Radiology: Artificial Intelligence

ORIGINAL RESEARCH

Landmark Detection in Cardiac MRI by Using a Convolutional Neural Network

Hui Xue, PhD • *Jessica Artico, MD* • *Marianna Fontana, MD* • *James C. Moon, MD* • *Rhodri H. Davies, MD, PhD* • *Peter Kellman, PhD*

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, National Health Service, London, England (J.A., J.C.M., R.H.D.); and National Amyloidosis Centre, Royal Free Hospital, London, England (M.F.). Received August 17, 2020; revision requested October 28; revision received April 28, 2021; accepted June 15. **Address correspondence to** H.X. (e-mail: *hui.xue@nih.gov*).

Supported in part by the Division of Intramural Research of the National Heart, Lung, and Blood Institute, National Institutes of Health (grants Z1A-HL006214-05 and Z1A-HL006242-02).

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

Radiology: Artificial Intelligence 2021; 3(5):e200197 • https://doi.org/10.1148/ryai.2021200197 • Content codes: CA MR

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

Materials and Methods: This retrospective study included cine, late gadolinium enhancement (LGE), and T1 mapping examinations from two hospitals. The training set included 2329 patients (34 089 images; mean age, 54.1 years; 1471 men; December 2017 to March 2020). A hold-out test set included 531 patients (7723 images; mean age, 51.5 years; 323 men; May 2020 to 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 (RV) insertion points and left ventricular (LV) center points were detected. Model outputs were compared with manual labels assigned by two readers. The trained model was deployed to MRI 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 images, detection rates were 96.6% for cine, 97.6% for LGE, and 98.7% for T1 mapping. The Euclidean distances between model-assigned and manually assigned labels ranged from 2 to 3.5 mm for different landmarks, indicating close agreement between model-derived landmarks and manually assigned labels. For all views and imaging sequences, no differences between the models' assessment of images and the readers' assessment of images were found for the anterior RV insertion angle or LV length. Model inference for a typical cardiac cine series took 610 msec with the graphics processing unit and 5.6 seconds with central processing unit.

Conclusion: A CNN was developed for landmark detection on both long- and short-axis CMR images acquired with cine, LGE, and T1 mapping sequences, and the accuracy of the CNN was comparable with the interreader variation.

Supplemental material is available for this article.

Cardiac MRI (CMR) is emerging as a mainstream modality for imaging the cardiovascular system for diagnosis and intervention. CMR has advanced beyond the scope of imaging the anatomy and can be used to acquire comprehensive quantitative measures of the myocardium. These include relaxometry T1, T2, and T2* measures (1,2) for the assessment of fibrosis, edema, and iron as well as for assessment of tissue composition for the fat fraction (3) and include physiologic measures for mapping of myocardial perfusion (4,5) and blood volume (6). These capabilities open new opportunities and simultaneously place new demands on image analysis and reporting. A fully automated solution brings increased objectivity and 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 for quantifying cardiac function, viability, and tissue properties (7). A recent study showed that 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 by using CNNs to measure the ejection fraction and other parameters with a performance level matching that of expert readers (9), and CNN measurements have demonstrated improved reproducibility in multicenter trials (8,10). Cardiac perfusion images have been successfully analyzed and reported on the MRI scanners (11) by 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 with use of the standard American Heart Association sector model (7). For long-axis views, the ventricular length can be measured if the valve and apical points can be delineated. Variation in

Abbreviations

A-RVI = anterior RV insertion point, C-LV = LV center point, CMR = cardiac MRI, CNN = convolutional neural network, LGE = late gadolinium enhancement, LV = left ventricular, MOLLI = modified Look-Locker inversion recovery, P-RVI = posterior RV insertion point, RV = right ventricular

Summary

A convolutional neural network (CNN) was developed for labeling landmarks on long- and short-axis cardiac MR images acquired by using cine, late gadolinium enhancement, and T1 mapping sequences, and the performance of CNN labeling was comparable with that of manual labeling.

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 was similar for different cardiac MRI views.
- Models were integrated with MRI scanners by using Gadgetron InlineAI, with less than 1 second of model inference time.

Keywords

Cardiac, Heart, Convolutional Neural Network (CNN), Deep Learning Algorithms, Machine Learning Algorithms, Feature Detection, Quantification, Supervised Learning, MR Imaging

the LV length is a useful marker and has been shown to be the principal component of LV pumping in patients with chronic myocardial infarction (14). Furthermore, cardiac landmark detection can be useful on its own for applications such as automated imaging section planning.

In this study, we developed a CNN-based solution for automatic cardiac landmark detection on CMR images. Detection was defined as the process of locating 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 on the image. The performance of the trained model was quantitatively evaluated by comparing the success rates between CNN labeling and manual labeling and by computing the Euclidean distance between manually derived and model-derived landmarks. To evaluate the feasibility of using models for CMR reporting, the model-derived and manually derived angle of the anterior RV insertion point (A-RVI) and LV length were used. To demonstrate their clinical feasibility, the trained CNN models were integrated with MRI scanners by using Gadgetron InlineAI (15) and were used to automatically measure the LV length from long-axis cine images. 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 long-axis series (two chamber, three chamber, and four chamber) and short-axis series (Fig 1). The following points were detected on different views: *(a)* short-axis view, the A-RVI, posterior RV insertion point (P-RVI), and LV center point (C-LV); *(b)* two-chamber view, the anterior and inferior points; *(c)* three-chamber view, the inferolateral and anteroseptal points; *(d)* four-chamber view, the inferoseptal and anterolateral points; and *(e)* long-axis view, the apical point. The trained CNN models were tested on cardiac cine images, late gadolinium enhancement (LGE) images, and T1 maps acquired by using 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 examinations were performed at the Barts Heart Centre (London, England), and all T1 MOLLI images were acquired at the Royal Free Hospital (London, England). Both long- and short-axis views were acquired for cine and LGE series. For T1 mapping, one to three short-axis sections were acquired per patient. The data used in this study were not used in prior publications.

Data were acquired with the required ethical and/or secondary audit use approvals or guidelines (as per each center), which permitted retrospective analysis of anonymized data without requiring written informed consent for secondary usage for the purpose of technical development, protocol optimization, and quality control. Institutions acquiring data were in the United Kingdom and were not subject to the Health Insurance Portability and Accountability Act. All data were anonymized and delinked for analysis, with approval being provided by the local Office of Human Subjects Research (exemption 13156). Appendix E1 (supplement) provides information about patient inclusion criteria.

Table 1 summarizes the training and test datasets. For training, a total of 34 089 images from 2329 patients (mean age, 54.1 years; 1471 men) were included—29 214 cine, 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 who underwent LGE imaging also underwent cine imaging. Data acquisition in every imaging period was consecutive. The test set consisted of 7723 images from 531 consecutive patients (mean age, 51.5 years; 323 men). The test data were acquired between May and June 2020. There was no overlap between the 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 by using both 1.5-T (four Magnetom Aera scanners, Siemens Healthineers) and 3-T (one Magnetom Prisma, Siemens Healthineers) MRI scanners. In the training set, 1790 patients were imaged with 1.5-T scanners and 539 patients were imaged with 3-T scanners. In the test set, 462 patients were imaged with 1.5-T MRI and 69 were imaged with 3-T MRI. Typically, 30 cardiac phases were reconstructed for each heartbeat for every cine scan. For training and testing purposes, the first phase (typically the end-diastolic phase) and the end-systolic phase were selected. Given that there was a large

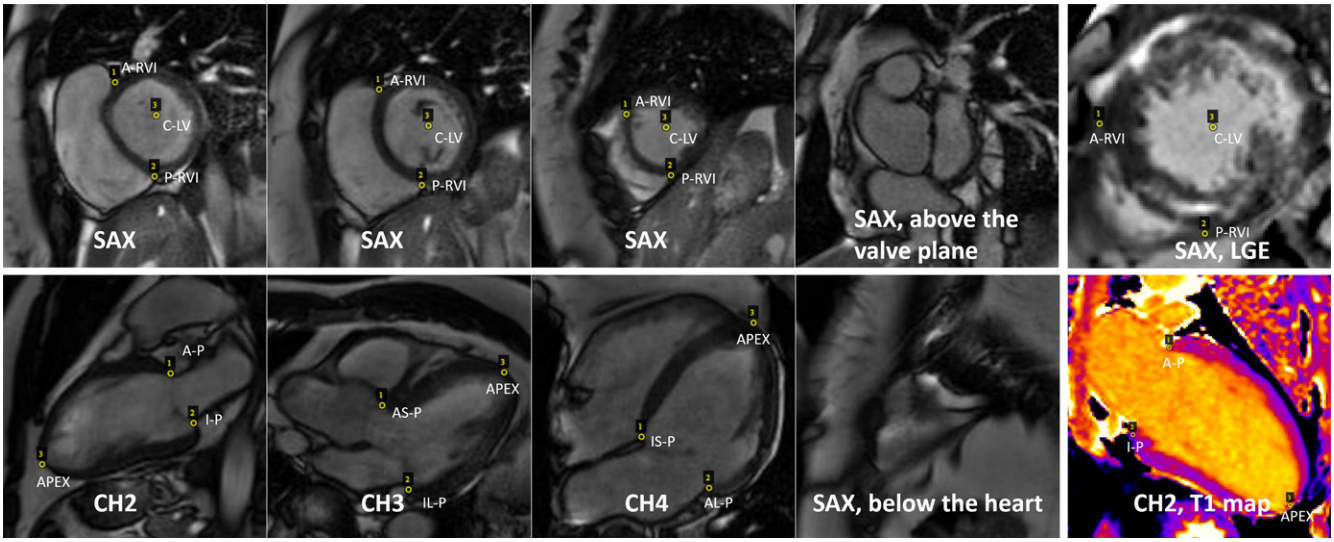

Figure 1: Example of cardiac MR images with landmarks. Three short-axis (SAX) views are shown on the top row. The first three images of the second row show examples of long-axis two-chamber (CH2), three-chamber (CH3), and four-chamber (CH4) views. The anterior point (A-P) and inferior point (I-P) were depicted on the two-chamber view. The inferolateral point (IL-P) and anteroseptal point (AS-P) were depicted on the three-chamber view, and the inferoseptal point (IS-P) and anterolateral point (AL-P) were depicted on the four-chamber view. The apical point (APEX) was depicted on all long-axis views. For the short-axis images, the anterior right ventricular (RV) insertion point (A-RVI), posterior RV insertion point (P-RVI), and left ventricular (LV) center point (C-LV) were depicted. Note that for some SAX sections (the rightmost column), no landmarks can be identified. The last column gives examples of late gadolinium enhancement (LGE) images and T1 maps. Transfer learning was applied to detect landmarks by using these imaging applications.

number of patients, these acquired cardiac phases represent a sufficiently broad variation. For those who underwent contrast-enhanced studies, the gadolinium-based contrast agent (gadoterate meglumine [Dotarem, Guerbet]) was administered at 4 mL/sec 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 imaging was performed with retrospective electrocardiographic gating (30 cardiac phases were reconstructed) and twofold parallel imaging acceleration by using generalized autocalibrating partially parallel acquisition, or GRAPPA (17). For the short-axis acquisition, eight to 14 sections were typically used in order to cover the LV area.

Phase-sensitive inversion recovery for LGE imaging.—Phase-sensitive inversion recovery LGE imaging was performed by using a free-breathing sequence (18) to enable coverage of the entire LV area while applying respiratory motion correction and averaging. Phase-sensitive LGE reconstruction (19) was used to achieve insensitivity to the inversion time. A previous study (20) showed that this free-breathing technique is more robust against respiratory motion and resulted in improved LGE image quality.

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

Data Preparation and Labeling

Because the acquired field of view may have varied among patients, all images were first resampled to a fixed 1-mm^2 pixel spacing and were 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 is large enough to cover the heart, as the MRI technicians generally positioned each patient so that the heart would be close to the center of the field of view. The use of cine MRI often causes a shadow across the field of view (Fig E1 [supplement]), as the tissue that is further away from receiver coils at the chest and spine will have a reduced signal intensity due to the inhomogeneity of the surface coil receiver sensitivity. To compensate for this shading, for every cine image in the dataset, a surface coil inhomogeneity correction algorithm (23) was applied to estimate the rate of the 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, and $P = .5$ was the probability of selecting the original version. This served as one data augmentation step. Additional details on other data augmentation procedures are found in Appendix E2 (supplement).

One reader (H.X., with 9 years of experience in CMR research and 3 years of experience in deep learning) manually labeled all images for training and testing. A second reader (J.A., with 3 years of experience in CMR clinical reporting) was invited to label part of the test dataset to assess interreader variation. J.A. labeled 1100 images (cine and LGE: 100 images for every long-axis view, 200 images for every short-axis view; 100 images for every T1 map).

The Visual Geometry Group Image Annotator software (*https://www.robots.ox.ac.uk/~vgg/software/via/*), or VIA, was used by both readers for the manual labeling of landmarks. The data labeling took approximately 150 hours in total. Table 1 shows the training and test datasets.

Table 1: Information for Training and Test Dataset Distribution and Acquisition

Imaging View	No. of Patients	No. of Images	Time Period
Training set			
All	2329	34089	…
Cine			December 18–29, 2017; January 2–28, 2018; January 2–April 19, 2020
CH2	2115	4232	
CH3	2102	4206	
CH4	2127	4256	
SAX	702	16520*	
LGE			January 2–February 29, 2020
CH2	599	599	
CH3	582	582	
CH4	599	599	
SAX	178	2018†	
T1 MOLLI			
SAX	202	1077	January 2–March 25, 2020
Test set			
All	531	7723	
Cine			May 1–July 3, 2020
CH2	347	694	
CH3	345	690	
CH4	347	692	
SAX	128	3008‡	
LGE			May 1–July 3, 2020
CH2	370	370	
CH3	370	370	
CH4	370	372	
SAX	96	1082§	
T1 MOLLI			
SAX	161	445	May 1–July 23, 2020

Note.—CH2 = two chamber, CH3 = three chamber, CH4 = four chamber, LGE = late gadolinium enhancement, MOLLI = modified Look-Locker inversion recovery, SAX = short axis.
* A total of 3803 images were acquired outside the left ventricle 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.

Model Development

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

Model Training

A variation of U-Net architecture was implemented (25,26) for heatmap detection. As shown in Figure 3, the network was organized as layers for different spatial resolutions. Specific details on the model architecture are described in Appendix E4 (supplement). The input to the model was a two-dimensional image (ie, to detect the landmarks from a time series of cine images, 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. The CNN output score tensor had dimensions of 400 × 400 × 4. To train the network, the Kullback-Leibler divergence was computed between the ground truth heatmap and the softmax tensor of the scores. Besides this entropy-based loss, the shape loss was further computed as the soft Dice ratio (27). The soft Dice ratio was computed as the product of two probability maps over their sum. The final loss was a sum of the entropy-based loss and the soft Dice ratio, which used both entropy-based information and region costs. This strategy of using a combined loss has previously been employed in deep learning segmentation and has been found to improve segmentation robustness (28,29).

For the long-axis views, all views were trained together as a multitask learning task. Because 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 two-chamber, three-chamber, or four-chamber images, and refined network weights.

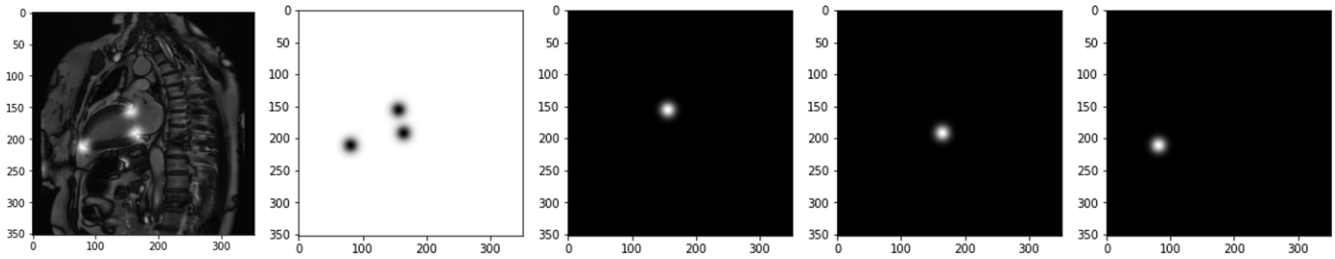

Figure 2: The landmark detection problem can be reformulated as a semantic segmentation problem. Every landmark point on the two-chamber image on the left can be convolved with a Gaussian kernel and converted into a spatial probability map or heatmap (upper row, from left to right: probability for background,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

The data for training were split, with 90% of all patients for training and 10% for validation. The training and validation datasets were split on a per-study basis, such that there was no mixing of patients between the two datasets. The Adam optimizer was used, and the initial learning rate was 0.001, the β values were 0.9 and 0.999, and the ε value was 1×10^{-8}. The learning rate was reduced by 2 whenever the cost function plateaued. Training lasted 50 epochs (approximately 4 hours), and the best model was selected as the one demonstrating the highest performance on the validation set. The CNN models were implemented by using PyTorch (30), and training was performed on a personal computer running Ubuntu 20.04 with four NVIDIA GeForce GTX 2080Ti graphics processing unit cards, each with 11 GB of random access memory. Data parallelization was used across multiple graphics processing unit cards to speed up training.

Because there were more cine images than LGE and T1 MOLLI images, a fine-tuning strategy was implemented by using transfer learning. For both long- and short-axis images, a model was first trained with the cine dataset and then fine-tuned with either the LGE or T1 training set. Transfer learning was implemented to first train the neural networks with the cine data as the pretrained model. The LGE or T1 data were used to fine-tune the pretrained model with a reduced learning rate (31). To perform the fine-tuning, the initial learning rate was set at 0.0005, and the models were trained for a total of 10 epochs. For each type of image, separate models were trained for landmark detection on short- and long-axis images, 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 with 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 the detected landmarks and assigned labels was computed and reported separately for different section views and different landmark points. Results from model detection and manual labeling were compared, and the Euclidean distance between the findings of the two readers was reported.

The detected key points were further processed to compute two derived measurements: the angle of the A-RVI to the C-LV for short-axis views and the LV length for long-axis views, the latter of which was computed as length from the detected apical point to the midpoint of two valve points (32). The model-derived results were compared with the manual labels. The results of the first reader were compared with those of the second reader to obtain references for interreader variation.

Results are presented as means ± standard deviations (instead of standard errors). A paired *t* test was performed, and $P < .05$ was considered to indicate a statistically significant difference (Matlab R2017b, MathWorks).

To test the sensitivity of detection performance in terms of the size of the Gaussian kernel used to generate the heatmap, two additional models were trained for long-axis cine images, with σ values equaling 6.0 and 2.0 pixels. Detection performance was compared across different kernel sizes for cine long-axis test images.

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

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

Model Deployment

To demonstrate the clinical relevance of landmark detection of CMR, an inline application was developed to automatically measure the LV length from long-axis cine images on the MRI scanner. The trained long-axis model was integrated with MRI scanners by using the Gadgetron InlineAI toolbox (15). Although 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 imaging. The resulting landmark detection and LV length measurements were displayed and available for immediate evaluation prior to the next image series being obtained. Figure E4 (supplement) provides more information for this landmark de-

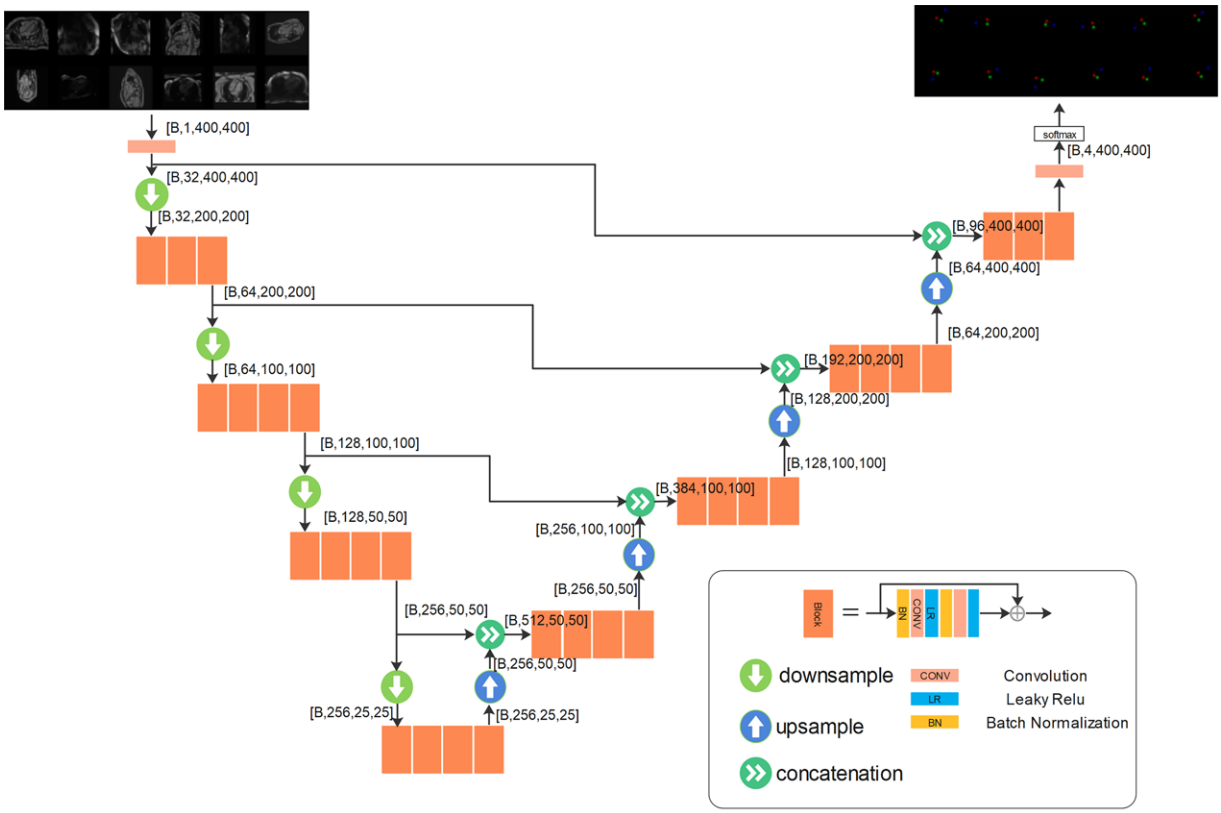

Figure 3: The backbone convolutional neural network developed for landmark detection has a U-Net structure. More layers can be inserted in both the downsampling branch and the upsampling branch, and more blocks can be inserted into each layer. The output layer outputs the per-pixel scores, which go through softmax function. For the landmark detection on long-axis images, data from three views were used together to train one model. As shown in the input, every minibatch was assembled by using randomly selected images from three views and was used for back propagation. A total of four layers with three or four blocks per layer were used in this experiment. The output tensor shapes were reported by using the format [B, C, H, W], where B is the size of the minibatch, and C is the number of channels, and H and W are the image height and width. Input images have one channel for image intensity, and the output has four channels for three landmarks and the background. The illustration for outputs plots three color-coded landmark channels and omits the background channel.

tection application. This example can be viewed in the Movie. Appendix E6 (supplement) provides additional information on model deployment and processing times.

The source files used 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 that 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, the model successfully detected landmarks on 99.8% (2072 of 2076 images; no false-positive findings) of the two-chamber, three-chamber, and four-chamber long-axis images and on 96.6% (2906 of 3008 test images; 24 false-positive findings) of the short-axis images. For the LGE, the model successfully detected landmarks on 99.4% (1105 of 1112 images; two false-positive findings) of all long-axis views and on 97.6% (1056 of 1082; 11 false-positive findings) of all short-axis views. For T1 mapping, the model successfully detected landmarks on 98.7% (439 of 445 images; no false-positive findings) of the images.

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

For the 102 mislabeled short-axis images acquired using cine imaging, the A-RVI was missed on 51, the P-RVI was missed on 25, and the C-LV was missed on 13. Half of the errors were found to be on the most basal and apical sections (defined as top two sections, or the last section for a short-axis series). For the 26 mislabeled short-axis images acquired by using LGE imaging, the A-RVI was missed on seven, the P-RVI was missed on one, and the C-LV was missed on two. A total of 11 errors were due to unnecessary landmarks being detected in sections outside the LV area. All detection failures on T1 MOLLI images (six of 445 test images) were failures to detect the P-RVI, which was due to unusual imaging planning for one patient. Examples of mislabeled short-axis cases can be found in Figure E3 (supplement).

Euclidean Distances between Readers and the CNNs

For all images on which detection was successful, the Euclidean distances between the model-assigned labels and the expert-assigned labels were computed. Tables 3 and 4 show the Euclidean distances and two derived measurements, which were reported separately for all imaging views and imaging se-

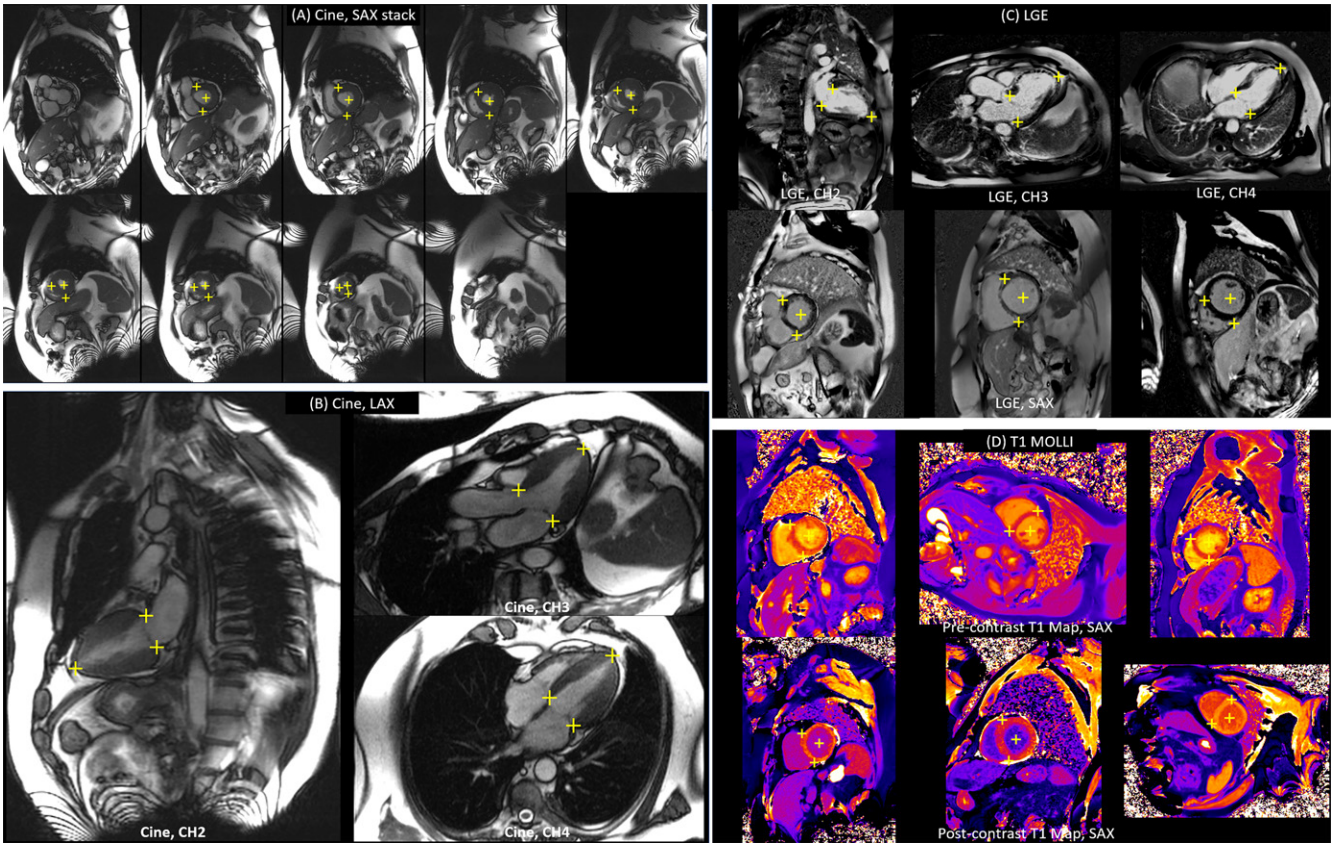

Figure 4: Examples of landmark detection. The left panels show landmark detection on **(A)** long-axis (LAX) and **(B)** short-axis (SAX) cine images. The right panels are examples of detection on **(C)** late gadolinium enhancement (LGE) and **(D)** T1 mapping modified Look-Locker inversion recovery (MOLLI) images. CH2 = two chamber, CH3 = three chamber, CH4 = four chamber.

Table 2: Detection Rate across Three Imaging Applications and All Tested CMR Views

Image Type	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	98.7 (439/445)

Note.—Data are percentages, with numbers of images in parentheses. CH2 = two chamber, CH3 = three chamber, CH4 = four chamber, CMR = cardiac MRI, LGE = late gadolinium enhancement, MOLLI = modified Look-Locker inversion recovery, SAX = short axis.

quences. The distances between the trained model and the first reader ranged from 2 to 3.5 mm. Figure 5 shows detection examples with model-derived and manually derived landmarks and their Euclidean distances, which demonstrate that the model-derived landmarks were in close proximity to the manually assigned labels. The mean Euclidean distances ± standard deviations for the long-axis cine and LGE images were 2.5 mm ± 1.9 and 3.0 mm ± 2.4. For the short-axis views, the mean Euclidean distance (across all landmarks) for cine, LGE, and MOLLI images were 2.5 mm ± 1.8, 2.4 mm ± 2.5, and 2.2 mm ± 2.0, respectively.

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

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-pixel design to determine the sensitivity to the kernel size. The mean distances to the manually assigned

Table 3: Landmark Detection on Two-, Three-, and Four-Chamber Views

Image Type and Landmark	Euclidean Distance (mm)		Left Ventricular Length Difference (%)			
	First vs CNN	First vs Second	First vs CNN	*P* Value	First vs Second	*P* Value
Cine						
CH2			2.0 ± 1.7	.42	1.9 ± 1.4	.95
A-P	2.1 ± 1.8	2.8 ± 1.9				
I-P	2.4 ± 2.0	3.0 ± 3.9				
APEX*	2.4 ± 1.8	4.1 ± 2.8				
CH3			1.5 ± 1.3	.79	2.0 ± 1.7	.97
IL-P	2.4 ± 1.7	2.8 ± 1.6				
AS-P*	2.2 ± 1.5	4.0 ± 2.4				
APEX*	3.2 ± 2.4	3.8 ± 2.1				
CH4			1.4 ± 1.2	.92	2.0 ± 1.4	.77
AL-P	3.4 ± 2.1	3.5 ± 2.0				
IS-P*	2.1 ± 1.7	2.6 ± 1.6				
APEX	2.8 ± 1.9	2.8 ± 1.6				
LGE						
CH2			2.7 ± 2.5	.16	2.5 ± 2.1	.82
A-P	2.9 ± 2.6	3.3 ± 2.0				
I-P	3.4 ± 2.7	3.4 ± 2.5				
APEX	3.1 ± 2.6	3.4 ± 2.5				
CH3			2.6 ± 2.6	.37	2.9 ± 2.2	.34
IL-P	3.4 ± 3.1	3.5 ± 2.1				
AS-P*	2.7 ± 2.1	3.6 ± 2.3				
APEX	3.3 ± 2.8	3.3 ± 2.5				
CH4			2.0 ± 1.4	.13	1.9 ± 1.9	.53
AL-P	3.1 ± 1.6	3.3 ± 2.2				
IS-P	2.0 ± 1.5	2.5 ± 2.3				
APEX	2.7 ± 1.2	2.1 ± 1.6				

Note.—"First vs CNN" indicates the comparisons of manual labels from the first reader with labels from the trained model, and "First vs Second" indicates the comparisons between the two readers for the test data labeled by both. Unless otherwise specified, data are means ± standard deviations. AL-P = anterolateral point, A-P = anterior point, APEX = apical point, AS-P = anteroseptal point, CH2 = two chamber, CH3 = three chamber, CH4 = four chamber, CMR = cardiac MRI, CNN = convolutional neural network, IL-P = inferolateral point, I-P = inferior point, IS-P = inferoseptal point, LGE = late gadolinium enhancement.
* Indicates $P < .05$ (paired *t* test) for the comparison of the distance between the "First vs CNN" and "First vs Second."

landmarks from the first reader were 2.3 mm ± 1.6 and 2.2 mm ± 1.6 for models trained with σ of 2.0 and 6.0 pixels, and no differences were observed when compared with a σ of 4.0. The LV length was estimated for σ of 2.0 and 6.0 and showed no differences compared with measurements performed by the experts ($P > .2$ for all views). Figure E5 (supplement) provides an example of landmark detection with computed probability maps for three models, which shows 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 training and ease deployment. Among images from the entire training dataset, the 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 MRI scanner.

Landmark detection by 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 computed the x–y coordinates of landmark points, and L2 regression loss was used for training. Second, landmark coordinates were implicitly

Table 4: Landmark Detection on CMR Short-Axis Views

Image Type and Landmark	Euclidean Distance (mm)		A-RVI Angle Difference (degrees)			
	First vs CNN	First vs Second	First vs CNN	*P* Value	First vs Second	*P* Value
Cine			1.3 ± 3.4	.14	−0.7 ± 4.1	.89
A-RVI	3.1 ± 1.8	3.5 ± 2.6				
P-RVI	2.4 ± 2.1	2.7 ± 1.6				
C-LV	2.0 ± 1.1	2.4 ± 1.2				
LGE			0.14 ± 2.9	.92	−2.0 ± 4.5	.62
A-RVI	3.0 ± 3.2	3.6 ± 3.1				
P-RVI	2.8 ± 2.6	3.3 ± 2.6				
C-LV*	1.5 ± 0.9	2.3 ± 1.1				
T1 MOLLI			1.6 ± 3.1	.31	1.7 ± 3.9	.41
A-RVI	2.5 ± 2.0	3.0 ± 2.8				
P-RVI	2.5 ± 2.6	2.5 ± 2.0				
C-LV	1.6 ± 1.0	2.0 ± 1.1				

Note.—"First vs CNN" indicates the comparisons of manual labels from the first reader with labels from the trained model, and "First vs Second" indicates the comparisons between the two readers for the test data labeled by both. Unless otherwise specified, data are means ± standard deviations. A-RVI = anterior right ventricular insertion point, C-LV = left ventricular center point, CMR = cardiac MRI, CNN = convolutional neural network, LGE = late gadolinium enhancement, MOLLI = modified Look-Locker inversion recovery, P-RVI = posterior right ventricular insertion point.
* Indicates $P < .05$ (paired t test) for the comparison of the distance between "First vs CNN" and "First vs Second."

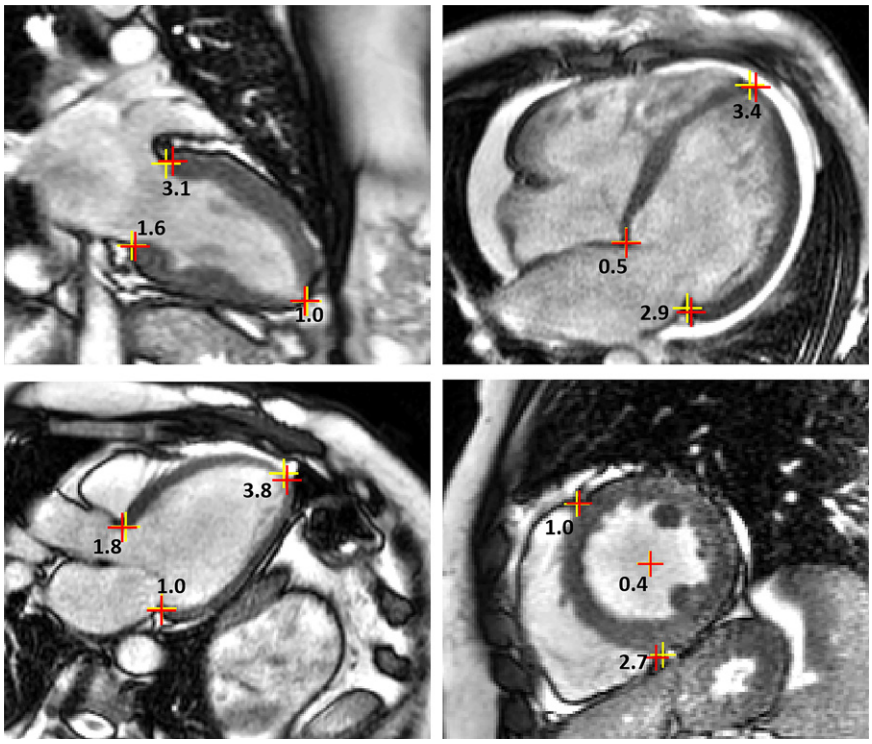

Figure 5: Examples of landmark Euclidean distances. For every pair of manually delineated (red crosses) and model-delineated (yellow crosses) landmarks, the distance (in millimeters) is labeled.

coded as heatmaps. In this context, the detection problem was reformulated as a segmentation problem. In human pose estimation, the segmentation-based models outperformed regression models (24,36). Here, fewer landmarks were detected and were more sparsely distributed spatially. The human pose images had much more variation than images of human faces, which often had been preprocessed as frontal position (37). It is easier for heatmap detection to handle landmark occlusion. For example, in Figure 1, some images may not include the targeted landmarks, which is represented by a 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 CMR (38). This study trained a heatmap detection model on 892 long-axis and 493 short-axis steady-state free precession cine images. The midvalve plane and apical points were automatically detected and compared with manual localization, with a mean distance of approximately 5–7 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 the RV insertion points.

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 the 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 that had substantially different contrasts, to enable automated reporting and measurement of global longitudinal shortening. Detection was slightly less accurate on basal and apical short-axis sections. In these regions, the "ambiguity" of anatomy increases, leading the model to demonstrate more variance in data labeling and more difficulties with providing the correct inference. Additional discussion can be found in Appendix E7 (supplement).

There were limitations to this study. First, a single reader labeled the entirety of the datasets. Because of limited research resources, the second reader only labeled a portion of the 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 anatomic landmark (eg, the RV center), imaging sequence, or cardiac view, more training data would be required. The 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 anatomic landmarks. Third, the data used in this study were collected from a single MRI vendor (Siemens). A recent study (41) reported that the performance of deep learning models trained on imagers from one vendor may decrease when used on imagers from different vendors, although augmentation was used to improve robustness. Further validation will be required to extend the proposed CNN models for use with CMR imagers from other vendors. It is very likely to require further data curation and training. Fourth, because of the availability of different imaging sequences, not all imaging sequences were performed across both of the included institutions, which limits the evaluation of generalizability across hospitals. We expect that the on-scanner deployment could enable the proposed models to be used in more hospitals and that further studies could provide more comprehensive datasets. Other limitations are related to 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 that an even larger configuration may be needed if the imaging planning is far off center. The model can be retrained with an even larger field of view, but the inline detection-result feedback could be used to alert readers to the need for adjustment or repeat acquisition.

In this study, a CNN-based solution for landmark detection on CMR 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 that the CNN's detection performance was comparable with the interreader variation. On the basis of the detected landmarks, the RV insertion points and LV length can be reliably measured.

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.

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. Bhuva 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 [abstract]. 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; 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, Puntil 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 Assist Interv 2009;12(Pt 2):741–749.
24. Belagiannis V, Zisserman A. Recurrent human pose estimation. In: Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 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, Cardoso MJ. 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. Vol 10553, Lecture Notes in Computer Science. 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: Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2018; 624–628.
29. Jadon S. A survey of loss functions for semantic segmentation. ArXiv 2006.14822 [preprint] http://arxiv.org/abs/2006.14822. Posted June 26, 2020. Accessed July 23, 2021.
30. Steiner B, Devito Z, Chintala S, et al. PyTorch: an imperative style. In: Proceedings of Advances in Neural Information Processing Systems 32 (NeurIPS 2019). San Diego, Calif: Neural Information Processing Systems Foundation, 2019.
31. Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. J Big Data 2016;3:9.
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 [preprint] http://arxiv.org/abs/1710.00977. Posted October 3, 2017. Accessed July 23, 2021.
34. Colaco S, Han DS. Facial keypoint detection with convolutional neural networks. In: Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2020; 671–674.
35. Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient object localization using Convolutional Networks. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2015; 648–656.
36. Pfister T, Charles J, Zisserman A. Flowing ConvNets for human pose estimation in videos. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2015; 1913–1921.
37. Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 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 [preprint] https://arxiv.org/abs/1804.04963. Posted April 13, 2018. Accessed July 23, 2021.
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.