

Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning—proof of concept in congenital heart disease

Andreas Hauptmann¹ | Simon Arridge¹ | Felix Lucka^{1,2} | Vivek Muthurangu^{3*} | Jennifer A. Steeden^{3*}

¹Department of Computer Science, University College London, London, United Kingdom

²Computational Imaging, Centrum Wiskunde and Informatica (CWI), Amsterdam, Netherlands

³UCL Centre for Cardiovascular Imaging, University College London, London, United Kingdom

Correspondence

Vivek Muthurangu, UCL Centre for Cardiovascular Imaging, Institute of Cardiovascular Science, 30 Guildford Street, London WC1N 1EH, United Kingdom.

Email: v.muthurangu@ucl.ac.uk

Funding information

Royal Society, Grant/Award Number: DH130079; H2020 LEIT Information and Communication Technologies, Grant/Award Number: No 732411; Engineering and Physical Sciences Research Council, Grant/Award Number: EP/N022750/1; Stichting voor de Technische Wetenschappen, Grant/Award Number: NWO 613.009.106/2383

Purpose: Real-time assessment of ventricular volumes requires high acceleration factors. Residual convolutional neural networks (CNN) have shown potential for removing artifacts caused by data undersampling. In this study, we investigated the ability of CNNs to reconstruct highly accelerated radial real-time data in patients with congenital heart disease (CHD).

Methods: A 3D (2D plus time) CNN architecture was developed and trained using synthetic training data created from previously acquired breath hold cine images from 250 CHD patients. The trained CNN was then used to reconstruct actual real-time, tiny golden angle (tGA) radial SSFP data (13× undersampled) acquired in 10 new patients with CHD. The same real-time data was also reconstructed with compressed sensing (CS) to compare image quality and reconstruction time. Ventricular volume measurements made using both the CNN and CS reconstructed images were compared to reference standard breath hold data.

Results: It was feasible to train a CNN to remove artifact from highly undersampled radial real-time data. The overall reconstruction time with the CNN (including creation of aliased images) was shown to be >5× faster than the CS reconstruction. In addition, the image quality and accuracy of biventricular volumes measured from the CNN reconstructed images were superior to the CS reconstructions.

Conclusion: This article has demonstrated the potential for the use of a CNN for reconstruction of real-time radial data within the clinical setting. Clinical measures of ventricular volumes using real-time data with CNN reconstruction are not statistically significantly different from gold-standard, cardiac-gated, breath-hold techniques.

KEYWORDS

convolutional neural network, deep artifact suppression, deep learning, real-time, ventricular volumes

*Vivek Muthurangu and Jennifer A. Steeden contributed equally to this work

© 2018 The Authors Magnetic Resonance in Medicine published by Wiley Periodicals, Inc. on behalf of International Society for Magnetic Resonance in Medicine.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1 | INTRODUCTION

Cardiovascular magnetic resonance (CMR) is the reference standard method of measuring ventricular volumes in patients with congenital heart disease (CHD). Conventionally, breath-hold (BH), multi-slice, short-axis, cardiac-gated, balanced steady-state-free precession (bSSFP) cine imaging is used to acquire this data.¹ Unfortunately, many patients find breath-holding difficult and a rapid free-breathing alternative is desirable.

Real-time imaging offers an alternative to BH imaging, but relies on high levels of data undersampling to ensure adequate spatial and temporal resolution. This means that sophisticated reconstruction techniques are necessary to produce artifact-free images. One example is compressed sensing (CS), which is increasingly used to reconstruct real-time CMR images.^{2–4} However, there are drawbacks to CS including: (1) Computationally intensive and time-consuming reconstruction, and (2) challenging optimization of reconstruction parameters, which can result in unnatural looking images. Therefore, alternative reconstruction approaches may be useful.

The removal of aliases in images from undersampled MR data can be thought of as a de-noising problem. This is particularly true if the sampling pattern produces incoherent artifacts, as is the case with tiny golden angle (tGA) radial spokes.⁵ Recently, convolutional neural networks (CNN) have been shown to be well suited to de-noising and artifact removal problems.^{6–8} Consequently, it may be possible to use CNNs to remove aliasing from undersampled real-time MR data. This approach is dependent on large amounts of high quality training data, and one potential source is previously acquired, retrospectively cardiac-gated BH-bSSFP cine imaging.

In this study, “synthetic” training data was created from a library of BH-bSSFP cine images and used to train a CNN to map between aliased and artifact-free images (deep artifact suppression). The trained CNN was then used to remove aliases from actual undersampled real-time MR data acquired in prospectively scanned patients with CHD. The aims of this study were to (1) investigate the effect of different radial sampling patterns on the effectiveness of CNN-based reconstruction, (2) acquire actual real-time data in patients with CHD using a tGA radial sampling pattern, and (3) compare reconstruction speed, image quality, and accuracy of ventricular volumes calculated from CNN reconstruction compared to a current state-of-the-art CS algorithm.

2 | METHODS

2.1 | Deep artifact suppression algorithm—general concept

The deep artifact suppression algorithm was based on a modified residual U-Net architecture.^{9,10} A residual U-Net

is a multi-scale CNN where images are sequentially down-sampled and then up-sampled to produce an updated and clean version of the artifact-contaminated input image. This particular structure is especially suitable for tomographic problems (such as MR imaging), where the normal operator (projection followed by back-projection) is a type of convolution.¹⁰ The effectiveness of this approach has been shown in several studies for 2D tomographic problems^{11,12} and in 2.5D,⁶ where the temporal information is processed as 2D channels. In our study, the input was a 3D data set (2D image through time). A 3D structure was used to enforce temporal consistency between frames and prevented the flickering artifact present if a 2D slice-by-slice approach is used (see Supporting Information Video S1).

The residual U-Net was trained with synthetic training data consisting of paired artifact-free “ground truth” magnitude images and corresponding undersampled, artifact-contaminated images. The trained network was then tested with synthetic test data to quantify performance. Finally, the same trained network was used to reconstruct actual real-time data acquired in patients with CHD. These steps are fully explained below. The use of retrospectively collected training and test data, as well as collection of prospective real-time data, was approved by the local research ethics committee, and written consent was obtained from all subjects and/or guardians (National Research Ethics Service Committee reference: 06/Q0508/124).

2.2 | Preparation of synthetic training data

The synthetic training data was created from 2D short axis stacks of Cartesian, retrospectively cardiac-gated, breath-hold, multi-slice, short-axis, bSSFP cine images (BH-bSSFP). The imaging parameters were: matrix = $\sim 240 \times 204$, FOV = $\sim 320 \times 265$ mm, phase partial-Fourier 6/8, spatial resolution = $\sim 1.3 \times 1.3$ mm, slice thickness = ~ 10 mm, temporal resolution = ~ 32 ms, reconstructed cardiac phases = 40. Training data was collected from 250 previously scanned children and adults with pediatric heart disease or CHD (mean age: 22.3 ± 12.6 y, male: 140). No data from patients with single ventricles or images with breathing or arrhythmia artifacts were included in the synthetic training data. Full details of patient diagnoses are supplied in Supporting Information Text S1. On average, each patient had ~ 9 2D slices, resulting in 2276 paired 3D (2D plus time) data sets for training.

The first step in creating training data was to convert the retrospectively cardiac-gated data into synthetic real-time data that mirrored the acquisition parameters of the real-time sequence (as stated below). To provide similar spatial resolution, images were linearly resampled onto a 192×192 matrix, reducing the spatial resolution from 1.3 mm to 1.7 mm. This was followed by linear resampling in time onto points that were 36.4 ms apart (temporal resolution of the real-time sequence). The total number of temporal frames equaled the floor of the R-R wave

divided by 36.4 ms. The resulting data formed the ground truth images necessary for training.

To create undersampled, artifact-contaminated training data, the synthetic real-time data was Fourier transformed and gridded onto a $13\times$ undersampled radial trajectory (see below). The undersampled data was then regridded and inverse Fourier transformed back into image space. Both the ground truth and artifact-contaminated images were cropped to a 128×128 matrix to constrain the learning problem to the anatomy of interest (heart). The ground truth and artifact-contaminated images were also linearly interpolated through time to give 20 frames, because all inputs to the network had to be the same dimensions. Finally, each 3D data set was also normalized to have signal intensities in the range $[0, 1]$. All processing required for creation of the synthetic training data was performed in MATLAB (The MathWorks, Natick, MA). A flow diagram of the process is included in Supporting Information Figure S1.

2.3 | Assessing the effect of sampling strategy

Radial sampling was chosen for this study, because it is known to produce less pronounced undersampling artifacts compared to Cartesian acquisitions.¹³ We further hypothesized that a continuously rotating tiny golden angle ($\sim 23.63^\circ$) approach may be optimal for a residual U-Net reconstruction because of the noise-like characteristics of the aliases. Although tiny golden angles spacing has been shown to reduce eddy current effects and optimize image quality,¹³ the effect of specific sampling patterns on deep artifact suppression had not been investigated. Therefore, we compared the continuously rotating tiny golden angle sampling (tGA_{rot}) to

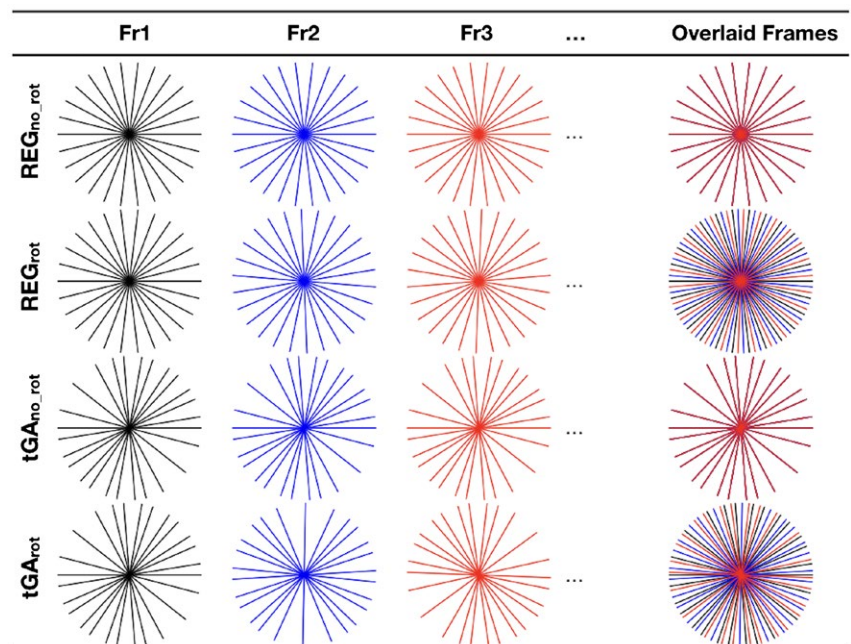
3 other sampling patterns (see Figure 1): (1) a tiny golden angle scheme with no rotation between frames (tGA_{no_rot}), (2) a regular angle scheme with regular rotation between frames (REG_{rot}), and (3) a regular angle scheme with no rotation between frames (REG_{no_rot}). Figure 2 shows both the sampling pattern and their corresponding artifacts.

2.4 | Residual U-Net architecture and training

Four residual U-Nets were trained, one for each of the radial sampling patterns. The synthetic training data for each U-Net consisted of paired artifact-free ground truth magnitude images and the corresponding undersampled, artifact-contaminated images (different for each sampling pattern). The architecture of the residual U-Net relied on a multi-scale decomposition of the input image and a skip connection at each scale (see Figure 3 for the chosen architecture). Each convolutional layer had a filter size of $3 \times 3 \times 3$ and was equipped with a rectified linear unit as nonlinearity, except the last layer that produced the residual update. The filters were equally weighted in spatial and temporal domain, and hence, no directions were favored in the training process. The output of the network was the input artifact-contaminated image with the residual update, and the result was projected to positive numbers by a rectified linear unit to enforce non-negativity.

Implementation and training of the residual U-Net was done in Python with TensorFlow.¹⁴ We minimized the 2-loss of the reconstructed volume to the desired ground truth. The training was done for 350 epochs with the adaptive moment estimation algorithm (ADAM),¹⁵ with an initial step size of

FIGURE 1 The 4 radial sampling patterns tested in this study. Fully sampled k-space requires 182 uniformly spaced radial spokes. An acceleration factor of 13 was used, resulting in 14 radial spokes acquired per frame, regardless of sampling pattern. REG_{no_rot} , regular angle scheme (equal angular spacing of $\sim 12.9^\circ$ between each profile) with no rotation between frames. REG_{rot} , regular angle scheme with regular rotation (of $\sim 0.99^\circ$) between frames so that 13 consecutive frames constitutes a fully sampled k-space. tGA_{no_rot} , tiny golden angle scheme with no rotation between frames. tGA_{rot} , continuously rotating tiny golden angle sampling



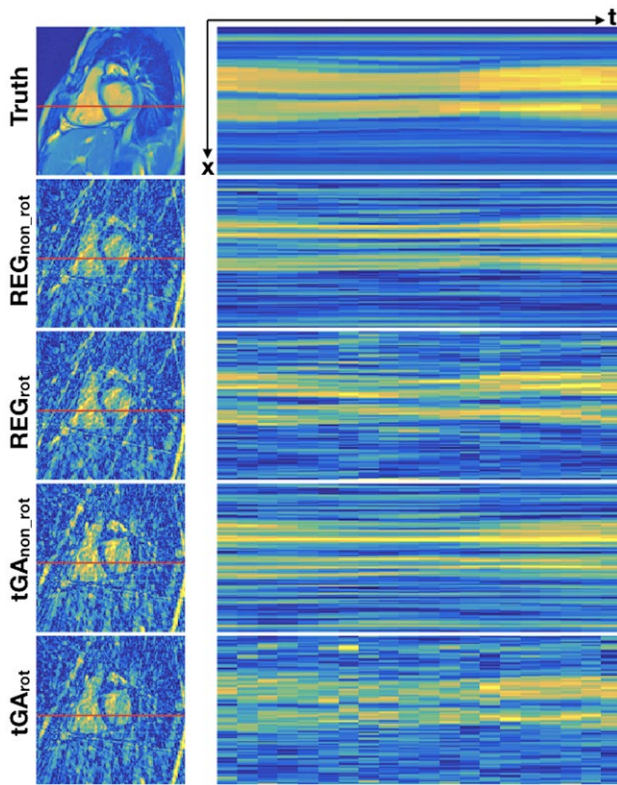


FIGURE 2 Typical artifact produced from the 4 radial sampling patterns tested in this study. Left: a single frame from the cine. Right: 1 line through the ventricles (shown in red) through time. REG_{no_rot} produces regular aliases with no variation through time. REG_{rot} produces regular aliases in space that change position over time. tGA_{no_rot} produces spatially incoherent aliases that do not vary through time. tGA_{rot} produces spatially incoherent aliases that vary through time

10^{-3} and batches of 8 samples. The total training time for each network took 12 h on a Titan XP GPU with 12 Gb memory.

2.5 | Evaluation of synthetic test data

The performance of the 4 separately trained networks (1 for each sampling pattern) were tested using a synthetic test data set. This data set was created from BH-bSSFP data, previously acquired in 25 children and adults with pediatric heart disease or CHD (mean age: 24.3 ± 13.3 y, range: 8–64 y, male: 13) that were not used to create the synthetic training data (full list of diagnoses in Supporting Information Text S1). Preparation of the synthetic test data was the same as for the training data. The artifact-contaminated magnitude data for each of the sampling patterns were input into their correspondingly trained networks. The output data were compared quantitatively and qualitatively to the ground truth images. Quantitative assessment was carried out using root-mean-square error (RMSE) and Structural Similarity Index (SSIM). The results of these analyses were averaged over the cardiac phases and slices for each patient. Qualitative assessment was performed using all cardiac phases from the middle slice of each patient. The output images for each sampling pattern were compared to the corresponding ground truth images by a CMR specialist (V.M., with 15 y experience) and evaluated for: (1) absence or deformation of papillary muscles or trabeculations, (2) poor motion fidelity, and (3) blurring of the endocardial border.

2.6 | Assessment of robustness

The chosen protocol parameters and the level of noise in the artifact-contaminated images may affect the performance

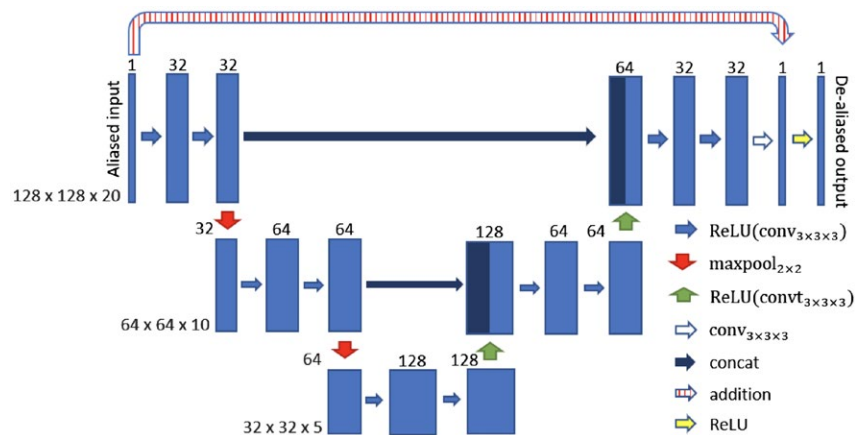


FIGURE 3 The chosen residual U-Net structure for spatio-temporal deep de-aliasing. The input is given by the aliased reconstruction from undersampled data. The numbers on top of the blue bars denote the number of channels for each layer. The resolution for each multilevel decomposition is shown in gray on the left. Each convolutional layer is equipped with a rectified linear unit as nonlinearity, given by $\text{ReLU}(x) = \max(x, 0)$

of the final trained network. The robustness of the resultant tGA_{rot} network was assessed by modifying the synthetic real-time test data, in terms of (1) SNR, (2) acceleration factor, and (3) image cropping.

The effect of SNR was assessed by artificially adding Gaussian noise to the artifact-contaminated tGA_{rot} synthetic test images, varying the resultant SNR between 20 dB to 10 dB. To assess the effect of acceleration factor, the synthetic real-time data was undersampled with acceleration from $10\times$ to $16\times$. The robustness to position of image cropping, was assessed by shifting the cropped region by $0/\pm 4/\pm 8/\pm 12$ pixels, in the x and/or y direction (giving 48 shifted regions for each subject). Resultant images were reconstructed using the previously trained network (tGA_{rot} with $13\times$ acceleration) and were compared to the ground truth images using RMSE and SSIM.

2.7 | Actual in vivo study

Actual real-time data was prospectively acquired in 10 children and adults with pediatric or CHD referred to our center for clinical CMR (mean age: 33.6 ± 16.8 y, range: 16–64 y, male: 3) (full list of diagnoses in Supporting Information Text S1). Exclusion criteria were (1) inability to breath-hold, (2) a single ventricle, and (3) arrhythmia. All patients were imaged on a 1.5T MR scanner (Avanto, Siemens Medical Solutions, Erlangen, Germany) with VCG gating. Real-time undersampled radial data was acquired, as described below. In addition, reference standard assessment of ventricular volumes was performed using standard BH-bSSFP cine imaging in the short axis (acquisition parameters same as parameters of the training data; matrix = $\sim 240 \times 204$, FOV = $\sim 320 \times 265$ mm, phase partial-Fourier = $6/8$, spatial resolution = $\sim 1.3 \times 1.3$ mm, slice thickness = ~ 10 mm, bandwidth = ~ 940 Hz/pixel, flip angle = $\sim 70^\circ$, TE/TR = $\sim 1.2/2.4$ ms, temporal resolution = ~ 32 ms, reconstructed cardiac phases = 40). Ten to 15 contiguous slices were acquired in the short axis to ensure coverage of the whole ventricular volume, with 1 slice acquired per breath hold (~ 6 s).

In addition, actual real-time data was prospectively acquired in 1 healthy volunteer to assess the effect of rapid or large respiratory motion on the performance of the final trained network. All imaging parameters were the same as above. Real-time data were acquired during breath-holding, normal breathing, and rapid, deep breathing.

2.7.1 | Real-time sequence

The new real-time sequence was based on an in-house single-shot multi-slice 2D radial bSSFP technique. Imaging parameters were: matrix = 192×192 , FOV = 320×320 mm, spatial resolution = 1.67×1.67 mm, slice thickness = 10 mm, bandwidth = 1240 Hz/pixel, flip angle = 67° , TE/TR

= $1.4/2.8$ ms. Consecutive spokes were separated by the tiny golden angle, and 14 spokes were used to reconstruct each frame (no temporal overlap between frames). This represented an undersampling factor of $13\times$, based on 182 spokes being required to fully sample k-space.¹⁶ The resultant temporal resolution was 36.4 ms. Acquisition of each slice required 2 R-R intervals; the first R-R interval being used to reach the steady state and the second for data acquisition. Ten to 15 contiguous slices were acquired in the short axis and after acquisition the imaging plane was moved to the next contiguous slice. Image acquisition was performed during free-breathing, using two spine coils and one body-matrix coil (giving a total of 12 coil elements).

2.7.2 | Reconstruction of real-time data

Creation of artifact-contaminated images from the undersampled radial data was performed online in Siemens reconstruction environment (ICE VB17, two AMD Opteron processors, clockspeed: 3.2 GHz, cache: 6 MB). The k-space data was gridded, Fourier-transformed into image space, followed by weighted sum over the coil elements. The resultant artifact-contaminated images were processed in MATLAB similarly to the test data; cropped to 128×128 matrix, followed by temporal interpolation to 20 frames. The resultant images subsequently had the artifact suppressed using the previously trained tGA_{rot} residual U-Net.

The undersampled radial data was also reconstructed using the CS algorithm, golden angle radial acquisition with temporal total variation sparsity regularization (GRASP).¹⁷ To provide an unbiased comparison in terms of image quality and reconstruction time, we used the Berkeley Advanced Reconstruction Toolbox (BART)¹⁸ as it supports parallel computation using GPUs. This toolbox uses the ADMM algorithm for temporal total variation constraints.¹⁹ The number of iterations was set to 50 as a good trade-off between reconstruction speed and reconstruction accuracy (determined from preliminary experiments described in Supporting Information Figure S2). Coil sensitivity information was estimated from the k-space centre.²⁰ The regularization level was selected empirically (to 0.025) based on the first 5 patients to minimize artifacts without producing too much temporal blurring. The GRASP algorithm was performed on the same computer as the residual U-Net (Titan XP GPU with 12 Gb memory).

2.8 | Analysis of in vivo data

Image quality and ventricular volumes were compared between the gold-standard BH-bSSFP images and the real-time radial data reconstructed with the GRASP algorithm and the deep artifact suppression residual U-Net.

2.8.1 | Image quality

Qualitative image scoring was performed by a CMR specialist (V.M.). For all patients, the mid-ventricular short-axis cine images from each technique were viewed in a random order. Each cine loop was graded on a 5-point Likert scale in 4 categories: sharpness of endocardial border (1=non-diagnostic, 2=poor, 3=adequate, 4=good, 5=excellent), temporal fidelity (or blurring) of wall motion (1=non-diagnostic, 2=poor, 3=adequate, 4=good, 5=excellent), and residual artifacts (1=non-diagnostic, 2=severe, 3=moderate, 4=mild, 5=minimal).

Quantitative edge sharpness was calculated by measuring the maximum gradient of the normalized pixel intensities across the border of the septum, as previously described.²¹ To reduce noise, which results in artificially high gradients (representing sharp edges), the pixel intensities were fit to a tenth order polynomial, before differentiation. Edge sharpness was calculated in 6 positions across the septum for all cardiac phases, and the average value was used for comparison.

2.8.2 | Ventricular function

Quantification of left and right ventricular volumes was calculated for each technique. The end-diastolic and end-systolic phases for each ventricle were chosen through visual inspection of the mid-ventricular cine. At end diastole and systole, all ventricular sections were manually segmented by a CMR specialist (V.M.). The endocardial border was traced using the OsiriX open source DICOM viewing platform (Osirix v.9.0, OsiriX foundation, Switzerland).²² The papillary muscles and trabeculae were included in the blood-pool.

End-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF) were calculated as previously described.²³

2.9 | Statistics

Statistical analyses were performed by using STATA software (STATA SE, v.14.2). Qualitative image scores for synthetic test data are expressed as percentage of cases where the artifact was visible. Comparisons of continuous variables was performed using one-way repeated measures analysis of variance (ANOVA) with post hoc testing using Bonferroni correction for significant results. Comparison of ratio data (image scoring of synthetic test data) was performed with the chi-squared test and comparison of ordinal data (image scoring of clinical data) was performed using the Wilcoxon signed-rank test. For assessment of agreement of ventricular volumes and function, the BH-bSSFP data was used as the reference standard for Bland-Altman analysis. A *P*-value of <0.05 indicated a significant difference.

3 | RESULTS

3.1 | Synthetic test data

Figure 4 shows image quality of the synthetic test data for the different sampling patterns (corresponding movie can be seen in Supporting Information Video S2). Table 1 shows the RMSE and SSIM results for the different radial sampling patterns. The tGA_{rot} had significantly lower RMSE (better reconstruction accuracy) than all of the other sampling patterns (*P* < 0.0001) and significantly higher SSIM (better reconstruction accuracy) than all the

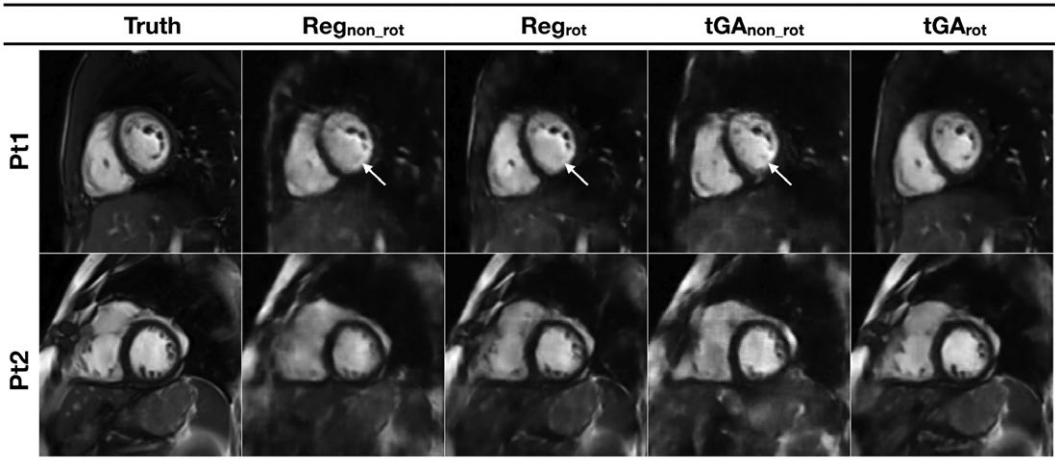


FIGURE 4 Examples images from 2 “synthetic real-time” test data sets reconstructed from the 4 different sampling patterns, using correspondingly trained residual U-Net’s. Pt1 shows total loss of some papillary muscles (indicated with arrow) compared to the truth data, with all trajectories except tGA_{rot}. Pt2 shows blurring across the ventricle, particularly seen with the non-rotating trajectories (see Supporting Information Video S2 for corresponding movie)

TABLE 1 Assessment of different radial sampling patterns

	REG _{no_rot}	REG _{rot}	tGA _{no_rot}	tGA _{rot}
RMSE ($\times 10^{-2}$)	8.0 ± 1.5^a	4.9 ± 1.0^a	8.4 ± 1.6^a	4.2 ± 1.3
SSIM	0.64 ± 0.04^a	0.83 ± 0.03^a	0.63 ± 0.05^a	0.87 ± 0.03
Pap (%)	92 ^a	20 ^a	88 ^a	0
Motion (%)	48 ^a	24 ^a	48 ^a	4
Blur (%)	100 ^a	68	100 ^a	52

Quantitate measures, RMSE and SSIM calculated with the “truth” images. Qualitative measures, percentage of case with visible artifacts. Pap, absence or deformation of papillary muscles or trabeculations. Motion, poor motion fidelity. Blur, blurring of the endocardial border.

^aValues are significantly different from tGA_{rot} ($P < 0.05$).

other sampling patterns ($P < 0.0001$). Qualitative measures of the images from the different sampling patterns are also shown in Table 1. The tGA_{rot} had significantly lower proportions of absent/deformed papillary/trabeculations, poor motion fidelity, and blurred endocardial borders compared to tGA_{no_rot} and REG_{no_rot} ($P < 0.0001$). In addition, tGA_{rot} had significantly lower proportions of absent/deformed papillary/trabeculations ($P = 0.018$) and poor motion fidelity ($P = 0.045$) than REG_{no_rot}.

3.1.1 | Assessment of robustness

For the tGA_{rot} network, as SNR decreased from 20 dB to 10 dB, the SSIM decreased by 6.8% (0.862 at 20 dB, 0.794 at 10 dB, and 0.868 with no additional noise), and the RMSE increased by 1.3% (0.043 at 20 dB, 0.056 at 10 dB, and 0.042 with no additional noise) (see Supporting Information Figure S3). As acceleration factor increased from 10 \times to 16 \times , SSIM decreased by to 6% (0.89 at 10 \times and 0.83 at 16 \times) and RMSE increased by 0.9% (0.041 at 10 \times and 0.050 at 16 \times) (see Supporting Information Figure S4). The 13 \times undersampled data appeared to perform better than this trend, probably because the network was trained on 13 \times undersampled data. Altering the position of image cropping had little effect on SSIM and RMSE. The lowest SSIM was 0.858 at a shift of +12 pixels in x and y, and was 1.04% less than with no shift applied. The highest RMSE was 4.53×10^{-2} also at a shift of +12 pixels in x and y, and was 0.34% greater than with no shift applied. See Supporting Information Figure S5 for all results.

3.2 | Prospective in vivo study

Real-time data was successfully acquired and reconstructed (GRASP and residual U-Net) in all 10 patients. On average, there were ~12 slices per patient, resulting in 122 3D data sets. Reference standard BH-bSSFP data was also successfully acquired in all patients. Total acquisition time for BH-bSSFP stack was 279 ± 65 s, and for the real-time radial bSSFP, it was 18 ± 3 s.

3.2.1 | Reconstruction time

Overall reconstruction times for all slices (~12 slices, from raw data file until completion of reconstruction) was ~5.6 \times faster for deep artifact suppression compared to compressed sensing (~22.0 s vs. ~111.4 s).

The compressed sensing reconstruction took ~3.4 s to read the raw k-space data file (containing all phases, for all slices, with 12 coil elements), and the GRASP algorithm (including calculation of trajectories and coil sensitivity information) took ~8.9 s per slice.

For the deep artefact suppression, creation of artifact-contaminated DICOM images took ~14.8 s for all slices (~1.3 s per slice, including reading of the raw k-space data file, gridding of data, coil combination, and saving to DICOM). Deep artifact suppression of these images took ~7.2 s for all slices, including ~5.4 s to read all DICOM files (all phases, for all slices), before initialization of variables for each slice and running the residual U-Net (~0.15 s per slice).

3.2.2 | Image quality

Representative images are shown in Figure 5 and the corresponding movie can be seen in Supporting Information Video S3. The residual U-Net reconstruction of the real-time data had better image quality than the same data reconstructed using GRASP. In particular, there is still significant residual aliasing seen in the GRASP reconstruction, as well as greater temporal blurring (as seen in the x-t plots of Figure 5). There are some subtle differences between the real-time reconstructions and the BH-bSSFP images (i.e., reduced pericardial fat). These are probably because of minor slice position variations between the free-breathing and breath-holding acquisitions. These differences between the U-Net and GRASP reconstructions are reflected in the qualitative image scoring (Table 2), with the residual U-Net reconstruction having similar myocardial delineation, motion fidelity, and artifact scores to the BH-bSSFP images and superior scores to the

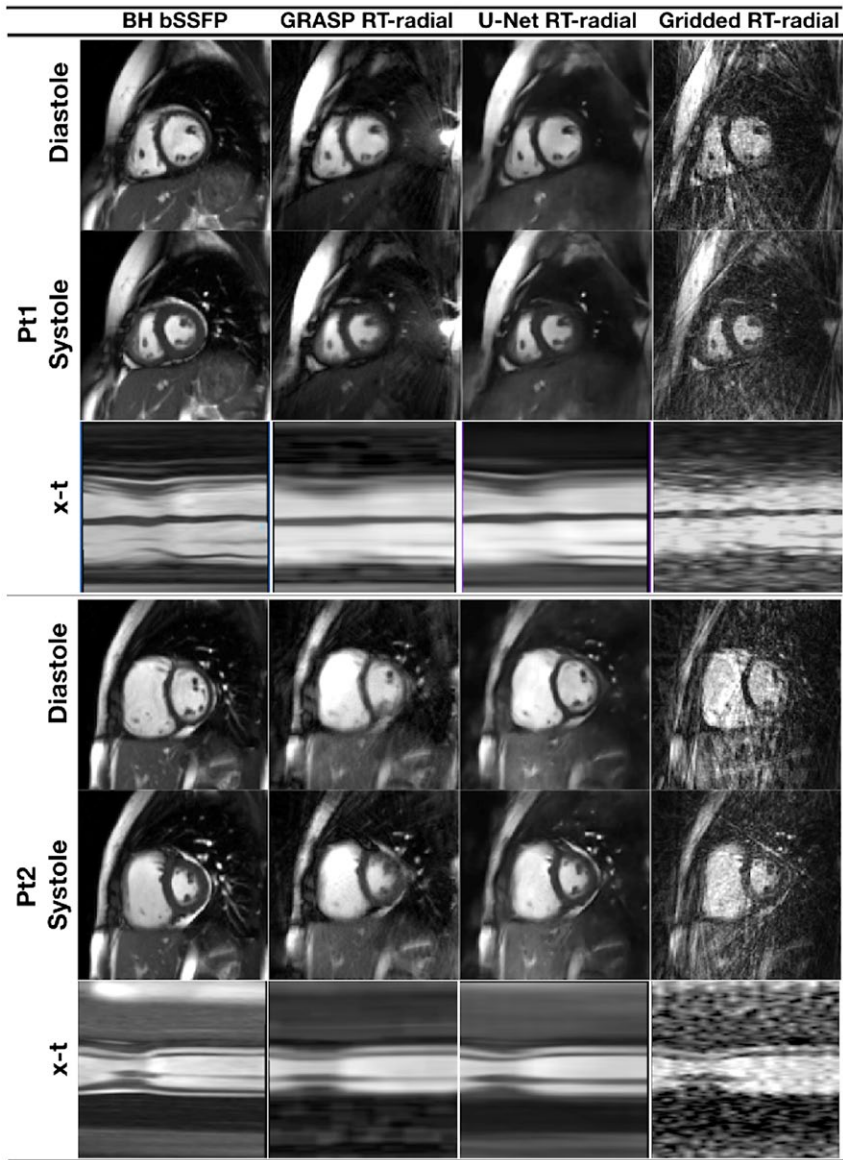


FIGURE 5 Example image quality in 2 prospective patients, shown in peak systole and peak diastole, as well as an x-t plot, from the BH-bSSFP sequence and the real-time radial sequence reconstructed with GRASP and the residual U-Net (see Supporting Information Video S3 for corresponding movie)

GRASP reconstructed data ($P < 0.05$). Quantitative edge sharpness (Table 2) was higher in the BH-bSSFP images, followed by the residual U-Net reconstruction and then GRASP reconstruction, with all comparisons being significant ($P < 0.007$).

Figure 6 shows images from the single volunteer under different respiratory conditions and the corresponding movie can be seen in Supporting Information Video S4. It can be seen that the U-Net reconstruction appears robust to increasing respiratory motion. This is in contrast to the GRASP reconstruction that is compromised during deep rapid breathing, particularly at end diastole.

3.2.3 | Ventricular volume quantification

Biventricular volumes and function are shown in Table 3 (with all Bland-Altman graphs for the LV in Supporting Information Figure S6 and for the RV in Supporting

TABLE 2 Quantitative and qualitative image quality results from prospective in vivo data comparing the BH-bSSFP sequence to the real-time radial sequence with GRASP and residual U-Net reconstructions

	BH-bSSFP	RT GRASP	RT U-Net
ES	0.66 ± 0.15	$0.38 \pm 0.08^{a,b}$	0.55 ± 0.14^a
Edge	4.7 ± 0.45	$3.4 \pm 0.49^{a,b}$	4.2 ± 0.75
Motion	4.8 ± 0.40	$3.4 \pm 0.49^{a,b}$	4.3 ± 0.46
Artifact	4.9 ± 0.30	$3.8 \pm 0.40^{a,b}$	4.6 ± 0.49

Quantitative image scoring: edge sharpness (ES) results. Qualitative image scoring (5-point Likert scale): edge, sharpness of endocardial border; motion, temporal fidelity (or blurring) of wall motion; artifact, residual artifacts.

^aValues are significantly different from BH-bSSFP ($P < 0.05$).

^bValues are significantly different from RT U-Net ($P < 0.05$).

Information Figure S7). The limits of agreement for both the residual U-Net and GRASP reconstructions were similar. However, there were several significant

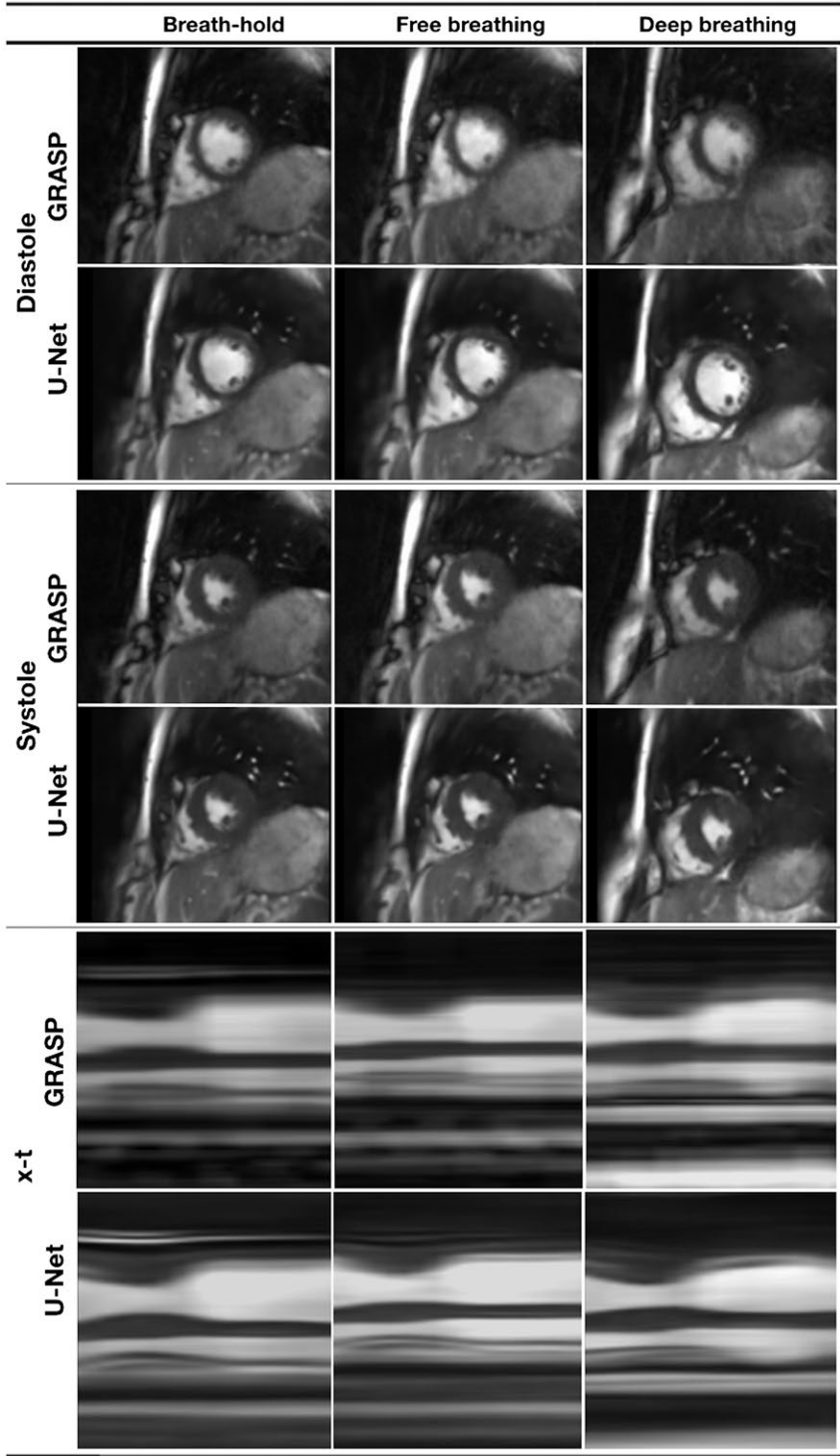


FIGURE 6 Image quality in a single volunteer from the real-time radial sequence reconstructed with GRASP and the residual U-Net under different respiratory conditions (breath-holding, free-breathing and deep, rapid breathing) (see Supporting Information Video S4 for corresponding movie)

biases with the GRASP reconstructed data including left ventricular ESV ($P = 0.04$) and EF ($P < 0.001$) and right ventricular EDV ($P < 0.001$). The residual U-Net reconstructed data was only associated with a significant bias for right ventricular EDV ($P = 0.004$), which was still less than the GRASP data ($P = 0.036$). The Bland-Altman graphs for LV ESV, LV EF, and RV EDV are shown in Figure 7.

4 | DISCUSSION

The main findings of this study included (1) it was possible to train a 3D residual U-Net to remove artifacts (deep artifact suppression) in synthetic radially undersampled data created from previously acquired Cartesian BH-SSFP images, (2) the ability to perform deep artifact suppression on synthetic test data was determined by the simulated undersampling pattern,

TABLE 3 Ventricular volumes from prospective in vivo data comparing the BH-bSSFP sequence, as well as the real-time radial sequence, with GRASP and residual U-Net reconstructions

	Mean \pm SD			Bias (limits of agreement)	
	BH-bSSFP	RT GRASP	RT U-Net	RT GRASP	RT U-Net
LV EDV (mL)	148 \pm 44	143 \pm 44	151 \pm 46	−4.6 (−19.2 to 10.0)	3.0 (−11.9 to 17.9)
LV ESV (mL)	56 \pm 27	60 \pm 29 ^a	58 \pm 29	4.1 (−8.0 to 16.2)	1.9 (−7.9 to 11.7)
LV EF (%)	64 \pm 10	60 \pm 11 ^{a,b}	63 \pm 11	−4.1 (−9.5 to 1.3)	−0.3 (−4.1 to 3.5)
RV EDV (mL)	213 \pm 97	198 \pm 89 ^{a,b}	204 \pm 92	−14.9 (−33.4 to 3.6)	−8.6 (−23.2 to 6.0)
RV ESV (mL)	92 \pm 49	89 \pm 48	91 \pm 47 ^a	−3.0 (−16.0 to 10.0)	−1.1 (−22.3 to 20.1)
RV EF (%)	58 \pm 7	57 \pm 6	57 \pm 6	−1.5 (−9.1 to 6.1)	−1.4 (−9.5 to 6.7)

Abbreviations: LV, left ventricular volume; RV, right ventricular volume; EDV, end diastolic volume; ESV, end systolic volume; EF, ejection fraction.

^aValues are significantly different from BH-bSSFP ($P < 0.05$).

^bValues are significantly different from RT U-Net ($P < 0.05$).

(3) deep artifact suppression of actual radial real-time data acquired in new patients was successful with a residual U-Net trained using synthetic data, (4) total reconstruction times were significantly shorter ($>5\times$) using a residual U-Net compared to a conventional CS reconstruction, and (5) image quality and measurement of ventricular volumes was superior using the residual U-Net compared to CS reconstructions of the same raw data.

4.1 | Deep artifact suppression and synthetic training data

Machine learning, especially deep learning, relies on large amount of training data. Most clinical CMR services have large amounts of images stored from previously scanned patients. However, there are 2 problems with using this type of data to train deep learning algorithms capable of reconstructing real-time data. The first is that the CMR data is usually stored as magnitude images rather than complex raw data. Simple magnitude data is not suitable for training the iterative deep learning methods that have previously been used for CMR reconstruction.^{7,8,24–26} Therefore, in this study, we chose to use a residual U-Net to remove aliasing from artefact-contaminated gridded images. This approach has previously been shown to be successful in removing aliasing in retrospectively undersampled dynamic Cartesian CMR data.⁶ We have built on this previous work by using a much larger training data set and implementing non-Cartesian undersampling. The second problem with using previously acquired data is that it does not consist of paired ground truth and undersampled real-time data. Therefore, retrospectively cardiac-gated, Cartesian, breath-hold bSSFP images were used to create both synthetic training and test data. The advantage of using synthetic data is that the ground truth is known. The disadvantage of using synthetic data is that errors during acquisition (such as gradient delays or trajectory errors) and respiratory motion (because of the free-breathing acquisition

for real-time imaging) cannot be corrected for. Nevertheless, we were able to show that this synthetic data could be used to train a residual U-Net to successfully perform deep artifact suppression.

Interestingly, the reconstruction accuracy of the synthetic test data was highly dependent on the radial undersampling pattern used. Non-rotating trajectories, where the same spokes were used in all frames, resulted in significant quantitative errors (as measured by RMSE and SSIM) and qualitative errors. Specifically, the images reconstructed for non-rotating trajectories were associated with missing or deformed papillary muscles and/or trabeculations in over 88% of synthetic test cases. Furthermore, the non-rotating trajectories also produced images with the worst motion fidelity and blurring of the endocardial border. In contrast, the continuously rotating strategies performed better, probably because of aliases having a less coherent structure through time. In particular, tGA_{rot} produced aliases that appeared noise-like and we believe this is the reason for the superior results seen with this sampling pattern. A full description of the association between the aliasing patterns and reconstruction accuracy are outside the scope of this article. Nevertheless, the superior quality of tGA_{rot} sampling for the synthetic test data confirmed that it was the right choice for the actual in vivo part of this study.

4.2 | In vivo study

We demonstrated clinical translation of the residual U-Net by successfully performing deep artifact suppression on actual real-time images acquired in new patients. Importantly, because a tGA_{rot} pattern was chosen, it was possible to reconstruct the same raw data using a compressed sensing algorithm (i.e., the GRASP technique). One of the hypothetical advantages of a deep learning approach over CS is quicker reconstruction times. In this study, we found that residual U-Net reconstruction was $>5\times$ quicker than the GRASP reconstruction, when

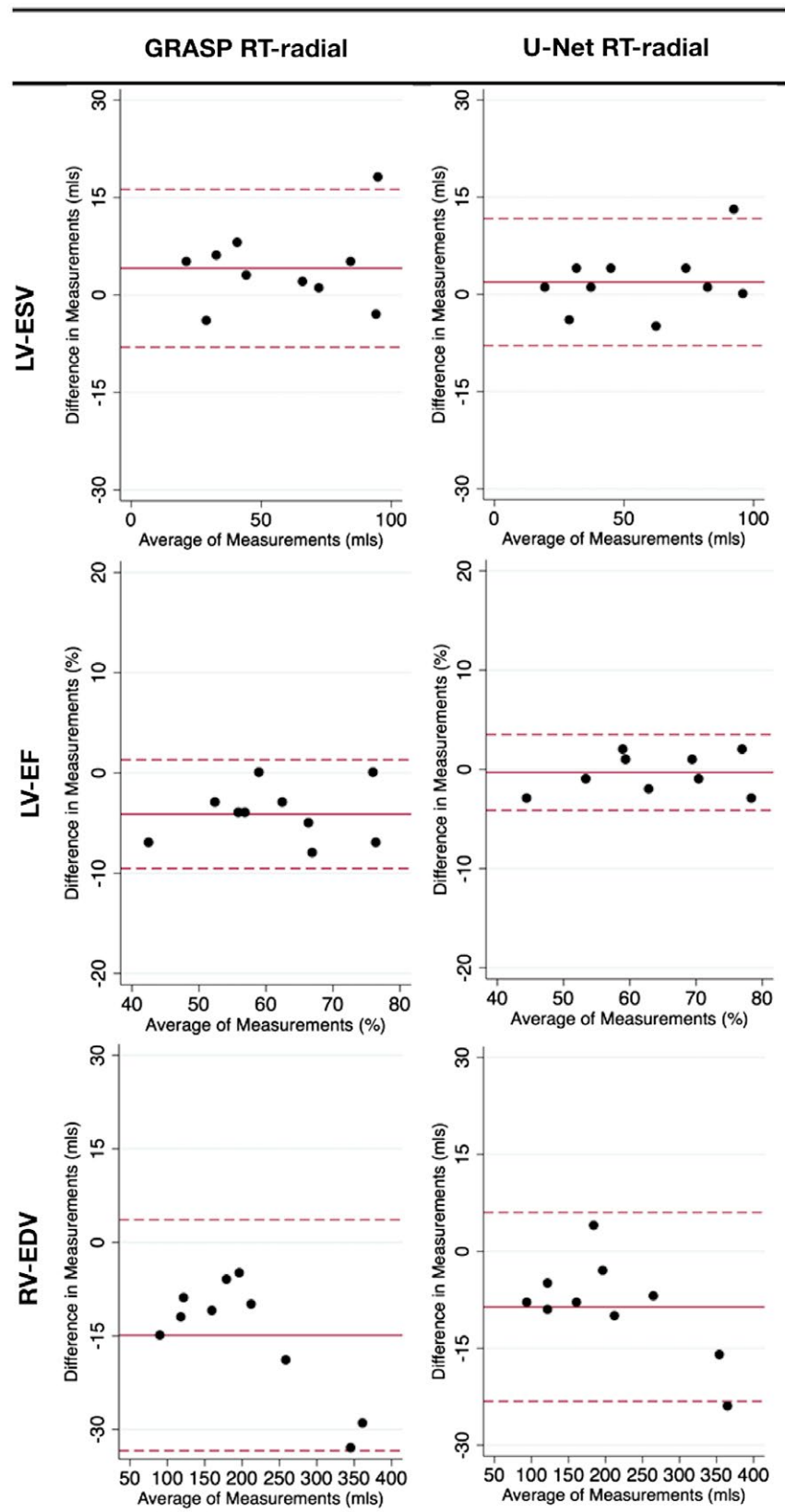


FIGURE 7 Bland-Altman plots showing ventricular volume results with significant biases. Comparison of ventricular volumes acquired with BH-bSSFP to real-time data reconstructed with GRASP and residual U-Net. Solid red line shows mean difference, dashed lines shows ± 2 SD

both were run on the same GPU. It should be noted that the reconstruction times for GRASP are almost linearly dependent on the number of iterations. Therefore, reconstruction time could be easily shortened by simply reducing the number of

iterations. However, this would reduce image quality of the GRASP reconstruction, which is already lower than the residual U-Net reconstruction was performed on cropped images, however we would not expect this to have significant

effect on the U-Net reconstruction. In addition, the U-Net residual U-Net reconstruction was performed on cropped images, however we would not expect this to have significant effect on the evaluation time. Consequently, we believe that by accelerating reconstruction, deep learning approaches could play an important role in clinical translation of heavily under-sampled real-time techniques.

In this study, we also found that the residual U-Net reconstructed images provided more accurate quantification of ventricular volumes compared to GRASP reconstructed images. The reason for this was probably better image quality, as reflected by better qualitative scores for endocardial border sharpness, temporal fidelity of wall motion, and residual artifacts. We believe that the superior image quality of the U-Net reconstruction was the result of using a high quality, large volume training data set. Conversely, the poorer CS reconstructions could be because of sub-optimal choice of the regularization parameter. However, the presence of both residual artifacts and poor edge sharpness and/or temporal fidelity suggests that changing regularization would be of little benefit. In fact, the bias and limits of agreement of the GRASP reconstructed data were better than previous studies⁴ using the same technique. This suggests that the GRASP implementation was adequately optimized. Therefore, we believe that the residual U-Net reconstruction is superior and may allow more accurate measurement of ventricular volumes.

4.3 | Study limitations

The main limitation for this study is that the training data was acquired during a breath-hold, while the real-time data was acquired during free-breathing. This means that the network cannot “learn” anything about respiratory motion and may not be able to correct for this additional motion. However, we did perform an experiment in a single volunteer that demonstrated that the U-Net reconstruction was relatively robust to even large respiratory perturbations. This suggests that the network learns the structure of the noise to be removed, rather than the expected motion of the underlying structures. Nevertheless, further work is necessary to fully understand the effect of motion not apparent in the training data.

In this study, we also did not investigate the full diversity of CHD. Specifically, no patients with single ventricles were included in the training data set or the in vivo study. Even though this first study shows that the trained network generalizes well, the full generalization capability of the network to varying motion and anatomies should be investigated in future work.

Another drawback of our implementation of a deep artifact suppression network was that all the training and subsequent data needed to be of the same size. However, when acquiring real-time data over 1 cardiac cycle, the number of frames is dependent on the patient's heart rate. This meant

that it was necessary to resample the aliased images through time, altering the temporal aliasing pattern of the acquired data. However, as both the training and actual real-time data were processed in the same way, we do not believe that it is a significant limitation.

A further limitation of our approach was that the training and actual input data consisted of coil-combined magnitude images, rather than raw multi-coil complex data. The main benefit of this approach was that previously acquired data that was easily retrievable from a conventional clinical image archive could be used for training. However, the absence of phase data in our approach may prevent optimum image restoration, in particular dealing with signal cancellation because of aliasing. Although this was not noticeable in the in vivo data, this issue warrants further investigation in future studies. In addition, we did not investigate the effect of dark band artefacts in either the training data or actual input data on image quality. This study was performed at 1.5T, and dark band artefacts were not a significant problem. Nevertheless, this issue must be investigated if our approach is to be considered for higher field strengths.

Finally, in this study, we choose to minimize the 2-loss and do not fully explore other loss functions such as 1-loss or SSIM-loss. The 1-loss function has been shown to provide superior training or improved results in image restoration,²⁷ super resolution,²⁸ and MR image reconstruction.²⁹ Therefore, we performed preliminary tests comparing the 1-loss and 2-loss using the tGA_{rot} trajectory. Assessment of image quality from the trained networks using the synthetic test data, described above, demonstrated <0.01% difference in RMSE and <0.5% difference in SSIM (further information in Supporting Information Figure S8). This is similar to previous deep learning studies, which have shown limited differences in RMSE and SSIM values using 1-loss or SSIM-loss functions, however, with improved visual quality compared to 2-loss.^{6,30} Nevertheless, further investigation of different loss functions and other optimizations of learning would be desirable in the future.

5 | CONCLUSION

This article demonstrates the potential of using a residual U-Net for deep artifact suppression of real-time radial data within a clinical setting. Once the networks have been trained, the reconstruction times are very short, making these techniques particularly appealing within busy clinical workflow. We have shown that ventricular volumes measured from images reconstructed using a residual U-Net are not statistically significantly different from the reference standard, cardiac gated, BH techniques. Therefore, we believe that this technique may help with full adoption of real-time CMR in clinical practice. This would be particularly useful in children and sick patients who are unable to hold their breath.

REFERENCES

1. Grothues F, Smith GC, Moon J, 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;90:29–34.
2. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med.* 2007;58:1182–1195.
3. Kido T, Kido T, Nakamura M, et al. Assessment of left ventricular function and mass on free-breathing compressed sensing real-time cine imaging. *Circ J.* 2017;81:1463–1468.
4. 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;79:2745–2751.
5. Wundrak S, Paul J, Ulrici J, Hell E, Rasche V. A small surrogate for the golden angle in time-resolved radial MRI based on generalized Fibonacci sequences. *IEEE Trans Med Imaging.* 2015;34:1262–1269.
6. Sandino CM, Dixit N, Cheng JY, Vasanawala SS. Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging. In Proceedings of Medical Imaging meets Neural Information Processing Systems Conference, Long Beach, CA, 2017. p. 19.
7. Yang G, Yu S, Dong H, Slabaugh G, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging.* 2018;37:1310–1321.
8. Qin C, Schlemper J, Caballero J, Price A, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *arXiv preprint arXiv* 2017:171201751.
9. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015. p. 234–241.
10. Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process.* 2017;26:4509–4522.
11. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys.* 2017;44:e360–e375.
12. 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;80:1189–1205.
13. Chan RW, Ramsay EA, Cheung EY, Plewes DB. The influence of radial undersampling schemes on compressed sensing reconstruction in breast MRI. *Magn Reson Med.* 2012;67:363–377.
14. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. In Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, Savannah, GA, 2016. p. 265–283.
15. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv* 2014:1412.6980.
16. Larson P, Gurney PT, Nishimura DG. Anisotropic field-of-views in radial imaging. *IEEE Trans Med Imaging.* 2008;27:47–57.
17. 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.
18. Tamir JI, Ong F, Cheng JY, 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.
19. Afonso MV, Bioucas-Dias JM, Figueiredo MA. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans Image Process.* 2011;20:681–695.
20. McKenzie CA, Yeh EN, Ohliger MA, Price MD, Sodickson DK. Self-calibrating parallel imaging with automatic coil sensitivity extraction. *Magn Reson Med.* 2002;47:529–538.
21. Steeden JA, Atkinson D, Hansen MS, Taylor AM, Muthurangu V. Rapid flow assessment of congenital heart disease with high-spatiotemporal-resolution gated spiral phase-contrast MR imaging. *Radiology.* 2011;260:79–87.
22. Rosset A, Spadola L, Ratib O. OsiriX: an open-source software for navigating in multidimensional DICOM images. *J Digit Imaging.* 2004;17:205–216.
23. Muthurangu V, Lurz P, Critchely J, Deanfield J, Taylor A, Hansen MS. Real-time assessment of right and left ventricular volumes and function in patients with congenital heart disease by using high spatiotemporal resolution radial k-t SENSE. *Radiology.* 2008;248:782–791.
24. Schlemper J, Caballero J, Hajnal JV, Price A, Rueckert D. A deep cascade of convolutional neural networks for MR image reconstruction. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap PT, Shen D, editors. Information processing in medical imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25–30, 2017, Proceedings. Cham: Springer International Publishing; 2017. p. 647–658.
25. 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.
26. Yu S, Dong H, Yang G, Slabaugh G, et al. Deep de-aliasing for fast compressive sensing MRI. *arXiv preprint arXiv* 2017:170507137.
27. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging.* 2017;3:47–57.
28. Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, 2017. p. 3.
29. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial networks for compressed sensing automates MRI. *arXiv preprint arXiv.* 2017:170600051.
30. Hammernik K, Knoll F, Sodickson DK, Pock T L2 or not L2: impact of loss function design for deep learning MRI reconstruction. In Proceedings of the 25th Annual Meeting of ISMRM, Honolulu, HI, 2017. Abstract 0687.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the on-line version of this article.

FIGURE S1 Flow diagram showing the steps taken to convert the retrospectively cardiac-gated, BH-bSSFP data to synthetic real-time data used to train the residual U-Net

FIGURE S2 Effect of number of iterations on GRASP reconstruction speed and GRASP reconstruction accuracy. Preliminary experiments showing (A) average reconstruction time per slice for different number of iterations (red line shows average time per slice for 50 iterations, as used in this study), (B) resultant image quality in 1 patient for different numbers of iterations, (C) plot of difference in resultant images between different numbers of iterations (red line shows difference in images between 50–60 iterations), and (D) difference images between numbers of iterations

FIGURE S3 Assessment of robustness: effect of SNR. (A) Average RMSE and SSIM results across all 222 test data sets at different levels of SNR. (B) Plot of RMSE at different SNR levels. (C) Plot of SSIM at different SNR levels. (B and C) Red line shows reference RMSE/SSIM with no additional noise. (D) Example of image quality from 1 patient with input data from different SNR levels

FIGURE S4 Assessment of robustness: effect of acceleration. (A) Average RMSE and SSIM results across all 222 test data sets at different acceleration factors. (B) Plot of RMSE at different acceleration factors. (C) Plot of SSIM at different acceleration factors. (B and C) Red line shows reference RMSE/SSIM at 13× (this is the acceleration factor the network was trained with). (D) Example of image quality from 1 patient

FIGURE S5 Assessment of robustness: effect of shifting the cropping region. (A) Average RMSE at different cropping positions in x and y across all 222 test data sets. (B) Plot of RMSE at different a cropping positions. (C) Average SSIM at different cropping positions in x and y across all 222 test data sets. (D) Plot of SSIM at different cropping positions. (E) Example of image quality from 1 patient

FIGURE S6 Bland-Altman plot comparing left ventricular volumes acquired with BH-bSSFP to real-time data reconstructed with GRASP and residual U-Net. Solid red line

shows mean difference, dashed lines shows ± 2 standard deviation

FIGURE S7 Bland-Altman plot comparing right ventricular volumes acquired with BH-bSSFP to real-time data reconstructed with GRASP and residual U-Net. Solid red line shows mean difference, dashed lines shows ± 2 SD

FIGURE S8 Images showing difference related to training with L1 and L2 cost function, and table showing RMSE and SSIM compared to the ground truth

VIDEO S1 Movie showing residual flickering artifact when a 2D U-Net is applied to dynamic data. Left: images with residual aliasing used as input to U-Net. Right: output from 2D U-Net. The flickering artifact is not seen with a 3D U-Net (2D plus time), because it is used in this study

VIDEO S2 Example movies from 2 “synthetic real-time” data sets reconstructed from the 4 different sampling patterns using correspondingly trained residual U-Nets. Pt1 shows total loss of some papillary muscles compared to the truth data, with all trajectories except tGA_{rot} . Pt2 shows blurring across the ventricle, particularly seen with the non-rotating trajectories

VIDEO S3 Example movies from 2 prospective patients from the BH-bSSFP sequence and the real-time radial sequence reconstructed with GRASP and the residual U-Net

VIDEO S4 Example movies from single volunteer acquired during different respiratory maneuvers: (left) breath-hold, (middle) normal breathing, (right) deep rapid breathing. Top row is GRASP reconstruction, and the bottom row is the U-net reconstruction

TEXT S1 Diagnosis of patients used for training of the residual U-Net, creating of the synthetic test data, and for the prospective in-vivo study

How to cite this article: 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. <https://doi.org/10.1002/mrm.27480>