

Landmark Detection in Cardiac MRI Using a Convolutional Neural Network

Hui Xue

Jessica Artico

Marianna Fontana

James C. Moon

Rhodri H. Davies

Peter Kellman

From the National Heart, Lung and Blood Institute, National Institutes of Health, 10 Center Dr, Bethesda, MD 20892 (H.X., P.K.); Barts Heart Centre, Barts Health NHS Trust, London, England (J.A., J.C.M., R.H.D.); and National Amyloidosis Centre, Royal Free Hospital, London, England (M.F.). Received XXX; revision requested XXX; revision received XXX; accepted XXX; final version accepted XXX. Supported by the National Heart, Lung and Blood Institute, National Institutes of Health by the Division of Intramural Research (grants Z1A-HL006214-05 and Z1A-HL006242-02). **Address correspondence to H.X.** (e-mail: hui.xue@nih.gov).

<https://doi.org/10.1148/ryai.2021200197>

Purpose: To develop a convolutional neural network (CNN) solution for landmark detection in cardiac MRI.

Materials and Methods: This retrospective study included cine, late-gadolinium enhancement (LGE), and T1 mapping scans from two hospitals. The training set included 2329 patients (34019 images; mean age 54.1 years; 1471 men; December 2017-March 2020). A hold-out test set included 531 patients (7723 images; mean age 51.5 years, 323 men; May 2020-July 2020). CNN models were developed to detect two mitral valve plane and apical points on long-axis images. On short-axis images, anterior and posterior right ventricular insertion points and left ventricle center were detected. Model outputs were compared with manual labels by two readers. The trained model was deployed to MR scanners.

Results: For the long-axis images, successful detection of cardiac landmarks ranged from 99.7% to 100% for cine images and from 99.2% to 99.5% for LGE images. For the short-axis, detection rates was 96.6% for cine, 97.6% for LGE, and 98.9% for T1-mapping. The Euclidean distances between model and manual labels ranged from 2 to 3.5 mm for different landmarks, indicating close agreement between model landmarks to manual labels. No differences were found for the anterior right ventricular insertion angle and left ventricle length by the models and readers for all views and imaging sequences. Model inference on MR scanner took 610 msec on the graphics processing unit and 5.6 sec on central processing unit, respectively, for a typical cardiac cine series.

Conclusion: A CNN was developed for landmark detection in both long and short-axis cardiac MR images for cine, LGE and T1 mapping sequences, with the accuracy comparable to the interreader variation.

Published under a CC BY 4.0 license.

A convolutional neural network was developed for labeling landmarks on long-and short-axis cardiac MR images for cine, late-gadolinium enhancement, and T1 mapping with a performance comparable to manual labeling.

Abbreviations

AL-P = anteroseptal point, A-P = anterior point, A-RVI = Anterior right ventricular insertion, AS-P = anterolateral point, C-LV = LV center, CMR = cardiac MRI, CNN = convolutional neural network, IL-P = inferoseptal point, I-P = inferior point, I-RVI = Inferior right ventricular insertion, IS-P = inferolateral point, LV = left ventricular, MOLLI = modified Look-Locker inversion recovery, RV = right ventricular, RVI = right ventricular insertion

Key Points

The developed model achieved a high detection rate for cardiac landmarks (ranging from 96.6% to 99.8%) on the test dataset.

Comparison of right ventricular insertion angle and left ventricular length measurements between the developed model and experts were similar on different cardiac MRI scan views.

Models were integrated on MR scanners using Gadgetron InlineAI with <1s inference time.

Author contributions:

Guarantor of integrity of entire study, H.X.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, H.X., J.A.; clinical studies, H.X., J.A., M.F., J.C.M.; experimental studies, H.X., J.A., R.H.D., P.K.; statistical analysis, H.X.; and manuscript editing, H.X., M.F., J.C.M., R.H.D., P.K.

Conflicts of interest are listed at the end of this article.

Cardiac MRI (CMR) is emerging as a main-stream modality to image the cardiovascular system for diagnosis and intervention. CMR imaging has advanced beyond the scope of imaging anatomy and can provide comprehensive quantitative measures of the myocardium. These include relaxometry T1, T2, and T2* (1,2) to assess fibrosis, edema, and iron, tissue composition for fat fraction (3) and physiologic measures such as myocardial perfusion (4,5) and blood volume (6) mapping. These capabilities open new opportunities and simultaneously place new demands on image analysis and reporting. A fully automated solution brings increased objectivity, reproducibility, and higher patient throughputs.

Research in the field of automated analysis and reporting of CMR is continuing to advance. In clinical practice, manual delineation by cardiologists remains the main approach to quantify cardiac function, viability, and tissue properties (7). A recent study showed a detailed manual analysis by an expert can take anywhere from 9 to 19 minutes (8). Thus, automated image delineation could help reduce the time needed for image assessment.

Deep learning models, convolutional neural networks (CNNs) in particular, have been developed to automate CMR analysis. Cardiac cine images can be automatically analyzed using CNNs to measure ejection fraction and other parameters to match the expert-level performance (9) and have demonstrated improved reproducibility in multicenter trials (8,10). Cardiac perfusion images have been successfully analyzed and reported on MR scanners (11) using CNNs. CNNs have also been developed to quantify left ventricular (LV) function in

multivendor, multicenter experiments (12). Additionally, deep learning CNNs have been developed for automatic myocardial scar quantification (13). Current research has focused on automating the time-consuming processes of segmenting the myocardium.

To achieve automated analysis and reporting of CMR, key landmark points must be located on the cardiac images. For example, right ventricular (RV) insertion points are needed to report quantitative maps using the standard American Heart Association sector model (7). For the long-axis views, ventricular length can be measured if valve and apical points can be delineated. The variation of LV length is a useful marker and shown to be the principal component of left ventricular pumping in patients with chronic myocardial infarction (14). Furthermore, cardiac landmark detection can be useful on its own for applications such as automated imaging slice planning.

In this study we developed a CNN-based solution for automatic cardiac landmark detection for CMR images. Detection was defined as the process to locate the key landmark points from CMR images acquired in both short and long-axis views. The proposed CNN model predicts the spatial probability of a landmark in the image. The performance of the trained model was quantitatively evaluated by comparing against manual labels for success rate and computing Euclidean distance between manual and model derived landmarks. To evaluate the feasibility of models for CMR reporting, two measures were computed from the model landmarks and compared with the manual values of the angle of anterior RVI point and length of the LV. To demonstrate clinical feasibility, the trained CNN models were integrated on MR scanners using Gadgetron InlineAI (15) and used to automatically measure the LV length from long-axis cine. The developed model has the potential to reduce the amount of time needed for CMR image assessment.

Materials and Methods

Study Design

The developed CNN was designed to detect landmarks on both the long-axis (two-chamber [CH2], three-chamber [CH3], and four-chamber [CH4]) and short-axis series (Fig 1). The following points were detected in different views: (a) short-axis view, anterior and inferior RV insertion (A-RVI and I-RVI) and LV center points (C-LV); (b) CH2 view, anterior and inferior points (A-P and I-P); (c) CH3 view, inferolateral and anteroseptal points (IS-P and AL-P); (d) CH4, inferoseptal and anterolateral points (IL-P, AS-P), and (e) long-axis view, apical point (APEX). The trained CNN models were tested on cardiac cine, late gadolinium enhancement (LGE), and T1 maps derived from a modified Look-Locker inversion recovery (MOLLI) imaging sequence (1,16).

Data Collection

In this retrospective study, a dataset was assembled from two hospitals. All cine and LGE scans were performed at the Barts Heart Centre, London, UK and all T1 MOLLI images were acquired from the Royal Free Hospital, London, UK. Both long-and short-axis views were acquired for cine and LGE. T1 mapping acquired one to three short-axis slices per patient. The data used in this study was not utilized in prior publications.

Data were acquired with the required ethical and/or secondary audit use approvals or guidelines (as per each center) that permitted retrospective analysis of anonymized data without

requiring written informed consent for secondary usage of the purpose of technical development, protocol optimization, and quality control. Institutions acquiring data were in the UK and not subject to HIPAA. All data were anonymized and delinked for analysis with approval by the local Office of Human Subjects Research (Exemption #13156). Appendix E1 (supplement) provides information about subject inclusion criteria.

Table 1 summarizes the training and test datasets. For training, a total of 34019 images were included from 2329 patients (mean age 54.1 years; 1471 men), with 29214 cine and 3798 LGE and 1077 T1 images. Cine training data were acquired from three time periods in 2017, 2018 and 2020, as listed in Table 1. All patients with LGE scans also had cine imaging. Data acquisition in every scan period was consecutive. The test set consisted of 7723 images from consecutively acquired 531 patients (mean age 51.5 years, 323 men). The test data were acquired between May 2020 to June 2020. There was no overlap between training and test sets. No test data were used in any way during the training process and was a completely held-out dataset.

CMR Acquisition

Images were acquired using both 1.5 T (four MAGNETOM Aera, Siemens AG Healthcare, Erlangen, Germany) and 3 T (one MAGNETOM Prisma, Siemens AG Healthcare) MR scanners. In the training set, 1790 patients were scanned with 1.5T scanners and 539 were scanned with 3T. In the test set, 462 patients were scanned at 1.5T MRI and 69 scanned at 3T. Typically 30 cardiac phases were reconstructed for each heartbeat for every cine scan. For the training and testing purpose, the first phase (typically end-diastolic) and the end-systolic phase were selected. Given that there was a large number of patients, these acquired cardiac phases would represent a sufficiently broad variation. For those underwent contrast study, the gadolinium-based contrast agent (gadoterate meglumine, Dotarem, Guerbet, Paris, France) was administered at 4 mL/s at a dose of 0.05 mmol/kg.

Imaging Sequences

The imaging parameters for each sequence are shown in Table E1 (supplement).

Balanced steady state free precession Cine imaging Cine acquisitions were performed with retrospective electrocardiogram gating (30 cardiac phases were reconstructed) and two-fold parallel imaging acceleration using GRAPPA (17). For the short-axis acquisition, a scan typically had 8 to 14 sections to cover the LV.

Phase sensitive inversion recovery for LGE imaging Phase sensitive inversion recovery (PSIR) LGE imaging was performed with a free-breathing sequence (18) for whole LV coverage with respiratory motion correction and averaging. The phase sensitive LGE reconstruction (19) was used to achieve insensitivity to inversion time. Previous studies (20) showed this free-breathing technique is more robust against respiratory motion and delivered improved LGE image quality.

T1 mapping using MOLLI T1 mapping used a previously published MOLLI protocol (1). The sampling strategy was 5s (3s)3s for precontrast T1 scans and 4s(1)3s(1s)2s for postcontrast scans. A retrospective motion correction algorithm (21) was applied to MOLLI images and then went through the T1 fitting (22) to estimate per-pixel maps.

Data Preparation and Labeling

Since the acquired field-of-view may have varied between patients, all images were first resampled to a fixed 1 mm^2 pixel spacing and padded or cropped to 400×400 pixels before input into the CNN. This corresponds to a processing field of view of 400 mm^2 which was large enough to cover the heart, since the MR technicians generally placed the heart close to the center of field of view. The cine MRI often causes a shadow across the field of view (Fig E1 [supplement]), as the tissue which is further away from receive coils on the chest and spine will have reduced signal intensity due to inhomogeneity of the surface coil receive sensitivity. To compensate for this shading, for every cine image in the dataset, a surface coil inhomogeneity correction algorithm (23) was applied to estimate slowly varying surface coil sensitivity which was used to correct this inhomogeneity. During training, either the original cine image or the corrected image was fed into the network with a probability $P = .5$ to pick original version. This served as a data augmentation step. Additional details on other data augmentation are found in Appendix E2 (supplement).

One reader (HX, 9 years of experience in CMR imaging research and 3 years of experience in deep learning) manually labeled all images for training and test (41742 images). A second reader (JA, 3 years of experience in CMR clinical reporting) was invited to label part of the test dataset to assess interreader variation. JA labeled 1,100 images (cine and LGE: 100 images for every long-axis view, 200 images for short-axis; T1 maps: 100 images). The VIA Image Annotator software (<http://www.robots.ox.ac.uk/~vgg/software/via/>) was used by both readers for manual labeling of landmarks. The data labeling took ~ 150 hours in total. Table 1 shows the training and test datasets.

Model Development

The landmark detection problem was formulated as a “heat map” (24). As shown in Figure 2, every landmark point was convolved with a Gaussian kernel (sigma was 4.0 pixels) and the resulting blurred distribution represents the spatial probability of this landmark. Detecting three landmarks was equivalent to a semantic segmentation problem for four classes (background class and one object class for each landmark). Class label for different landmarks was represented as channels in probability maps; thus, if there are three landmarks to be detected, it will have four heat maps (three maps for three landmarks and one for background). Additional information on the heat maps are described in Appendix E3 (supplement).

Model Training

A variation of U-net architecture was implemented (25,26) for heat-map detection. As shown in Figure 3, the network was organized as layers for different spatial resolution. Specific details on model architecture are described in Appendix E4 (supplement). The input to model was a two-dimensional image (ie, to detect the landmarks from a time series, of cine image, the model was applied to each two-dimensional image using the current model configuration).

In the data preparation step, all images were resampled and cropped to 400×400 pixels square. The CNN output score tensor had dimensions $400 \times 400 \times 4$. To train the network, the KL divergence was computed between ground-truth heat-map and SoftMax tensor of scores. Besides this entropy-based loss, the shape loss was further computed as the soft Dice ratio (27). Soft Dice ratio was computed as the production of two probability maps over their sum. The final loss was a sum of entropy-based loss and soft Dice ratio, which used both entropy-based

information and region costs. This strategy to use a combined loss has been previously used in deep learning segmentation and found to improve segmentation robustness (28,29).

For the long axis, all views were trained together as a multitask learning task. Since the number of images for each long-axis view was roughly equal, no extra data rebalancing strategy was applied. Instead, every minibatch randomly selected from CH2, CH3 or CH4 images and refined network weights.

The data for training was split with 90% of all patients for training and 10% for validation. The training and validation datasets were split in a per-study basis, such that there was no mixing of patients between the two datasets. The Adam optimizer was used with an initial learning rate of 0.001, betas were 0.9 and 0.999 and epsilon was 1e-8. The learning rate was reduced by 2 whenever the cost function plateaued. Training lasted 50 epochs (~ 4 hours) and the model was selected as the one giving the highest performance on the validation set. The CNN model was implemented using PyTorch (30) and training was performed on an Ubuntu 20.04 PC with four NVIDIA GTX 2080Ti GPU cards, each with 11GB RAM. Data parallelization was used across multiple GPU cards to speedup training.

Since there were more cine images than LGE and T1 MOLLI, a fine-tuning strategy was implemented using transfer learning. For both long-and short-axis images, a model was first trained with cine dataset and then fine-tuned with either LGE or T1 training sets. The transfer learning was implemented to first train the neural networks with cine data as the pretrained model. The LGE or T1 data were used to fine-tune the pretrained model with reduced learning rate (31). To perform the fine tuning, the initial learning rate was set to be 0.0005 and a total of 10 epochs were trained. For each type of image, separate models were trained for short-and long-axis detection, respectively.

Performance Evaluation and Statistical Analysis

The trained model was applied to all test samples. All results were first visually reviewed to determine whether landmarks were missed or unnecessarily detected (further details are described in Appendix E5 [supplement]).

The detection rate or success rate was computed as the percentage of samples which had landmarks that were correctly detected. This rate was the ratio between the number of images with all landmarks detected and the total number of tested images. For all samples with successful detection, the Euclidean distance between detected landmarks and labels were computed and reported separately for different slice views and different landmark points. Results from model detection and manual labels were compared and Euclidean distance between two readers were reported.

The detected key points were further processed to compute two derived measurements: (a) the angle of anterior RV insertion point to LV center for short-axis views; (b) the length of LV for long-axis views, computed as length from detected apical point to the middle point of two valve points (32). The model derived results were compared between manual labels. The results of the first reader were compared with the second reader to give references for interreader variation.

Results were presented as mean \pm SD, instead of standard error. Paired t test was performed and a P value less than 0.05 was considered statistically significant (Matlab R2017b, Mathworks Inc., MA, USA).

To test the sensitivity of detection performance to the size of the Gaussian kernel used to generate the heat map, two additional models were trained for long-axis cine images with sigma being 6.0 and 2.0 pixels. The detection performance was compared across different kernel sizes for cine long-axis test images.

To visualize the characteristics of what trained models learned from the image, a saliency map was computed as the derivative of the CNN loss function with respect to the input image. Higher magnitude in the saliency map indicates the corresponding image content has more impact on the model loss and indicates the CNN model learned to put more weight in those regions.

The cine long-axis test datasets were further split according to the scanner field strength. The Euclidean distance were compared for 3T and 1.5T.

Model Deployment

To demonstrate the clinical relevance of CMR landmark detection, an inline application was developed to measure LV length from long-axis cine images automatically on the MR scanner. The trained long-axis model was integrated onto MR scanners using the Gadgetron InlineAI toolbox (15). While the imaging was ongoing, the trained model was loaded and after the cine images were reconstructed, the model was applied to the acquired images as part of the image reconstruction workflow (inline processing) at the time of scan. The resulting landmark detection and LV length measurements were displayed and available for immediate evaluation prior to the next image series. Figure E4 (supplement) provides more information for this landmark detection application. A movie of this example can be found in Movie 1. Appendix E6 provides additional information on model deployment and processing times.

The source file to train the model are shared at https://github.com/xueh2/CMR_LandMark_Detection.git.

Results

Model Landmark Detection Rates

The trained model was applied to the test datasets. Examples of landmark detection for different long-axis and short-axis views (Fig 4) demonstrate the trained model was able to detect the specified landmarks. Table 2 summarizes the detection rate for all views and sequences on the test dataset. For the cine, 99.8% (2072 of 2076; 0 false-positive) of CH2, CH3, CH4 long-axis images and 96.6% (2906 of 3008 test images; 24 false-positives) of short-axis images were successfully detected. For the LGE, the detection rates were 99.4% (1105 of 1112; two false-positives) for all long-axis views and 97.6% (1056 of 1082; 11 false-positives) for short axis. For the T1 mapping, the detection rate was 98.9% (439 of 445; 0 false-positive).

The few failed detections in long-axis test images were due to incorrect imaging planning, or unusual shapes of LV or poor image quality. Examples of misdetections long-axis images and discussion can be found in Figure E2 (supplement).

For the 102 misdetections short-axis images in cine, 51 missed the A-RVI and 25 missed the P-RVI and 13 missed LV center. Half of the errors were found to be on the most basal and apical slices (defined as top two sections or the last section for a short-axis series). For the 26 failed short-axis cases in LGE, seven missed the A-RVI, one missed the P-RVI, and two missed

LV center. A total of 11 errors were due to unnecessary landmarks detected in slices outside LV. All T1 MOLLI failures (six of 445 test images) missed P-RVI, due to unusual imaging planning for one patient. Examples of misdetections of short-axis cases can be found in Figure E3 (supplement).

Euclidean Distances between Readers and the CNNs

For all images where detection was successful, the Euclidean distances between model detection and expert labels were computed. Tables 3 and 4 show the Euclidean distances and two derived measurements, reported separately for all imaging views and imaging sequences. The distances between the trained model and the first reader ranged between 2 to 3.5 mm. Figure 5 shows detection examples with model derived and manual landmarks and their Euclidean distance reported, showing model landmarks were in close vicinity to the manual labels. The mean Euclidean distance for long-axis cine and LGE images were 2.5 ± 1.9 mm and 3.0 ± 2.4 mm. For the short-axis, the mean Euclidean distance (across all landmarks) for cine, LGE, and MOLLI were 2.5 ± 1.8 mm, 2.4 ± 2.5 mm and 2.2 ± 2.0 mm.

Tables 3 and 4 listed Euclidean distances between two readers for the labeled portion of tested data. The Euclidean distances between two human readers were comparable to model distances. No evidence of differences were found for the A-RVI angle and LV length measurement between the trained models and the first reader for all imaging applications and imaging views. For the test data labeled by both readers, no differences were found between two readers for both measures. The long-axis cine test images were split according to the acquired field strength (1.5T, 1668 images and 3T, 408 images). The mean distance for 1.5T images was 2.5 ± 1.6 mm and for 3T, 2.3 ± 1.5 mm ($P < .001$).

The model was retrained with two more different Gaussian kernel sizes (2.0 and 6.0 pixels) for the long-axis cine datasets, bracketing the 4.0 pixels design to determine the sensitivity to kernel size. The mean distances to the manual landmarks from the first reader was 2.3 ± 1.6 mm and 2.2 ± 1.6 mm for models trained with sigma being 2.0 and 6.0 pixels, which showed no differences compared with sigma 4.0. The LV length was estimated for sigma 2.0 and 6.0, showing no differences than human measurement ($P > .2$ for all views). Figure E5 (supplement) provides an example of landmark detection with computed probability maps for three models, showing that the detection was insensitive to Gaussian kernel sizes.

Discussion

This study presents a CNN-based solution for landmark detection in CMR. Three CMR imaging applications (cine, LGE, and T1 mapping) were tested in this study. A multitask learning strategy was used to simplify the training and ease deployment. Among the whole training dataset, a majority (86%) were cine images. As a result, a transfer learning strategy with fine tuning was applied to improve the performance of the LGE and T1 mapping detection. The resulting models had high detection rates across different imaging views and imaging sequences. An inline application was built to demonstrate the clinical usage of landmark detection to automatically measure and output LV length on the MR scanner.

Landmark detection using deep learning has not been extensively studied for CMR, but has been investigated for computer vision applications, such as facial key point detection (33,34) or human pose estimation (24,35). In these studies, two categories of approaches were explored

for key point detection. First, the output layer of a CNN explicitly computes the x-y coordinates of landmark points and L2 regression loss was used for training. Second, landmark coordinates were implicitly coded as heat-maps. In this context, the detection problem was reformulated as a segmentation problem. In the human pose estimation, the segmentation-based models outperformed regression models (24,36). Here fewer landmarks were detected and were more spatially sparse distributed. The human pose images had much more variation, compared with human faces which often had been preprocessed as front position (37). It is easier for heat-map detection to handle landmark occlusion. For example, in Figure 1, some images may not include targeted landmarks, which is represented by low probability of detection outputs. For these reasons, this study adopted the segmentation model for CMR.

A recent study used heatmap landmark detection in the context of automated image plane prescription for cardiac MRI (38). This study trained heatmap detection model on 892 long-axis and 493 short-axis SSFP cine images. The mid valve plane and apical points were automatically detected and compared with manual localization with the mean distance of $\sim 5\text{--}7$ mm. A recurrent U-net architecture was used in another study to perform myocardial segmentation and detection of mitral valve and RV insertion points from cardiac cine images in one forward pass (39). This neural network was trained on 6961 long-axis images and 670 short-axis images. The detection distance was 2.87 mm for the mitral valve points and 3.64 mm for RV insertions.

Another study developed a patched fully convolutional neural network to detect six landmarks from cardiac CT volume (40). The training was performed on 198 CT scans and resulting average Euclidean distances to the manual label were 1.82–3.78 mm. Compared with previous studies of cardiac landmark detection, the current study curated larger datasets and detected more landmarks in cine, as well as LGE and T1 maps which had substantially different contrasts, to enable automated reporting and measurement of global longitudinal shortening. Detection was slightly less accurate on basal and apical imaging short-axis slices. In these regions, the “ambiguity” of anatomy increases, leading to more variant in data labeling and more difficulties for model to give correct inference. Additional discussion can be found in Appendix E7 (supplement).

There are limitations to this study. First, a single reader labeled entire datasets. Due to the limitation of research resources, the second reader only labeled a portion of test set to measure interreader variation. Second, three imaging applications were tested in this study. If the model were to be applied to the detection of a new anatomy (eg, RV center), imaging sequence, or a different cardiac view, more training data will be required. Use of transfer learning would reduce the amount of new data needed. The development process would have to be iterative to cover more imaging sequences and anatomy. Third, the data used in this study was collected from a single MR vendor (Siemens). A recent study (41) reported performance of deep learning models trained on one vendor may drop for different vendors although augmentation was used to improve robustness. Further validation will be required to extend proposed CNN models to CMR images from other vendors. It is very likely to require further data curation and training. Fourth, due to the availability of different imaging sequences, not all imaging sequences were acquired in between the two included institutions, which limits the evaluation of across hospital generalization. We expect the on-scanner deployment could enable the proposed models to be used in more hospitals and further studies can provide more comprehensive datasets. Other limitations are on the preprocessing. Although the selected processing field of view of 400 mm^2 has been large enough to cover the heart in our imaging experience, it is possible even larger

configuration may be needed if the imaging planning is far off the center. The model can be retrained with even larger field of view, but the inline feedback of detection results could be used to flag readers to adjust or repeat acquisition.

In this study, a CNN-based solution for landmark detection on cardiac MRI was developed and validated. A large training dataset of 2329 patients was curated and used for model development. Testing was performed on 531 consecutive patients from two centers. The resulting models had high landmark detection rates across different imaging views and imaging sequences. Quantitative validation showed the CNN detection performance was comparable to the interreader variation. Based on the detected landmarks, RV insertion and LV length can be reliably measured.

Disclosures of Conflicts of Interest: H.X. disclosed no relevant relationships. J.A. disclosed no relevant relationships. M.F. disclosed no relevant relationships. J.C.M. disclosed no relevant relationships. R.H.D. disclosed no relevant relationships. P.K. disclosed no relevant relationships.

References

1. Kellman P, Hansen MS. T1-mapping in the heart: accuracy and precision. *J Cardiovasc Magn Reson* 2014;16(1):2.
2. Giri S, Chung YC, Merchant A, et al. T2 quantification for improved detection of myocardial edema. *J Cardiovasc Magn Reson* 2009;11(1):56.
3. Kellman P, Hernando D, Arai AE. Myocardial Fat Imaging. *Curr Cardiovasc Imaging Rep* 2010;3(2):83–91.
4. Xue H, Brown LAE, Nielles-Vallespin S, Plein S, Kellman P. Automatic in-line quantitative myocardial perfusion mapping: Processing algorithm and implementation. *Magn Reson Med* 2020;83(2):712–730.
5. Kellman P, Hansen MS, Nielles-Vallespin S, et al. Myocardial perfusion cardiovascular magnetic resonance: optimized dual sequence and reconstruction for quantification. *J Cardiovasc Magn Reson* 2017;19(1):43.
6. Nickander J, Themudo R, Sigfridsson A, Xue H, Kellman P, Ugander M. Females have higher myocardial perfusion, blood volume and extracellular volume compared to males - an adenosine stress cardiovascular magnetic resonance study. *Sci Rep* 2020;10(1):10380.
7. Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update : Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing. *J Cardiovasc Magn Reson* 2020;22(1):19.
8. Bhuvana AN, Bai W, Lau C, et al. A Multicenter, Scan-Rescan, Human and Machine Learning CMR Study to Test Generalizability and Precision in Imaging Biomarker Analysis. *Circ Cardiovasc Imaging* 2019;12(10):e009214.
9. Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* 2018;20(1):65.

10. Bernard O, Lalande A, Zotti C, et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans Med Imaging* 2018;37(11):2514–2525.
11. Xue H, Davies RH, Brown LAE, et al. Automated Inline Analysis of Myocardial Perfusion MRI with Deep Learning. *Radiol Artif Intell* 2020;2(6):e200009.
12. 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(1):81–88.
13. Fahmy AS, Neisius U, Chan RH, et al. Three-dimensional deep convolutional neural networks for automated myocardial scar quantification in hypertrophic cardiomyopathy: A multicenter multivendor study. *Radiology* 2020;294(1):52–60.
14. Asgeirsson D, Hedström E, Jögi J, et al. Longitudinal shortening remains the principal component of left ventricular pumping in patients with chronic myocardial infarction even when the absolute atrioventricular plane displacement is decreased. *BMC Cardiovasc Disord* 2017;17(1):208.
15. Xue H, Davies R, Hansen D, et al. Gadgetron Inline AI : Effective Model inference on MR scanner [abstr]. In: *Proceedings of the Twenty-Seventh Meeting of the International Society for Magnetic Resonance in Medicine*. Berkeley, Calif: International Society for Magnetic Resonance in Medicine, 2019; ISMRM, 2019; 4837.
16. Messroghli DR, Walters K, Plein S, et al. Myocardial T1 mapping: application to patients with acute and chronic myocardial infarction. *Magn Reson Med* 2007;58(1):34–40.
17. Breuer FA, Kellman P, Griswold MA, Jakob PM. Dynamic autocalibrated parallel imaging using temporal GRAPPA (TGRAPPA). *Magn Reson Med* 2005;53(4):981–985.
18. Kellman P, Larson AC, Hsu LY, et al. Motion-corrected free-breathing delayed enhancement imaging of myocardial infarction. *Magn Reson Med* 2005;53(1):194–200.
19. Kellman P, Arai AE, McVeigh ER, Aletras AH. Phase-sensitive inversion recovery for detecting myocardial infarction using gadolinium-delayed hyperenhancement. *Magn Reson Med* 2002;47(2):372–383.
20. Piehler KM, Wong TC, Punttil KS, et al. Free-breathing, motion-corrected late gadolinium enhancement is robust and extends risk stratification to vulnerable patients. *Circ Cardiovasc Imaging* 2013;6(3):423–432.
21. Xue H, Shah S, Greiser A, et al. Motion correction for myocardial T1 mapping using image registration with synthetic image estimation. *Magn Reson Med* 2012;67(6):1644–1655.
22. Xue H, Greiser A, Zuehlsdorff S, et al. Phase-sensitive inversion recovery for myocardial T1 mapping with motion correction and parametric fitting. *Magn Reson Med* 2013;69(5):1408–1420.
23. Xue H, Zuehlsdorff S, Kellman P, et al. Unsupervised inline analysis of cardiac perfusion MRI. *Med Image Comput Comput Assist Interv* 2009;12(Pt 2):741–749.

24. Belagiannis V, Zisserman A. Recurrent Human Pose Estimation. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, May 30–June 3, 2017. Piscataway, NJ: IEEE, 2017; 468–475.
25. Zhang Z, Liu Q, Wang Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci Remote Sens Lett* 2018;15(5):749–753.
26. Xue H, Tseng E, Knott KD, et al. Automated detection of left ventricle in arterial input function images for inline perfusion mapping using deep learning: A study of 15,000 patients. *Magn Reson Med* 2020;84(5):2788–2800.
27. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Cardoso J, Arbel T, Carneiro G, et al, eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2017, ML-CDS 2017. Lecture Notes in Computer Science*, vol 10553. Cham, Switzerland: Springer, 2017; 240–248.
28. Shvets A, Rakhlin A, Kalinin AA, Iglovikov V. Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, December 17–20, 2018. Piscataway, NJ: IEEE, 2018; 624–628.
29. Jadon S. A survey of loss functions for semantic segmentation. <http://arxiv.org/abs/2006.14822>. Posted 2020. Accessed DATE.
30. Steiner B, Devito Z, Chintala S, et al. PyTorch: An Imperative Style. In: *High-Performance Deep Learning Library*. NeuroIPS. LOCATION: PUBLISHER, 2019.
31. Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. *J. Big Data*. LOCATION: Springer International, 2016.
32. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr* 2015;28(1):1–39.e14.
33. Agarwal N, Krohn-Grimberghe A, Vyas R. Facial Key Points Detection using Deep Convolutional Neural Network - NaimishNet. arXiv:1710.00977v1 [csCV]. 2017;1–7. <http://arxiv.org/abs/1710.00977>. Posted 2017. Accessed DATE.
34. Colaco S, Han DS. Facial Keypoint Detection with Convolutional Neural Networks. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Fukuoka, Japan, February 19–21, 2020. Piscataway, NJ: IEEE, 2020; 671–674.
35. Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient object localization using Convolutional Networks. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 7–12, 2015. Piscataway, NJ: IEEE, 2015; 648–656.
36. Pfister T, Charles J, Zisserman A. Flowing ConvNets for Human Pose Estimation in Videos. In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December 7–13, 2015. Piscataway, NJ: IEEE, 2015; 1913–1921.

37. Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011, Colorado Springs, CO, June 20–25, 2011. Piscataway, NJ: IEEE, 2011; 529–534.
38. Blansit K, Retson T, Masutani E, Bahrami N, Hsiao A. Deep Learning-based Prescription of Cardiac MRI Planes. *Radiol Artif Intell* 2019;1(6):e180069.
39. van Zon M, Veta M, Li S. Automatic cardiac landmark localization by a recurrent neural network. In: Angelini E, Landman BA, eds. Proceedings of SPIE: medical imaging 2019—image processing. Vol 10949. Bellingham, Wash: International Society for Optics and Photonics, 2019; 1094916.
40. Noothout JMH, de Vos BD, Wolterink JM, Leiner T, Išgum I. CNN-based landmark detection in cardiac CTA scans. arXiv:1804.04963. <https://arxiv.org/abs/1804.04963>. Posted 2018. Accessed DATE.
41. Yan W, Huang L, Xia L, et al. MRI Manufacturer Shift and Adaptation: Increasing the Generalizability of Deep Learning Segmentation for MR Images Acquired with Different Scanners. *Radiol Artif Intell* 2020;2(4):e190195.
42. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ICML* 2016;10(6):730–743.
43. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. *ICML* 2013;28.
44. Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Comput Surv* 2020;53(3):1–34.
45. Fries JA, Varma P, Chen VS, et al. Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences. *Nat Commun* 2019;10(1):3111.
46. Reyes M, Meier R, Pereira S, et al. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiol Artif Intell* 2020;2(3):e190043.
47. Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. <https://arxiv.org/abs/1312.6034>. Posted 2013. Accessed DATE.

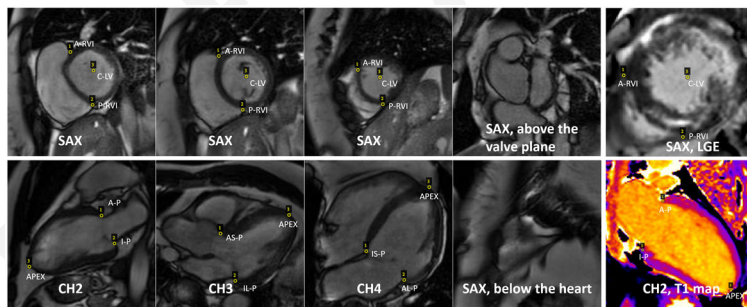


Figure 1: Example of cardiac MRI with landmarks. Three short-axis (SAX) views are shown on the top row. The first three images at the second row shows example of long axis views for two chamber (CH2), three chambers (CH3) and four chamber (CH4). The anterior and inferior points were detected on CH2 view. The inferolateral and

anteroseptal points were detected on the CH3 view, and inferoseptal and anterolateral points were detected on CH4. Apical points were detected for all LAX views. For the SAX images, the anterior and inferior right ventricular insertion and left ventricular center points were detected. Note for some SAX slices (the rightmost column), no landmarks can be identified. The last column gives examples of late gadolinium enhancement images and T1 maps. Transfer learning was applied to detect landmarks from these imaging applications.

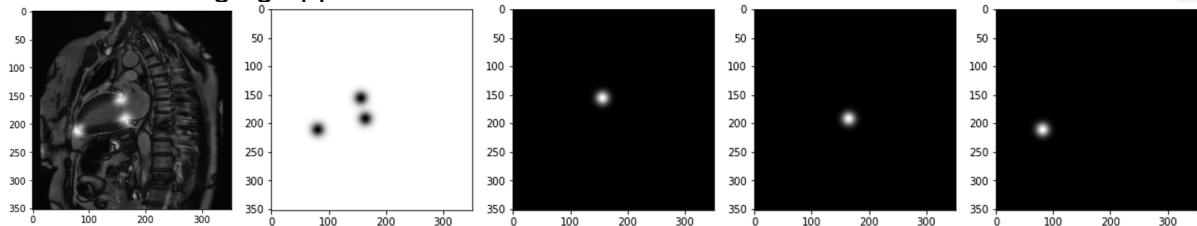


Figure 2: The landmark detection problem can be reformulated as a semantic segmentation problem. Every landmark point in this two-chamber image on the left can be convolved with a Gaussian kernel and converted into a spatial probability map or heat map (upper row, from left to right, probability for background, anterior valve point, inferior valve point and apex). Unlike the binary detection task with target being one-hot binary mask, loss functions working on continues probability such as the KL divergence are needed.

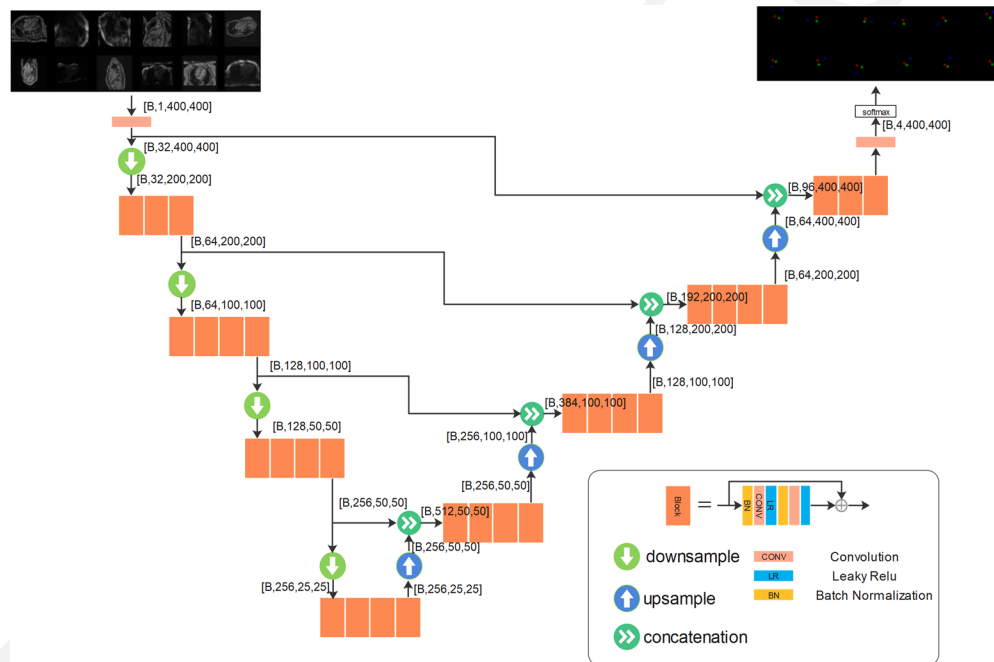


Figure 3: The backbone convolutional neural network developed for landmark detection had a U-net structure. More layers can be inserted to both downsampling and upsampling branches and more blocks can be inserted into each layer. The output layer outputs the per-pixel scores which goes through Softmax function. For the long-axis detection, data from three views were trained together for one model. As shown in the input, every minibatch was assembled by randomly selected images from three views

and used for back propagation. A total of four layers were used in this experiment with 3 or 4 blocks per layer. The output tensor shapes were reported in the figure, in the format of [B, C, H, W]. B is the size of minibatch and C is the number of channels. H and W are the image height and width. Input images have one channel for image intensity and output has four channels for three landmarks and background. The illustration here for outputs plots three landmark channels color-coded and omits the background channel.

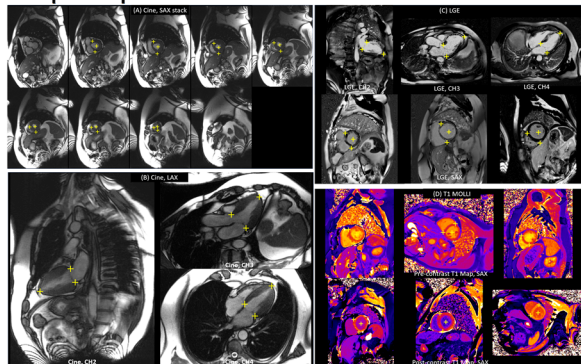


Figure 4: Examples of landmark detection. The left panel are cine detection examples for (a) long and (b) short-axis images. The right panel are (c) late-gadolinium enhanced (LGE) and (d) T1 examples.

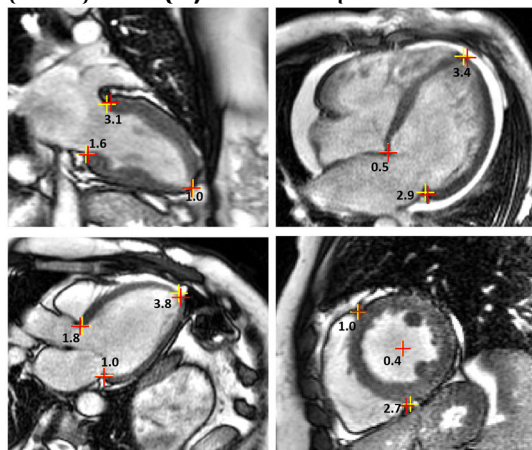


Figure 5: Examples of Euclidean distance of landmarks. For every pair of manual and model delineated landmarks, the distance (in mm) is labeled. Red indicates manually labeled landmarks, and yellow indicates landmarks generated from the model.

Table 1

Information for Training and Test Dataset Distribution and Acquisition

Imaging View	No. Patients	No. Images	Time period
A. Training			
All	2,329	34,019	
Cine			
CH2	2,115	4,232	12/18/17–12/29/17
CH3	2,102	4,206	1/2/18–1/28/18
CH4	2,127	4,256	1/2/20–4/19/20

SAX	702	16,520*	
LGE			
CH2	599	599	1/2/20–2/29/20
CH3	582	582	
CH4	599	599	
SAX	178	2,018†	
T1 MOLLI			
SAX	202	1,077	1/2/20–3/25/20
B. Testing			
All	531	7,723	
Cine			
CH2	347	694	5/1/20–7/3/20
CH3	345	690	
CH4	347	692	
SAX	128	3,008‡	
LGE			
CH2	370	370	5/1/20–7/3/20
CH3	370	370	
CH4	370	372	
SAX	96	1,082§	
T1 MOLLI			
SAX	161	445	5/1/20–7/23/20

Note.— CH2 = two-chamber, CH3 = three-chamber, CH4 = four-chamber, SAX = short axis, LGE = late-gadolinium enhancement, MOLLI = modified look-locker inversion recovery.

*A total of 3803 images were acquired outside the left ventricle (LV) and contained no landmarks.

†A total of 371 images did not contain landmarks.

‡A total of 813 images did not contain landmarks.

§A total of 222 images did not contain landmarks.

Table 2

Detection Rate for Three Imaging Applications at all Tested CMR Views

Imaging	Detection rate
Cine	
CH2	99.7% (692/694)
CH3	99.7% (688/690)
CH4	100% (692/692)
SAX	96.6% (2906/3008)
LGE	
CH2	99.5% (368/370)
CH3	99.5% (368/370)
CH4	99.2% (369/372)
SAX	97.6% (1056/1082)
T1 MOLLI	
SAX	98.9% (439/445)

Note.— CH2 = two-chamber, CH3 = three-chamber, CH4 = four-chamber, SAX = short axis, LGE = late-gadolinium enhancement, MOLLI = modified look-locker inversion recovery

Table 3

CMR Landmark Detection on CH2, CH3, and CH4 Views

Imaging and landmark	Euclidean distance		LV length difference in %			
	first versus CNN	first versus second	first versus CNN	<i>P</i> value	first versus second	<i>P</i> value
A. Cine						
CH2						
A-P	2.1 ± 1.8	2.8 ± 1.9	2.0 ± 1.7	0.42	1.9 ± 1.4	0.95
I-P	2.4 ± 2.0	3.0 ± 3.9				
APEX*	2.4 ± 1.8	4.1 ± 2.8				
CH3						
IL-P	2.4 ± 1.7	2.8 ± 1.6	1.5 ± 1.3	0.79	2.0 ± 1.7	0.97
AS-P*	2.2 ± 1.5	4.0 ± 2.4				
APEX*	3.2 ± 2.4	3.8 ± 2.1				
CH4						
AL-P	3.4 ± 2.1	3.5 ± 2.0	1.4 ± 1.2	0.92	2.0 ± 1.4	0.77
IS-P*	2.1 ± 1.7	2.6 ± 1.6				
APEX	2.8 ± 1.9	2.8 ± 1.6				
B. LGE						
CH2						
A-P	2.9 ± 2.6	3.3 ± 2.0	2.7 ± 2.5	0.16	2.5 ± 2.1	0.82
I-P	3.4 ± 2.7	3.4 ± 2.5				
APEX	3.1 ± 2.6	3.4 ± 2.5				
CH3						
IL-P	3.4 ± 3.1	3.5 ± 2.1	2.6 ± 2.6	0.37	2.9 ± 2.2	0.34
AS-P*	2.7 ± 2.1	3.6 ± 2.3				
APEX	3.3 ± 2.8	3.3 ± 2.5				
CH4						
AL-P	3.1 ± 1.6	3.3 ± 2.2	2.0 ± 1.4	0.13	1.9 ± 1.9	0.53
IS-P	2.0 ± 1.5	2.5 ± 2.3				
APEX	2.7 ± 1.2	2.1 ± 1.6				

Note.—“1st vs AI” indicates the comparisons of manual labels from the first reader to the trained model and “1st vs 2nd” indicates the comparisons between the two readers for the test data labeled by both. The distances reported are mean ± SD.

*Indicates *P* < .05 (paired *t* test) for the comparison of the distance between the “1st vs. AI” and “1st vs. 2nd.”

Table 4

CMR Landmark Detection on Short-Axis Views

Imaging	Euclidean distance		A-RVI angle difference in degree			
	first versus AI	first versus second	first-AI	<i>P</i> value	first-second	<i>P</i> value
Cine						
A-RVI	3.1 ± 1.8	3.5 ± 2.6	1.3 ± 3.4	0.14	-0.7 ± 4.1	0.89
P-RVI	2.4 ± 2.1	2.7 ± 1.6				
C-LV	2.0 ± 1.1	2.4 ± 1.2				
LGE						
A-RVI	3.0 ± 3.2	3.6 ± 3.1	0.14 ± 2.9	0.92	-2.0 ± 4.5	0.62
P-RVI	2.8 ± 2.6	3.3 ± 2.6				
C-LV*	1.5 ± 0.9	2.3 ± 1.1				
T1 MOLLI						
A-RVI	2.5 ± 2.0	3.0 ± 2.8	1.6 ± 3.1	0.31	1.7 ± 3.9	0.41

P-RVI	2.5 ± 2.6	2.5 ± 2.0				
C-LV	1.6 ± 1.0	2.0 ± 1.1				

Note.—“1st vs AI” indicates the comparisons of manual labels from the first reader to the trained model and “1st vs 2nd” indicates the comparisons between the two readers for the test data labeled by both. The distances reported are mean ± SD.

*Indicates $P < .05$ (paired t test) for the comparison of the distance between the “1st vs. AI” and “1st vs. 2nd.”

Appendix E1. Study Criteria

We included all consecutive patients who can successfully complete a routine CMR study with repeated breath-holds, without intentionally excluding structural abnormalities. Exclusion criteria included patients with a cardiac implantable electronic device, significant arrhythmia (atrial fibrillation or ectopy) during the scan, claustrophobia or inability to breath-hold. Since the purpose is to test the clinical performance of CNN models, patients who did not cooperate and failed to complete study was excluded. As a result, the datasets reflect the real distribution of different anatomical and pathological distribution.

Appendix E2. Data Augmentation

Other data augmentation included adding random gaussian noise (prob. to add noise is 0.5, noise sigma was uniformly picked from 10-30% of the mean image value) and adding blurring with a Gaussian kernel applied randomly to images (prob. to apply filtering $P = .5$, filter sigma was uniformly picked from [0.5, 1.0, 2.0] pixel). Often the cine images from the MR scanner had already been corrected for surface coil inhomogeneity, using pre-scan calibration data. However, there are instances where images are not corrected, depending on the reconstruction workflow and imaging applications. The training set used these augmentations to encourage the resulting model to work robustly independent of whether correction has been applied.

Appendix E3. Heatmaps

The heat maps were normalized to be probability maps. That is, for every pixel in the field-of-view, the sum over all classes including the background is 1.0. The pixel value in a heat map is the probability that this pixel belongs to corresponding landmark class or background. With the probability heat maps as the target, the training process will optimize the network parameters to minimize the distance between network outputs (after SoftMax) to the per-pixel probability distribution. For comparison, in the binary segmentation, the mask is either 0 or 1 (hard map). But in the current heat map setting, it uses “soft” probability. The entropy-based cost function (such as KL divergence which computes distance between two probability distributions) can be applied to both scenarios. In this way, landmark detection is formulated as a semantic segmentation problem.

Compared to the typical semantic segmentation setup where the segmentation targets are input as the binary mask, the “heat map” formulation replaced the binary masks as the probability maps. The highest probability exists at the location of landmark points. By optimizing the cost function such as KL divergence between the label probability and model

outputs, the optimization reduces the distance between spatial probability maps. The advantage of this method is that it considers the location of landmarks in context with the surrounding image structures.

Appendix E4. Model Architecture

Each layer can contain several blocks. Each block had two convolution layers with batch normalization (42) and LeakyReLU activation functions (43). The network can be made deeper by inserting more resolution layers or by inserting more blocks. Going down the downsampling branch, the image spatial resolution was reduced by $\times 2$ for every layer with the number of filters increased. Going up the upsampling branch, the spatial resolution was restored with a reduced number of filters. All convolution layers had filter size 3×3 with stride 1 and padding 1. The final convolution layer outputs a per-pixel score tensor which is converted to a probability tensor using a SoftMax.

Appendix E5. Landmark Review

To determine if landmarks were missing or unnecessarily detected, the detection images and results as jpg images were saved which were rapidly reviewed in image viewers to check whether all landmarks were detected. For example, if a mid-SAX slice was marked as three landmark points (see Fig 1) and only two points were detected by model, this case was reported as a failed detection case of false-negative. The false-positive failure was the case where a landmark location was not actually included in the image, but the model output a detection.

Appendix E6. Model Deployment and Processing Times

The deployed model was tested on the MR scanner for measurement of processing speed. On a tested server (2x Intel Xeon E5-2640 v3@3.400GHz, without GPU), it took ~ 74 ms to load the model and ~ 5.6 s to apply the model on all 30 phases of a cine series on central processing unit. When tested on a server with graphics processing unit (2x Intel Xeon Gold 6152@2.101GHz, 1x NVIDIA RTX 2080Ti), model loading took 66 ms and applying the model took 610 ms.

Appendix E7. Supplemental Discussion

While different neural network architecture or loss function may be optimized for higher accuracy, the limit of accuracy may be on the data labelling. Overall, the models had higher performance on LAX views than SAX slices. The reason is the less imaging and anatomical variation in long-axis acquisition. For a correctly prescribed LAX imaging slice, occlusion does not happen. A related finding is the detection of mid-cavity SAX slices was very robust. Therefore, future improvement in data labelling shall focus on the basal and apical SAX slices. There are a few failed detections due to unusual anatomy, inferior image quality and bad slide planning. More specific data collection for these “long-tail” scenario is needed to further improve models. One plausible strategy is to deploy models and monitor performance regularly and collect corner cases. Results also indicated the LV center detection had lower distance, compared to RV insertion points for SAX slices, which was consistent for AI vs. both human

readers. We assume that akin to the human, the network was learning the SAX LV center was equidistant from the well-defined myocardium border feature although the blood pool lacked texture. While the RV insertion points were in general well delineated, however occasionally the locations can become less well defined, e.g. due to the epicardial fat bearing bright intensity or when the RV wall was thin and less visible towards more apical slices.

The comparison between 3T and 1.5T cine LAX images showed 3T detection was slightly more accurate with lower distance. This can be attributed to the high SNR of 3T acquisition, which can help both data labelling and CNN detection. Although detection performance differed slightly across field strengths, this result may indicate model performance drifting can happen if acquisition conditions vary, which can require further data curation and retraining.

The results showed AI model was in closer agreement with the first annotator, which is not surprising since the CNN was trained on the labels delineated by this reader. In this study, the second annotator had labelled a much smaller number of images for the comparison purpose. Although the model generated landmarks were in ~2-4mm distance to both annotators in tested images, ideally labelled data from multiple experts can be beneficial, but this is often hard to achieve due to the limitation of data labelling resource. This problem may be mitigated by using machine learning methods requiring a smaller size of labelled data, such as few-shot learning (44) or weak supervision (45). In this study, the trained model was deployed to MR scanners, also encouraging broader clinical usage and collecting more feedback from multiple clinicians.

The current model processed every two-dimensional image to detect required landmarks. For the cine images, it could be beneficial to further use the temporal consistency across cardiac phases to improve the detection. The neural network will be modified to accept 2D+T image series and output landmarks for every input phase. The current strategy detected multiple landmarks in the field-of-view in one forward-pass, but treated every image separately, because labelled data did not exist for entire cine series. Future research will explore the temporal consistency to improve detection performance.

For the CNN models aiming for clinical deployment, it is crucial to ensure the models function correctly and can be trusted by the users. As discussed in a recent study (46), AI models can gain interpretability through semantics, visualization, counter examples, or gradient-based saliency maps. Current study deployed trained AI models on the MR scanners and provided explainability of AI models, to certain level, through the visualization of images with superimposed landmarks. The images with detected landmarks were sent over together with derived measurements, where the missed or incorrect landmarks became visible to end-user. Furthermore, the saliency map (47) was computed for representative test examples and given in Figure E6, where higher magnitude indicates the image features which mostly activated the network.

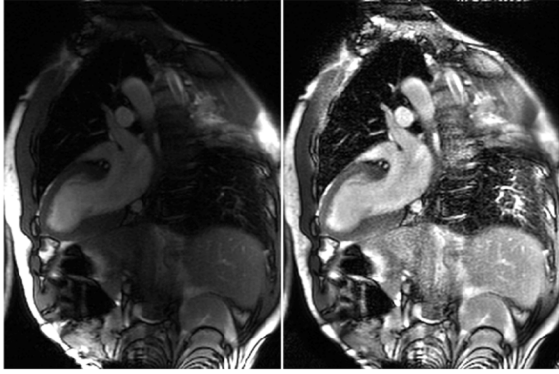


Figure E1: Example of surface coil inhomogeneity correction. A two-chamber slice before and after correcting the surface coil inhomogeneity. In this study, a copy of original image (left) and its corrected version (right) was kept in the dataset and randomly picked to feed into the CNN as a data augmentation step.

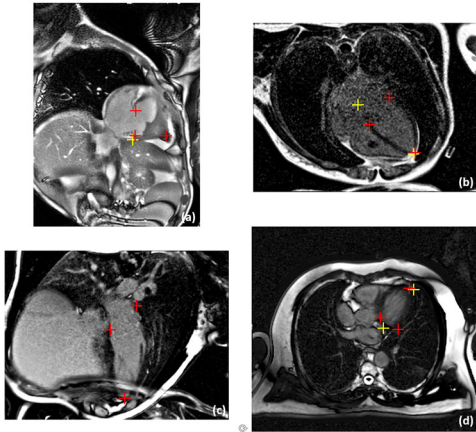


Figure E2: Examples of failed detection for LAX views. **(a)** This CH2 cine image contains unusual anatomy, due to congenital heart abnormality of this patient. Model missed two landmarks on this image. **(b)** This LGE image had very low signal-noise-ratio. The model correctly detected apical point but missed other landmarks. **(c)** An LGE image had severe aliasing artifacts, causing models to miss all three landmarks. **(d)** The acquisition plane of this CH4 LGE image was imperfectly placed, causing the model to miss landmarks. Red: manual landmarks; Yellow: model landmarks.

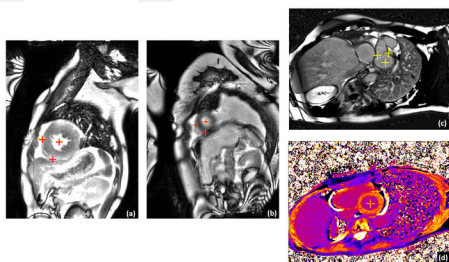


Figure E3: Examples of failed detection for SAX views. **(a)** Detection failed to find the P-RVI point on this cine image, due to the very small RV cavity. **(b)** Both RVI points were missed in this very apical cine slice. **(c)** This LGE image was acquired outside the LV, but model incorrectly outputted landmarks. **(d)** The P-RVI point was missed in this precontrast T1 map, likely due to nonstandard imaging plane subscribed. Red: manual landmarks; Yellow: model landmarks.

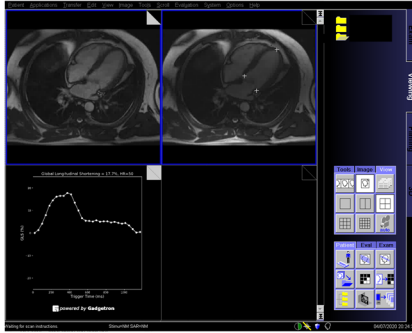


Figure E4: Model deployment and scanner integration. The trained landmark detection models can be useful for many CMR analysis tasks. As an example, the model for LAX detection was integrated on MR scanner and used to measure LV length for long-axis cine image series. The global longitudinal shortening ratio can be computed from the AI measurement as:

$$100 \times (LV_length_{ED} - LV_length_{ES}) / LV_length_{ED} .$$

In this example, a scanner screen snapshot shows a four-chamber cine processed with proposed landmark detection algorithm. The LV length for every cardiac phase was measured and longitudinal shortening ratio was computed. This approach was fully automated. The corresponding movie of this example can be found in Movie 1.

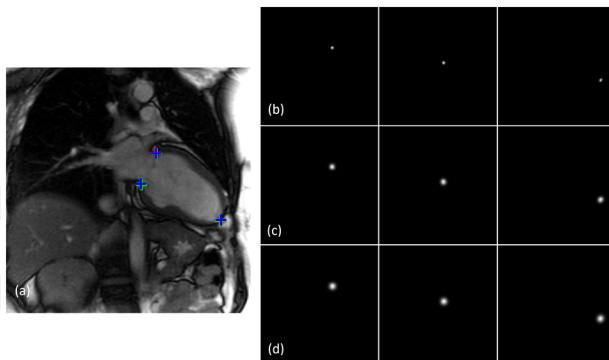


Figure E5: Landmark detection with different Gaussian kernel sizes. Three Gaussian kernels (sigma = 2.0, 4.0 and 6.0 pixels) were used to train three models for heat map detection. These models were applied to cine images to investigate the detection against different kernel size. **(a)** A two-chamber cine image with detected landmarks marked for sigma 2.0 (blue), 4.0 (red) and 6.0 (green). The probability maps of three landmarks for sigma 2.0 **(b)**, 4.0 **(c)** and 6.0 **(d)** were given in the right panel.

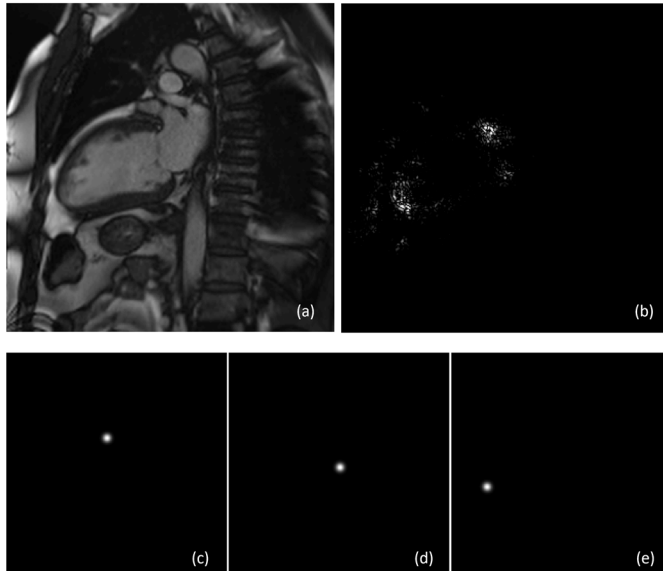
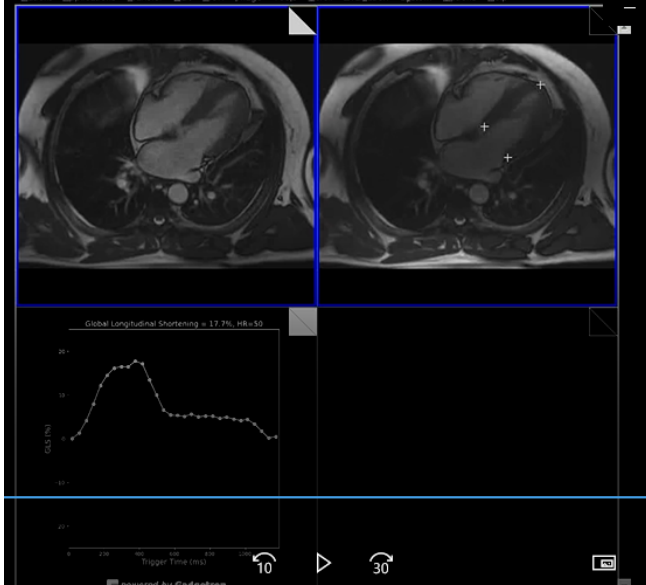


Figure E6: Saliency maps for landmark detection. **(a)** A two-chamber cine image tested for landmark detection. **(b)** The computed saliency map, where high intensity region in the map indicates the image content mostly activating the neural network. The region around apical and valve points are indeed noticed by the CNN, which leads to the model output of probability maps for three landmarks. **(c–e)** Background probability maps are not shown here, since pixel-wise probabilities summed over all four channels must be 1.0.



Movie 1: Screen snapshot of inline landmark detection from a MR scanner. Original cine image series and detected landmarks are shown in the upper row. The estimated global longitudinal shortening ratio is estimated for every cardiac phase and plotted as a curve for reporting on the scanner.

Table E1. Imaging Parameters for Cine, LGE, and MOLLI

	Cine	LGE	T1 Mapping
Imaging	bSSFP	PSIR	MOLLI
TR, msec	2.7	2.7	2.7
TE, msec	1.2	1.1	1.1
Matrix	256 × 144	256 × 144	256 × 144
Flip angle	50°	50°	35°
FOV, mm ²	360 × 270	360 × 270	360 × 270
Slice thickness, mm	8	8	8
Gap, mm	2	2	2
Bandwidth, Hz/pixel	1085	977	1085

Landmark Detection on Cardiac MRI using Convolutional Neural Networks

Key Result

A model was developed to detect landmarks on cardiac MRI and detection distances between the model and expert readers were similar to inter-reader variation.

Patients:

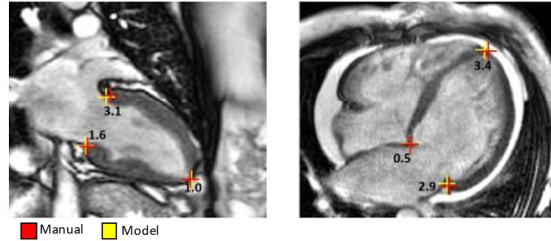
- Training: 2329 (34,019 images)
- Testing: 531 (7,723 images)

Methods:

- Three different cardiac MR image sets were acquired: bSSFP for cine imaging, PSIR for LGE imaging, and T1 mapping using MOLLI
- Separate models were trained for short and long-axis landmark detection

Results:

- The developed model had high detection rates (96.6% to 99.8%) for cardiac landmarks
- Landmark detection by the model was similar to human experts



Xue H et al. Published Online: July 14, 2021
<https://doi.org/10.1148/ryai.2021200197>

Radiology: Artificial Intelligence