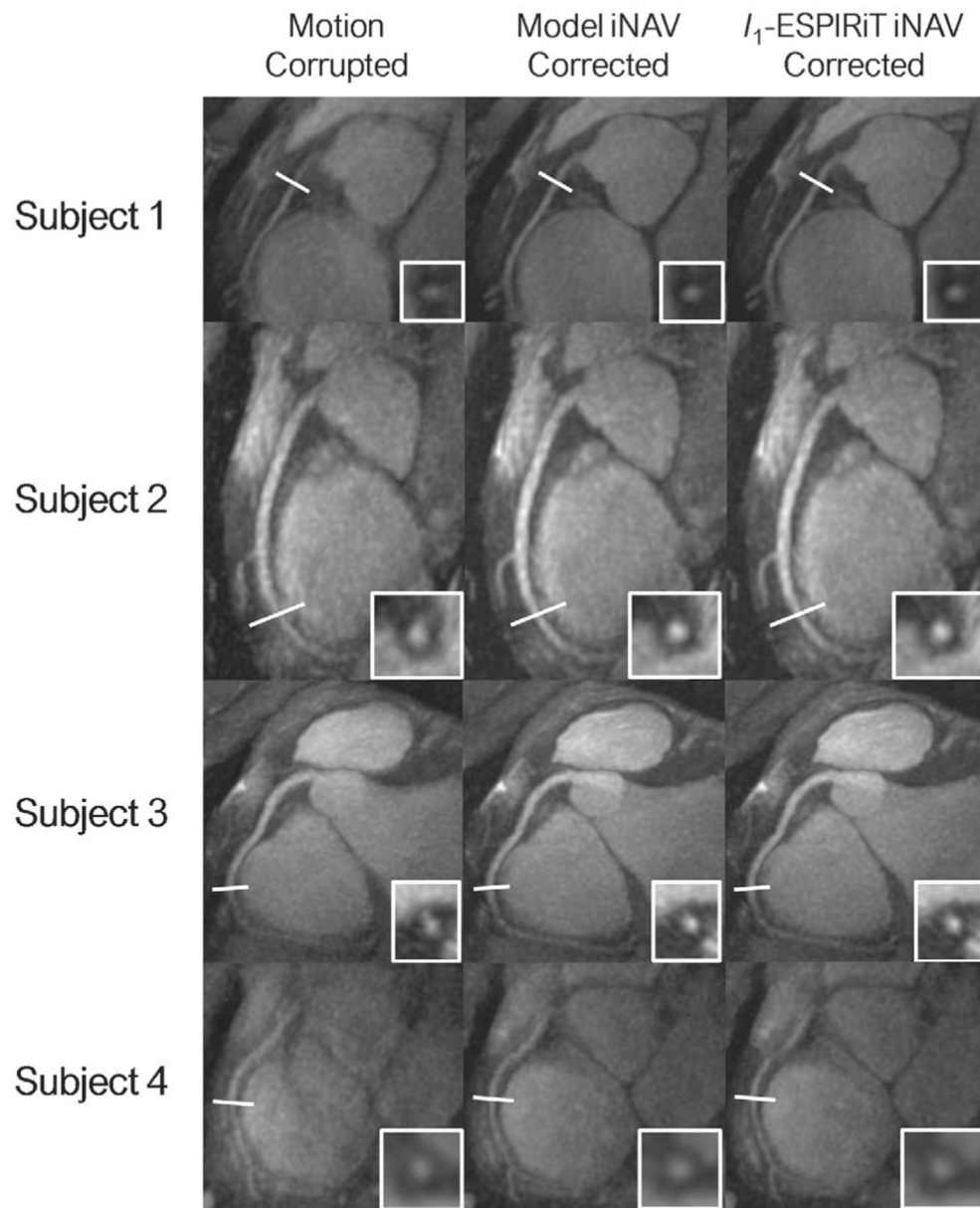
**FIGURE 3.**

The axial, sagittal, and coronal slices are shown from one heartbeat. 3D iNAV inputs (gridded images using the NUFFT operator), outputs using CNNs without data-consistency (DC), outputs using the unrolled model with DC and CNNs, and ground truths ( $I_1$ -ESPIRiT) are shown, respectively, for subjects 3–4 (test datasets). For each method, the differences are highlighted using white arrows which show how the unrolled model more closely matches the ground truth compared to using a CNN without DC

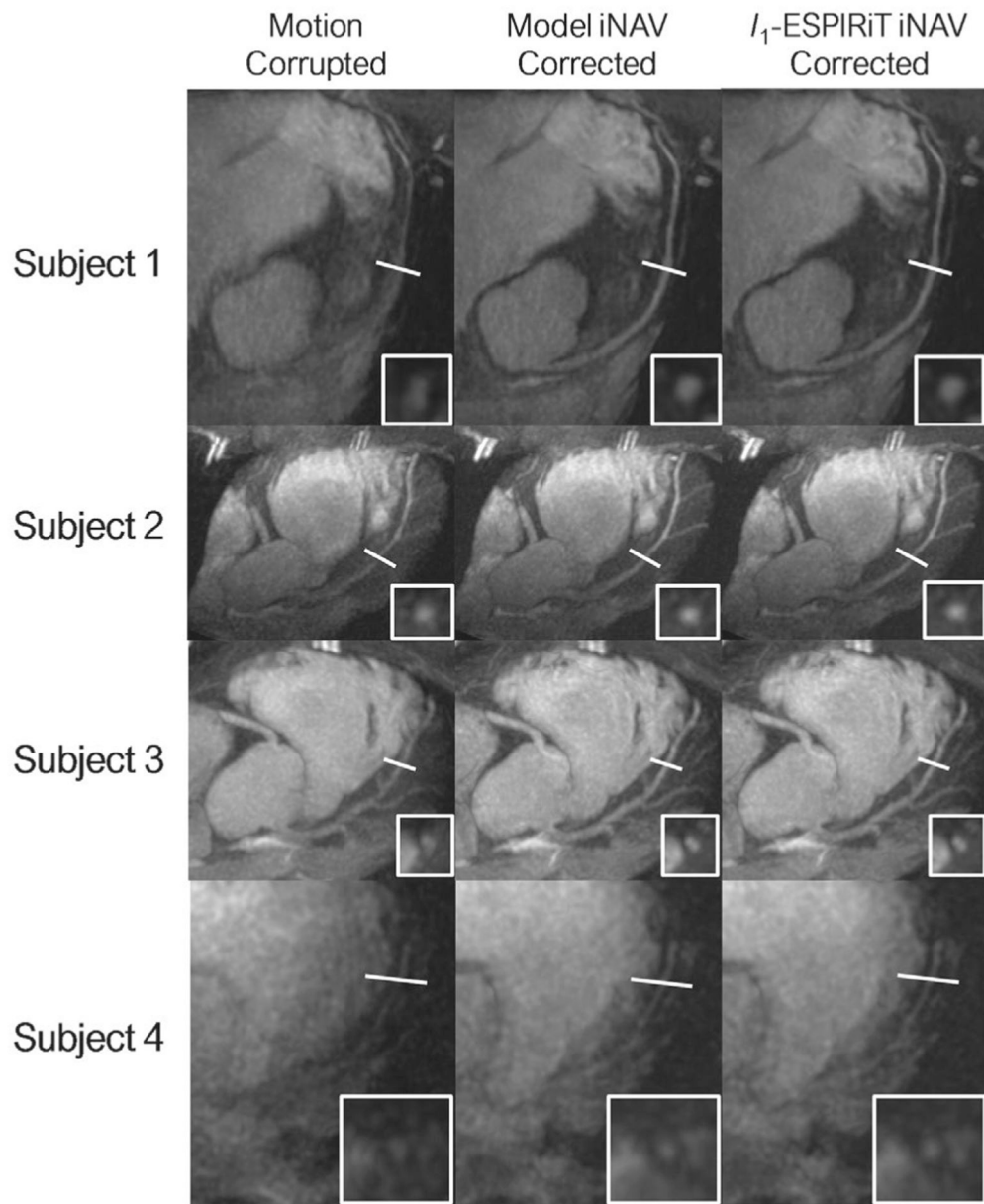


**FIGURE 4.**

The global motion estimates for subjects 1-4 (A-D) are shown in three different scatter plots (A/P, L/R, and S/I) with the corresponding correlation coefficients (R) for  $I_1$ -ESPIRiT versus the model

**FIGURE 5.**

A, Reformatted maximum intensity projection images of the RCA for four healthy volunteers are shown before and after motion correction. The cross-sectional views demonstrate similar improvements in the distal regions of the RCA when using both the  $I_1$ -ESPIRiT and DL model-based 3D iNAVs for motion correction

**FIGURE 6.**

A, Reformatted maximum intensity projection images of the LCA for four healthy volunteers are shown before and after motion correction with the corresponding cross-sectional views. The LCA sharpness increased in the medial and distal regions after motion correction and exhibited similar improvements when using  $l_1$ -ESPIRiT and DL model-based 3D iNAV's for motion correction

**TABLE 1**

Reconstructed image  $l_1$  loss for all four subjects when applying the CNN model without data-consistency (DC) and the unrolled model (DC and CNN)

| <b>Model Reconstructed <math>l_1</math> loss</b> |            |                 |
|--|------------|-----------------|
| <b>Subject</b>                                   | <b>CNN</b> | <b>DC + CNN</b> |
| Subject 1  | 2.44E+03   | 1.16E+03        |
| Subject 2  | 2.43E+04   | 2.16E+04        |
| Subject 3  | 2.20E+03   | 1.76E+03        |
| Subject 4  | 1.91E+03   | 9.09E+02        |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Average reconstruction times per 3D iNAV on different CPU and GPU devices using  $I_1$ -ESPIRiT, CNN (without data-consistency), and the unrolled model

**TABLE 2**

| Average reconstruction times (sec) |                |                  |                |      |                  |  |
|------------------------------------|----------------|------------------|----------------|------|------------------|--|
| CPU                                | $I_1$ -ESPIRiT | GPU              | $I_1$ -ESPIRiT | CNN  | Model (DC + CNN) |  |
| Xeon E5-2650 v4                    | 10.49          | NVIDIA Titan XP  | 1.68           | 0.36 | 0.83             |  |
| Intel i7-8700k                     | 10.10          | NVIDIA Titan RTX | 1.55           | 0.34 | 0.52             |  |