

A data-driven semantic segmentation model for direct cardiac functional analysis based on undersampled radial MR cine series

Tobias Wech¹  | Markus Johannes Ankenbrand^{2,3}  | Thorsten Alexander Bley¹ | Julius Frederik Heidenreich¹ 

¹Department of Diagnostic and Interventional Radiology, University Hospital Würzburg, Würzburg, Germany

²Comprehensive Heart Failure Center, University Hospital Würzburg, Würzburg, Germany

³Center for Computational and Theoretical Biology, University of Würzburg, Würzburg, Germany

Correspondence

Tobias Wech, Department of Diagnostic and Interventional Radiology, University Hospital Würzburg, Oberdürrbacher Straße 6, 97080 Würzburg, Germany.
Email: wech_t@ukw.de

Funding information

The German Federal Ministry of Education and Research (BMBF grant no. 05M20WKA)

Purpose: Image acquisition and subsequent manual analysis of cardiac cine MRI is time-consuming. The purpose of this study was to train and evaluate a 3D artificial neural network for semantic segmentation of radially undersampled cardiac MRI to accelerate both scan time and postprocessing.

Methods: A database of Cartesian short-axis MR images of the heart (148,500 images, 484 examinations) was assembled from an openly accessible database and radial undersampling was simulated. A 3D U-Net architecture was pre-trained for segmentation of undersampled spatiotemporal cine MRI. Transfer learning was then performed using samples from a second database, comprising 108 non-Cartesian radial cine series of the midventricular myocardium to optimize the performance for authentic data. The performance was evaluated for different levels of undersampling by the Dice similarity coefficient (DSC) with respect to reference labels, as well as by deriving ventricular volumes and myocardial masses.

Results: Without transfer learning, the pretrained model performed moderately on true radial data [maximum number of projections tested, $P = 196$; $DSC = 0.87$ (left ventricle), $DSC = 0.76$ (myocardium), and $DSC = 0.64$ (right ventricle)]. After transfer learning with authentic data, the predictions achieved human level even for high undersampling rates ($P = 33$, $DSC = 0.95$, 0.87 , and 0.93) without significant difference compared with segmentations derived from fully sampled data.

Conclusion: A 3D U-Net architecture can be used for semantic segmentation of radially undersampled cine acquisitions, achieving a performance comparable with human experts in fully sampled data. This approach can jointly

This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

accelerate time-consuming cine image acquisition and cumbersome manual image analysis.

KEYWORDS

cardiovascular magnetic resonance (CMR), deep learning, radial, semantic segmentation, undersampling

1 | INTRODUCTION

Cardiovascular magnetic resonance (CMR) imaging has been established as a noninvasive reference standard for diagnosis and management of cardiac diseases in clinical practice. The assessment of ventricular structure and function is an essential part of CMR. While left ventricular (LV) function is assessed in almost any imaging protocol independently of the underlying cardiac disease, specific diseases such as arrhythmogenic ventricular cardiomyopathy may also require assessment of right ventricular (RV) function.¹ Imaging is being performed in short axis orientation from the base of the left ventricle to the apex. Using a slice thickness of 6 to 8 mm, imaging therefore requires a significant amount of time. Ultimately, functional parameters such as end-diastolic volume/end-systolic volume (EDV/ESV), ejection fraction (EF), and myocardial mass are derived depending on the clinical need.²

For the acquisition and reconstruction part, a variety of applications of CMR have excellent preconditions for an acceleration by exploiting model-based imaging techniques like compressed sensing.³ The dynamic nature of, for example, cine or perfusion imaging leads to high redundancy in the temporal domain of image series, which can be modeled within a constraint of an iterative reconstruction. This has been used for accelerating procedures⁴⁻⁹ with respect to classical approaches that are ignoring these characteristics. In recent years, corresponding reconstruction approaches exploiting data-driven acceleration gained more importance.¹⁰⁻¹² All these methods exploit the fact that the aimed depiction is highly compressible and thus can be fully determined with less parameters than usually acquired during a fully sampled cine CMR exam, which suggests that an appropriately undersampled and accelerated image acquisition might carry sufficient information for adequate assessment of the ventricular structure and function.

Regarding postprocessing, databases, comprising cine CMR data alongside respective segmentation labels set by experienced operators, were used to train artificial neural networks to perform semantic segmentation fully automatically with the same accuracy as human operators.¹³⁻¹⁹

If now moderately undersampled, non-Cartesian sampling patterns like radial projections or spiral arms are used

for cine imaging, it becomes apparent that the human eye is usually capable of distinguishing between the anatomy (and its dynamics) and the superimposed undersampling artifacts. Especially if the sub-Nyquist patterns are varying throughout cine frames, dynamic depictions clearly reveal the contours of interest and relevant organ borders. This motivates deep learning-based approaches,^{20,21} which circumvent costly image reconstructions and directly estimate the volumes on undersampled data. In this article, we describe a 3D artificial neural network that was trained for semantic segmentation of the LV, the RV, and the myocardium with radially undersampled cine data as input. Radial acquisitions were first simulated based on a larger data set holding Cartesian cine images²² and used for a pretraining of the model (see Figure 1). For translation into a realistic setting, the network was fine-tuned using a smaller data set with genuine radial cine MRI data²³ and ultimately applied to unseen native radial cine MRI data with different levels of undersampling. The approach

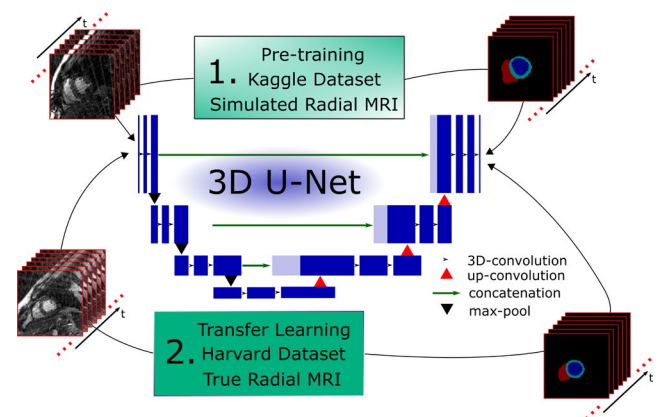


FIGURE 1 Schematic workflow for the presented training of a segmentation model. A, 3D neural network with U-Net architecture was pretrained using Cartesian cine images from the publicly available Kaggle data, which were radially undersampled by simulation, using 2D short axis slices from base to apex and the temporal domain as third dimension, respectively. B, Transfer learning was then performed using radially acquired cine images in midventricular short axis orientation from the Harvard data set, which were reconstructed with different rates of undersampling. The final network was ultimately evaluated using radial cine images from the Harvard data set, which were not used for training of the network before

aims at a fast evaluation of cardiac functional data, which were acquired in short scan times. Although this accelerates the typically lengthy and thus uncomfortable investigation; it also shortens the cumbersome and costly post processing, the manual image segmentation, which is applied to determine cardiac function so far. All codes and data to reproduce the results of this study are publicly available at https://github.com/expRad/Segment_Under sampled_cMRI.

2 | METHODS

2.1 | Assembly of database and simulation of radial data

2.1.1 | The Kaggle data

The publicly available Data Science Bowl Cardiac Challenge Data by the online community Kaggle (referred to as Kaggle data) provided 1140 cardiac cine MRI examinations,²⁴ which were complemented by numerical volume information of end-systolic and end-diastolic ventricular volume in the aftermath of the competition. No “ground truth” segmentation labels were provided. Images were stored in DICOM format resolving slices from base to apex with 30 phases of the cardiac cycle. Examinations were performed both on 1.5T and 3.0T systems (Siemens Magnetom Aera and Skyra) using FLASH and balanced steady-state free precession (bSSFP) sequences with a Cartesian acquisition trajectory. An overview of the complete data set and its variation in patient data and sequence parameters is given in Table 1.

2.1.2 | Data curation and creation of ground-truth segmentation labels

The Kaggle data are a compilation of real, clinical data from multiple sites and as such, subject to inconsistencies within individual examinations. A data curation was performed as described by Ankenbrand et al,²² to remove these inconsistencies as much as possible. More detailed information and examples regarding data curation can be found in this online repository: <https://github.com/chfc-cmi/cmr-seg-tl#data-curation-and-conversion>.

For the Kaggle data, no expert ground truth segmentation was available. Bai et al,¹³ however, developed and published a segmentation model based on a neural network, which achieved a performance similar to human experts.¹³ To prepare labels for our study, this openly available model (https://github.com/baiwenjia/ukbb_cardiac) was applied to each individual 2D image of the curated

TABLE 1 Data composition and measurement parameters of the Kaggle data

Metric	Count
Sex	
Male	670
Female	470
Age	
0-17 y	202
18-30 y	173
31-50 y	298
51+ y	467
Max age, y	88
Min age, y	0.04
1.5T	1025
3.0T	115
Metric	Range
Echo time, ms	1.04-1.54
Repetition time, ms	14-54.72
Bandwidth, Hz/pixel	915-1235
Slice thickness, mm	5-8
Matrix size	120-608 × 160-736
Resolution, mm	0.59-1.95
Phases	112-416

Kaggle data, each representing one cardiac phase of one slice in one subject. The created label comprised information about RV blood pool, LV blood pool, and LV myocardium. ESV and EDV were derived from the obtained segmentation and compared with the real ground-truth values provided by Kaggle. The accuracy of the network predictions was determined, and a confidence set was created where the predicted values were in the range $\pm 15\%$ of the true value. The final data set comprised 148,500 images from 484 examinations. All scores (label vs ground truth) for ESV and EDV values are listed in the above-mentioned online repository. A limited manual validation of the labels obtained by Bai’s network is available at: https://github.com/chfc-cmi/cmr-seg-tl/blob/master/code/kaggle/compare_manual_contours.ipynb.

2.1.3 | Simulation of radial raw data

To simulate cine series based on undersampled radial k-space trajectories, the images and labels were first organized in 3D matrices of matching x-y-t series (#4950 = 148,500 images/30 frames per slice). Each real-valued image series was superimposed with complex valued simulated coil sensitivities (#12, complex valued), determined by using Biot-Savart’s law,²⁵ which

was initialized by coil geometries, randomly chosen within typical ranges. The resulting 4D matrices (x - y - t coils) were Fourier transformed along the spatial domains (k_x - k_y - t coils) and each k_x - k_y space was masked by a pseudoradial pattern on the Cartesian grid (see Figure 2).

The radial trajectory was chosen as it results in comparably subtle artifacts incoherently superimposing the image in the case of undersampling. For each sample of the training data, the number of projections per frame (t) was randomly chosen from the set $P \in \{21, 34, 55, 89, 144, 233, 377\}$ during run time, leading to a random undersampling rate for each individual sample (varying throughout different epochs of the training). The number of projections was thus equal for all frames of one sample; however, the pattern was rotated by a randomly chosen angle for each subsequent frame to increase the incoherency of the resulting undersampling artifacts. The angle increment between projections within one frame was set to the “golden angle”²⁶; therefore, the elements of P were chosen as Fibonacci numbers. The obtained series were transformed back to image space (x - y - t coils) and combined via a root mean sum of squares along the coil dimension. The image spaces of the resulting matrix I consequently represented reconstructions of undersampled k -spaces: measurements violating the Nyquist-sampling criterion.

An exemplary image (x - y) created with 55 projections can be seen in Figure 2C.

2.2 | The 3D neural network architecture

A 3D U-Net as described by Çiçek et al²⁷ was implemented in MATLAB (Deep Learning Toolbox, v2020a; MathWorks) and trained for semantic segmentation of radial undersampled cine series. The input of the network was represented by the simulated cine series as described in the previous section. Subsequent to the simulation of radial undersampling, all real-valued x - y - t matrices I were cropped to a size of $160 \times 160 \times 24$ to provide a consistent input for the neural network. The t domain was cropped to 24 frames by discarding the final frames, which represent the highly redundant end-diastolic phases. The x - y domain of the cropped matrices was reviewed manually to confirm that trimming of the heart was avoided.

The network consisted of 55 layers (19,073,860 trainable parameters) with two blocks in each stage. One block featured a 3D convolution layers with a kernel size of $3 \times 3 \times 3$, followed by batch normalization and a rectified linear unit. The network was ultimately concluded by a Dice pixel classification layer, which provides a series of

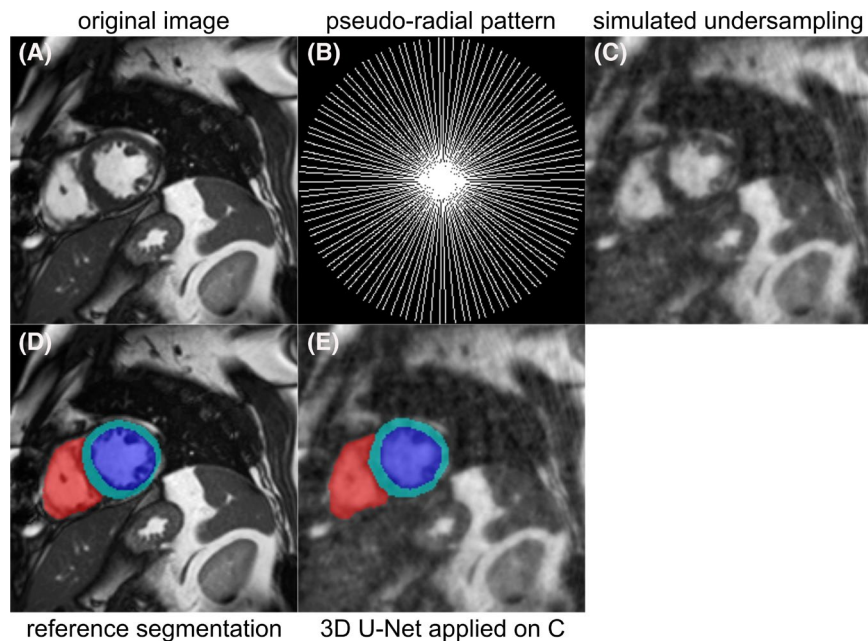


FIGURE 2 Simulation of radially undersampled data from fully sampled Cartesian data: The Kaggle data comprised 148,500 fully sampled cardiac cine images in short axis orientation acquired with Cartesian trajectories. One frame of one slice is depicted in (A). A 3D matrix was formed from matching x - y - t series. Images were superimposed with artificial coil sensitivities and Fourier transformed along the spatial domains. A pseudoradial pattern (B) was used to mask the Cartesian grid with a random number of projections as described in the Methods section. C, Shows the result of applying the simulated undersampling to (A). The 3D U-Net was pretrained for segmentation of the left ventricle (dark blue), the myocardium (light blue), and the right ventricle (red) using the undersampled reconstructions. (E) Shows the result of applying the trained network to (C) in comparison with the ground truth in (D)

segmentation masks \mathcal{S} (same size as \mathcal{I}), with labels for background, LV, RV, and myocardium. The output layer is based on the generalized Dice loss^{28,29} that weights the contribution of classes inversely to their volume and was chosen to suppress the influence of large background regions during training. The network architecture and training is illustrated in Figure 1. The magnitude MR input images were normalized by subtracting the mean value of all pixels and dividing by the SD.

2.3 | Pretraining of a 3D U-Net for segmentation of x-y-t series of cine MRI

The 3D U-Net was first pretrained using a major part of the simulated radial acquisitions as described in subsection 2.1. For every epoch, 4300 x-y-t inputs \mathcal{I} were processed as follows: For each individual \mathcal{I} , a random number of projections was chosen from P “on-the-fly,” that is a new undersampling pattern was applied each time the data set was loaded in each epoch. In addition, a different, randomly chosen set of coil sensitivities was superimposed as described in subsection 2.1. As the k-space trajectory and the coil sensitivities thus vary throughout different epochs, the data set was virtually augmented. Further data augmentation incorporated random applications of rotations of 90°, horizontal or vertical flips, or combinations of the aforementioned.

The simulated radially undersampled data set from subsection 2.1 was separated in 4300 cine series for training, 150 for validation, and 500 for testing. This represented a patient leave-out training: Slices in the test set were not from any patient used in the training set. Adaptive moment estimation (Adam) was used as optimizer of the loss function.³⁰ The initial learning rate was set to 1e-4; training was performed for 10 epochs until the loss of the training data reached a minimum without further decrease, and validation loss did not indicate overfitting. For epoch 6-10, the learning rate was reduced to 0.8e-4.

For testing, the trained model was subsequently applied to 500 hitherto unseen cine series. Basal and apical slices, which are not showing parts relevant for the quantification of cardiac function, were excluded by an expert radiologist (leaving 380 series for quantification). A separate segmentation was determined for each series using simulations with varying number of projections P as used within the training. The performance of the pretrained 3D U-Net was evaluated by determination of the Dice similarity coefficient (DSC) in comparison with the “ground truth” labels obtained from Cartesian fully sampled data by the model of Bai et al¹³ as described in subsection 2.1. EFs were derived for each patient and undersampling

factor, and were subsequently also compared with the ground truth.

2.4 | Transfer learning with truly radially sampled acquisitions from the Harvard data

To fine-tune the obtained model towards a realistic application, the pretrained network from subsection 2.3 was subjected to transfer learning using a second data set comprising radially sampled cine acquisitions. In conjunction with their publication on deep learning-based image reconstruction in cardiac MRI,³¹ the authors kindly made their MR raw data (multicoil, complex valued) openly available (<https://doi.org/10.7910/DVN/CI3WB6> referred to as Harvard data). In contrast to the Kaggle data, where radial projections were simulated, these data were actually acquired by a radial bSSFP sequence at 3T, and thus represent even more realistic samples for the aimed method. The data set comprised 108 cine series acquired in breath-hold, each in one midventricular slice orientation. Further acquisition parameters were reported as follows: TR = 3.1 ms, TE = 1.4 ms, in-plane resolution = 1.8 mm × 1.8 mm, slice thickness = 8 mm, FA = 48°, number of channels = 16 ± 1. Retrospective ECG-triggering was used to determine 25 cardiac phases in a segmented fashion, each consisting of 196 linearly ordered projections. A total scan time of 14 heartbeats on average was reported.

2.4.1 | Label creation (ground truth) and data curation

Using the fully sampled radial data, segmentation labels were again automatically determined for each frame of each cine series using the model of Bai et al.¹³ Results were checked for artifacts and inadequate segmentation by a trained expert with 4 years of experience in cardiac MRI, resulting in a final data set of 83 subjects with matching cine series and label series. The data set was split up for training, validation, and testing (61, 5, and 17 subjects, respectively).

2.4.2 | Radially undersampled data

Ultimately, undersampled reconstructions were determined using only a part of the radial projections provided with each data set. P equidistant projections per frame were used with P corresponding to {196, 98, 49, 33, 25}, representing samples with different undersampling factors R . $P = 196$ was chosen as maximum because it

represented the maximum number of readouts per frame contained in the provided data set. For the matrix size of 160×160 , which was processed to the network, this represents a slight undersampling ($R = \frac{160 \cdot \pi}{196 \cdot 2} = 1.3$), which was, however, not impairing the centric circular region of interest depicting the heart. Accordingly, the acceleration factors R for $P = \{196, 98, 49, 33, 25\}$ correspond to $R = \{1.3, 2.6, 5.1, 7.6, 10.1\}$.

2.4.3 | Training and evaluation

For each specimen, respective magnitude reconstructions using the five different sampling rates (different number of radial projections P as indicated above) were determined and paired with the “ground truth” labels obtained from the fully sampled counterpart. Subsequently, all samples were pooled in a joint data store used for training. For transfer learning, the initial learning rate was adapted to 0.002. A learning-rate schedule with a drop factor of 0.5 every five epochs was applied. The model was trained for 50 epochs until the training loss reached a minimum.

2.5 | Statistical analysis

The final model was ultimately applied to the 17 test data sets and evaluated by determination of DSC in comparison with the ground-truth segmentations for LV, RV, and myocardium obtained by the model suggested by Bai et al from the fully sampled radial counterparts.¹³ The evaluation was performed separately for each level of undersampling. To test the impact of the transfer-learning step, the same analysis was performed using the model that was derived from the Kaggle data only. A one-way analysis of variance (ANOVA) was performed separately

for each compartment (LV, RV, myocardium), to compare the results for the different numbers of projections per frame. A p value $< .05$ was considered statistically significant. Box plots were generated to complement the results of the ANOVA.

In addition, “volumes” were derived from predicted and ground-truth labels for the midventricular slice in accordance with functional assessment typically performed in clinical routine for the entire heart. For each subject, ESVs and EDVs were assessed for the left and the right ventricle and myocardial mass, EFs were calculated.

3 | RESULTS

3.1 | Kaggle data: pretraining

The pretraining using the Kaggle data set took 43 h on an NVIDIA Titan XP graphics processing unit (GPU). Applying the model subsequently to test data took only 0.3 s for one spatiotemporal cine series. Therefore, the entire segmentation process for one patient can be performed within approximately 5 s. Figure 2 illustrates the fully sampled Cartesian data and the simulated undersampled radial data, exemplarily with 55 projections per frame. The ground truth for fully sampled Cartesian data obtained by the model of Bai et al¹³ is shown as an overlay in Figure 2D. The predicted segmentation for simulated undersampled data (Figure 2C) is exemplarily shown in Figure 2E.

Table 2 shows the mean DSC between predicted data for each level of undersampling compared with the ground truth. The predicted segmentations of the LV, the myocardium, and the RV agree well with the ground truth even with only 55 projections, whereas performance decreases slightly for 34 and more strongly when

TABLE 2 Performance of the pretrained 3D U-Net for segmentation of simulated radially undersampled cardiac cine images from the Kaggle data set

Projections P	Acceleration factor R	LV	Myocardium	RV
377	fully sampled	0.88	0.80	0.75
144	1.7	0.89	0.80	0.77
89	2.8	0.88	0.80	0.77
55	4.6	0.86	0.77	0.75
34	7.4	0.81	0.71	0.67
21	11.9	0.64	0.52	0.40

Note: Data are shown as mean Dice similarity coefficient DSC. For each different scheme of undersampling (number of projections) the performance of the network was compared with the ground truth labels determined by the model proposed by Bai et al¹³ to fully sampled Cartesian cine images. The performance was evaluated individually for left ventricle blood pool (LV), the left ventricle myocardium, and the right ventricle blood pool (RV). Factor of acceleration (R) is indicated. Cases without segments in both ground-truth and predicted segmentation were not counted. Test data were augmented as described for training and measures were averaged across five repetitions of the evaluation.

applying the network to a data set with only 21 projections. Additional Bland-Altman plots for evaluation of the LV EF with each level of acceleration are provided in Supporting Information Figure S1, which essentially confirm this trend.

3.2 | Harvard data: transfer learning

After transfer learning using the Harvard data (6 h on an NVIDIA Titan XP GPU), applying the resulting model to test data resulted in mean DSC as presented in Table 3. The numbers given in parentheses show the according values if the model after the pretraining was applied, that is the neural network which was trained using the simulated images based on the Kaggle data only. First, the DSC values are significantly higher for the model, which was additionally subjected to the transfer learning for all areas and acceleration factors. Scores for the LV, the myocardium, and the RV remain constantly high when reducing the number of projections from 196 (“fully sampled”) down to 49 per frame. Still for only 33 projections, just a slight drop in performance can be observed for all regions, which is then more pronounced when the number is further reduced down to 25.

The statistical test indicated no significant variation within the groups of different undersampling factors for the LV and the RV. For the segmentation of the myocardium, however, ANOVA showed a significant difference with a p value of .0062. Subjecting the results to a post hoc analysis (multicomparison test with Bonferroni correction), a significant difference resulted between $R = 10.1$ and the specimen of $R = 1.3, 2.6, 5.1$, respectively. Boxplots (Figure 3) confirm these results, whereas also the values obtained for 10.1-fold acceleration in the RV appear with lower scores; however, no significant difference was detected by the ANOVA ($p = .73$). In Figure 4, the results of one test example are presented for the end-systolic and the end-diastolic cardiac phase. In accordance with the results

from the statistical test, segmentations remain stable up to a 7.6-fold acceleration, whereas larger areas of false segmentation start to manifest for $R = 10.1$ (see red arrows). Dynamic views of two further examples can be observed in Supporting Information Videos S1 and S2. In addition, mean values for specificity, sensitivity, accuracy, and precision are presented in Supporting Information Table S3. With a fairly constant high specificity throughout all acceleration factors, sensitivity shows comparable trends as observed by the DSC.

Figure 5 shows Bland-Altman analyses of the EF determined in both ground-truth and segmentation results in undersampled cine data. The average agreements for the left ventricular EF between the model predictions and ground truth were high, up to an acceleration factor of $R = 5.1$, whereas a small, but statistically not significant difference was observed at $R = 7.6$ (coefficient of variation [CV] = 5.6 %, $p = .13$ according to the Wilcoxon signed rank test). With accelerations of $R = 10.1$, the predicted EF was acceptable but significantly different as compared with the ground truth (CV = 8.4 %, $p = .04$). Median relative errors between the accelerated acquisitions and ground truth were small for quantitative parameters of the left ventricle (see Table 4). Although predictions for the RV-derived parameters were in good agreement with the ground truth for accelerations up to $R = 5.1$, higher undersampling led to significantly impaired values. Finally, median relative errors of the LV myocardial mass were acceptable up to a 7.6-fold acceleration.

4 | DISCUSSION

In this study, we present a 3D U-Net trained for semantic segmentation of radially undersampled cardiac cine MRI. For pretraining of the network, a large, open database of Cartesian cine data was used to simulate radial cine acquisitions with different levels of undersampling. The pretrained network was subsequently subjected to

Projections P	Acceleration			
	factor R	LV	Myocardium	RV
196	1.3	0.97 (0.87)	0.90 (0.76)	0.94 (0.64)
98	2.6	0.97 (0.85)	0.90 (0.74)	0.94 (0.62)
49	5.1	0.96 (0.81)	0.89 (0.66)	0.94 (0.53)
33	7.6	0.95 (0.68)	0.87 (0.43)	0.93 (0.26)
25	10.1	0.93 (0.36)	0.80 (0.16)	0.91 (0.12)

Note: Data are shown as mean Dice similarity coefficient. The performance was tested for different undersampling schemes (numbers of projections) compared with the ground truth. The performance was evaluated individually for segmentation of the left ventricle blood pool (LV), the left ventricular myocardium, and the right ventricle blood pool (RV). The data in parentheses show the performance for segmentation of undersampled data when using only the pretrained network without transfer learning.

TABLE 3 Performance of the transfer-learned 3D U-Net for the segmentation of non-Cartesian cardiac cine images from the Harvard data reconstructed with different undersampling schemes

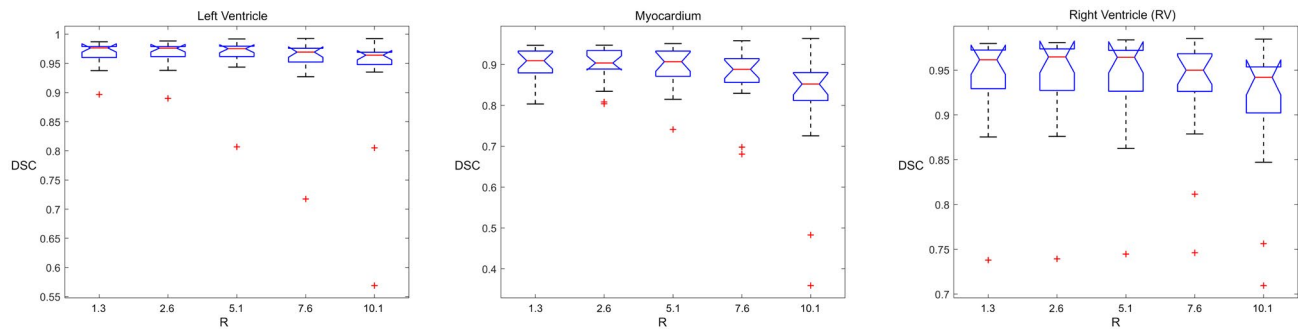


FIGURE 3 Box plots show the performance of the transfer-learned 3D U-Net for segmentation of radially undersampled cine images. Data are shown as Dice scores (DSC). Median is indicated by a red line within the box of the upper and lower quartile. Whiskers are indicated as black bars. Outliers are indicated as red plus signs. Data are shown individually for the left ventricle, the left ventricle myocardium, and the right ventricle. R represents the acceleration factor

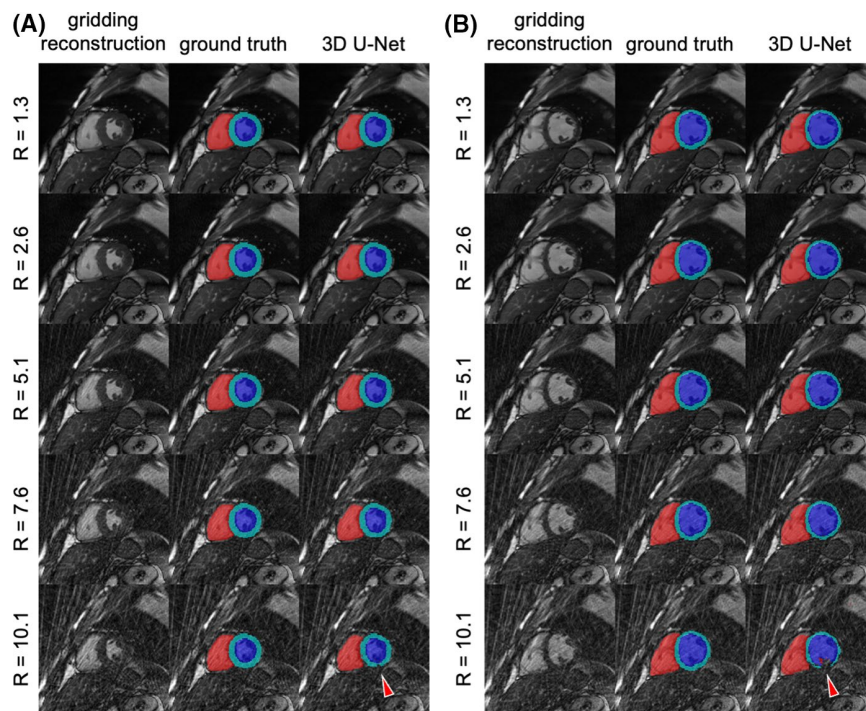


FIGURE 4 Performance of the transfer-learned 3D U-Net on test data on end-systolic (A) and end-diastolic (B) phases. Exemplary images for each scheme of undersampling (R = acceleration factor). For comparison, the ground truth determined by the model of Bai et al¹³ in fully sampled images is shown in the middle section. The model's prediction for left ventricle blood pool (blue), the left ventricle myocardium (light blue), and the right ventricle blood pool (red) are indicated. Note that the visual performance is well until $R = 7.6$. Performance is acceptable for undersampling with $R = 10.1$, where segmentation errors mainly occur in the myocardium (arrow). Additional movie files representing all frames of two further examples are Supporting Information Videos S1 and S2

transfer learning using the raw data of an open database of radially sampled cine data, again for different levels of undersampling. Ultimately, the final model was evaluated by means of a separate set of 17 subjects, which resulted in overall high DSC scores for the segmentation of LV, RV and myocardial tissue, which were stable even for only 33 projections per frame. Slice-specific “volumes” and “masses” confirmed this minimum number of readouts for the quantification of left ventricular function with only slightly increased

differences in EF compared with reconstructions from lower R (see Figure 5).

This means a clear acceleration with respect to the procedures usually performed in clinical routine, where Cartesian trajectories are used, which typically require 150 to 250 readouts to obey the Nyquist criterion. As assumed, the neural network was able to learn the human skill of “seeing through the artifacts” and discriminating the endo- and epicardial contours in dynamic cine series when radial undersampling strategies are used. Routine

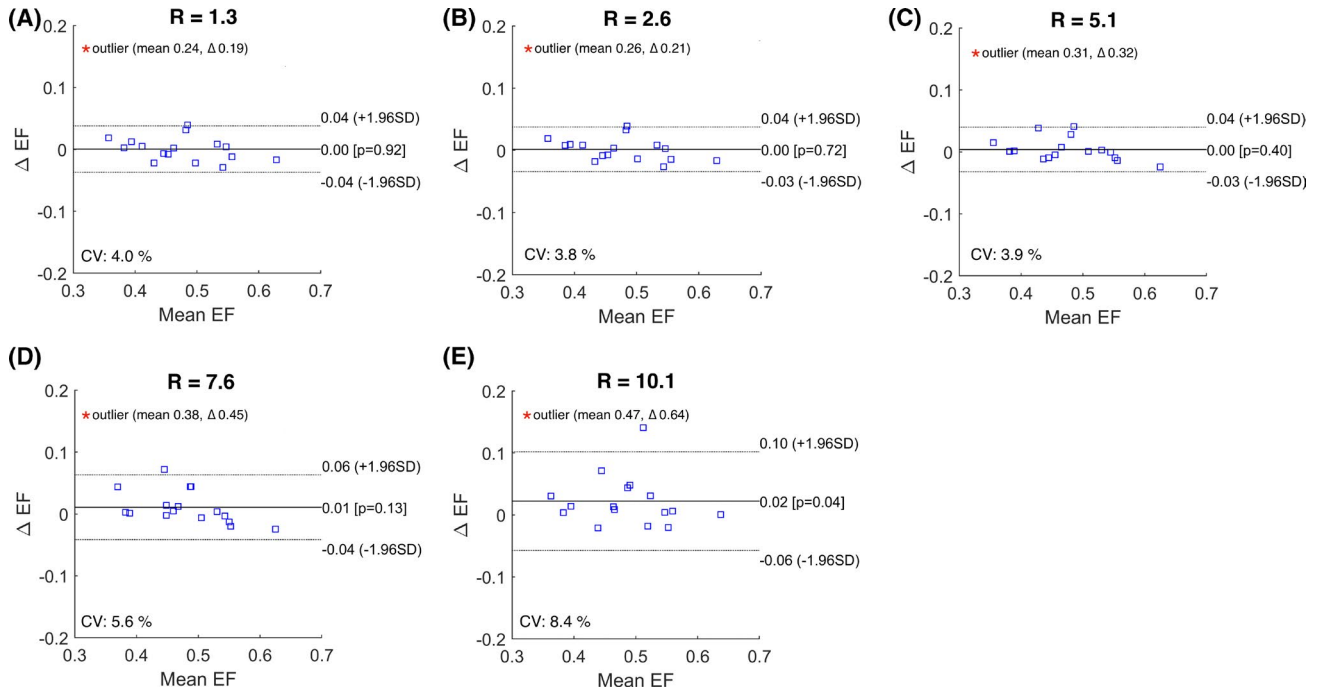


FIGURE 5 A-D, Bland-Altman analyses show high average agreement of ejection fractions (EFs) between the model predictions and ground truth up to 7.6-fold acceleration. For each undersampling scheme (R = acceleration factor) data are shown per subject for the “left ventricular ejection fraction” of the midventricular slice. Average agreement, 95% CI (1.96 SD), and coefficient of variation are shown. Statistical significance is indicated by p value (Wilcoxon signed rank test). The model failed to predict left ventricular volumes adequately in one subject; this specimen is marked by a red asterisk and was excluded from statistical analysis. Respective mean and Δ of the EFs are indicated

TABLE 4 Relative errors for cardiac functional analysis as compared with the ground truth

R	LV			Myocardium	RV		
	ESV	EDV	EF	mass	ESV	EDV	EF
1.3	4.1 (6.1)	3.5 (4.2)	2.7 (3.7)	8.9 (8.7)	4.0 (12.0)	3.1 (5.0)	6.1 (21.5)
2.6	2.9 (5.2)	3.4 (4.3)	2.6 (3.1)	9.0 (9.8)	3.4 (10.9)	2.9 (4.8)	5.1 (22.0)
5.1	2.2 (3.4)	2.4 (4.1)	2.1 (4.2)	7.5 (9.7)	4.4 (9.5)	1.9 (6.9)	7.6 (21.0)
7.6	5.6 (5.5)	2.8 (3.7)	2.6 (8.8)	6.7 (11.1)	3.2 (10.5)	3.3 (8.0)	10.2 (13.2)
10.1	6.1 (5.2)	3.3 (3.7)	3.6 (8.0)	11.5 (21.4)	9.8 (11.3)	5.4 (6.3)	11.2 (27.9)

Note: Data are shown as median and interquartile range (parentheses) of the relative errors in % for each acceleration factor R as compared to the ground truth. Abbreviations: EDV, end-diastolic volume; EF, ejection fraction; ESV, end-systolic volume; LV, left ventricle; mass, myocardial mass; RV, right ventricle.

postprocessing for the analysis of ventricular structure and function conventionally, in contrast, relies on manual segmentation by a trained expert, which is time-consuming and prone to subjective errors. Application of the 3D U-Net took only 0.3 s per cine series.

Previous studies have already demonstrated automated models for the segmentation of cardiac MRI, even with performances on the human level.¹³ The majority of these models, however, relies on fully sampled data and thus requires lengthy data acquisition or an additional sophisticated method to first reconstruct undersampled data (eg, compressed sensing). Only a few publications

report on directly training the model for the segmentation of undersampled data.^{20,21} These studies, however, only use Cartesian or simulated non-Cartesian data, which is an issue when transferring the model to real applications such as applying the model to data that was indeed acquired by radial projections. The latter point has been specifically addressed in our study: The pretrained model, which was trained with simulated radial data only (Kaggle), was applied to the MRI data in fact acquired by non-Cartesian radial projections (see values in parentheses in Table 3). Significantly decreased segmentation performances and corresponding DSC scores were observed,

hindering adequate quantification of the cardiac function. For translation into routine imaging, the pretrained model was fine-tuned using a comparatively small set of MRI data acquired with a non-Cartesian radial trajectory, which was ultimately tested by means of unseen radial cardiac cine series, reconstructed with different under-sampling schemes.

Considering the Bland-Altman analyses of the derived LV EF from radial cine MRI (see Figure 5), a 7.6-fold acceleration appears applicable. To be noted, the performance was higher for parameters of the LV (ESV, EDV) compared with the RV, which might be explained by a more heterogeneous appearance of the right ventricular shape. Considering these slightly lower performances, an acceleration factor of 5.1 would be the method of choice for combined quantitative analysis of LV, RV, and myocardial mass.

Although the pretrained model (Kaggle data) did not perform sufficiently on the genuine radial data (Harvard), the transfer learning with a comparatively small data set ultimately allowed predictions with high performance. Reversely, transfer learning of the model significantly decreased its performance when applied to the simulated data (see Supporting Information Table S2); however, when considering the different contrast and different appearance between simulated and genuine radial data, this is not surprising.

4.1 | Limitations

Even though transfer learning the model achieved high performance levels, the real non-Cartesian data set is still relatively small, and all data sets were acquired at a single site on a single scanner. This is a clear disadvantage with respect to the generalization ability of the model. Nevertheless, the focus of this work was to translate the principle of semantic segmentation to an authentic under-sampled data set and to prove the relevance of this step, and not the presentation of an ultimate model for clinical application.

Furthermore, the transfer learning and evaluation of the neural network was only performed on real data of midventricular slices, as the public database used unfortunately does not comprise the whole heart and no other open databases with radially sampled cine series were known to us. As can be deduced from the results of the pretraining, the overall performance is typically best at the midventricular slice, where the image presentation is most homogeneous without open contours, such as at the base and partial volume effects like at the apex. Thus, more training data are certainly necessary to introduce the presented method into clinical routine.

The neural network architecture used in our method (U-Net) represents a robust and, to a certain extent, established model. Nevertheless, a variety of new semantic segmentation techniques is presented continuously,^{19,32} such that there might already exist an even more suitable approach for the presented task. In particular, automatic machine-learning methods, for example, as presented in Isensee et al,³³ have proven to render most of the cumbersome manual adjustments, which are still necessary, obsolete; furthermore, they show potential to push the performance based on an improved goodness of the underlying model. This very general approach is undoubtedly also attractive for the further advancement of the method presented in this study. Furthermore, a 4D architecture capturing all dimensions of the cine exam at once (x-y-z-t) can potentially model the overall information even more efficiently¹¹; however, this also places high demands on the GPU memory.

5 | CONCLUSION

A 3D U-Net can be trained to accurately perform semantic segmentation in radially undersampled cine acquisitions. The final model was able to predict the common cardiac functional parameters up to 5.1-fold accelerated acquisition and up to 7.6-fold acceleration when deriving parameters from the LV only. This circumvents costly image reconstructions and enables fast postprocessing (0.3 s per cine series) immediately after accelerated and thus fast cine acquisitions. A combination of pretraining with a larger number of simulated data and subsequent transfer learning using data, which were actually sampled by radial trajectories, resulted in high DSC for realistic data sets, thereby not only proving the opportunity of eased clinical workflows, but also demonstrating the high importance of data authenticity when exploiting deep learning.

ACKNOWLEDGMENTS

The project underlying this report was funded by the German Federal Ministry of Education and Research (BMBF grant no. 05M20WKA). We thank Wenjia Bai and coauthors,¹³ and Hossam El-Rewaidy and coauthors,³¹ as well as the Kaggle community for providing their data for further scientific studies. We further thank M. Guerquin-Kern for providing open-source code for the simulation of MR coil sensitivities. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The Department of Radiology receives a research grant from Siemens Healthcare. The grant is not specifically directed towards any of the authors. The authors declare


that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

All source codes and data paired with detailed instructions of how to reproduce the results of this work are publicly available at: https://github.com/expRad/Segment_undersampled_cMRI. The Kaggle data originated from an openly accessible data set from the Data Science Bowl Challenge Data 2016. It can be accessed via the following link: <https://www.kaggle.com/c/second-annual-data-science-bowl/data>. The herein used data were accessed on July 29, 2019.²⁴ As the Kaggle data are subject to inconsistencies, we performed a data curation as described by Ankenbrand et al.²² Detailed information and examples regarding data curation can be found in this online repository: <https://github.com/chfc-cmi/cmr-seg-tl#data-curation-and-conversion>. The number of examinations in the repository within the 15% confidence set is higher than in this article. This is because of an update introducing an additional normalization of the Kaggle data before applying the model presented by Bai et al,¹³ which ultimately improved the overall performance. The Harvard data origins from an openly accessible data set from the Harvard dataverse, provided by Hossam El-Rewaady²³ and can be accessed via <https://doi.org/10.7910/DVN/CI3WB6>.

ORCID

Tobias Wech  <https://orcid.org/0000-0002-2813-7100>

Markus Johannes Ankenbrand  <https://orcid.org/0000-0002-6620-807X>

Julius Frederik Heidenreich  <https://orcid.org/0000-0002-0683-7982>

REFERENCES

- Kramer CM, Barkhausen J, Bucciarelli-Ducci C, Flamm SD, Kim RJ, Nagel E. Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. *J Cardiovasc Magn Reson*. 2020;22:17.
- Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post processing in cardiovascular magnetic resonance: society for cardiovascular magnetic resonance (SCMR) board of trustees task force on standardized post processing. *J Cardiovasc Magn Reson*. 2013;15. <https://doi.org/10.1186/1532-429X-15-35>
- Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182-1195.
- Gamper U, Boesiger P, Kozierke S. Compressed sensing in dynamic MRI. *Magn Reson Med*. 2008;59:365-373.
- Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. *Magn Reson Med*. 2016;75:775-788.
- Frahm J, Voit D, Uecker M. Real-time magnetic resonance imaging: radial gradient-echo sequences with nonlinear inverse reconstruction. *Invest Radiol*. 2019;54:757-766.
- Ong F, Zhu X, Cheng JY, et al. Extreme MRI: large-scale volumetric dynamic imaging from continuous non-gated acquisitions. *Magn Reson Med*. 2020;84:1763-1780.
- Eirich P, Wech T, Heidenreich JF, et al. Cardiac real-time MRI using a pre-emphasized spiral acquisition based on the gradient system transfer function. *Magn Reson Med*. 2021;85:2747-2760.
- Küstner T, Bustin A, Jaubert O, et al. Isotropic 3D Cartesian single breath-hold CINE MRI with multi-bin patch-based low-rank reconstruction. *Magn Reson Med*. 2020;84:2018-2033.
- Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2018;37:491-503.
- Küstner T, Fuin N, Hammernik K, et al. CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Sci Rep*. 2020;10. <https://doi.org/10.1038/s41598-020-70551-8>
- Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2019;38:280-290.
- Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson BioMed Central*. 2018;20:65.
- Tao Q, Yan W, Wang Y, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. *Radiology*. 2019;290:81-88.
- Tran PV. A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI. arXiv. 2017. 1604.00494.
- Baumgartner CF, Koch LM, Pollefeys M, Konukoglu E. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2018;10663:111-119.
- Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med Image Anal*. 2019;51:21-45.
- Fahmy AS, El-Rewaady H, Nezafat M, Nakamori S, Nezafat R. Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks. *J Cardiovasc Magn Reson*. 2019;21.
- Chen C, Qin C, Qiu H, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. 2020;7. <https://doi.org/10.3389/fcvm.2020.00025>
- Schlemper J, Oktay O, Bai W, et al. Cardiac MR segmentation from undersampled k-space using deep latent representation learning. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2018;11070:259-267.
- Qin C, Bai W, Schlemper J, et al. Joint motion estimation and segmentation from undersampled cardiac MR image. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2018;11074:55-63.
- Ankenbrand MJ, Lohr D, Schlötelburg W, Reiter T, Wech T, Schreiber LM. Deep learning-based cardiac cine segmentation:

- transfer learning application to 7T ultrahigh-field MRI. *Magn Reson Med.* 2021;86:2179-2191.
23. El-Rewaify H. Replication Data for: Multi-Domain Convolutional Neural Network (MD-CNN) For Radial Reconstruction of Dynamic Cardiac MRI. 2020. <https://doi.org/10.7910/DVN/CI3WB6>
 24. Kaggle Data Science Bowl Cardiac Challenge Data. 2016. <https://www.kaggle.com/c/second-annual-data-science-bowl/data>
 25. Guerquin-Kern M, Lejeune L, Pruessmann KP, Unser M. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Trans Med Imaging.* 2012;31:626-636.
 26. Winkelmann S, Schaeffter T, Koehler T, Eggers H, Doessel O. An optimal radial profile order based on the golden ratio for time-resolved MRI. *IEEE Trans Med Imaging.* 2007;26:68-76.
 27. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics).* 2016;9901:424-432.
 28. Crum WR, Camara O, Hill DLG. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging.* 2012;31:626-636.
 29. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge CM. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics).* 2017;10553:240-248.
 30. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv. 2014. 1412.6980.
 31. El-Rewaify H, Fahmy AS, Pashakhanloo F, et al. Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI. *Magn Reson Med.* 2021;85:1195-1208.
 32. Ma J. Cutting-edge 3D Medical Image Segmentation Methods in 2020: Are Happy Families All Alike? arXiv. 2021. 2101.00232.
 33. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2020;18:203-211.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of the article at the publisher's website.

FIGURE S1 Bland-Altman analyses show high average agreement of ejection fractions (EF) between the pre-trained model predictions and ground truth of the kaggle-dataset up to 4.6-fold acceleration (A – D). For each undersampling scheme (R = acceleration factor) data are shown per subject for the left ventricular ejection fraction of the mid-ventricular slice. Average agreement, 95 % confidence interval (1.96 standard deviation) and coefficient of variation (CV) are shown

TABLE S1 The listed tables hold the mean specificity, sensitivity, accuracy and precision for applying the pre-trained model to the test set of the Kaggle data

TABLE S2 Performance of the transfer-learned 3D U-Net for segmentation of simulated radially undersampled cardiac cine images from the kaggle data test split. Data are shown as mean DSC. For each different scheme of undersampling (number of projections) the performance of the network was compared to the ground truth labels determined by the model proposed by Bai et al. to fully-sampled Cartesian cine images. The performance was evaluated individually for left ventricle blood pool (LV), left ventricle myocardium and right ventricle blood pool (RV). Factor of acceleration (R) is indicated. Cases without segments in both ground truth and predicted segmentation were not counted. Test data were augmented as described for training and measures were average across five repetitions of the evaluation

TABLE S3 The listed tables hold the mean specificity, sensitivity, accuracy and precision for applying the final model to the test set of the Harvard data

TABLE S4 Performance of Bai's 2D model (trained on fully sampled data only) for the segmentation of non-Cartesian cardiac cine images from the Harvard data reconstructed with different undersampling schemes. Data are shown as mean DSC. The performance was tested for different undersampling schemes (numbers of projections) compared to the ground truth (i.e. application of Bai's model on P = 196). Scores were determined individually for segmentation of the left ventricle blood pool (LV), the left ventricular myocardium and the right ventricle blood pool (RV). As only fully sampled cine series were used to train Bai's model, DSC values drop rapidly for increasing R. The transfer-learned 3D U-net, however, showed a stable performance up to R = 7.6 and only a slightly reduced mean DSC for R = 10.1 (see table 3 of the main manuscript)

VIDEO S1 Dynamic view demonstrating the performance of the transfer-learned 3D U-Net on one example of the test data. Acceleration factors correspond to those of Figure 4

VIDEO S2 Dynamic view demonstrating the performance of the transfer-learned 3D U-Net on another example of the test data. Acceleration factors correspond to those of Figure 4

How to cite this article: Wech T, Ankenbrand MJ, Bley TA, Heidenreich JF. A data-driven semantic segmentation model for direct cardiac functional analysis based on undersampled radial MR cine series. *Magn Reson Med.* 2022;87:972–983. <https://doi.org/10.1002/mrm.29017>