Deep complex convolutional network for fast reconstruction of 3D late gadolinium enhancement cardiac MRI

Hossam El-Rewaidy1,2 | Ulf Neisius1 | Jennifer Mancio1 | Selcuk Kucukseymen1 | Jennifer Rodriguez1 | Amanda Paskavitz1 | Bjoern Menze2 | Reza Nezafat1

1Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts
2Department of Computer Science, Technical University of Munich, Munich, Germany

Correspondence
Reza Nezafat, PhD, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215, USA.
Email: rnezafat@bidmc.harvard.edu

Funding information
American Heart Association, Grant/Award Number: 15EIA22710040; National Institutes of Health, Grant/Award Number: 5R01HL129185

Several deep-learning models have been proposed to shorten MRI scan time. Prior deep-learning models that utilize real-valued kernels have limited capability to learn rich representations of complex MRI data. In this work, we utilize a \textit{complex-valued} convolutional network (\textsc{C}Net) for fast reconstruction of highly under-sampled MRI data and evaluate its ability to rapidly reconstruct 3D late gadolinium enhancement (LGE) data. \textsc{C}Net preserves the complex nature and optimal combination of real and imaginary components of MRI data throughout the reconstruction process by utilizing complex-valued convolution, novel radial batch normalization, and complex activation function layers in a U-Net architecture. A \textit{prospectively} under-sampled 3D LGE cardiac MRI dataset of 219 patients (17 003 images) at acceleration rates $R = 3$ through $R = 5$ was used to evaluate \textsc{C}Net. The dataset was further retrospectively under-sampled to a maximum of $R = 8$ to simulate higher acceleration rates. We created three reconstructions of the 3D LGE dataset using (1) \textsc{C}Net, (2) a compressed-sensing-based low-dimensional-structure self-learning and thresholding algorithm (LOST), and (3) a real-valued U-Net (realNet) with the same number of parameters as \textsc{C}Net. LOST-reconstructed data were considered the reference for training and evaluation of all models. The reconstructed images were quantitatively evaluated using mean-squared error (MSE) and the structural similarity index measure (SSIM), and subjectively evaluated by three independent readers. Quantitatively, \textsc{C}Net-reconstructed images had significantly improved MSE and SSIM values compared with realNet (MSE, 0.077 versus 0.091; SSIM, 0.876 versus 0.733, respectively; \textit{p} < 0.01). Subjective quality assessment showed that \textsc{C}Net-reconstructed image quality was similar to that of compressed sensing and significantly better than that of realNet. \textsc{C}Net reconstruction was also more than 300 times faster than compressed sensing. Retrospective under-sampled images demonstrate the potential of \textsc{C}Net at higher acceleration rates. \textsc{C}Net enables fast reconstruction of highly accelerated 3D MRI with superior performance to real-valued networks, and achieves faster reconstruction than compressed sensing.

Abbreviations: BN, batch normalization; \textsc{CReLU}, complex rectified linear unit; LGE, late gadolinium enhancement; LOST, low-dimensional-structure self-learning and thresholding algorithm; MSE, mean-squared error; realNet, real-valued network; SSIM, structural similarity index measure.
KEYWORDS
complex convolutional network, deep learning, image reconstruction, late gadolinium enhancement, MRI

1 | INTRODUCTION

Long MRI scan time remains a challenge, impacting patient throughput and limiting spatial and temporal resolution. Over the past two decades, numerous acceleration techniques such as parallel imaging, constrained reconstruction, dictionary-based reconstruction, compressed sensing, and magnetic resonance fingerprinting have been developed to reduce scan time. Data-driven techniques such as deep learning are being explored to accelerate image acquisition without explicit assumptions on data. Recent deep learning reconstruction techniques are based on multiple convolution operations using pre-trained kernels with additional non-linearity in the form of activation functions. Different convolutional neural network schemes are utilized to reconstruct MR images from under-sampled acquisitions. The iterative optimization process of compressed sensing formulated as a variational model was recently unfolded in the form of a variational network, in which each layer acts as one gradient descent iteration. Similarly, the iterative reconstruction scheme was represented as deep-cascaded convolutional networks with interleaved data consistency layers that maintain acquired data fidelity after each convolutional network. Instead of cascading networks, Qin et al. developed a convolutional recurrent neural network that shares reconstruction information across different network iterations to simulate the iterative optimization process. Deep generative networks have also been utilized to remove aliasing artifacts from magnitude MR images where the generator is trained to produce artifact-free images.

Both magnitude and complex imaging data are used in deep learning-based MR image reconstruction. Readily available magnitude images are appealing for developing and testing new deep-learning-based techniques. However, the phase also carries important information in MR reconstruction that cannot be ignored. While several recent studies have used complex image/k-space information as network input, these approaches process real and imaginary components of data separately, concatenated to each other, or fed into a deep convolutional network as different input channels, analogous to RGB information of a color image. However, these approaches only use a kernel in which components of the kernel are real values, which have limited capability to learn naturally acquired complex MRI data representations. Thus, there is an unmet need for deep-learning reconstruction models that process MRI data in the complex domain to enable learning richer representations of the complex data.

A convolutional network framework that utilizes complex-valued convolutional kernels to learn complex representations from data was recently proposed. These kernels have complex weights of real and imaginary components that are adjusted with a convolutional operation incorporating both real and imaginary components of the input data and output feature maps. This family of neural networks exhibits specific characteristics in its learning, self-organization, and processing dynamics. The complex manifolds learned by complex networks are easier to optimize and faster to learn, have longer noise-immunity memory, and exhibit better generalization characteristics. Considering the complex nature of MRI signals and the necessity for an optimal combination of real and imaginary parts of the k-space data, a complex neural network may be beneficial in MRI reconstruction. Complex neural networks were previously utilized to estimate tissue parametric maps in MRI fingerprinting and MR image reconstruction.

In this work, we utilize and evaluate a complex convolutional neural network (CNNet) for fast reconstruction of highly under-sampled 3D late gadolinium enhancement (LGE) MR data. Novel batch normalization (BN) is proposed to normalize complex-valued feature maps and MR data without magnitude or phase distortions. CNNet performance is qualitatively and quantitatively evaluated using a prospectively under-sampled cardiac MRI dataset of 3D LGE images. In this study, we used a compressed-sensing based low-dimensional-structure self-learning and thresholding algorithm (LOST) as a reference for training and evaluating our model.

2 | METHODS

We consider a 3D complex-valued zero-filled MR image, \( Y \in \mathbb{C}^N \), such that

\[
Y = SF^{-1} \Gamma X + \varepsilon,
\]

where \( X \in \mathbb{C}^{N_4} \) represents fully sampled k-space data from \( N_c \) coils, \( N = N_x N_y N_z \), where \( N_x, N_y, \) and \( N_z \) are the numbers of acquired k-space samples in the x, y, and z directions, respectively; \( F^{-1} \) is the inverse Fourier coefficient matrix (defined as the Kronecker product between an \( N \times N \) DFT matrix and an \( N_c \times N_c \) identity matrix), \( \varepsilon \) is modeled as an additive white Gaussian noise, \( \Gamma \) is an under-sampling mask that controls the acceleration rate during MR acquisition, and \( S \) is a matrix contains the complex conjugate of coil sensitivity maps from \( N_c \) coils of size \( N \times NN_c \). In the case
of accelerated acquisition (ie beyond the Nyquist criterion), Equation 1 results in aliased image data \( Y \), where the characteristics of artifacts in \( Y \) are largely dependent on the acceleration rate and the under-sampling pattern.

A priori knowledge of image and artifact characteristics can be employed to regularize an optimization function to produce an acceptable solution for the reconstruction problem. In deep learning, image and artifact characteristics can be exploited as trainable parameters from the same domain as the under-sampled images. Training these parameters can be achieved by solving the following optimization problem:

\[
\min_{\Theta} \| \hat{X} - \Psi(Y(\Theta)) \|^2_2
\]

where \( \hat{X} \) represents the single-channel LOST-reconstructed data in the image domain, and \( \Psi \) is a domain transform that maps the corrupted under-sampled image manifold to an artifact-free image manifold using learnable parameters \( \Theta \). Each of the trained parameters \( \Theta \) acts as an operator in a nonlinear system that can be generalized to reconstruct new artifact-free images from under-sampled MR acquisitions.

In this work, we aim to exploit both image and artifact characteristics in their original complex form by learning the complex domain transform, \( \Psi \), and parameters, \( \Theta \). Each step of the proposed reconstruction algorithm is described below.

### 3 | NETWORK ARCHITECTURE

\( \mathcal{C} \)Net is a fully convolutional network with a U-net architecture (Figure 1) that propagates complex image data in contractive and expansive paths for multi-scale artifact removal and multi-resolution de-noising.\(^{17,20,24,34} \) In the contractive path, complex image input data are fed into two complex convolutional layers each with 64 kernels to extract basic features and noise patterns. The resulting feature maps are down-sampled using a complex convolution layer of 64 kernels and a stride of 2. This process is repeated through three down-sampling stages; at each stage, higher-scale features are extracted using complex convolutional layers with twice the previous number of kernels (ie 64, 128, and 256 over the three down-sampling stages). The resulting feature maps pass through two convolution layers of 512 and 256 complex kernels. The expansive path then maps the feature maps at each down-sampling stage onto an analogous stage of similar map size and kernel number (Figure 1B). The up-sampling layers are utilized 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. Corresponding feature maps in the up-sampling and down-sampling stages are concatenated.

At all previous steps, each convolutional layer is followed by radial BN and complex activation layers, where all kernels are 2D complex valued consisting of real and imaginary components of size \( 3 \times 3 \) (ie, the equivalent tensor size is \( 3 \times 3 \times 2 \)). The last feature maps are combined using a complex convolutional layer with kernel size \( 1 \times 1 \) to reconstruct the final output image. The following sections describe various components of the proposed network.

For a 2D complex image, the corresponding \( \mathcal{C} \)Net data point is represented by a real-valued tensor of size \( (N_x, N_y, 2) \) that includes both real and imaginary components of the complex image. This tensor flows through multiple complex operational points in the network as follows.

#### 3.1 | Complex convolutional layer

In the proposed network, the complex input image, \( I = x + iy \), is convolved with a complex filter, \( k = a + ib \),\(^{26} \) such that

\[
I \circ k = (a \circ x - b \circ y) + i(b \circ x + a \circ y)
\]

where \( \circ \) represents a convolution operation. This operation can be accomplished in the real-valued arithmetic by convolving the same kernels, \( a \) (or \( b \)), with each of the real and imaginary parts of the image, \( x \) and \( y \), separately (Figure 2). The gradients of \( a \) (or \( b \)) should be calculated in the backwards direction considering both operations on \( x \) and \( y \).

#### 3.2 | Radial BN layer

BN is important for accelerating and stabilizing the convergence of neural networks.\(^{35} \) Although the original BN was proposed for real-valued networks, a generalization to the complex data was proposed by Trabelsi et al.\(^{26} \) In their complex version, the distribution of both real and imaginary components was independently shifted to have zero mean, then scaled by the covariance matrix of the real and imaginary components to ensure equal variance for the two components. However, separately shifting the distribution of each component towards zero mean induces distortion in the phase and magnitude (Figure 3).
To address this challenge, we propose radial BN in which phase information is maintained and magnitude is scaled such that relative differences are preserved between complex quantities. In radial BN, the input complex data are transformed to polar coordinates, \( Z = R e^{i \theta} \). Standard BN is applied to the magnitude data, \( R \), to have a mean of \( \tau \) and standard deviation of 1; thus, the normalized magnitude can be calculated as

\[
R_{bn} = \left( \frac{R - \mu_R}{\sqrt{\sigma_R^2 + \epsilon}} \right)^\gamma + \beta + \epsilon. \tag{4}
\]
where $\mu_R$ and $\sigma_R^2$ are the respective mean and variance of $R$, $\epsilon$ is a constant added to the variance for numerical stability, and $\beta$ and $\gamma$ are trainable parameters for shifting and scaling data distribution. We introduce a new constant, $\tau$, to ensure a positive value for the normalized $R$ ($\tau = 1$ was empirically used in our experiments). The normalized complex data are transformed back to Cartesian coordinates using the normalized magnitude and the same phase, $Z = R_{\text{norm}} e^{i\theta}$.

3.3 | Down-sampling and up-sampling layers

To collect artifact patterns on multiple scales and allow multi-resolution artifact removal, complex feature maps generated by the convolutional layers were down-sampled in some parts of the network then up-sampled again to retain the original resolution of the output images. To down-sample the feature maps, a complex convolutional layer with stride greater than 1 was utilized. The up-sampling layer generates complex feature maps with a larger size than the input feature maps by bilinear interpolation of the real and imaginary components.

3.4 | Complex activation function

Real-valued activation functions can be extended in the complex domain to separately activate real and imaginary parts, eg a complex rectified linear unit (CReLU):

$$\text{CReLU} = \text{ReLU}(\Re(Z)) + i \text{ReLU}(\Im(Z)),$$

where $\Re(Z)$ and $\Im(Z)$ are the real and imaginary components of the complex-valued image $Z$. CReLU is holomorphic (ie complex differentiable in neighborhood points) when both real and imaginary components are strictly positive or negative.

4 | IMAGE ACQUISITION AND PRE-PROCESSING

To assess the performance of the proposed reconstruction techniques, we utilized a dataset of 219 patients (145 males, mean 55 years) referred for a clinical cardiac MRI examination for viability assessment. These patients were recruited prospectively as part of our previous study.
Informed consent was obtained from each subject and the imaging protocol was approved by the institutional review board and institutional human subjects committee. 3D LGE images were acquired using a 1.5 T Philips Achieva system (Philips Healthcare, Best, The Netherlands) with a 32-channel cardiac coil. Images were acquired in the axial direction to cover the whole heart using a gradient echo imaging sequence with the following parameters: repetition time/echo time = 5.2/2.6 ms, field of view = 320 × 320 × 100-120 mm³, flip angle = 25°, and spatial resolution = 1.0-1.5 mm³. A free-breathing ECG-triggered navigator-gated with inversion-recovery gradient echo imaging sequence was used. The 3D \( k \)-space was prospectively under-sampled in all patients using a pseudorandom mask, where the \( k \)-space was fully sampled within 15-20% in the \( k_y \) direction and 25% in the \( k_z \) direction around the center of \( k \)-space, and randomly sampled elsewhere\(^{37,38} \) (Figure 1). The acceleration rate was randomly prospectively chosen by the operator between \( R = 3 \) and \( R = 5 \) (130 patients at \( R_p = 3 \), 25 patients at \( R_p = 4 \), and 64 patients at \( R_p = 5 \)). \( k \)-space data from all 32 coil channels were exported and used for evaluation. All images were previously reconstructed using the LOST algorithm\(^{39} \) for training and evaluation.
5 | NETWORK TRAINING

The 3D k-space data from 32 coils were zero-filled and transformed into the complex image domain by 3D inverse Fourier transformation. Coil sensitivity information was then utilized to combine data from different coils into single complex-valued 3D volume data per patient in the image domain using $B_1$-weighted reconstruction. The corresponding LOST-reconstructed image data were used as a reference for loss calculations during network training and evaluation of results during testing. All images were normalized using the same radial BN process.

We divided the data into training (153 patients; 11 726 slices) and testing datasets (66 patients; 5277 slices). The network was trained until convergence with a fixed number of epochs (ie 50) via an Adam optimizer with a learning rate function, $0.1^{\text{epoch}/20}+1$, where the learning rate exponentially decreased with the number of epochs. The batch number was 100 complex images of size $256 \times 256$. The mean-squared error (MSE) loss ($\ell$) function was applied to minimize the error between LOST and network prediction images, such that $\ell = \|x_{\text{UST}} - x_{\text{URUS}}\|_2^2$, where $x_{\text{UST}}$ represents LOST-reconstructed magnitude images and $x_{\text{URUS}}$ is the magnitude of CNNet complex predictions.

This model was implemented using the open-source Python-based PyTorch library Version 0.41 (code available at https://github.com/hossam-elrewaidy/urus-mri-recon). The $B_1$-weighted reconstruction was performed on MATLAB (MathWorks, Natick, Massachusetts). All models in this work were trained and tested on an NVIDIA DGX-1 system equipped with eight Tesla V100 GPUs (each of 32 GB memory and 5120 cores), and a CPU of 88 cores: Intel Xeon 2.20 GHz each, and 504 GB RAM memory. CNNet takes an average of 4 h to train on this machine. CPU-based LOST reconstruction was performed on a computing cluster of 20 cores and 100 GB of RAM.

6 | EXPERIMENTAL EVALUATION

CNNet was evaluated by two means: initially, the prospectively under-sampled data ($R_p = 3-5$) were reconstructed, then we retrospectively under-sampled the original data to simulate higher accelerations beyond the prospective under-sampling rates in separate experiments. Three new datasets ($D_1$, $D_2$, and $D_3$) were generated from the originally acquired dataset ($D_0$) by increasing the acceleration rate by 1 in each new dataset. Since $D_0$ includes patient data acquired at different acceleration rates, $R_p = [3, 4, 5, 6]$, $D_1$ includes retrospective acceleration rates, $R_r = [4, 5, 6]$, $D_2$ includes rates $R_r = [5, 6, 7]$, and $D_3$ includes rates $R_r = [6, 7, 8]$. A pseudorandom mask was used to retrospectively under-sample the acquired k-space, where 16-21% of the k-space lines at the center were fully sampled and the rest randomly under-sampled.

For comparison, a real-valued network (realNet) of the same U-net architecture was built using the standard real-valued ReLU, BN, and convolutional layers. RealNet has the same number of convolutional layers as CNNet but has an increased number of kernels at each layer as illustrated in Figure S1. The total number of trainable parameters in realNet is 12 475 073, versus 12 472 578 in CNNet. RealNet has two input channels containing real and imaginary components of zero-filled complex images and an output of a magnitude reconstructed image. The optimal hyper-parameters were experimentally determined for both CNNet and realNet.

Deep residual networks (ResNet) were also investigated in this work. Complex ResNet was built with five standard residual blocks of 64, 128, 256, 128, and 64 3 x 3 complex kernels, respectively (Figure S2). Two convolutional layers were added at the network input and output with 64 and 7 x 7 complex kernels, respectively. The total number of trainable parameters in this network was 13 089 x 10^2. Similar to CNNet, ResNet has complex-valued 2D images at the input and output layers.

7 | DATA ANALYSIS

For quantitative assessment of CNNet reconstruction performance, the MSE and structural similarity index measure (SSIM) of the magnitude of CNNet-reconstructed and LOST-reconstructed images were calculated. While MSE calculates the unobserved mean intensity difference between the predicted and reference images, SSIM quantifies human visual perception qualities by combining luminance, contrast, and structural differences between predicted and reference images. Per-slice and per-patient LGE scar percentages were calculated using a conventional semi-automatic thresholding-based scar quantification method with three standard deviations of signal from the remote normal myocardium. For this purpose, the left ventricular myocardium was manually segmented from the entire heart in all patients with a scar in the testing dataset.

For qualitative assessment, three independent readers (with 10, 6, and 6 years of experience in cardiac MRI, respectively) graded the reconstructed images using LOST, CNNet, and realNet on a five-point score to evaluate overall quality per patient (1—poor quality with large artifacts, 2—fair quality with moderate artifacts, 3—good quality with small artifacts, 4—excellent quality with no artifacts, and 5—spectacular quality as fully sampled images). Readers were blinded to the method of reconstruction. In addition, each reader identified patients with a left ventricular scar. Finally, all three reconstructed images from each patient were shown to the readers simultaneously, and each reader selected the best imaging dataset for best overall image quality.
A two-tailed Student t-test was performed for comparison of continuous variables between reconstruction methods. To compare categorical data, the Chi-squared test was used. Significance was declared at two-sided p-values less than 0.05. For pairwise comparisons following a three-group inferential test that was significant, a Bonferroni correction was used.

8 | RESULTS

Sample LGE images reconstructed using realNet, CNet, and LOST from different patients acquired at prospective acceleration rates $R_p = 3$ and 5 are shown in Figure 4. CNet-reconstructed images maintain scar-blood contrast and are visually comparable to those of LOST. Figure 5 shows two different patient LGE images with scar reconstructed by LOST, CNet, and realNet. In both images, CNet preserves better fine details of the scar than realNet with respect to the reference images, as indicated by the error maps.

Figure 6 shows CNet-reconstructed images scanned at prospective acceleration rate $R_p = 3$ and corresponding retrospectively under-sampled versions at $R_r = 4, 5,$ and 6 from D1, D2, and D3 datasets, respectively, compared with realNet images. Noisy, blurry zero-padded images were restored by CNet reconstruction at all acceleration rates. In addition, CNet reconstruction recovered scar-blood contrast and scar shape compared with zero-padded images with respect to reference images at different acceleration rates. CNet was also able to perform at higher acceleration rates (up to $R_r = 8$) and maintain image details (Figure 7).

The performance of CNet was compared with realNet and zero-filled images (Table 1). CNet images showed significantly higher SSIM than zero-filled images and significantly lower MSE for all datasets (ie D0, D1, D2, and D3) ($p < 0.01$ for all). The visual perceptual-based SSIM values of CNet images were significantly higher than those of realNet for all datasets ($p < 0.01$). However, CNet-based MSE values were significantly lower than those of realNet in D0 and D1 only ($p < 0.01$).

CNet showed faster convergence during training and more stable testing results at different training epochs when radial BN layers were included (Figure S3). CNet reported an MSE of 0.083 and SSIM of 0.862 without radial BN, compared with 0.077 and 0.876 when radial BN was utilized. The U-net-based CNet showed slightly better performance compared with the complex ResNet. The MSE and SSIM of ResNet-reconstructed images were 0.079 and 0.869 when radial BN was utilized, and 0.087 and 0.858 without BN, respectively.

CNet performance at different epochs throughout network training is shown in Figure 8. SSIM increased and MSE decreased as the number of epochs increased in the prospectively (ie D0) and retrospectively (ie D1, D2, and D3) under-sampled datasets. The consistent evolution of SSIM
and MSE between training and testing for all datasets indicates minimal overfitting of the model. Sample magnitude and phase parts of feature maps captured after the first complex convolutional layer showed a variety of image-specific features and noise patterns within the layer (Figure S4).

There was no difference between CNet and LOST (3.54 ± 0.92 and 3.51 ± 1.05, respectively, p = 0.54) in the overall assessment of image quality by three independent readers (Table 2). CNet-reconstructed images had better image quality than those using realNet (3.54 ± 0.92 versus 3.06 ± 0.99; p < 0.01) (Table 2). In the testing dataset, readers identified on average 20 patients with a scar in LGE images reconstructed by LOST and CNet, but only 19 patients with a scar on LGE images reconstructed by realNet. CNet, LOST, and realNet 3D LGE were ranked as the best method for image quality in 24, 24, and 17 patients, respectively (Table 2).

There was an excellent correlation in scar extent between CNet and LOST (R² > 0.97 and R² > 0.99 for per-slice and per-patient scar percentages, respectively) (Figure 9). A Bland-Altman plot showed a small bias (0.29%) and narrow limits of agreements (6.86%) for per-slice scar percentage error by CNet with respect to LOST reference images. The per-patient scar percentage error was 0.17 ± 1.49% by CNet with respect to LOST and increased consistently with increasing acceleration rates (Figure 9D).

For a typical 3D LGE dataset of size 256 pixels × 256 pixels, 100 slices, and 32 channel coils, the reconstruction time was about 15 s for CNet and 89 min for LOST, which constitutes a 310-fold decrease in reconstruction time.

9 | DISCUSSION

In this work, we utilize a deep complex convolutional network to improve MR image reconstruction from under-sampled acquisitions. This network exploits a priori knowledge of MR image and artifact characteristics in the reconstruction process by learning complex image representations in the form of complex-valued kernels. The efficient utilization of phase information via dual complex components (ie real and imaginary) of the MRI data throughout the reconstruction pipeline allows efficient artifact removal. To stabilize the convergence of CNet, a novel radial BN method that maintains relative differences between data points without magnitude or phase distortions was proposed.
The U-net architecture used in our study allows multi-scale artifact removal. The network receptive field increases after each down-sampling layer such that MR images are filtered at different resolution levels and up-sampled to provide clean versions of the artifact-contaminated images at each level. Unlike conventional U-net networks, the down-sampling was performed using complex convolution layers with a stride of 2 instead of the traditional pooling schemes. Although the convolution-based down-sampling adds more trainable parameters to the network, it was utilized in this work because pooling schemes are not well investigated for complex networks. In addition, the bypass connections at each down-sampling stage create shortcuts for the gradient flow in shallow layers and offer better convergence characteristics. Further studies are warranted to investigate the performance of other complex network architectures.

The large prospectively under-sampled dataset considered for the training and evaluation of our models covers the whole heart in the axial direction and provides a wide range of heart structures, slice locations, and scar shapes. This heterogeneous dataset allowed efficient training of our model with minimal overfitting without the need for data augmentation. The large testing dataset also enabled a comprehensive evaluation of our model performance and generalizability in a clinical setting.

**FIGURE 6** Representative LGE images at prospective acceleration rate $R_p = 3$ and retrospective acceleration rates $R_r = 4$, 5, and 6 reconstructed by CNet compared with LOST reference, realNet, and zero-filled images at each acceleration rate. Image quality is restored with minimal noise and blurring artifacts in CNet-reconstructed images, similar to that of the reference images. Scar-blood contrast is also maintained at higher acceleration rates.
We demonstrate the importance of incorporating phase information in the reconstruction process by comparing the complex-valued CNet with the real-valued realNet. CNet shows superior image quality compared with realNet in both quantitative and qualitative evaluations, despite their similar architectures and an equal number of trainable parameters. However, CNet performs twice the number of convolutional operations compared with realNet, since each complex-valued convolution operation was implemented by four real-valued convolution operations. Although our LGE dataset was acquired with relatively short echo time (2.6 ms), which reduces phase errors (mainly caused by field inhomogeneity), the
phases in our input complex data carry valuable information related to the \( k \)-space under-sampling pattern (i.e., pseudorandom trajectory) and hence can help remove under-sampling artifacts.

Image quality using CNet was very similar to that of LOST in both subjective and objective assessments. However, CNet reconstruction time was over 300 times shorter, which is important when considering its adoption in a clinical setting. CNet yielded 3D LGE images without artifact and showed potential for acceleration beyond what compressed sensing can achieve. Additional efforts to increase the acceleration rate in a prospectively under-sampled imaging sequence are needed to evaluate the performance of CNet at high acceleration rates.

In many MRI applications, collecting a fully sampled 3D dataset that can be used as a reference dataset is not possible. A fully sampled high-resolution dataset typically requires a very long scan time, and image quality is impacted by factors such as respiratory or cardiac motion. For post-contrast MRI sequences such as LGE, a change in the underlying signal such as contrast washout adds additional complexity. Therefore, calculating the loss function in such sequences is challenging. In our study, we relied on LOST-reconstructed images for learning and evaluation of CNet. While not ideal, this is the only solution that allowed us to evaluate the potential of CNet in high-resolution 3D LGE imaging.

### TABLE 1

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MSE</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>CNet</td>
<td>realNet</td>
</tr>
<tr>
<td>D0</td>
<td>0.246</td>
<td>0.077</td>
</tr>
<tr>
<td>D1</td>
<td>0.263</td>
<td>0.098</td>
</tr>
<tr>
<td>D2</td>
<td>0.278</td>
<td>0.117</td>
</tr>
<tr>
<td>D3</td>
<td>0.290</td>
<td>0.131</td>
</tr>
</tbody>
</table>

### TABLE 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Overall image quality</th>
<th>Presence of LGE</th>
<th>Method preference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOST</td>
<td>CNet</td>
<td>realNet</td>
</tr>
<tr>
<td></td>
<td>LOST</td>
<td>CNet</td>
<td>realNet</td>
</tr>
<tr>
<td></td>
<td>LOST</td>
<td>CNet</td>
<td>realNet</td>
</tr>
<tr>
<td>Reader 1</td>
<td>3.0 ± 0.9</td>
<td>3.14 ± 0.7</td>
<td>2.89 ± 0.8</td>
</tr>
<tr>
<td>Reader 2</td>
<td>3.29 ± 1.0</td>
<td>3.21 ± 0.9</td>
<td>2.5 ± 0.9</td>
</tr>
<tr>
<td>Reader 3</td>
<td>4.24 ± 0.84</td>
<td>4.26 ± 0.71</td>
<td>3.77 ± 0.87</td>
</tr>
</tbody>
</table>

Image quality using CNet was very similar to that of LOST in both subjective and objective assessments. However, CNet reconstruction time was over 300 times shorter, which is important when considering its adoption in a clinical setting. CNet yielded 3D LGE images without artifact and showed potential for acceleration beyond what compressed sensing can achieve. Additional efforts to increase the acceleration rate in a prospectively under-sampled imaging sequence are needed to evaluate the performance of CNet at high acceleration rates.

In many MRI applications, collecting a fully sampled 3D dataset that can be used as a reference dataset is not possible. A fully sampled high-resolution dataset typically requires a very long scan time, and image quality is impacted by factors such as respiratory or cardiac motion. For post-contrast MRI sequences such as LGE, a change in the underlying signal such as contrast washout adds additional complexity. Therefore, calculating the loss function in such sequences is challenging. In our study, we relied on LOST-reconstructed images for learning and evaluation of CNet. While not ideal, this is the only solution that allowed us to evaluate the potential of CNet in high-resolution 3D LGE imaging.
During CNet training, the loss function was calculated between the magnitude of CNet-derived complex images and reference LOST magnitude images. While mapping the complex input image to a magnitude reference forces the network to produce an optimal result for magnitude images without consideration of errors in the phase, complex-to-complex mapping of both input and output images was not investigated in this work due to unavailability of complex-valued reference images, since the complex output of LOST is not normally saved in our clinical workflow.

In this study, we used a complex neural network in a U-net type architecture. The concept of complex neural networks can potentially be adopted in other architectures such as cascaded networks. In LGE imaging, the magnitude of reconstructed images is used to assess the presence of the scar, and phase images are often discarded. However, in several other imaging sequences such as phase-contrast MRI, recovering the phase information is important. Future studies should assess both magnitude and phase recovery to determine if a complex network can provide better image phase reconstruction.

For scar quantification, we used a three-standard-deviation thresholding method that utilizes the distribution of pixel intensities within the myocardium to assess the scar volume. Although the distribution of the myocardial intensities could be altered due to the non-linear processing performed by convolutional networks, an excellent correlation \( R^2 > 0.97 \) was reported between the quantified scar volumes from CNet-reconstructed data and the reference images. An optimal threshold for scar quantification may depend on the reconstruction algorithm and warrants further studies.

Our study has several limitations. We did not have a fully sampled 3D k-space dataset as a reference for comparison, and images were compared with a compressed-sensing reconstruction as the reference standard. While our imaging datasets are relatively large, there was only a small subset of patients with a scar in our training datasets. The performance of the complex convolutional network was assessed only by 3D LGE. In this work, we investigated a complex network solely with 2D convolutional layers. However, 3D convolutional networks have the potential to learn spatial correlations in 3D and further studies are warranted to assess the performance of 3D networks. We compared the performance of CNet with a real-valued network that takes real and imaginary channels as input and produces single-channel magnitude image as output; however, a better comparison would be with a real-valued network that has real and imaginary channels in both input and output. We tested the performance of a complex convolutional network using the widely used U-net architecture only without intermediate data-consistency steps. Including data-consistency steps that reconstruct coil-combined 3D acquired k-space data in our 2D network was challenging, and future studies are warranted to utilize more advanced models that incorporate data-consistency steps.

**10 | CONCLUSION**

CNet enables fast reconstruction of large 3D MRI datasets with superior performance compared with real-valued kernel networks. Our results demonstrate that CNet can reconstruct 3D LGE images with acceleration rates up to 8 with a more than 300-fold speed-up in reconstruction time compared with compressed sensing.
REFERENCES


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: El-Rewaidy H, Neisius U, Mancio J, et al. Deep complex convolutional network for fast reconstruction of 3D late gadolinium enhancement cardiac MRI. NMR in Biomedicine. 2020;e4312. https://doi.org/10.1002/nbm.4312