

RAPID COMMUNICATION

Rapid reconstruction of highly undersampled, non-Cartesian real-time cine *k*-space data using a perceptual complex neural network (PCNN)

Daming Shen^{1,2} | Sushobhan Ghosh³ | Hassan Haji-Valizadeh^{1,2} |
Ashitha Pathrose²  | Florian Schiffrers³ | Daniel C. Lee^{2,4} | Benjamin H. Freed⁴ |
Michael Markl^{1,2} | Oliver S. Cossairt³ | Aggelos K. Katsaggelos⁵ | Daniel Kim^{1,2} 

¹Biomedical Engineering, McCormick School of Engineering and Applied Science, Northwestern University, Evanston, Illinois

²Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, Illinois

³Department of Computer Science, McCormick School of Engineering and Applied Science, Northwestern University, Evanston, Illinois

⁴Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

⁵Department of Electrical and Computer Engineering, McCormick School of Engineering and Applied Science, Northwestern University, Evanston, Illinois

Correspondence

Daniel Kim, PhD, Department of Radiology, Northwestern University, 737 N. Michigan Avenue, Suite 1600, Chicago, IL 60611, USA. Email: daniel.kim3@northwestern.edu

Funding information

National Heart, Lung, and Blood Institute, Grant/Award Numbers: R01HL116895, R01HL138578, R01HL151079; National Institute of Biomedical Imaging and Bioengineering, Grant/Award Number: R21EB024315; National Institute on Aging, Grant/Award Number: R21AG055954; American Heart Association, Grant/Award Number: 19IPL0134760317; National Institutes of Health, Grant/Award Numbers: R01HL151079, R21AG055954, R21EB024315, R01HL138578, R01HL116895

Highly accelerated real-time cine MRI using compressed sensing (CS) is a promising approach to achieve high spatio-temporal resolution and clinically acceptable image quality in patients with arrhythmia and/or dyspnea. However, its lengthy image reconstruction time may hinder its clinical translation. The purpose of this study was to develop a neural network for reconstruction of non-Cartesian real-time cine MRI *k*-space data faster (<1 min per slice with 80 frames) than graphics processing unit (GPU)-accelerated CS reconstruction, without significant loss in image quality or accuracy in left ventricular (LV) functional parameters. We introduce a perceptual complex neural network (PCNN) that trains on complex-valued MRI signal and incorporates a perceptual loss term to suppress incoherent image details. This PCNN was trained and tested with multi-slice, multi-phase, cine images from 40 patients (20 for training, 20 for testing), where the zero-filled images were used as input and the corresponding CS reconstructed images were used as practical ground truth. The resulting images were compared using quantitative metrics (structural similarity index (SSIM) and normalized root mean square error (NRMSE)) and visual scores (conspicuity, temporal fidelity, artifacts, and noise scores), individually graded on a five-point scale (1, worst; 3, acceptable; 5, best), and LV ejection fraction (LVEF). The mean processing time per slice with 80 frames for PCNN was 23.7 ± 1.9 s for pre-processing (Step 1, same as CS) and 0.822 ± 0.004 s for dealiasing (Step 2, 166 times faster than CS). Our PCNN produced higher data fidelity metrics (SSIM = 0.88 ± 0.02 , NRMSE = 0.014 ± 0.004) compared with CS. While all the visual scores were significantly different ($P < 0.05$), the median scores were all 4.0 or higher for both CS and PCNN. LVEFs measured from CS and PCNN were strongly correlated ($R^2 = 0.92$) and in good agreement (mean difference = -1.4% [2.3% of mean]; limit of agreement = 10.6% [17.6% of mean]). The proposed PCNN is capable of rapid reconstruction (25 s per slice with 80 frames) of non-Cartesian real-time cine MRI *k*-space data, without significant loss in image quality or accuracy in LV functional parameters.

Abbreviations: AF, atrial fibrillation; CNN, convolutional neural network; CS, compressed sensing; CV, coefficient of variation; DICOM, Digital Imaging and Communications in Medicine; DL, deep learning; FOV, field of view; GPU, graphics processing unit; LOA, limit of agreement; LV, left ventricular; LVEDV, LV end-diastolic volume; LVEF, LV ejection fraction; LVESV, LV end-systolic volume; LVSV, LV stroke volume; MSE, mean squared error; NRMSE, normalized root mean square error; NUFFT, non-uniform fast Fourier transform; PCNN, perceptual complex neural network; ReLU, rectified linear unit; RING, radial intersections; SSIM, structural similarity index; SVS, summed visual score; VGG, Visual Geometry Group.

KEYWORDS

compressed sensing (CS), convolutional neural network (CNN), deep learning (DL), perceptual complex neural network (PCNN), perceptual loss, real-time cine MRI

1 | INTRODUCTION

While electrocardiogram-gated, breath-hold cine MRI with balanced steady-state free precession readout is the reference test for evaluation of cardiac function,^{1,2} its diagnostic yield may be limited in patients with arrhythmia and/or dyspnea due to severe image artifacts. One approach to overcome this limitation is to perform highly accelerated real-time cine MRI using compressed sensing (CS)³ with Cartesian^{4,5} or radial *k*-space sampling.^{6,7} The three key components of CS are sparsity, incoherent aliasing artifacts, and nonlinear optimization with L1-norm. Despite promising results using CS-accelerated real-time cine MRI, its lengthy image reconstruction time may hinder its clinical translation, including interventional or stress testing MRI where real-time support is critical. Thus, there is an unmet need to develop highly accelerated image reconstruction methods that support accelerated, real-time cine MRI acquisitions.

One solution to accelerating CS reconstruction is using a graphics processing unit (GPU); however, the acceleration is limited since CS remains iterative and nonlinear. To circumvent the problem of computation-intensive iterations, feed-forward deep learning (DL)⁸ has emerged as a promising alternative for solving inverse problems compared with iterative approaches.^{9–11} DL-based image reconstruction is roughly categorized into agnostic, decoupled physics-based, and post-processing learners. Agnostic solvers learn a direct mapping from the input to the output domain without any knowledge of the forward model at any point in training or testing.¹² Agnostic solvers require a huge number of training data and are hard to optimize. Decoupled approaches first learn a comprehensive representation of the image space independent of the imaging problem at hand, eg from a large set of reconstructed MR images. This knowledge is then used to guide the image reconstruction.¹³ Physics-based learners incorporate a differentiable version of the imaging operator (eg the Fourier transform in MRI) into the training process and reduce the number of required training data drastically.^{14,15}

Finally, there remain learners that focus on post-processing to remove possible artifacts that arise in non-iterative algorithms.^{16,17} The key advantage of this approach is that it is simpler to implement. The basic strategy is to train a network to learn the weights (convolutional kernels) for dealiasing undersampled MR images from a large dataset containing pairs of aliased and dealiased images. In the testing phase, the network applies the “learned” model to dealias images from a separate testing dataset. While the training phase is computationally intensive due to backpropagation of gradients and often requires GPU computing, the testing phase is significantly faster than CS, and could even be transferred with CPU computing, thereby making it a good vehicle to reduce the processing time of reconstructing accelerated real-time cine MRI.

To date, two proof-of-concept studies have used DL to reconstruct real-time cine MR data, with each study having advantages and disadvantages.^{18,19} In this study, we sought to develop a novel DL approach that goes beyond the two prior studies.^{18,19} Our main contributions are the following:

- implementing a complex neural network that is capable of learning correlated and uncorrelated (ie noise) information contained in real and imaginary components of complex MRI signal detected in quadrature (ie 90° phase offset between real and imaginary components, as shown in Figure 1);
- incorporating a perceptual loss term to maintain high-level features better than per pixel loss, as previously described²⁰;
- training and testing the proposed network with multi-slice data from a larger group (40 in total; 20 for training, 20 for testing) of patients with atrial fibrillation (AF), which has not been addressed by previous DL-based image reconstruction studies;
- handling highly accelerated (15-fold) cine data. Our approach to handling complex data in the image domain is different from that of prior studies that handled complex data, either in the image domain as magnitude and phase²¹ or in the *k*-space domain.^{12,22} The key advantage of the

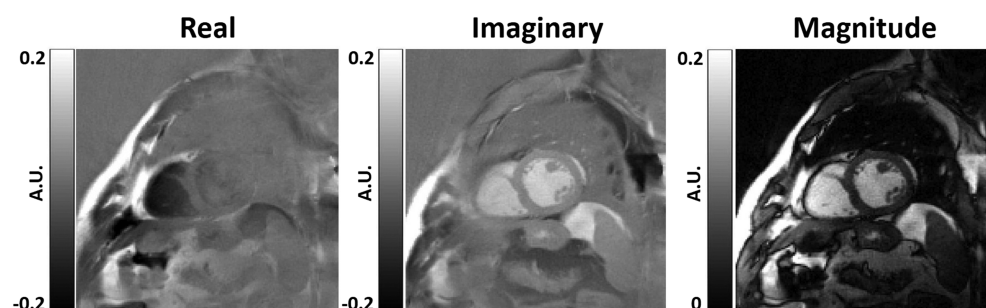


FIGURE 1 Real, imaginary, and magnitude parts of a real-time cine complex MR image, illustrating correlated and uncorrelated (noise) information detected using a quadrature RF receiver system

proposed approach over previous approaches is that it does not require extensive GPU memory (ie fast processing), because it handles coil-combined complex data without requiring fully connected layers or fidelity layers.

The purpose of this study was to implement a perceptual complex neural network (PCNN) for reconstruction of non-Cartesian real-time cine MRI k -space data faster (<1 min per slice with 80 frames) than GPU-accelerated CS reconstruction, without significant loss in image quality or accuracy in left ventricular (LV) functional parameters. We compare the proposed PCNN with previously proposed CNN network architecture¹⁹ for completeness.

2 | MATERIALS AND METHODS

2.1 | Patient demographics

This study was conducted in accordance with protocols approved by our institutional review board and was Health Insurance Portability and Accountability Act (HIPAA) compliant. All subjects provided informed consent in writing. We prospectively enrolled 40 patients with prior history of AF (mean age = 68.1 ± 9.6 years; 31 males; 9 females). In 8 out of 20 patients (mean age = 68.6 ± 10.6 years; 16 males; 4 females) used for training, MRI was repeated within two weeks to evaluate test-retest reproducibility for a separate study, such that 28 sets of multi-slice, multi-phase cine k -space datasets were used for training. Multi-slice, multi-phase datasets from the remaining 20 patients (mean age = 67.6 ± 8.7 years; 15 males; 5 females) were used to test the trained neural networks. We elected to reserve data from 20 patients for testing in order to achieve high power for our statistical analysis. For basic demographics information of our patients, including age, sex, AF type, and resting heart rate, see Table 1. Other clinical characteristics were considered irrelevant for this study and thus omitted due to space constraints. To estimate the arrhythmia burden during MRI, we calculated the coefficient of variation (CV) of heartbeat duration, which was extracted from the raw data header of real-time cine running continuously for multiple heartbeats per slice, for multiple slices (total scan time was ~ 60 s).

2.2 | MRI Hardware

Real-time cine scans were conducted on a 1.5 T whole-body MRI scanner (MAGNETOM Aera, Siemens Healthcare, Erlangen, Germany). The scanner was equipped with a gradient system capable of achieving a maximum gradient strength of 45 mT/m and maximum slew rate of 200 T/m/s. A body coil was used for RF excitation. Both body matrix and spine coil arrays (30-34 elements in total) were used for signal reception.

2.3 | Pulse sequence

Relevant imaging parameters of real-time cine MRI using radial k -space sampling included field of view (FOV) = $288 \text{ mm} \times 288 \text{ mm}$, matrix size = 160×160 , spatial resolution = $1.8 \text{ mm} \times 1.8 \text{ mm}$, slice thickness = 8 mm, $T_E = 1.4 \text{ ms}$, $T_R = 2.7 \text{ ms}$, receiver bandwidth = 975 Hz/pixel, 11 radial spokes per cardiac frame, tiny golden angle sequence = 23.62814° ,²³ effective acceleration factor = 15 (with respect to Cartesian equivalent), temporal resolution = 29.7 ms, 12-17 short-axis planes, and flip angle 50° . Although each 2D plane was scanned for 5 s during free-breathing, only the first 80 out of 166 cardiac frames were used from each patient due to GPU memory limitation.

2.4 | Computer hardware

For training and testing on undersampled raw k -space data, we used a GPU workstation (Tesla V100 32GB memory, NVIDIA, Santa Carla, CA, USA; 32 Xeon E5-2620 v4 128 GB memory, Intel, Santa Clara, CA, USA) equipped with Python (Version 3.7, Python Software Foundation),

TABLE 1 Summary of baseline patient characteristics ($N = 40$). M, males; F, females

	Total	Training	Testing
Age (years)	68.1 ± 9.6	68.6 ± 10.6	67.6 ± 8.7
Sex	31 M/9 F	16 M/4 F	15 M/5 F
Resting heart rate (bpm)	66.7 ± 12.0	68.2 ± 12.9	65.3 ± 11.1
AF type	32 paroxysmal/8 persistent	15 paroxysmal/5 persistent	17 paroxysmal/3 persistent
Arrhythmia burden (%)	24.6 ± 9.3	26.5 ± 9.8	22.8 ± 8.6

PyTorch (Version 1.4, Berkeley Software Distribution), and MATLAB (R2017b, MathWorks, Natick, MA, USA) running on a Linux operating system (Ubuntu 16.04).

2.5 | GPU-accelerated CS reconstruction as obtainable ground truth in patients with arrhythmia

In patients with AF, standard electrocardiogram-gated breath-hold cine MRI produces poor image quality with considerable ghosting and blurring artifacts. Thus, it was not feasible to obtain a fully sampled reference for this study. Instead, we used the CS reconstruction as obtainable ground truth.

For reference, the same undersampled k -space data were reconstructed using the same GPU workstation. We adapted our previously described radial CS reconstruction code implemented in MATLAB⁷ with two modifications: (a) GPU based non-uniform fast Fourier transform (NUFFT)²⁴ and (b) coil compression using principal component analysis (PCA)²⁵ to produce eight virtual coils. In the preprocessing step (gradient delay correction + gridding + coil combination), we performed self-calibrated gradient delay correction using the radial intersections (RING) method,²⁶ GPU-based NUFFT to convert the radial k -space data to zero-filled images in Cartesian space, and additional processing on time average image to derive auto-calibrated coil sensitivity profiles using the method described by Walsh et al.,²⁷ followed by a weighted sum over the coil elements. Coil-combined, zero-filled cine images (initial solution), multi-coil raw k -space data, k -space sampling masks, and coil sensitivity maps were used as inputs to the previously described iterative CS algorithm,⁷ which enforced sparsity along the time dimension using temporal finite difference (temporal total variation) as the sparsifying transform and nonlinear conjugate gradient with back-tracking line search as the optimization algorithm with 30 iterations. The cost function used is described in

$$\hat{x} = \operatorname{argmin}_x \|FSx - y\|_2^2 + \lambda \|Tx\|_1 \quad (1)$$

where F is the undersampled FFT operator, S is the estimated coil sensitivities in x - y space, x is the image series to be reconstructed in x - y - t space, y is the acquired multi-coil k -space data, T is the temporal finite difference operator, and λ is the normalized regularization weight that controls the tradeoff between data consistency and sparsity terms. We incorporated back-tracking line search to ensure high data fidelity, at the expense of computational efficiency. Normalized regularization weight was set as 0.1 of the maximum signal of time average image. We established 0.1 (relative to the maximum value) as the optimal regularization weight by sweeping over a range from 0.001 to 0.1 (steps of 0.05) and identifying an optimal regularization weight that achieves a good balance between suppression of aliasing artifacts and temporal blurring of myocardial wall motion. We determined this optimal regularization weight based on visual inspection of six training datasets.

2.6 | Network architecture

We implemented a reconstruction pipeline that performs pre-processing in MATLAB and dealias coil-combined images in PyTorch. We elected to work with coil-combined images due to GPU memory limitation. After the same pre-processing step as described for CS, coil-combined, zero-filled, complex images were used as input. Our network was trained on 398 2D + time sets of zero-filled, real-time cine images obtained from 20 patients (eight in whom we obtained another set of cine data), corresponding to 31 840 2D images in total. The trained network was tested on 275 2D + time sets of zero-filled, real-time cine images obtained from the remaining 20 patients, corresponding to 22 000 2D images in total.

By modifying the network and loss function, we explored three different ways of processing the complex data to achieve optimal image quality, as shown in Figure 2. Three different residual 3D (2D + time) U-Nets^{28–31} with identical architecture but different loss functions were tested: (1) a magnitude network with mean squared error (MSE) loss alone, which uses traditional operations (convolution, rectified linear unit (ReLU), etc) to process the absolute value of the complex data; (2) a complex network with MSE loss alone, which uses complex operations to process the complex data (see below and Figure 3 for more details); (3) PCNN using both complex operations and MSE and perceptual loss terms. For complex networks (2 and 3), the batch normalization and ReLU layers had separate weights for the real and imaginary feature maps, while the pooling layers were the same. For PCNN, instead of using MSE loss alone,^{32–34} we added a perceptual loss^{20,31} using the first 15 layers of a pre-trained Visual Geometry Group (VGG)-16 network³⁵ to maintain high-level features better than pixel-wise MSE loss. Only the first 15 layers of the VGG-16 network were used to extract features, since the last layer is used for classifying features. The total loss function can be described by

$$\text{Loss}_{\text{total}} = \frac{1}{N} \|\varphi(x) - y\|_2^2 + \frac{1}{N} \|f_{\text{vgg}}(x) - f_{\text{vgg}}(y)\|_2^2 \quad (2)$$

FIGURE 2 A, The U-Net architecture used for all three networks; B, the pipeline for PCNN training. The complex U-Net and PCNN used the same complex convolution operations as shown in Figure 3. While PCNN uses both perceptual loss and pixel-wise MSE loss functions, conventional magnitude and complex U-Net used only the pixel-wise MSE loss function. For visual display of the outcome of the VGG network for CS and DL, see Figure S1 in Supporting Information

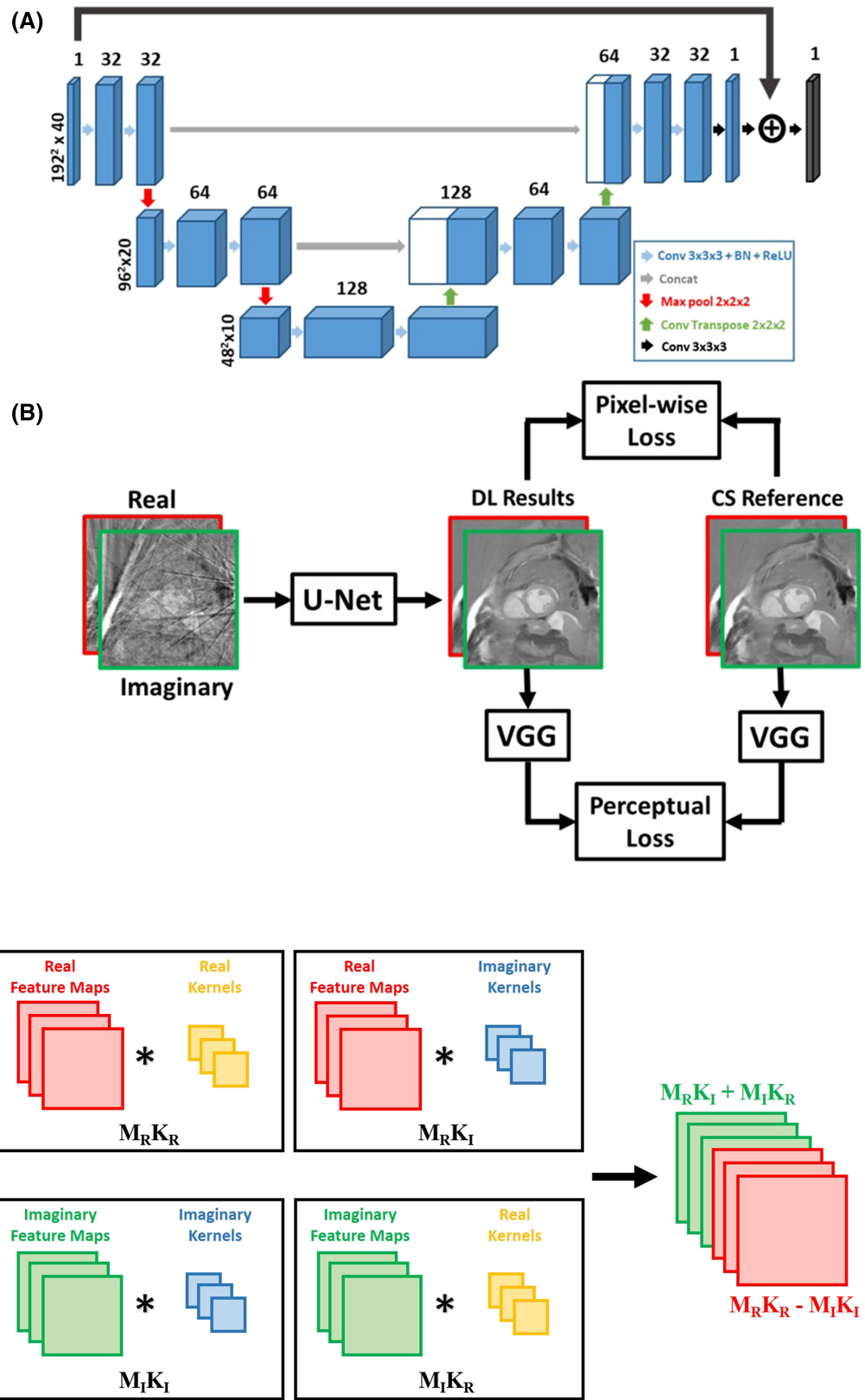


FIGURE 3 Complex convolution operation used in complex U-Net and PCNN. The real (M_R) and imaginary (M_I) feature maps are separated by creating an extra dimension and convolved with real (K_R) and imaginary (K_I) kernels as shown. The results are sorted and separated in the next layer with $(M_R K_R - M_I K_I)$ as the real part and $(M_R K_I + M_I K_R)$ as the imaginary part

where φ is the U-Net, f_{vgg} is the VGG network, N is the total number of voxels, x is the zero-filled images (either real or imaginary), and y is the reference images (either real or imaginary) reconstructed with CS. For a visual display of the outcome of the VGG network for CS and DL, see Figure S1 in Supporting Information. The following training parameters were used: batch size = 1, ADAM optimizer, 50 epochs, and learning rate = 0.0001 with a decay rate of 0.95 for each epoch. Training for the magnitude network took approximately 8 h, whereas training for the complex network and PCNN took approximately 20 h and 24 h, respectively.

As shown in Figure 3, we performed complex convolution³⁶ on the complex data. To support this complex operation, we created one additional dimension for the feature maps to carry both the real (M_R) and imaginary (M_I) parts and used two separate kernels (K_R and K_I) to perform the complex convolution as described by

$$(M_R + iM_I) * (K_R + iK_I) = (M_R * K_R - M_I * K_I) + i(M_R * K_I + M_I * K_R). \quad (3)$$

2.7 | Quantitative metrics of image quality

Given that images reconstructed with different methods are perfectly registered, we calculated the structural similarity index (SSIM)³⁷ and normalized root mean square error (NRMSE) to infer image quality with respect to reference images reconstructed with CS. For both SSIM and NRMSE calculations, we focused on a smaller region of interest (central FOV with 80×80 voxels) that encapsulates the heart region. To evaluate image blurring, we calculated the blurring metric³⁸ on a 0 to 1 continuous scale, where 0 is defined as sharp and 1 is defined as blurred.

2.8 | Visual metrics of image quality

To evaluate the diagnostic confidence produced by the proposed PCNN, two non-invasive cardiologists (DCL and BHF with 17 and 8 years of experience, respectively) graded the CS reconstructed images (reference) and best DL reconstructed images, where best among magnitude, complex, and PCNN was determined by quantitative metrics (SSIM, NRMSE, blur metric). For efficient analysis, evaluation was limited to three short-axis planes (base, mid, apex) only. In total, 40 cine datasets (20 sets each for DL and CS), grouped as a set of three short-axis planes, were randomized and de-identified for dynamic display. Prior to visual evaluation, the two readers were given training datasets to calibrating their scores together, where a score of three is defined as clinically acceptable. Following training, the readers were blinded to image acquisition type (CS and DL), each other, and clinical history. Each set of three short-axis planes was graded on a five-point Likert scale: conspicuity of endocardial border at end diastole (1, non-diagnostic; 2, poor; 3, adequate; 4, good; 5, excellent), temporal fidelity (blurring or ghosting or lack thereof) of wall motion (1, non-diagnostic; 2, poor; 3, adequate; 4, good; 5, excellent), any visible artifact on the heart (1, non-diagnostic; 2, severe; 3, moderate; 4, mild; 5, minimal), and apparent noise throughout (1, non-diagnostic; 2, severe; 3, moderate; 4, mild; 5, minimal). The summed visual score (SVS) was calculated as the sum of conspicuity, temporal fidelity, artifact, and noise scores, with 12 defined as clinically acceptable.

2.9 | LV function assessment

In total, 40 cine datasets (20 patients $\times$ 2 (CS and best DL) sets per patient) were analyzed by another reader (AP) with 2 years of experience as a medical research fellow, using standard methods on a workstation equipped with commercial software (CVi42, Cardiovascular Imaging, Calgary, Canada). Functional parameters included LV ejection fraction (LVEF), LV end-systolic volume (LVESV), LV end-diastolic volume (LVEDV), and LV stroke volume (LVSV). For consistency, the most basal slice was defined as the plane that has 50% or more of the blood pool surrounded by myocardium, and the most apical slice was defined as the plane showing blood pool at end diastole. The reader repeated the analysis with a 2 weeks gap between analyses to determine whether inter-reconstruction variability is similar to intra-observer variability.

2.10 | Statistical analysis

The statistical analyses were conducted by one investigator (DS) using MATLAB. Using average reader scores, we used the Wilcoxon signed-rank test to detect differences between two groups. For continuous variables (SSIM, NRMSE, blur metric), we used analysis of variance to detect differences between multiple groups, with Bonferroni correction to compare each DL reconstruction to CS as reference. For cardiac functional parameters, we performed Pearson's correlation and Bland-Altman analysis to examine association and agreement. Reported continuous variables represent mean $\pm$ standard deviation. $P < 0.05$ was considered significant for all statistical tests.

3 | RESULTS

The mean CV of the R-R interval for the entire cohort was $24.6 \pm 9.3\%$, while the corresponding R-R intervals for the training and testing cohorts were $26.5 \pm 9.8\%$ and $22.8 \pm 8.6\%$, respectively, indicating moderate levels of arrhythmia. The mean processing time per slice with 80 frames along the proposed pipeline for PCNN was 23.7 ± 1.9 s for pre-processing (Step 1) and 0.822 ± 0.004 s for dealiasing (Step 2). The corresponding processing time along the GPU-accelerated CS pipeline was 23.7 ± 1.9 s for pre-processing (Step 1) and 136.4 ± 2.4 s for dealiasing (Step 2). The reconstruction time including the identical pre-processing step for DL was 6.5 times faster than that for CS, whereas the dealiasing processing time (excluding the pre-processing step) for DL was 166 times faster than that for CS.

Figure 4 shows representative real-time cine reconstructed MR images obtained with the following methods: (1) CS as reference; (2) zero-filled image immediately after NUFFT; (3) magnitude network with MSE alone; (4) complex network with MSE loss alone; (5) PCNN. The corresponding difference images with respect to CS are also shown, where the PCNN showed the smallest number of residual artifacts. For dynamic display of Figure 4, see Video S1 in Supporting Information.

Summarizing the result over 20 patients (see Table 2), compared with CS, the PCNN produced the best image quality metrics (SSIM = 0.88 ± 0.02 , NRMSE = 0.014 ± 0.004), significantly ($P < 0.05$) better than the magnitude and complex networks (SSIM < 0.75, NRMSE > 0.020). Relative to CS, the blur metrics were not significantly ($P > 0.05$) different for the magnitude network and PCNN, whereas they were significantly ($P < 0.05$) lower for the zero-filled and complex networks. Given that PCNN produces the best results in two out of three categories, we elected to use PCNN throughout.

Figure 5 shows representative images of three patients reconstructed with PCNN and CS, highlighting similar image qualities. For dynamic display of images shown in Figure 5, see Videos S2-4 in Supporting Information. Table 3 summarizes average reader scores for CS and PCNN. While all the scores were significantly different ($P < 0.05$) between the two groups, all individual visual scores and SVS were well above the clinically acceptable cut points of 3.0 and 12.0, respectively.

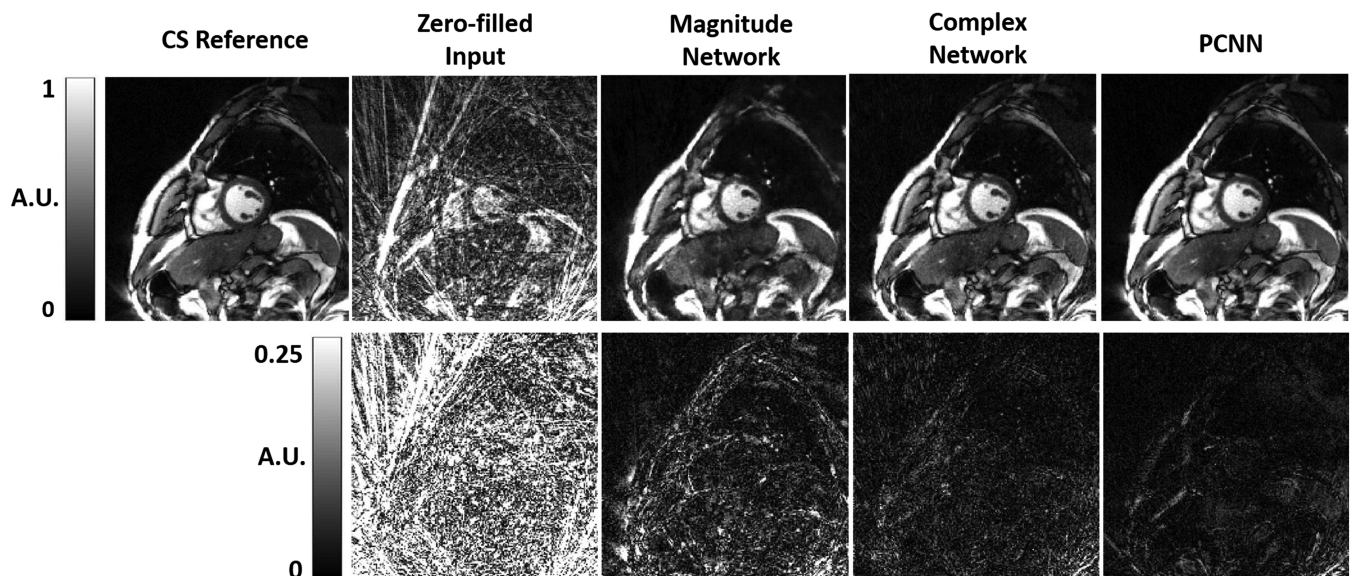


FIGURE 4 Top row: representative images of CS reference (first column), zero-filled image immediately after NUFFT (second column) and reconstruction results by three different networks—magnitude network (third column), complex network with MSE loss term only (fourth column), and PCNN (fifth column), displayed in 0–1.0 arbitrary units (A.U.). Bottom row: The corresponding difference images with respect to CS reference, displayed in 0–0.25 A.U. to bring out differences. For dynamic display of top row, see Video S1 in Supporting Information

TABLE 2 Summary of quantitative metrics ($N = 20$). NRMSE and SSIM for zero-filled input images and reconstruction results by three different networks compared with CS reference. For NRMSE and SSIM, $^*P > 0.05$ corresponds to a non-significant difference in the pair. For the blur metric, $^{+ \#}P > 0.05$ corresponds to a non-significant difference in the pair. Note that the blur metric scores for the reconstructions with the zero-filled and complex network with MSE loss term only are artificially better, because they contained substantial amounts of streaking artifacts, which have sharp edges

	CS reference	Zero-filled	Magnitude	Complex	PCNN
NRMSE		0.082 ± 0.011	$0.025 \pm 0.005^*$	$0.020 \pm 0.006^*$	0.014 ± 0.004
SSIM		0.232 ± 0.025	0.663 ± 0.056	0.742 ± 0.069	0.884 ± 0.023
Blur metric	$0.338 \pm 0.015^{+ \#}$	0.188 ± 0.008	$0.340 \pm 0.017^{+ \#}$	0.314 ± 0.019	$0.337 \pm 0.013^{\leq \#}$

Figure 6 shows scatter plots resulting from linear regression analysis illustrating strong correlation between CS and DL analyses ($R^2 \geq 0.92$) and between repeated DL analyses ($R^2 \geq 0.93$) for all four LV functional categories. Figure 7 shows Bland-Altman plots, illustrating good agreement between CS and DL analyses for LVEDV (mean = 98.0 mL; mean difference = -0.54 mL [0.5% relative to mean], limit of agreement (LOA) = 14.5 mL [14.8% relative to mean]), LVESV (mean = 38.6 mL; mean difference = 1.0 mL [2.6% relative to mean]; LOA = 11.3 mL [29.3%

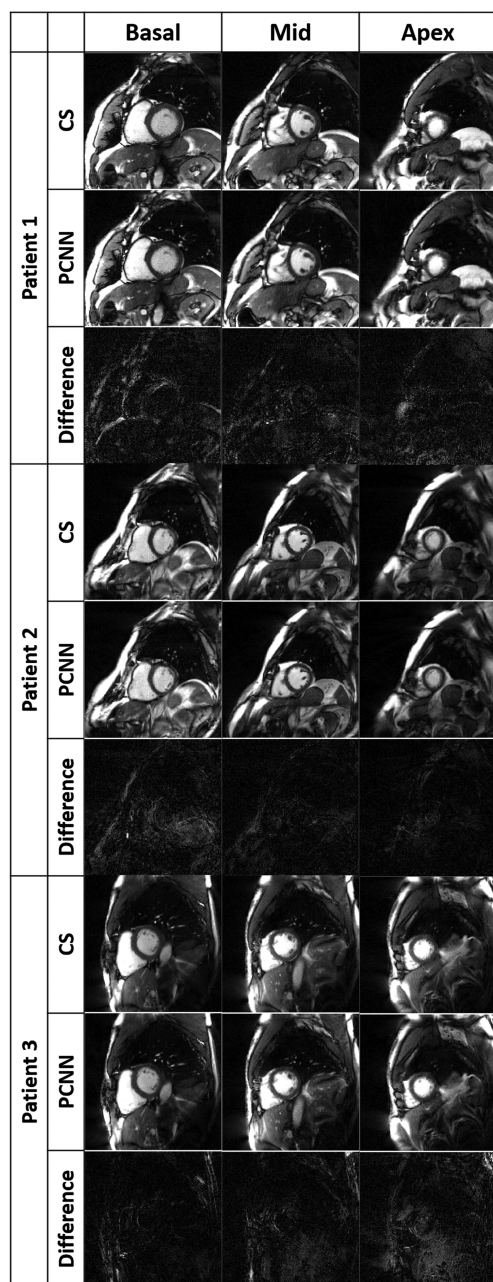


FIGURE 5 Representative images of three different patients reconstructed with CS and PCNN and a difference image displayed with fourfold narrow grayscale to bring out differences: basal plane, mid-ventricular plane, and apical plane. For dynamic display, see Videos S2-4 in Supporting Information

TABLE 3 Summary of average reader visual scores. Reported values represent median and 25th to 75th percentiles (parentheses). * $P < 0.05$ corresponds to significant difference

	CS	PCNN
Myocardial edge definition	5.0* (4.5-5.0)	4.5* (4.0-4.5)
Temporal fidelity	4.75* (4.5-5.0)	4.0* (3.5-4.5)
Artifact level	4.5* (4.0-5.0)	4.25* (4.0-4.5)
Noise level	4.5* (4.5-5.0)	4.5* (4.0-4.5)
SVS	18.75* (17.5-19.5)	17.0* (16.0-18.0)

relative to mean)), LVSV (mean = 59.4 mL; mean difference = -1.6 mL [2.6% relative to mean]; LOA = 14.9 mL [25.0% relative to mean]), and LVEF (mean = 61.6%; mean difference = -1.4 % [2.3% relative to mean]; LOA = 10.9% [17.6% relative to mean]). Figure 7 also shows good agreement between repeated DL analyses for LVEDV (mean difference = -0.9 mL [0.9% relative to mean]; LOA = 8.7 mL [8.9% relative to mean]), LVESV (mean difference = 0.9 mL [2.4% relative to mean]; LOA = 10.3 mL [26.0% relative to mean]), LVSV (mean difference = -1.8 mL [3.2% relative to mean]; LOA = 9.3 mL [16.0% relative to mean]), and LVEF (mean difference = -1.4 % [2.3% relative to mean]; LOA = 10.0% [16.7% relative to mean]).

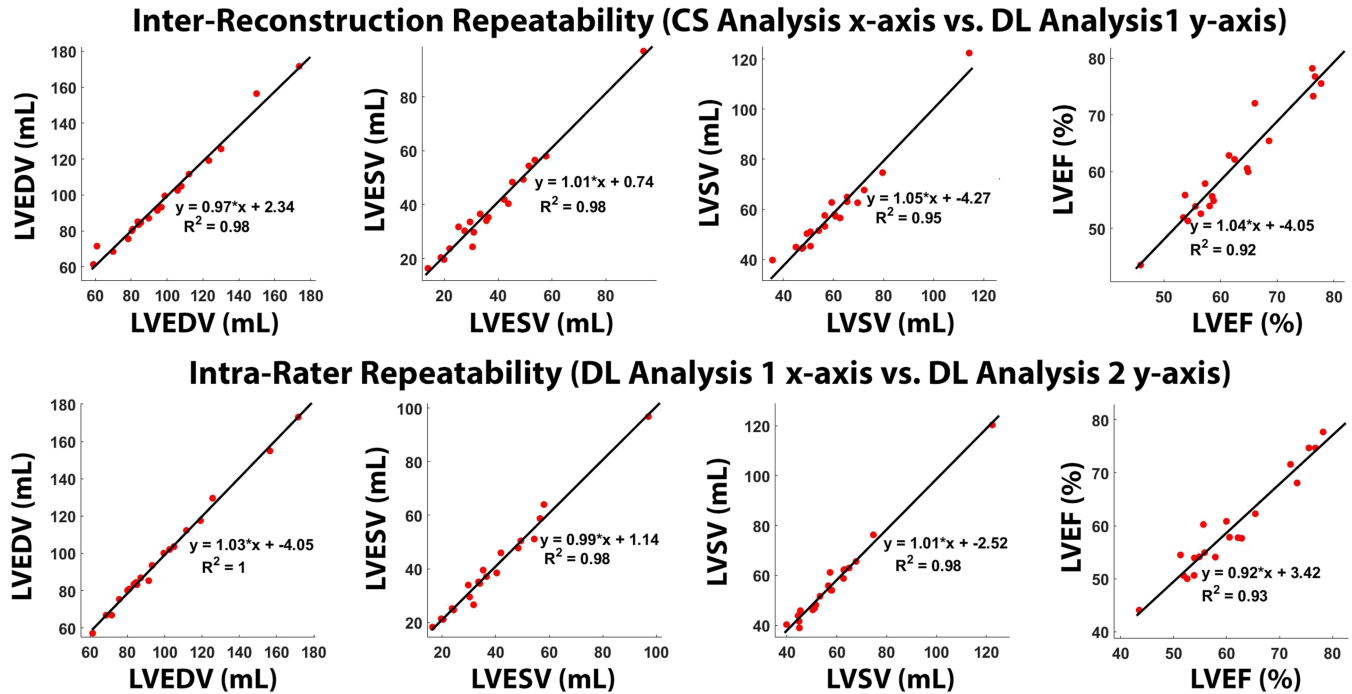


FIGURE 6 Linear regression plots illustrating strong correlation between reconstruction methods ($R^2 \geq 0.92$) and between repeated DL analyses ($R^2 \geq 0.93$) for all four LV functional parameters

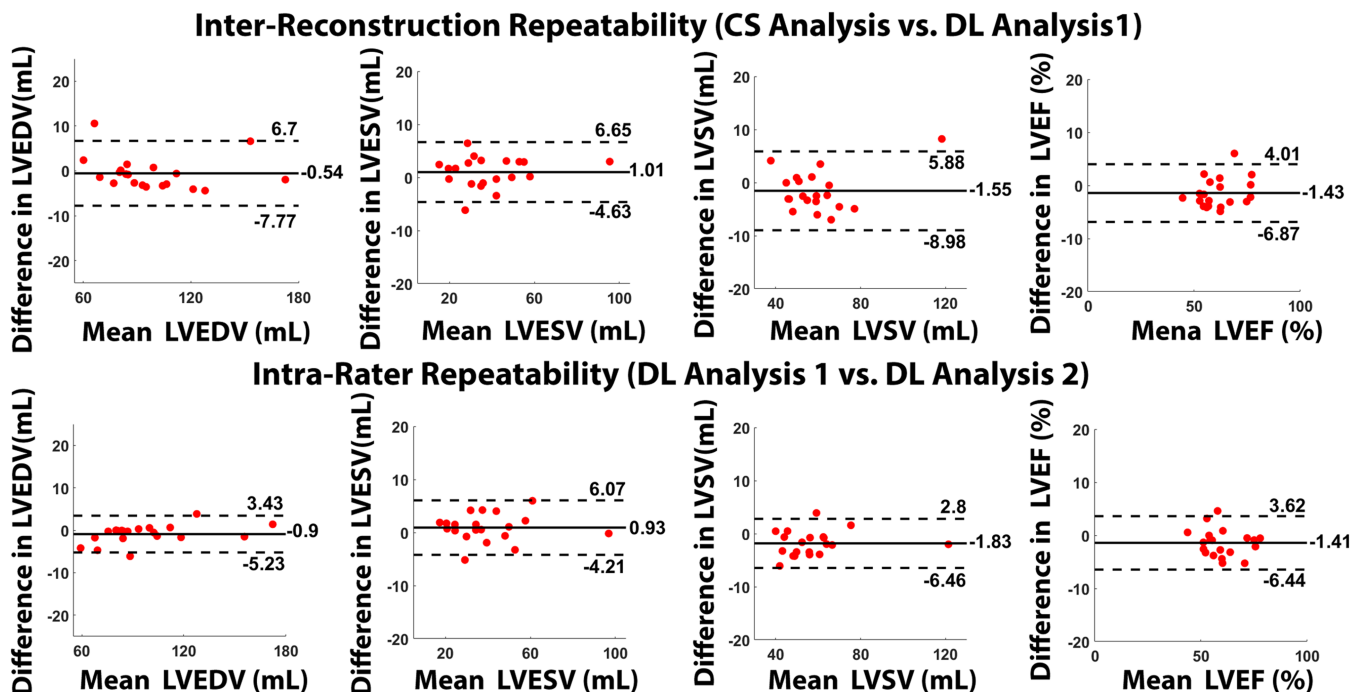


FIGURE 7 Bland-Altman plots illustrating good agreement between reconstruction methods and between repeated DL analyses for all four LV functional parameters

4 | DISCUSSION

This study describes the implementation of a rapid DL reconstruction pipeline for reconstruction of non-Cartesian real-time cine complex data faster (25 s per slice with 80 frames) than GPU-accelerated CS (2 min 40 s per slice with 80 frames), without significant loss in image quality metrics (SSIM = 0.88 ± 0.02 , NRMSE = 0.014 ± 0.004), SVS, or LV functional parameters. By optimally learning different information contained in the real and imaginary parts of complex data and adding a perceptual loss term to suppress incoherent image features, the proposed PCNN outperformed the other two architectures (magnitude with MSE loss term alone, complex network with MSE loss term alone) and successfully produced clinically acceptable image quality. Our engineering approach is based on MR physics, where the real and imaginary components contain correlated and uncorrelated (noise) information detected using a quadrature RF receiver system. Therefore, learning image features in both real and imaginary components enables more faithful image reconstruction than magnitude learning only. While the MSE loss is commonly used in DL image reconstruction, it may produce perceptually incoherent image details. By taking the perceptual loss into account, our PCNN produced better image quality compared with a complex network with an MSE loss term alone.

This study compares and contrasts with prior proof-of-concept DL studies for reconstructing real-time cine MR data^{18,19} as follows. The study by Schlemper et al¹⁸ used a cascade of convolutional neural networks (CNNs) to train on retrospectively undersampled Cartesian k -space cine data derived from fully sampled raw k -space acquired during breath-holding. The strengths of this study are that it incorporated a k -space data fidelity term and maintained multi-coil information to ensure faithful reconstruction. However, it had the following limitations: (a) data from only 10 patients in total (five for training and five for testing); (b) it did not evaluate performance on non-Cartesian k -space data; (c) the network did not learn respiratory motion because training data were acquired during breath-holding; (d) good results up to ninefold acceleration only. The study by Hauptmann et al¹⁹ used a residual U-Net to train on synthetically undersampled non-Cartesian k -space data derived from magnitude (ie Digital Imaging and Communications in Medicine (DICOM)) images acquired during breath-holding. The strength of this study is that testing was evaluated on zero-filled images derived from prospectively acquired 13-fold accelerated radial k -space data. However, it had the following limitations: (a) deriving synthetic radial undersampled k -space data from DICOM (magnitude) files is analytically incorrect, since the signal phase information is lost following the magnitude operation; (b) the network did not learn respiratory motion because training data were obtained during breath-holding; (c) data tested from only 10 patients; (d) dealiasing performed on coil-combined, zero-filled magnitude images.

This study has several interesting points worth emphasizing. First, the proposed PCNN avoids complex value calculation that is not officially supported by PyTorch and minimizes loss of information when performing a magnitude operation to complex images. Our method provides an engineering solution to the current challenge of complex value optimization with CNNs. Second, both complex networks outperformed the magnitude network. This may be because of the fact that real and imaging components contain both correlated and uncorrelated (noise) image contents (Figure 1). Third, we used a GPU-based NUFFT in the pre-processing step to accelerate the gridding process. Despite best efforts, the pre-processing steps including gradient delay correction, gridding, and coil combination (23.7 s) were 29 times longer than the CNN filtering time (0.82 s). A future study is warranted to implement a more efficient NUFFT in PyTorch (<https://github.com/mmuckley/torchkbnufft>) to further reduce the pre-processing time. Fourth, the proposed PCNN pipeline produced clinically acceptable image quality, despite not having a k -space data fidelity term, by optimally learning both imaginary and real components and incorporating both MSE and perceptual loss terms. This is an efficient strategy for faithfully reconstructing non-Cartesian data, because performing NUFFT would undoubtedly slow down the processing. Fifth, while the blur metric appears to be better for the reconstructions with the zero-filled and complex network with MSE loss term only, those scores were artificially boosted by a substantial number of streaking artifacts. Thus, the blur metric values for these two reconstructions need to be interpreted with caution. Sixth, we used the industry standard L2 loss to train our network. Several studies have shown that L1 loss may produce better results than L2 loss.^{39–41} A future study is warranted to compare the performance between L1 and L2 loss functions for training our data with PCNN. Seventh, PCNN was trained on CS as reference. As such, it was not designed to outperform CS in terms of image quality, but to outperform CS in terms of computational speed.

This study has several limitations that warrant further discussion. First, we used CS reconstructed real-time cine images as practical ground truth, because it was not possible to obtain fully sampled data in patients with AF. On one hand, we do not have access to ground truth, so the best we can do is treat CS reconstructions as ground truth. On the other hand we have demonstrated a neural network implementation that we can confidently say has successfully learned the CS algorithm, as verified by the results and analysis presented in this paper. Second, we did not incorporate a k -space data consistency term into our model because NUFFT and inverse NUFFT are time consuming operations for non-Cartesian data. Another practical reason for not including a data consistency layer is GPU memory requirement, since such an operation would also necessitate multi-coil information. A future study is warranted to incorporate a data consistency layer for non-Cartesian data using a GPU server with very high memory capacity. Third, our training (multi-slice 2D + time) data were obtained from 20 patients. While the total number of patients may appear to be small, we used 31 840 2D images (or 1 173 749 760 voxels) and 2 547 072 parameters for PCNN for paired supervised learning with $3 \times 3 \times 3$ kernels. Note that our training data size (20 patients) is four times larger than the training data size (five patients) used by Schlemper et al.¹⁸ Fourth, while PCNN produced clinically acceptable visual scores for all four individual categories, its lowest score was temporal fidelity of myocardial wall motion. Subtle blurring of myocardial wall motion was visible in some slices, which may have contributed to small (1.4%) underestimation in LVEF. Fifth, we designed PCNN based on a U-Net. It may be possible to achieve better results using more sophisticated

unrolled network architectures^{41,42} with more powerful GPU and more training data, at the expense of greater computational demand and processing time. From a practical point of view, fast processing is essential for our clinical application, and access to a high-end GPU server with very high memory may be limited at most centers. Nonetheless, a future study is warranted to compare the performance between the proposed PCNN and more sophisticated networks.

In summary, this study describes implementation, training, and testing of an image reconstruction pipeline including a PCNN architecture for reconstruction of non-Cartesian real-time cine complex MR data faster than GPU-accelerated CS reconstruction, without significant loss in quantitative metrics of image quality, SVS, or LV functional parameters, thereby verifying clinical translatability.

ACKNOWLEDGEMENTS

The authors thank the National Institutes of Health (R01HL116895, R01HL138578, R21EB024315, R21AG055954, R01HL151079) and American Heart Association (19IPLOI34760317) for funding support.

CONFLICTS OF INTEREST

None of the authors have relationships with industry related to this study.

FUNDING INFORMATION

National Institutes of Health (R01HL116895, R01HL138578, R21EB024315, R21AG055954, R01HL151079) and American Heart Association (19IPLOI34760317).

ORCID

Ashitha Pathrose  <https://orcid.org/0000-0003-1761-0780>

Daniel Kim  <https://orcid.org/0000-0003-2660-8973>

REFERENCES

1. American College of Cardiology Foundation Task Force on Expert Consensus Documents. 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. *J Am Coll Cardiol*. 2010;55(23):2614-2662. <https://doi.org/10.1016/j.jacc.2009.11.011>
2. Grothues F, Smith GC, Moon JC, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. *Am J Cardiol*. 2002;90(1):29-34. [https://doi.org/10.1016/s0002-9149\(02\)02381-0](https://doi.org/10.1016/s0002-9149(02)02381-0)
3. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58(6):1182-1195. <https://doi.org/10.1002/mrm.21391>
4. 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(1):64-74. <https://doi.org/10.1002/mrm.24440>
5. 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. <https://doi.org/10.1186/1532-429X-15-79>
6. 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(3):707-717. <https://doi.org/10.1002/mrm.24980> Epub 2013/10/22.
7. 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*. 2018;79(5):2745-2751. <https://doi.org/10.1002/mrm.26918>
8. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. <https://doi.org/10.1038/nature14539>
9. Adler J, Oktem O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*. 2017;33(12):124007. <https://doi.org/10.1088/1361-6420/aa9581>
10. Lucas A, Iliadis M, Molina R, Katsaggelos AK. Using deep neural networks for inverse problems in imaging beyond analytical methods. *IEEE Signal Processing Mag*. 2018;35(1):20-36. <https://doi.org/10.1109/Msp.2017.2760358>
11. Ongie G, Jalal A, Metzler CA, Baraniuk RG, Dimakis AG, Willett R. Deep learning techniques for inverse problems in imaging. *IEEE J Sel Areas Inform Theory*. 2020;1(1):39-56. <https://doi.org/10.1109/jsait.2020.2991563>
12. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018;555(7697):487-492. <https://doi.org/10.1038/nature25988>
13. Bora A, Jalal A, Price E, Dimakis AG. Compressed sensing using generative models. In: Doina P, Yee Whye T, eds. *Proceedings of the 34th International Conference on Machine Learning*. Vol 70. Sydney, Australia: JMLR.org; 2017:537-546.
14. Maier AK, Syben C, Stimpel B, et al. Learning with known operators reduces maximum training error bounds. *Nat Mach Intell*. 2019;1(8):373-380. <https://doi.org/10.1038/s42256-019-0077-5>
15. Gilton D, Ongie G, Willett R. Neumann networks for linear inverse problems in imaging. *IEEE Trans Comput Imaging*. 2020;6:328-343. <https://doi.org/10.1109/tci.2019.2948732>
16. Haji-Valizadeh H, Shen D, Avery RJ, et al. Rapid reconstruction of four-dimensional MR angiography of the thoracic aorta using a convolutional neural network. *Radiol Cardiothorac Imaging*. 2020;2(3):e190205. <https://doi.org/10.1148/ryct.2020190205>

17. Fan L, Shen D, Haji-Valizadeh H, et al. Rapid dealiasing of undersampled, non-Cartesian cardiac perfusion images using U-Net. *NMR Biomed*. 2020;33(5):e4239. <https://doi.org/10.1002/nbm.4239>
18. 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(2):491-503. <https://doi.org/10.1109/TMI.2017.2760978>
19. 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(2):1143-1156. <https://doi.org/10.1002/mrm.27480>
20. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: Leibe B, Matas J, Sebe N, Welling M, eds. *European Conference on Computer Vision—ECCV 2016. Lecture Notes in Computer Science 9906*. Cham, Switzerland: Springer; 2016:694-711. https://doi.org/10.1007/978-3-319-46475-6_43
21. Lee D, Yoo J, Tak S, Ye JC. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Trans Biomed Eng*. 2018;65(9):1985-1995. <https://doi.org/10.1109/TBME.2018.2821699>
22. Han Y, Sunwoo L, Ye JC. *k*-space deep learning for accelerated MRI. *IEEE Trans Med Imaging*. 2020;39(2):377-386. <https://doi.org/10.1109/TMI.2019.2927101>
23. Wundrak S, Paul J, Ulrici J, et al. Golden ratio sparse MRI using tiny golden angles. *Magn Reson Med*. 2016;75(6):2372-2378. <https://doi.org/10.1002/mrm.25831>
24. Knoll F, Schwarzl A, Diwok C, Sodickson DK. gpuNUFFT—an open source GPU library for 3D regridding with direct Matlab interface. *Proc Int Soc Magn Reson Med*. 2014;22:4297.
25. Huang F, Vijayakumar S, Li Y, Hertel S, Duensing GR. A software channel compression technique for faster reconstruction with many channels. *Magn Reson Imaging*. 2008;26(1):133-141. <https://doi.org/10.1016/j.mri.2007.04.010>
26. Rosenzweig S, Holme HCM, Uecker M. Simple auto-calibrated gradient delay estimation from few spokes using Radial Intersections (RING). *Magn Reson Med*. 2019;81(3):1898-1906. <https://doi.org/10.1002/mrm.27506>
27. Walsh DO, Gmitro AF, Marcellin MW. Adaptive reconstruction of phased array MR imagery. *Magn Reson Med*. 2000;43(5):682-690. [https://doi.org/10.1002/\(sici\)1522-2594\(200005\)43:5<682::aid-mrm10>3.0.co;2-g](https://doi.org/10.1002/(sici)1522-2594(200005)43:5<682::aid-mrm10>3.0.co;2-g)
28. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Lecture Notes in Computer Science Vol 9351. Cham, Switzerland: Springer; 2015:234-241.
29. Hyun CM, Kim HP, Lee SM, Lee S, Seo JK. Deep learning for undersampled MRI reconstruction. *Phys Med Biol*. 2018;63(13):135007. <https://doi.org/10.1088/1361-6560/aac71a>
30. Kyong Hwan J, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process*. 2017;26(9):4509-4522. <https://doi.org/10.1109/TIP.2017.2713099>
31. Yang G, Yu S, Dong H, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. 2018;37(6):1310-1321. <https://doi.org/10.1109/TMI.2017.2785879>
32. Eo T, Jun Y, Kim T, Jang J, Lee HJ, Hwang D. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magn Reson Med*. 2018;80(5):2188-2201. <https://doi.org/10.1002/mrm.27201>
33. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med*. 2018;80(5):2139-2154. <https://doi.org/10.1002/mrm.27178>
34. Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magn Reson Med*. 2018;80(3):1189-1205. <https://doi.org/10.1002/mrm.27106>
35. Simonyan K, Zisserman A. *Very deep convolutional networks for large-scale image recognition*. Paper presented at: International Conference on Learning Representations; May 7-9, 2015; San Diego, CA. <https://arxiv.org/abs/1409.1556>
36. Trabelsi C, Bilaniuk O, Zhang Y, et al. Deep complex networks. 2017. <https://arxiv.org/abs/1705.09792>
37. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612. <https://doi.org/10.1109/tip.2003.819861>
38. Crete F, Dolmiere T, Ladret P, Nicolas M. The blur effect: perception and estimation with a new no-reference perceptual blur metric. *Proc SPIE*. 2007; 6492:64920I.
39. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging*. 2017;3(1):47-57. <https://doi.org/10.1109/Tci.2016.2644865>
40. Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced deep residual networks for single image super-resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE; 2017. <https://doi.org/10.1109/CVPRW.2017.151s>
41. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging*. 2019;38(1):167-179. <https://doi.org/10.1109/TMI.2018.2858752>
42. Lonnig K, Putzky P, Sonke JJ, Reneman L, Caan MWA, Welling M. Recurrent inference machines for reconstructing heterogeneous MRI data. *Med Image Anal*. 2019;53:64-78. <https://doi.org/10.1016/j.media.2019.01.005>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Shen D, Ghosh S, Haji-Valizadeh H, et al. Rapid reconstruction of highly undersampled, non-Cartesian real-time cine *k*-space data using a perceptual complex neural network (PCNN). *NMR in Biomedicine*. 2021;34:e4405. <https://doi.org/10.1002/nbm.4405>