FULL PAPER

Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI

Hossam El-Rewaidy[1,2] 🔟 | Ahmed S. Fahmy[1] | Farhad Pashakhanloo[1] | Xiaoying Cai[1,3] | Selcuk Kucukseymen[1] | Ibolya Csecs[1] | Ulf Neisius[1] | Hassan Haji-Valizadeh[1] | Bjoern Menze[2] | Reza Nezafat[1]

[1]Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, USA

[2]Department of Computer Science, Technical University of Munich, Munich, Germany

[3]Siemens Medical Solutions USA, Inc. Cary, North Carolina, USA

Correspondence
Reza Nezafat, Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Ave, Boston, MA 02215, USA.
Email: rnezafat@bidmc.harvard.edu

Purpose: Cardiac MR cine imaging allows accurate and reproducible assessment of cardiac function. However, its long scan time not only limits the spatial and temporal resolutions but is challenging in patients with breath-holding difficulty or non-sinus rhythms. To reduce scan time, we propose a multi-domain convolutional neural network (MD-CNN) for fast reconstruction of highly undersampled radial cine images.

Methods: MD-CNN is a complex-valued network that processes MR data in k-space and image domains via k-space interpolation and image-domain subnetworks for residual artifact suppression. MD-CNN exploits spatio-temporal correlations across timeframes and multi-coil redundancies to enable high acceleration. Radial cine data were prospectively collected in 108 subjects (50 ± 17 y, 72 males) using retrospective-gated acquisition with 80%:20% split for training/testing. Images were reconstructed by MD-CNN and k-t Radial Sparse-Sense(kt-RASPS) using an undersampled dataset (14 of 196 acquired views; relative acceleration rate = 14). MD-CNN images were evaluated quantitatively using mean-squared-error (MSE) and structural similarity index (SSIM) relative to reference images, and qualitatively by three independent readers for left ventricular (LV) border sharpness and temporal fidelity using 5-point Likert-scale (1-non-diagnostic, 2-poor, 3-fair, 4-good, and 5-excellent).

Results: MD-CNN showed improved MSE and SSIM compared to kt-RASPS (0.11 ± 0.10 vs. 0.61 ± 0.51, and 0.87 ± 0.07 vs. 0.72 ± 0.07, respectively; $P < .01$). Qualitatively, MD-CCN significantly outperformed kt-RASPS in LV border sharpness (3.87 ± 0.66 vs. 2.71 ± 0.58 at end-diastole, and 3.57 ± 0.6 vs. 2.56 ± 0.6 at end-systole, respectively; $P < .01$) and temporal fidelity (3.27 ± 0.65 vs. 2.59 ± 0.59; $P < .01$).

Conclusion: MD-CNN reduces the scan time of cine imaging by a factor of 23.3 and provides superior image quality compared to kt-RASPS.

1 | INTRODUCTION

Cardiac MR cine imaging with balanced steady-state free precession (bSSFP) allows accurate and reproducible measurement of cardiac function.[1] Cine images are frequently collected using electrocardiograph (ECG)-gated segmented acquisition during multiple breath-holds. For patients with breathing difficulties or non-sinus rhythms, real-time cine imaging at lower temporal and spatial resolutions is commonly acquired. A single-shot acquisition scheme is frequently used but the spatial and temporal resolution are often limited. Accelerated acquisition and advanced reconstruction can be used to increase temporal and spatial resolutions for both segmented and real-time single-shot cine imaging. Over the past two decades, there have been continuous innovations in image acceleration techniques for cardiac cine imaging, such as view-sharing,[2,3] partial-Fourier imaging,[4] parallel imaging,[5,6] and compressed sensing (CS).[7-12] Parallel imaging with rate 2 accelerated cine imaging is commonly adopted in clinical routines. Recently, CS imaging has become available by vendors to reduce scan time. However, the acceleration has been limited and the reconstruction remains to be time-consuming, which hinders the clinical workflow. In addition, CS reconstruction generally requires additional state-of-the-art hardware to accommodate the higher demand for reconstruction engines.[12]

Recent advances in deep learning-based image reconstruction techniques provided yet another opportunity to achieve higher acceleration rates and improve spatial and temporal resolution in cardiac MRI.[13-18] For cine imaging, both convolutional and recurrent neural networks have been used to suppress image domain aliasing artifacts in cartesian[13-15] and radial[16,19] sampling schemes. To enable high acceleration rates, temporal correlations were exploited using 3D convolutional kernels to learn spatio-temporal features in the image domain.[13,16] In cartesian acquisitions, shared information among neighboring time frames was exploited to fill in missing k-space lines.[13,15] Schlemper et al. used a simple k-space data-sharing strategy before applying their cascaded network by filling the missing k-space lines directly from the nearest acquired line in the neighboring frames.[13] Although aggregating k-space lines from different time frames reduces image domain aliasing, it may increase temporal blurring at higher acceleration rates. Convolutional networks have also been used to share information across time in a multi-supervised network on cartesian k-space acquisition.[15] However, the additional hyperparameters associated with the multiple

loss functions led to challenges in training, and the performance of this network degrades at high acceleration rates of ≥8. To achieve higher acceleration rates, radial sampling can be used to allow more efficient coverage of k-space and offer an enhanced view-sharing strategy among time frames. However, view sharing in k-space using convolutional neural networks (CNN) has not been investigated for radial acquisitions.

In this work, we propose a multi-domain convolutional neural network (MD-CNN) for cine imaging with radial k-space sampling that processes the radial MR data in k-space, image, and time domains from multiple coils by end-to-end training. MD-CNN performance is evaluated on a retrospective gated cine imaging dataset for image quality using both quantitative and qualitative assessments.

2 | METHODS

2.1 | Network architecture

MD-CNN is a fully complex-valued CNN with two main components: k-space and image-domain subnetworks (Figure 1). Both subnetworks are trained end-to-end from scratch. The input to MD-CNN is a 4D complex-valued matrix of size $N_c \times N_{tw} \times N_x \times N_y$, representing the gridded k-space of a 2D cross-sectional image from multiple coils and across multiple consecutive time frames. N_{tw} denotes the number of time frames, and N_c denotes the number of coils.

2.1.1 | k-space subnetwork

The function of this subnetwork is to refine estimations of the 2D gridded k-space data using convolutional kernels. Missing k-space data are estimated by interpolation of the acquired MR data using multiple convolutional kernels. Correlations across time frames are also exploited in the form of view-sharing via 3D convolutional kernels applied on the 2D gridded k-space and a window of neighboring time frames (N_{tw}). The k-space subnetwork consists of two identical residual blocks of cascaded convolutional layers. Each k-space block consists of three 3D complex-valued convolutional layers (XConv)[20] with $2N_c$, $2N_c$, and N_c kernels each of size $3 \times 5 \times 5$; where N_c is the number of coils. A complex rectified linear unit (XReLU) follows each convolutional layer. A residual connection is added after each k-space block to reduce the
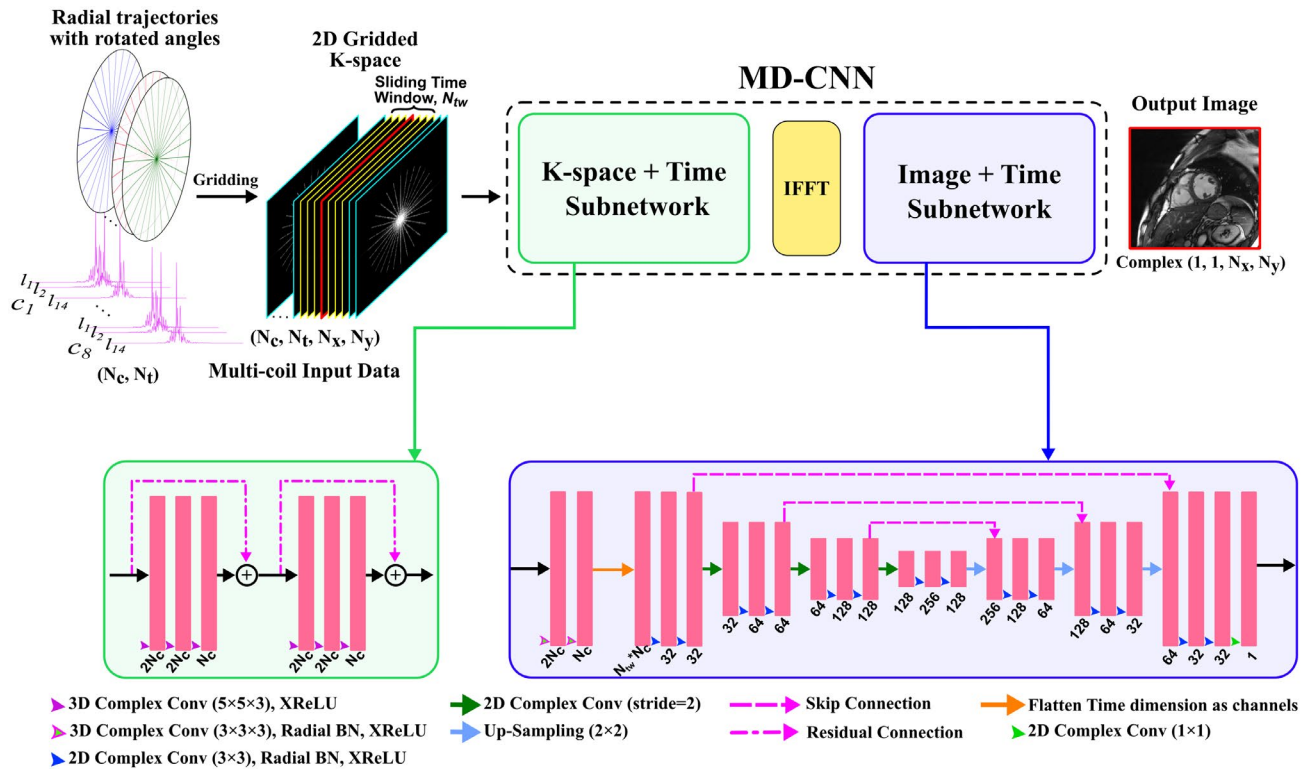
FIGURE 1 Proposed multi-domain convolutional neural network (MD-CNN) pipeline. The k-space data from N_c coils acquired with radial trajectories are gridded onto the cartesian grid using NUFFT. To reconstruct the frame (red), multi-coil 2D k-space data from N_{tw} adjacent time frames (yellow) are fed into the MD-CNN. The output of the MD-CNN is a coil-combined reconstructed image for the target frame. MD-CNN consists of two subnetworks: k-space and image-domain. The k-space subnetwork takes complex-valued 2D k-space input data of size (N_c, N_{tw}, N_x, N_y) to be processed by two identical residual blocks. Each block consists of three complex convolutional layers with 3D kernels of size $(3 \times 5 \times 5)$. Each convolutional layer is followed by a complex rectified linear unit (XReLU) for activation. The resulting k-space data is transformed into the image domain by inverse fast Fourier transform (IFFT). In the image subnetwork, data are fed into two convolutional layers with 3D kernels of size $(3 \times 3 \times 3)$ to exploit spatio-temporal correlations. The time dimension in the resulting 3D feature maps are flattened into channels to produce 2D feature maps of size $(N_c N_{tw}, N_x, N_y)$. A 2D U-net is then used to process the 2D feature maps from N_C coils and N_{tw} frames to reconstruct a coil-combined reconstructed image

large dynamic range of k-space values. The resulting k-space data are then transformed into the image domain using inverse fast Fourier transform (FFT).

2.1.2 | Image domain subnetwork

In the image domain, shared structural information among neighboring time frames is exploited by two 3D XConv layers consisting of $2N_c$ and N_c kernels each of size $3 \times 3 \times 3$ with radial batch normalization (RBN) and XReLU.[20] The resulting 3D maps (ie, 2D+time; size of $N_c \times N_{tw} \times N_x \times N_y$) are projected in the 2D space, where time frames are flattened into channels to produce 2D feature maps.

A 2D U-Net architecture is then used to remove residual artifacts and combine data from different time frames and coils into a coil-combined output image. The U-Net subnetwork consists of contracting and expanding paths that can remove image artifacts at multiple resolutions. In the contracting path, three stages of spatial down-sampling are

applied consisting of two 2D XConv, RBN, and XReLU layers at each stage. The number of kernels doubles after each down-sampling stage (32, 64, and 128, respectively). The resulting feature maps pass through two XConv layers of 256 and 128 kernels of size 3×3, respectively. The expansive path maps the output at each down-sampling stage to an analogous stage of similar map size and kernel number using skip connections. Up-sampling layers are used to increase the size of feature maps by a factor of 2 at each up-sampling stage to provide a clean version of the artifact-contaminated images at different resolution levels. The final feature maps are combined using the XConv layer of one 1×1 kernel to generate a coil-combined complex-valued reconstructed image.

2.2 | Data acquisition

The study was approved by the Beth Israel Deaconess Medical Center Institutional Review Board (IRB number 2001P000793). All imaging was performed on a 3T

system (MAGNETOM Vida, Siemens Healthcare, Erlangen, Germany) using body and spine phased-array coils. Radial bSSFP cine data were acquired in 101 patients and 7 healthy subjects (108 total subjects; age 50 ± 17 y, 72 males, heartbeats 72 ± 15 beats/min, and weight 80.6 ± 18 kg). Participants provided written informed consent to use cardiac MRI studies for research purposes. The indications for clinical CMR scans for our cohort are reported in Supporting Information Table S1, which is available online. A mid-ventricular slice was imaged during breath-holding with the following imaging parameters: repetition time/echo time (TR/TE) = 3.06/1.4 ms, field of view = 380 × 380 mm^2, matrix size = 208 × 208, in-plane resolution = 1.8 × 1.8 mm^2, slice thickness = 8 mm, flip angle = 48°, number of channels = 16 ± 1, and retrospective ECG-triggering with 25 cardiac phases calculated. In each patient, 196 radial views were acquired on average per cardiac phase with a breath-hold duration of ~14 heartbeats.

2.3 | Data preparation

The dataset was randomly divided into training (87 subjects) and testing (21 subjects) subsets. The complex-valued reference images at each time frame were reconstructed with all acquired 196 radial views per frame using non-uniform fast Fourier transform (NUFFT) implemented on the scanner,[21] where the complex-valued images were saved before writing the DICOM images. An undersampled dataset was synthesized by selecting a small set (eg, N_v =14 views) from the acquired 196 radial views at each time frame; such that an equivalent acquisition duration of one heartbeat per slice

is achieved. The effective acceleration rate is 23.3 based on the Nyquist sampling requirement (208 × $\frac{\pi}{2}$ = 327 views).[22] To create an undersampled dataset, radial views across a 14-spokes radial trajectory (T_{14}) were selected from the acquired highly sampled radial data (ie, 196 spokes). First, the uniform angles of T_{14} are calculated and matched to the nearest 14 angles of the 196-spokes trajectory (Figure 2). The selected 14 angles are considered as the undersampled trajectory.

To create rotated angles trajectory among different frames, T_{14} was rotated in each frame, p, with an angle $\theta_p = \frac{p\,\pi}{N_v\,N_{tw}}$, where p is the frame index (p = 0, 1, … 24), N_{tw} is a window of neighboring time frames (eg, N_{tw} = 7), and N_v is the total number of acquired views (eg, 196) (Figure 2). To enable optimized k-space coverage for data sharing among consecutive N_{tw} frames, the rotating angle θ_p also maintains uniform angular distance among all views from the N_{tw} neighboring frames. Gradient delay induced errors in the radial trajectory were automatically corrected using the RING method[23] implemented from the Berkeley Advanced Reconstruction Toolbox (BART).[24] Gradient delays were corrected in each time frame using only the selected 14 radial views.

The radial k-space of each frame and each coil was reconstructed using inverse NUFFT. Coil compression was applied to obtain a fixed number of coils (N_c = 8) per frame using BART.[25] Complex-valued gridded data from consecutive time frames of window size, N_{tw} = 7, with eight coils per frame were normalized in the image domain using RBN. Considering the periodic nature of the dynamic cardiac imaging, information can be shared from the last frames of the dynamic cine imaging to reconstruct the first frames and vice versa. For example, to reconstruct the first frame, a window
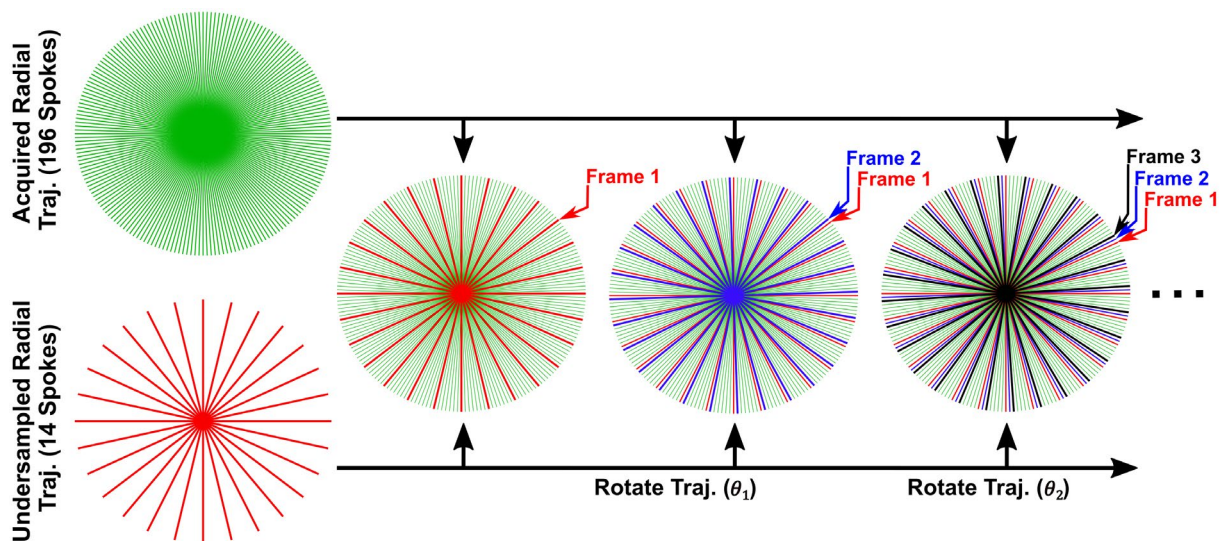

FIGURE 2 Radial undersampling using uniform-angle rotating radial trajectories. An undersampled uniform radial trajectory of 14 spokes is selected from the acquired highly sampled radial data (196 spokes). The radial trajectory in a frame p is rotated with an angle $\theta_p = \frac{p\pi}{N_v N_{tw}}$, where N_v is the number of views per frame (eg, 14 spokes), and N_{tw} is the number of neighboring time frames (eg, seven frames). The undersampled radial lines from N_{tw} consecutive time frames have a uniform angular-spacing trajectory that allow optimal k-space coverage for data sharing

of neighboring frames at indices {23, 24, 25, 1, 2, 3, and 4} is included at the network input. Next, FFT was applied to obtain 2D k-spaces fed into MD-CNN to reconstruct the coil-combined image at a single time frame (ie, corresponding to the middle frame of the input window $\left\lfloor \frac{N_{tw}}{2} \right\rfloor$).

2.4 | Network training

MD-CNN was trained with 250 epochs via a stochastic gradient descent SGD optimizer with a momentum of 0.9 and learning rate function $0.01^{\lfloor epoch/50 \rfloor + 1}$, where the learning rate exponentially decreases with the number of epochs. The batch number was 15 complex-valued data points of size $8 \times 7 \times 208 \times 208$. The mean-squared error loss function was applied to minimize reconstruction error between MD-CNN predictions and reference images at any time frame, t, such that:

$$\min_{\Theta_k, \Theta_m} \left\| M_{xy}(t) - \Psi \left(A\Phi \left(K_{xyw}(t) | \Theta_k \right) | \Theta_m \right)_{xy} \right\|_2^2,$$

where $M_{xy}(t)$ is a complex-valued, coil-combined reference image at a single time point, t, reconstructed using all 196 acquired radial views, $t \in [0, 1, \ldots 24]$; $K_{xyw}(t)$ is a complex-valued, multi-coil 2D spatial k-space gridded from undersampled radial acquisitions at N_{tw} time frames (ie, 2D k-space + time); A is the Fourier encoding matrix of the 2D FFT; $\Phi(.)$ is a nonlinear interpolation function controlled by trainable parameters, Θ_k; and $\Psi(.)$ is a non-linear mapping function controlled by trainable parameters, Θ_m, which take N_{tw} time frames as input to reconstruct a single frame at the output. The same network parameters, Θ_k and Θ_m, were trained using all time frames. During MD-CNN training, time frames from all patients in the training dataset were randomized and 15 (ie, batch size) frames from different patients were selected to be reconstructed in each training iteration.

MD-CNN was implemented in Python using the PyTorch library version 0.41.[26] All models in this work were trained and tested on an NVIDIA DGX-1 system equipped with 8T V100 graphics processing units (GPUs; each of 32 GB memory and 5120 cores), central processing unite (CPU) of 88 core: Intel Xeon 2.20 GHz each, and 504 GB RAM. Only four GPUs were used to train MD-CNN. The total MD-CNN training time was 24 h.

2.5 | Performance evaluation

To investigate the impact of the window size N_{tw} of input time frames, MD-CNN performance was evaluated on testing dataset reconstructed at three different window sizes (N_{tw}

= 5, 7, and 9 frames), where MD-CNN was trained from scratch at each window size. The performance of MD-CNN was compared with a compressed-sensing based k-t Radial Sparse-Sense (kt-RASPS) reconstruction method, which exploits the temporal sparsity of the data.[11,27] The same undersampled dataset prepared for MD-CNN was used for kt-RASPS reconstruction where 14 views, corresponding to a temporal resolution of 42.8 ms, were selected per frame with a uniform-angle rotating among different frames. The kt-RASPS algorithm was implemented using the parallel imaging and compressed sensing tools in BART with GPUs and parallel processing. Coil sensitivities were calculated using the ESPIRiT method.[25] kt-RASPS parameters were optimized on 10 randomly selected subjects from the training dataset, where images were evaluated to determine the optimal value of the regularization level and number of iterations. kt-RASPS performance was evaluated over a range of regularization levels (λ) from 0.01-0.1 with a step of 0.005 and number of iterations ranging from 25 to 100, with 25 steps. MD-CNN performance was also compared to the CS-based method proposed by Miao et al,[28] in which a combination of locally low rank (LLR) and temporal finite differences (FD) were used as regularization terms. This method was used to reconstruct the same undersampled dataset of 14 views/frame with regularization parameters $\lambda_{LLR} = 0.06$ and $\lambda_{FD} = 0.006$.

To evaluate MD-CNN performance compared to other deep learning frameworks, a 3D U-net network[16] was used to reconstruct the same undersampled data (ie, 14 spokes). This network consisted of contracting and expanding paths with two stages of spatio-temporal down-sampling and up-sampling in each path, respectively. The radial cine data from all time frames was gridded using NUFFT, coil-combined, and fed into the U-net as a 3D volume in the image domain of size $208 \times 208 \times 24$ (ie, 2D + time) with real and imaginary components as different channels. To allow down-sampling and up-sampling of the temporal dimension inside the network, the temporal dimension of the input and reference data was reduced to 24 frames (compared to the original data of 25 frames) using linear interpolation of all time frames. The mean-squared error loss function was used to train this network using the stochastic gradient descent optimizer with a learning rate of 0.01 and 350 epochs.

MD-CNN and kt-RASPS performance was evaluated by quantitative and qualitative measures. Mean squared error (MSE) and structural similarity index measure (SSIM) were calculated for MD-CNN, U-net and kt-RASPS with respect to the reference images. The magnitude values of all images are normalized by the 90th percentile before assessment. Edge sharpness of left ventricular (LV) borders was quantitatively evaluated across six LV myocardial segments (based on the AHA 16-segment model) in all time frames. In each segment, edge sharpness was calculated as the maximum gradient of the normalized intensity profile (between 0 and 1) across this

segment. To avoid misleading gradients produced by noise, each intensity profile was fit to a high-order polynomial function of order 10.

For qualitative assessment, three readers (U.N., S.K., and I.C. with > 5 y of experience in cardiovascular imaging) blinded to the reconstruction methods independently assessed images reconstructed by MD-CNN, kt-RASPS, and the reference based on four different metrics: (a-b) the sharpness of LV borders at end-diastolic (ED) and end-systolic (ES) phases, (c) temporal fidelity of LV wall motion using a 5-point Likert scale (1-non-diagnostic, 2-poor, 3-fair, 4-good, and 5-excellent), and (d) residual image artifacts in the entire field-of-view on a 5-point scale (1-minimal, 2-mild, 3-moderate, 4-severe, and 5-non-diagnostic).

Endocardial and epicardial LV myocardial borders from MD-CNN, kt-RASPS, and reference images in ED and ES frames were manually delineated by all readers. LV myocardium (ie, areas between endocardial and epicardial contours) and LV blood pool cavity were calculated from the extracted contours for the three datasets (ie, reference, kt-RASPS, and MD-CNN). To further investigate the effect of LV border sharpness on manual delineation, we calculated the DICE index between kt-RASPS vs. reference and MD-CNN vs. reference for the LV myocardium and LV blood cavity at the ED and ES phases.

2.6 | Data analysis

Normal data distributions were expressed in the form of mean $\pm$ SD. The two-sided Student's t-test compared continuous variables between different reconstruction methods. Analysis of variance or Kruskal-Wallis tests were used as appropriate for comparison of multiple groups. For comparison of categorical data, the Chi-squared test was used. Significance was declared at two-sided P-values < .05. For pairwise comparisons following a three-group inferential test, a Bonferroni correction was used. Bland-Altman plots assessed agreement between different readers. Statistical analyses were conducted using MATLAB (2017a, The MathWorks Inc., Natick, Massachusetts, United States).

3 | RESULTS

MD-CNN performance increased as window size increased, as indicated by the decreasing MSE (0.125 $\pm$ 0.09, 0.11 $\pm$ 0.1, and 0.106 $\pm$ 0.1) and increasing SSIM (0.85 $\pm$ 0.072, 0.87 $\pm$ 0.067, and 0.87 $\pm$ 0.062) for N_{tw} = 5, 7, and 9, respectively. MD-CNN reconstructed data also showed similar temporal fidelity for the three window sizes with increased flickering artifacts at N_{tw} = 5 compared to seven and nine frames (Supporting Information Video S1). A window size

of seven frames was, therefore, used in the rest of our experiments due to its similar performance to N_{tw} = 9 with lower reconstruction time and memory requirements, and improved performance compared to N_{tw} = 5. Supporting Information Figure S1 shows the MSE for MD-CNN images in the training and testing datasets at each of the 250 training epochs. Both training and testing MSE consistently decreased as the number of training epochs increased, indicating a robust performance and minimal overfitting of the MD-CNN. Parameters tuning for kt-RASPS yielded minimal streaking artifacts and maximal temporal fidelity at λ = 0.025 and #iterations = 50 (Supporting Information Videos S2 and S3).

Figure 3 shows examples from two subjects and compares the images (and their corresponding k-space) after different stages of the MD-CNN reconstructions. Comparing the input and output of the 1st and 2nd blocks of the k-space subnetwork (columns 1–3), the k-space convolutional subnetwork gradually estimated the missing data at more k-space locations by interpolating values near the acquired k-space data and neighboring time frames. K-space interpolation reduced streaking artifacts in the corresponding images while maintaining image sharpness. The image subnetwork further reduced streaking artifacts and improved image sharpness and contrast (column 4). The final MD-CNN-reconstructed images demonstrated a similar contrast as the reference images (column 5).

Cine images from undersampled radial data were reconstructed by NUFFT, kt-RASPS, U-net, and MD-CNN compared to the reference images (Figure 4). MD-CNN, U-net, and kt-RASPS reconstructions showed less pronounced streaking artifacts than NUFFT images. Similar to reference images, MD-CNN images preserved anatomical sharpness and contrast of the myocardium compared to kt-RASPS, and less residual artifacts compared to U-net. Cine imaging movies for two additional cases reconstructed by LLR+FD, kt-RASPS, MD-CNN, and reference showed improved temporal fidelity of MD-CNN network compared to CS methods (Supporting Information Videos S4 and S5) and reduced residual streaking artifacts compared to the 3D U-net. Figure 5 shows MD-CNN reconstructions from one subject at different undersampling rates: 14, 16, and 20 spokes per frame compared to the corresponding reference image. X-t plots and Supporting Information Video S6 showed increased temporal fidelity as undersampling rate decreases indicated by less blurring in the LV wall at 20 spokes compared to reconstructions with 14 spokes per frame.

MD-CNN showed lower MSE and improved SSIM compared to kt-RASPS and U-net (P < .01 for both) (Table 1). The intensity profiles of MD-CNN reconstructed images also showed higher sharpness across the blood-myocardium at different segments than kt-RASPS at ED and ES frames (Figure 6, and Supporting Information Figure S2, respectively). MD-CNN maintained the blood-myocardium
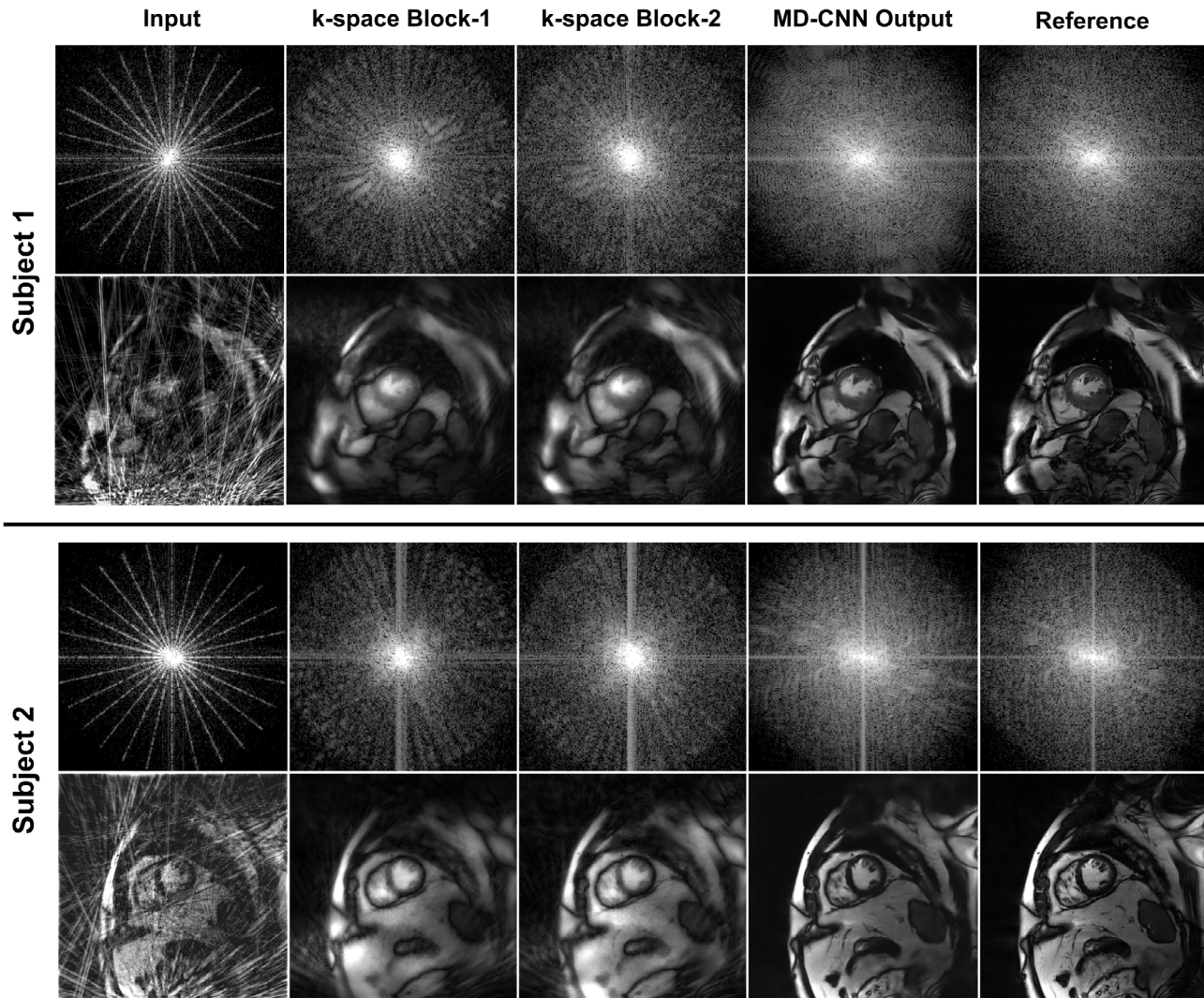
FIGURE 3 Performance of the k-space subnetwork in two subjects. The k-space data are shown at different stages of the network: the gridded input k-space, after the 1st and 2nd residual blocks of the k-space subnetwork, the k-space of MD-CNN reconstruction, and the k-space of the reference image (rows 1 and 3). The corresponding images are also shown (rows 2 and 4). The k-space gaps are gradually filled and the corresponding streaking artifacts of input images are suppressed after k-space interpolation in 1st and 2nd k-space blocks (columns 2 and 3). The final MD-CNN reconstructed images have further reduced artifacts and improved image quality similar to the reference images. The k-space of the input, k-space block-1, and kspace block-2 are taken from a single coil. All k-space data are presented on a logarithmic scale

contrast of small anatomic structures such as papillary muscles, as indicated by the intensity profiles in segment 4 of Figure 6. Per-segment LV sharpness is illustrated in the bull's-eye plot with mean $\pm$ SD for each segment calculated over all time frames in the testing dataset. MD-CNN images had better LV sharpness in all segments compared to kt-RASPS ($P < .01$) (Table 1).

Qualitative assessments of LV border sharpness, temporal fidelity, and residual artifacts are reported in Table 2. The sharpness of the LV myocardium at end-diastole and end-systole was significantly improved in MD-CNN reconstruction compared to kt-RASPS reconstruction (3.87 ± 0.66 vs. 2.71 ± 0.58, and 3.57 ± 0.58 vs. 2.56 ± 0.60 for end-diastole and end-systole, respectively; P-value $< .01$). LV myocardium sharpness was higher in ED frames compared to ES

MD-CNN and kt-RASPS, potentially due to less cardiac motion at end-diastole. MD-CNN reconstructed images exhibited less temporal LV wall blurring compared to kt-RASPS (P-value $< .01$). Residual artifacts were reduced in MD-CNN images indicating the network's ability to recover image sharpness while suppressing artifacts.

The percentage error of the extracted LV myocardial area in MD-CNN images was lower than in kt-RASPS images, with a 9% and 4% decrease at ED and ES frames relative to reference images, respectively (P-value $< .01$) (Table 3). The LV cavity area extracted from MD-CNN also showed improved DICE index and reduced percentage error compared to that of the kt-RASPS images at both ED and ES frames with respect to the reference images. Figure 7 and Supporting Information Figure S3 shows agreement between
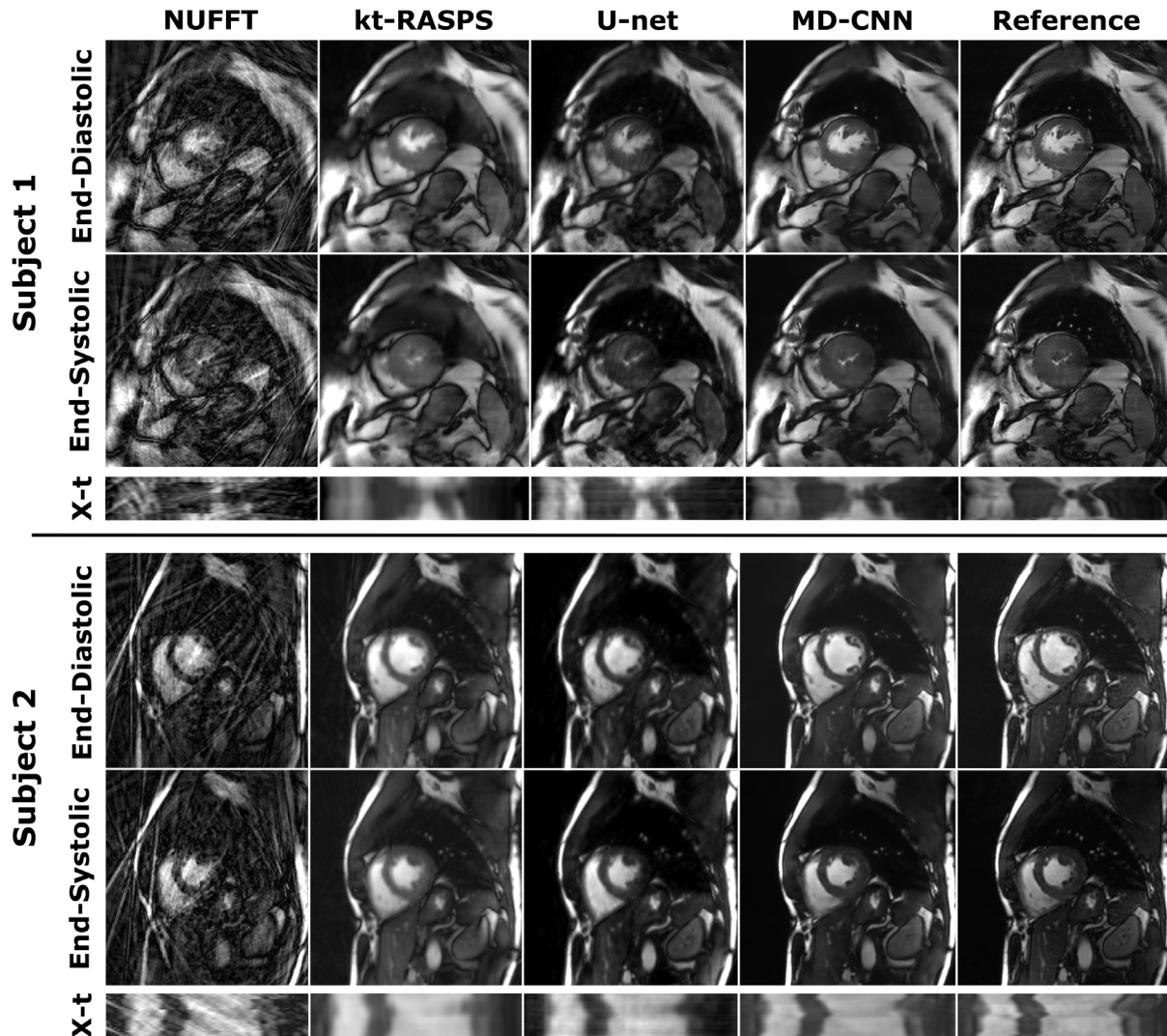
FIGURE 4 Representative reconstructions by NUFFT, kt-RASPS, U-net, and MD-CNN compared to the reference images for two subjects at end-diastolic and end-systolic phases, as well as x-t plots across all time frames
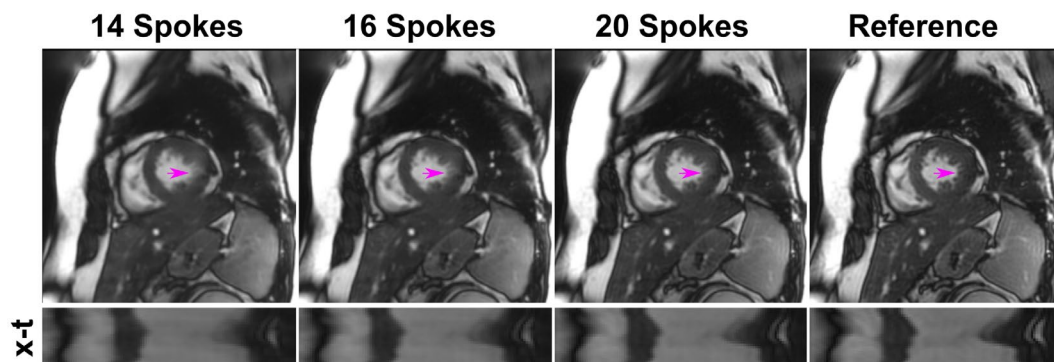

FIGURE 5 Representative reconstruction by MD-CNN at different undersampling rates: 14, 16, and 20 spokes per frame compared to reference image. X-t plots of the reconstructed images at each undersampling rate (bottom row) shows the temporal fidelity across all time frames. Temporal fidelity increases as undersampling decreases (magenta arrows)

readers segmenting the LV myocardial area from kt-RASPS, MD-CNN, and reference images. MD-CNN showed higher agreement between readers (ie, smaller limits of agreement) similar to that of reference images (Supporting Information Table S2). MD-CNN takes 1.6 ± 0.35 s to reconstruct a set of 25 frames of one slice compared to 5.4 ± 1.1 s for kt-RASPS.

TABLE 1 Quantitative assessment of image quality by mean-squared error (MSE), structural-similarity index measure (SSIM), and LV myocardial sharpness for kt-RASPS, 3D U-net, and MD-CNN reconstructed images with respect to the reference

Method	kt-RASPS	U-net	MD-CNN	Reference	*P*-value
MSE	0.61 ± 0.51*	0.13 ± 0.16	0.11 ± 0.10	–	<.01
SSIM	0.72 ± 0.07*	0.84 ± 0.08*	0.87 ± 0.067	–	<.01
Quantitative LV myocardial sharpness	0.12 ± 0.04*§	0.16 ± 0.06§	0.17 ± 0.06§	0.21 ± 0.07	<.01

All values are reported as mean ± SD.

§*P*-value < .01 compared to Reference; *P*-value < .01 compared to MD-CNN.
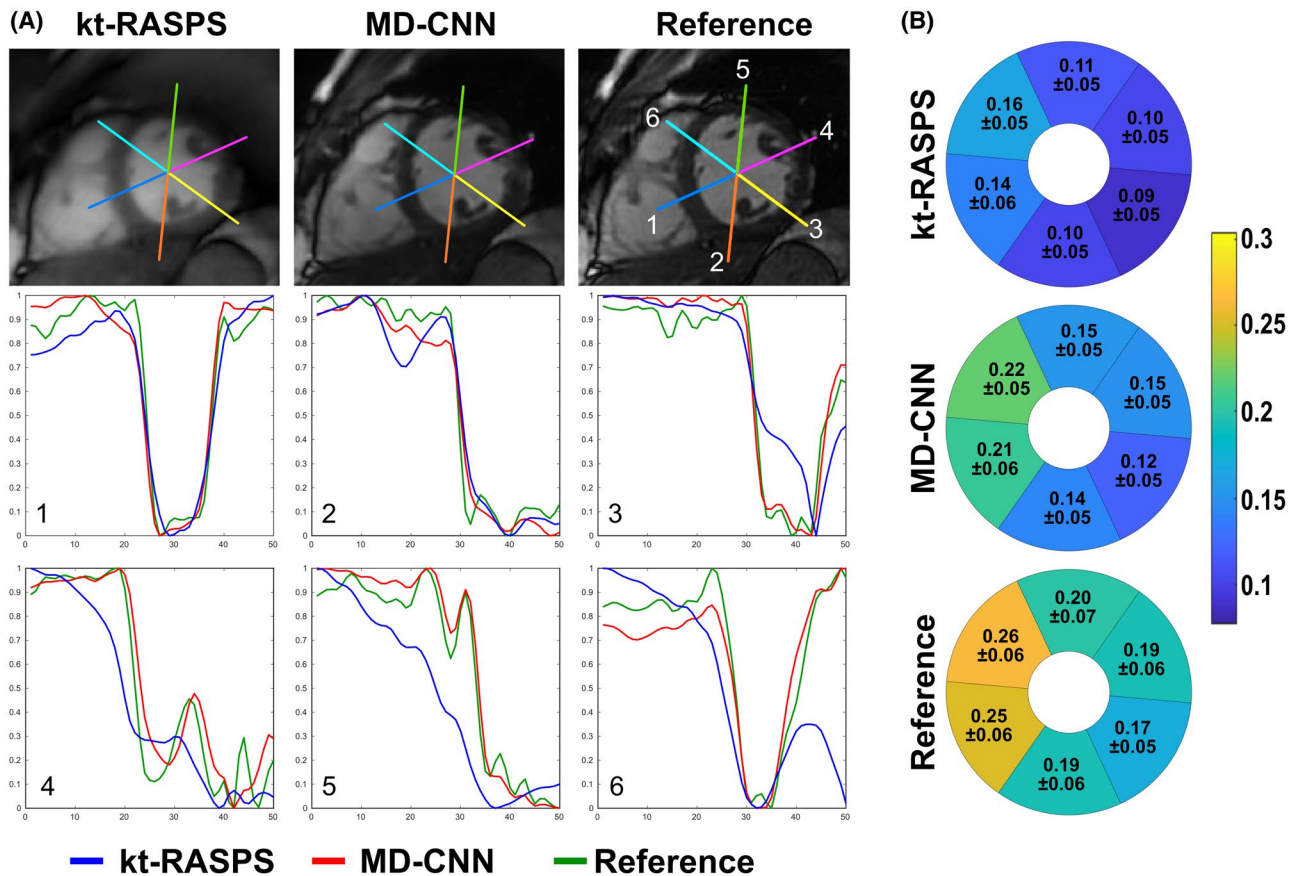

FIGURE 6 Quantification of the left ventricular (LV) edge sharpness at six myocardial segments in GRAS, MD-CNN, and reference images. A, Sample reconstructed images at end-systole by the three methods (kt-RASPS, MD-CNN, and Reference) showing the six segments at which LV sharpness was quantified and the intensity profiles corresponding to each segment of the sample images. B, The bull's eye plot shows the value of the quantified LV edge sharpness at the six different segments of the myocardium with mean ± SD calculated over all time frames of the testing dataset

4 | DISCUSSION

In this study, we developed a multi-domain network that processes MR data in both k-space and image domains with an end-to-end training process. In general, all k-space operations can be performed in the image domain and vice versa; for example, convolving a kernel to an image is equivalent to multiplying this kernel's response to the k-space of that image. However, in the context of CNN, all operations are based on convolutional kernels only; hence, certain operations are

feasible to be performed in the k-space and others in the image domain using CNN. The k-space subnetwork performs an interpolation of the input 2D k-space using shared information among adjacent frames to provide an enhanced estimation of missing k-space data, particularly at locations far from the acquired radial lines (ie, gaps among radial lines). Filling such k-space gaps is crucial for suppressing image domain's streaking artifacts. Although conventional gridding techniques use single hand-crafted kernel to interpolate MR signals at all k-space locations, the performance of such kernels degrades

Method	kt-RASPS	MD-CNN	Reference
Sharpness of LV borders at end-diastole	2.71 ± 0.58*§	3.87 ± 0.66§	4.71 ± 0.52
Sharpness of LV borders at end-systole	2.56 ± 0.60*§	3.57 ± 0.58§	4.60 ± 0.58
Temporal fidelity of wall motion	2.59 ± 0.59*§	3.27 ± 0.65§	4.65 ± 0.51
Residual artifacts	2.63 ± 1.09§	2.38 ± 0.81§	1.81 ± 0.82

TABLE 2 Qualitative assessment of LV border sharpness at end-diastolic and end-systolic cardiac phases, temporal fidelity of myocardial wall motion, and residual artifacts in kt-RASPS, MD-CNN, and reference images

§P-value < .01 compared to Reference;

*P-value < .01 compared to MD-CNN.

Method	Phase	kt-RASPS	MD-CNN	P-value
DICE index of LV myocardium	Diastole	0.77 ± 0.12	0.84 ± 0.05	<.01
	Systole	0.83 ± 0.06	0.88 ± 0.07	<.01
DICE index of LV endocardium	Diastole	0.97 ± 0.01	0.98 ± 0.03	.011
	Systole	0.96 ± 0.01	0.97 ± 0.01	<.01
Percentage error of LV myocardial area (%)	Diastole	21.4 ± 19.1	12.0 ± 13.1	<.01
	Systole	13.8 ± 13.3	9.4 ± 8.0	<.01
Percentage error of LV endocardial area (%)	Diastole	4.5 ± 4.0	3.6 ± 3.3	.15
	Systole	7.3 ± 6.2	5.2 ± 4.3	<.01

TABLE 3 DICE index and percentage error of LV myocardial and endocardial areas at end-diastolic and end-systolic phases extracted from kt-RASPS and MD-CNN images with respect to reference images

rapidly at higher acceleration rates, leaving large k-space gaps and hence more streaking artifacts. The amount and shape of such artifacts are largely determined by the design of the gridding kernel and the undersampling rate. CNN, on the other hand, offers an efficient k-space interpolation strategy that optimizes multiple convolutional kernels on many training samples to determine the proper interpolation weights specific to the undersampling scheme used. Additionally, the non-linear activation functions used after each convolutional layer in CNN may allow a spatially variant interpolation process, in which the high- and low-frequency k-space content can be interpolated differently.

Considering that successive time frames have similar structural information, k-space interpolation across the time dimension (using 3D convolution kernels) also allows spoke sharing among neighboring frames with similar image content. Data sharing in turn allows for better estimation of k-space values at more locations on k-space gaps, thereby reducing streaking artifacts. To exploit shared information across the time dimension in the form of spokes sharing, the radial trajectory was rotated from frame to frame at a small angle (1.8°) that maintained a uniform angular spacing among neighboring frames.[29,30] The importance of k-space processing in our network was demonstrated by assessing the resulting k-space of each residual block in the k-space subnetwork. The blocks of the k-space subnetwork gradually filled the missing values in k-space and led to significantly reduced streaking artifacts in the reconstructed images while maintaining fine image details and sharpness.

The refined k-space data are transformed into the image domain using inverse FFT for further processing via the image-subnetwork. To guarantee a proper flow of gradients between the two subnetworks for the back-propagation algorithm during training, a differentiable FFT operation was implemented using native PyTorch functions. Maintaining complex-valued MR data throughout the MD-CCN intermediate layers was also necessary to allow a proper transformation of data between the k-space and image domain. To process complex-valued data, a complex convolutional network was used to maintain the proper combination of real and imaginary components of complex MR data throughout all network layers.

In the image-subnetwork, the improved myocardium-blood contrast, artifact suppression, and sharp edges in the MD-CNN output images are likely associated with the ability of the image-subnetwork to compensate for variable k-space density and suppress more residual artifacts. Considering the incoherent residual (streaking) artifact patterns among different frames, the 3D convolution layers in this subnetwork enable further removal of residual artifacts by sharing spatio-temporal correlations among neighboring frames, similar to compressed sensing.[12,16]

MD-CNN images had superior image quality than kt-RASPS and U-net on all quantitative measures. MD-CNN was trained to produce images that minimally differ from reference images in pixel value, such that quantitative measures are implicitly biased toward the MD-CNN-reconstructed images compared to kt-RASPS images. To alleviate this
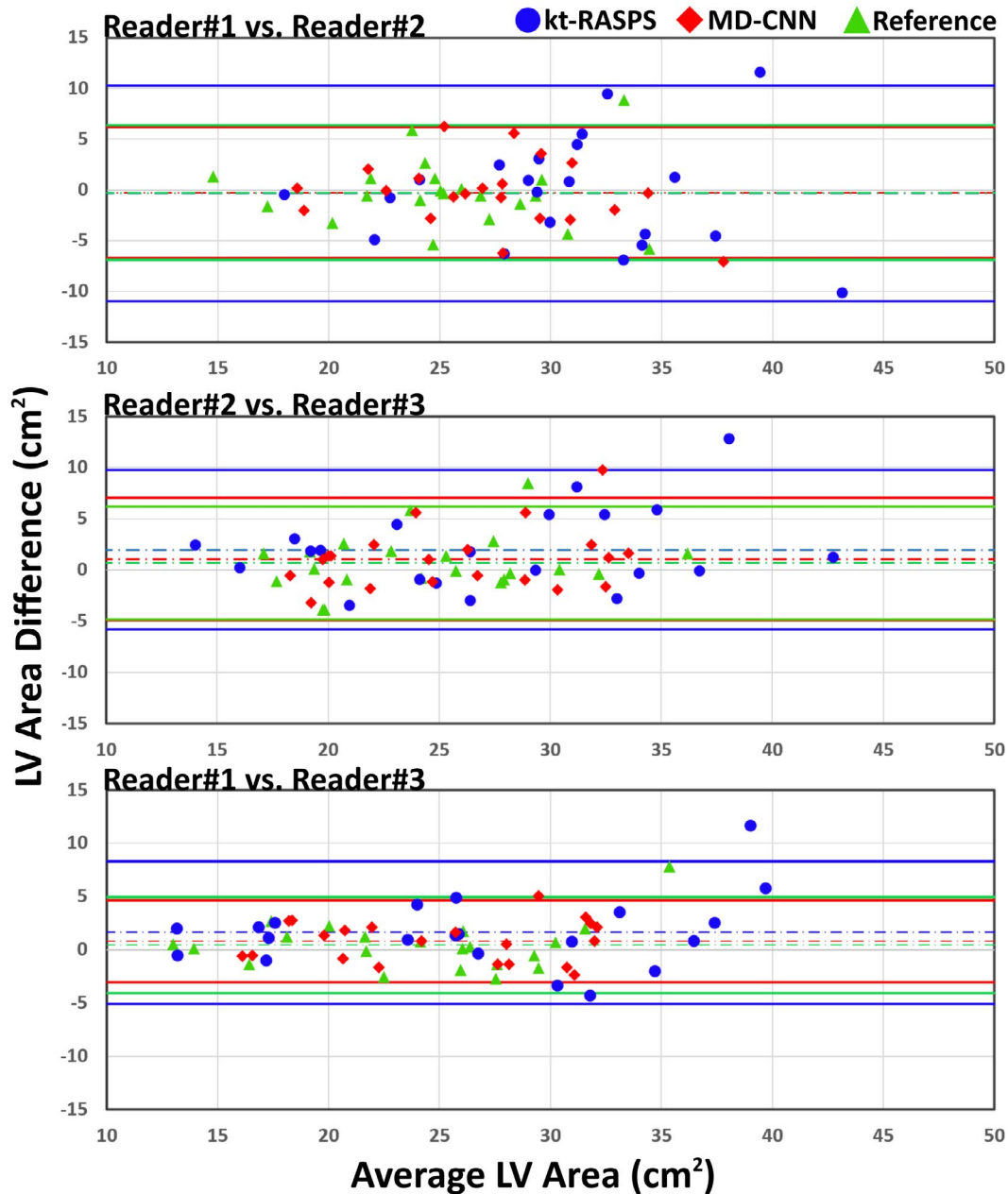
FIGURE 7 Bland-Altman plots showing inter-observer agreement of three comparisons: Reader 1 vs. Reader 2, Reader 2 vs. Reader 3, and Reader 1 vs. Reader 3 of the extracted LV myocardial areas in kt-RASPS, MD-CNN, and reference images. Biases and limits of agreement for each comparison are reported in Supporting Information Table S2

bias, kt-RASPS, MDC-CNN, and reference images were normalized by the 90th percentile before MSE and SSIM calculations. In addition, to alleviate the impact of myocardium-blood contrast differences on LV sharpness quantifications, each intensity profile across the LV myocardium was standardized separately from 0 to 1 before the sharpness calculations. In addition, MD-CNN allows further suppression of the residual streaking artifacts compared to the 3D U-net. This is mainly due to the ability of the k-space interpolation subnetwork to provide cleaner images to the image-domain subnetwork, unlike the U-net that processes cine data in the image domain only.

In the qualitative assessment, MD-CNN images also demonstrated improved LV edge sharpness at ED and ES phases and reduced temporal blurring compared to kt-RASPS images. This was also indicated by greater agreement among readers assessing the extracted LV areas from MD-CNN vs. kt-RASPS reconstruction. These results also reflect higher confidence among readers to accurately segment the sharp LV borders from MD-CNN images. Although the temporal fidelity of MD-CNN reconstruction at 14 views was assessed as fair by the readers, higher temporal fidelity can be obtained by including more views (optimally observed at 20 views per frame). Both methods

showed similar residual artifact content, but the artifacts were different in nature. Kt-RASPS endorsed streaking artifacts, while flickering artifacts were observed in MD-CNN.

A few considerations may have contributed to inferior reconstruction by kt-RASPS in this study. The kt-RASPS method is more suitable for real-time imaging applications where a large number of frames are available (>>25), so that sufficient sparsity along the time dimension is achieved. However, in our reconstruction problem, only 25 phases were available for reconstruction, which may have led to additional blurring in kt-RASPS images. The limited number of frames in our data also led to faster kt-RASPS reconstruction times compared to the real-time applications.[12] On the other hand, the proposed MD-CNN method is able to exploit redundancy in a small number of adjacent frames (seven in the experiments) and thereby efficiently reduce undersampling artifacts. In addition, the rotation angle in this study is uniform, as opposed to the golden-angle rotation that is widely used in real-time applications.[12,22] However, the uniform trajectories exhibit better characteristics that allow equidistant spacing between views when compared to golden-angle trajectories.

Our study has some limitations. Only one slice at mid-LV was acquired per patient, which was not sufficient to calculate LV functional parameters (eg, LV volumes and ejection fraction). All patients studied were in sinus rhythm, and generalization of the reconstruction techniques in patients with irregular rhythms was not studied.

5 | CONCLUSION

The proposed multi-domain network uses k-space and image domain processing units to efficiently reconstruct highly undersampled radial cine MR data. MD-CNN exploits the spatio-temporal correlations among neighboring time frames and multiple coils to enable higher acceleration rates.

CONFLICT OF INTEREST
Xiaoying Cai is a Siemens Employee.

DATA AVAILABILITY STATEMENT
The code and data that support the findings of this study are openly available in Harvard dataverse at (https://dataverse.harvard.edu/dataverse/cardiacmr), reference number (https://doi.org/10.7910/DVN/E9FFZK).

ORCID
Hossam El-Rewaidy (iD) https://orcid.org/0000-0002-5266-8702

REFERENCES

1. American College of Cardiology Foundation Task Force on Expert Consensus Documents; Hundley WG, Bluemke DA, et al. ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: A report of the american college of cardiology foundation task force on expert consensus documents. *Circulation*. 2010;121:2462-2508.
2. Rehwald WG, Kim RJ, Simonetti OP, Laub G, Judd RM. Theory of high-speed MR imaging of the human heart with the selective line acquisition mode. *Radiology*. 2001;220:540-547.
3. Song HK, Dougherty L. k-Space weighted image contrast (KWIC) for contrast manipulation in projection reconstruction MRI. *Magn Reson Med*. 2000;44:825-832.
4. McGibney G, Smith MR, Nichols ST, Crawley A. Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. *Magn Reson Med*. 1993;30:51-59.
5. Kellman P, Chefd'hotel C, Lorenz CH, Mancini C, Arai AE, McVeigh ER. Fully automatic, retrospective enhancement of real-time acquired cardiac cine MR images using image-based navigators and respiratory motion-corrected averaging. *Magn Reson Med*. 2008;59:771-778.
6. Kellman P, Chefd'hotel C, Lorenz CH, Mancini C, Arai AE, McVeigh ER. High spatial and temporal resolution cardiac cine MRI from retrospective reconstruction of data acquired in real time using motion correction and resorting. *Magn Reson Med*. 2009;62:1557-1564.
7. Feng L, Srichai MB, Lim RP, et al. Highly accelerated real-time cardiac cine MRI using k-t SPARSE-SENSE. *Magn Reson Med*. 2013;70:64-74.
8. Haji-Valizadeh H, Rahsepar AA, Collins JD, et al. Validation of highly accelerated real-time cardiac cine MRI with radial k-space sampling and compressed sensing in patients at 1.5T and 3T. *Magn Reson Med*. 2017;79:2745-2751. https://doi.org/10.1002/mrm.26918.
9. Kido T, Kido T, Nakamura M, et al. Compressed sensing real-time cine cardiovascular magnetic resonance: Accurate assessment of left ventricular function in a single-breath-hold. *J Cardiovasc Magn Reson*. 2016;18:50.
10. Voit D, Zhang S, Unterberg-Buchwald C, Sohns JM, Lotz J, Frahm J. Real-time cardiovascular magnetic resonance at 1.5 T using balanced SSFP and 40 ms resolution. *J Cardiovasc Magn Reson*. 2013;15:79.
11. Feng L, Grimm R, Block KT, et al. Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn Reson Med*. 2014;72:707-717.
12. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. *Magn Reson Med*. 2016;75:775-788.
13. 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*. 2018;37:491-503.
14. 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*. 2019;38:280-290.

15. Wang S, Ke Z, Cheng H, et al. DIMENSION: Dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. *NMR Biomed*. 2019:e4131. https://doi.org/10.1002/nbm.4131.

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.

17. 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.

18. Akçakaya M, Moeller S, Weingärtner S, Uğurbil K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magn Reson Med*. 2019;81:439-453.

19. Kofler A, Dewey M, Schaeffter T, Wald C, Kolbitsch C. Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data. *IEEE Trans Med Imaging*. 2020;39:703-717.

20. El-Rewaidy H, Neisius U, Mancio J, et al. Deep complex convolutional network for fast reconstruction of 3D late gadolinium enhancement cardiac MRI. *NMR Biomed*. 2020:e4312.

21. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans Signal Process*. 2003;51:560-574.

22. Feng L, Grimm R, Block KT, et al. Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn Reson Med*. 2014;72:707-717.

23. Rosenzweig S, Holme HCM, Uecker M. Simple auto-calibrated gradient delay estimation from few spokes using Radial Intersections (RING). *Magn Reson Med*. 2019;81:1898-1906.

24. Tamir J, Ong F, Cheng J, Uecker M, Lustig M. Generalized magnetic resonance image reconstruction using the Berkeley advanced reconstruction toolbox. In: Proceedings of ISMRM Workshop on Data Sampling and Image Reconstruction, Sedona, AZ; 2016. p. 33.

25. 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.

26. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. In: *Neural Information Processing Systems Workshop*. California, USA; 2017.

27. Feng L, Chandarana H, Xu J, Block KT, Sodickson D, Otazo R. K-t radial SPARSE-SENSE: Combination of compressed sensing and parallel imaging with golden angle radial sampling for highly accelerated volumetric dynamic MRI. In: International Society for Magnetic Resonance in Medicine. Melbourne, Australia; 2012. p. 81.

28. Miao X, Lingala SG, Guo Y, et al. Accelerated cardiac cine MRI using locally low rank and finite difference constraints. *Magn Reson Imaging*. 2016;34:707-714.

29. Rasche V, De Boer RW, Holz D, Proksa R. Continuous radial data acquisition for dynamic MRI. *Magn Reson Med*. 1995;34:754-761.

30. Zhang S, Block KT, Frahm J. Magnetic resonance imaging in real time: Advances using radial FLASH. *J Magn Reson Imaging*. 2010;31:101-109.

SUPPORTING INFORMATION

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

VIDEO S1 MD-CNN sample reconstruction of all time frames at three different neighboring window sizes (N_{tw} = 5, 7, and 9 frames) from the left side to the right side, respectively. The flickering artifacts are suppressed at larger window sizes (eg, 7 (middle) and 9 (right)) while the temporal fidelity did not degrade at these large window sizes when compared to a window size of 5

VIDEO S2 K-t radial sparse-sense (kt-RASPS) sample reconstruction of all time frames at three different values of regularization weights (λ = 0.01, 0.025, and 0.1) from the left side to the right side, respectively. The reconstructed data with low regularization values (λ = 0.01) showed a high level of streaking artifacts, and increased temporal blurring at higher regularization levels (λ = 0.1), while data at λ = 0.025 showed minimal streaking artifacts with relatively high temporal fidelity

VIDEO S3 K-t radial sparse-sense (kt-RASPS) sample reconstruction of all time frames at 3 different values of regularization weights (#iterations = 25, 50, and 100) from the left side to the right side, respectively. The image quality increases at #iterations = 50 compared to that at #iterations = 25, while image quality was maintained at #iterations = 50 and 100

VIDEO S4 Representative reconstruction at 14 spokes/frame of undersampled data from one subject at all cardiac phases by different reconstruction methods: the locally low rank and temporal finite differences compressed sensing method (LLR+FD), k-t radial sparse-sense (kt-RASPS), 3D U-net, the proposed MD-CNN network, and reference data, from left to right, respectively

VIDEO S5 Representative reconstruction at 14 spokes/frame of undersampled data from another subject at all cardiac phases by different reconstruction methods: the locally low rank and temporal finite differences compressed sensing method (LLR+FD), k-t radial sparse-sense (kt-RASPS), 3D U-net, the proposed MD-CNN network, and reference data, from left to right, respectively

VIDEO S6 Representative reconstruction at 14, 16, and 20 spokes/frame of undersampled data from one subject at all cardiac phases by the proposed MD-CNN network. Temporal fidelity increases as number of spokes increase

FIGURE S1 Mean squared error (MSE) for MD-CNN images in the training and testing datasets at each of the 250 training epochs

FIGURE S2 Sample reconstructed images at end-systole by the three methods (kt-RASPS, MD-CNN, and Reference) showing the six segments at which LV sharpness was quantified and the intensity profiles corresponding to each segment of the sample images

FIGURE S3 Bland-Altman plots showing inter-observer agreement of each comparison: Reader#1 vs. Reader#2, Reader#2 vs. Reader#3, and Reader#1 vs. Reader#3 of the extracted LV myocardial areas in kt-RASPS, MD-CNN, and reference images, separately

TABLE S1 Clinical indications for clinical cardiac MR scans of our patient cohort

TABLE S2 Biases and limits of agreement of the Bland-Altman plots for inter-observer agreement of 3 comparisons: Reader#1 vs. Reader#2, Reader#2 vs. Reader#3, and Reader#1 vs. Reader#3 of the extracted LV myocardial areas in kt-RASPS, MD-CNN, and reference images