UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Advancing Cardiovascular MRI Acquisition Through Deep Convolutional Neural Network-Based Localization

Permalink
https://escholarship.org/uc/item/0j43m8dm

Author
Blansit, Kevin

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

**Advancing Cardiovascular MRI Acquisition Through
Deep Convolutional Neural Network-Based Localization**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioinformatics and Systems Biology
with a Specialization in Biomedical Informatics

by

Kevin William Blansit

Committee in charge:

Professor Albert Hsiao, Chair
Professor Francisco Contijoch
Professor Elliot McVeigh
Professor Olivier Harismendy
Professor Michael Hogarth

2021

The dissertation of Kevin Blansit is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego
2021

DEDICATION

To my Father that has always inspired me,
To my Mother that has always been proud of me,
And to my Fiancé that has always believed in me.

I moreover would not be the student I am today if it were not for my peers and friends I have met along the way throughout graduate school. Christine, Sam, Ileana, Isaac, and Daniel have all stood out in their generosity and helping my make the last 7 years in Southern California my home.

EIPGRAPH

Measure what is measurable,
and make measurable what is not.

Galileo

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMNENTS

Chapter 4 is coauthored with Blansit, Kevin; Bahrami, Naeim; Rubel, Ame; Hsiao, Albert. The dissertation author was the primary investigator and author of this paper.

VITA

2011 — University of California Davis
    Bachelor of Science, Biochemistry and Molecular Biology

2016 — University of California Los Angeles
    Master of Science, Health Services Research

2021 — University of California San Diego
    Doctor of Philosophy, Bioinformatics and Systems Biology with a Specialization in Biomedical Informatics

ABSTRACT OF THE DISSERTATION

Advancing Cardiovascular MRI Acquisition Through
Deep Convolutional Neural Network-Based Localization

By

Kevin William Blansit

Doctor of Philosophy in Bioinformatics and Systems Biology
with a Specialization in Biomedical Informatics

University of California San Diego, 2021

Professor Albert Hsiao, Chair

Cardiac MRI is the gold standard for quantification of cardiac volumetry, function, and blood flow. Despite the wealth of information that may be gleaned from these acquisitions, its use has been limited primarily to academic and specialty clinics due to the need for specialty trained physicians and technologists required for planning of these scans.

Recently, deep convolutional neural networks (DCNNs) have shown promise in automating various aspects of radiology workflows in cardiac MRI, including tasks such as segmentation, disease prognosis, and tissue characterization. In addition to these tasks, DCNNs have been employed for anatomic localization. However, an often-cited limitation of clinically focused DCNNs is uncertainty of how well an algorithm will perform outside of its original training environment. Additionally, these systems are

often seen as "black boxes", which fail to provide an explanation of how the underlying system functions. Designing mechanisms that provide a rationalization for predictions may increase the confidence of clinical end-users in these systems is therefore essential for the adoption of any medically focused DCNN system.

With these concerns in mind, I propose automating the planning of cardiac MR imaging planes using a system of DCNNs. In the first chapter, I explore the potential of automating the prescription of long and short-axis imaging planes. To ensure the DCNN retained a rationalization for the plane prediction, we subdivided this challenge into two separate problems: 1) to identify the position of the apex and valve landmarks on either short or long-axis images, and 2) to use these inferred localizations to plan the short and long-axis image planes.

In my second chapter, I further extend this technique of planning the cardiac MR imaging planes by planning the dedicated axial and vertical long-axis imaging planes and explore the potential of this algorithm to perform across a varying range of clinical environments. Moreover, I assess the potential of integrating information from multiple image views to overcome the inherent anisotropic nature to cardiac MR.

Finally, in my third chapter, I explore the potential of efficiently adapting a DCNN to a new clinical environment by integrating uncertainty sampling and transfer learning methods. This includes the development of two novel uncertainty metrics which can be used to assess a DCNN's confidence in prediction.

Chapter 1: Introduction

1.1 Background

Cardiac MRI is the quantitative gold standard for measuring cardiac volumetry, function, and blood flow [1]. The human heart forms a natural axis, from which one can define both short and long-axes. To provide cardiovascular radiologist standardized 2-dimesntional views of the heart, cardiac MR imaging planes are defined by these axes [2]. Acquiring these long and short-axis images is an iterative process. Once an initial view has been acquired, subsequent views are then planned on the basis of the image position of the left ventricular apex, mitral valve, tricuspid valve, and aortic valve (Figure 1). Given the complex anatomy of the heart, cardiac MRI acquisition requires specialized training among MRI operators, limiting this technique mostly to academic and sub-specialty clinics [3].

Prior authors have attempted to automate the planning of cardiac MR using a variety of machine learning methods. For example, Jackson et al proposed using a semiautomated approach utilizing expectation maximization for planning the short and long-axis imaging planes [4]. This however required a technologist to label blood pools, requiring specialized knowledge of cardiac anatomy. Other methods utilized mesh-based segmentations based on canonical geometry or an anatomic atlas [5], [6]. These methods, however, relied on additional volumetric images not typically acquired for a cardiac MRI.
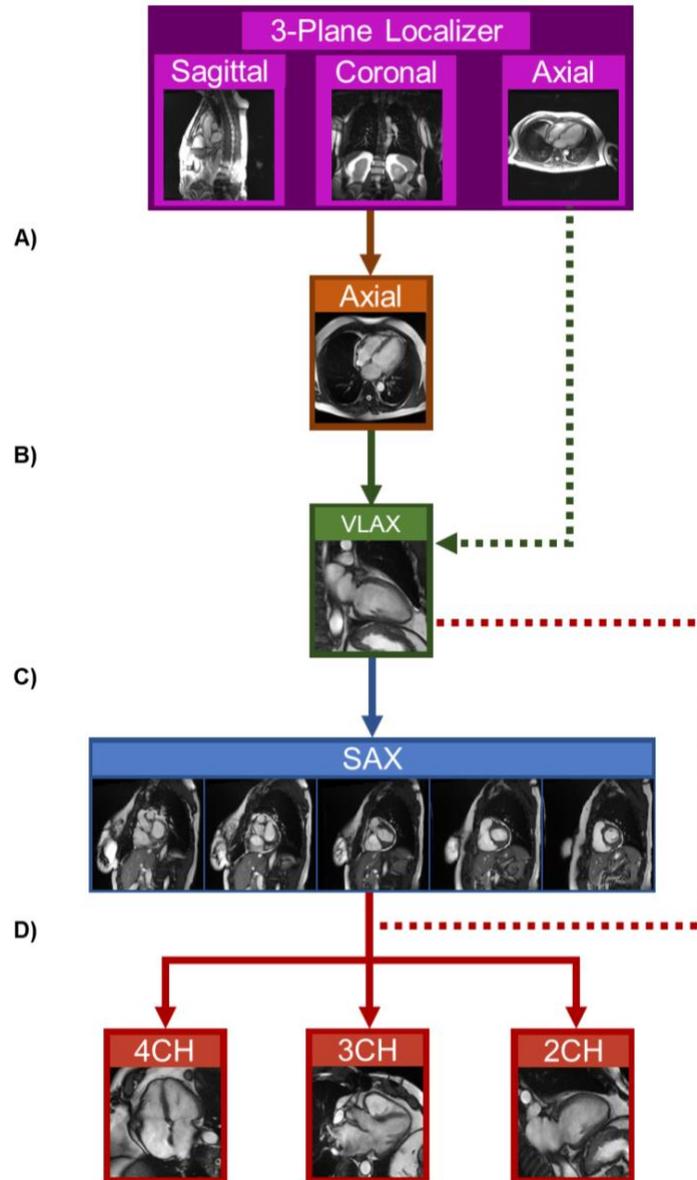
Figure 1*: Schematic for prescribing cardiac*
A) A 3-plane localizer series is acquired, consisting of sagittal, coronal, and axial views. B). From the coronal images, a dedicated axial stack of images can be prescribed to ensure adequate cardiac coverage. C) From either a 3-plane or dedicated axial stack of images, the vertical-long axis is prescribed by identifying the image locations of the left ventricular apex and mitral valve. D) The short-axis stack of images is prescribed from a vertical long-axis by identifying the image location of the apex and mitral valve. E). Dedicated long-axis images for the 4-, 3-, and 2-chamber images are acquired by identifying the image location of the mitral, tricuspid, and aortic valves. Cross referencing a previously acquired long-axis image (such as the vertical long-axis in dotted lines) may be performed to enforce acquiring these prescriptions along the long axis of the heart.

More recently, deep convolutional neural networks (DCNNs) have shown promise for automating visual tasks in cardiac MR imaging, including disease stratification, anatomic segmentation, and tissue characterization [7]–[10]. In addition to these tasks, DCNNs have been employed for anatomic localization of cardiac landmarks [11]. While there exists a number of ways to localize landmarks, one popular method for landmark localization is heatmap regression [12]–[16]. Prior authors have proposed other localization strategies, such as directly regressing the image position coordinates with a DCNN [17]–[19]. This strategy however requires the use of fully connected layers to learn the non-linear relationship between imaging features and the coordinate position. Therefore, fully convolutional networks (such as U-Net) are not able to preserve spatial information throughout the DCNN, potentially diminishing their capacity to recognize spatial information [20], [21]. More recently, Duan et al proposed a localization strategy in which pixels use binary classification for each landmark [11]. While this strategy is compatible with fully convolutional networks, most pixels will not contain a landmark, creating a highly imbalanced class problem [22]. In contrast to these other proposed localization methods, heatmap regression creates pseudoprobability heatmaps centered around each landmark [23]. This strategy may be more advantageous as it provides an attention map that the DCNN can refine with training, while additionally providing a direct attention map that is relevant to understanding how the DCNN functions. It however remains unclear if this approach of landmark localization can accurately localize the apex and valve landmarks and use these to prescribe the long and short-axis views.

Despite previous successes of DCCNs to improve clinical workflow in cardiac MRI, only a handful of proposed DCNNs have been implemented into clinical practice [10], [24]. Currently, the main limitations of implementing DCNNs into clinical practice are as follows: First, DCNNs are oftentimes trained on a limited dataset, which may limit their generalizability [25], [26]. One often cited limitation of machine learning algorithms is uncertainty around generalizability, where the performance of an algorithm may vary outside the environment where it was trained. For example, multiple studies have shown poorer DCNN performance on external vs. internal test data [27], [28]. This is particularly true in the context of MRI, where different image acquisition parameters may alter the image appearance, limiting the generalizability across clinical sites [25], [29]–[31]. Other studies have shown that curating a diverse dataset can help with external generalizability [32], [33]. It may therefore be possible to provide wide generalizability to a DCNN by training over a wide range of visually distinct data, such as MR images acquired with multiple field strengths. Despite the promise of this approach, however, curating a diverse biomedical imaging dataset may be challenging due to legal, ethical, or logistical obstacles [34]–[37]. Given these constraints, there has been prominent attention given to improving the data-efficiency of DCNN training – where one is able to achieve a similar level of performance with a reduced quantity of data [38]. Comparatively less attention, however, has been paid to improving DCNN training data-efficiency within the context of different MRI parameters.

Secondly, DCNN based systems are often criticized for their lack of explainability in how the system determines a result [39]–[42]. Given the complex structure of DCNNs, it is often challenging to determine the source of model error, which decreases

4

the confidence in clinical end-users of the system. Prior studies have highlighted the importance of providing clinical end users information on how these DCNN systems operate [43], [44]. Moreover, in the case of an incorrect DCNN, important design considerations must be given to design an appropriate failure mechanism that does not impede or halt clinical workflows. Failure to provide appropriately designed failsafes could provide distrust in the system, limiting the adoption of machine learning based systems [45]. Therefore, it is an important design and implementation consideration of how to appropriately design clinical DCNN-based systems.

Third and finally, most DCNNs often operate on a single image view. However, a limiting factor to cardiac MRI is that stacks of images are often highly anisotropic, and may have poor through-plane slice resolution [46]. This may result in a diminished ability to visualize and analyze apical slices of a short-axis image –necessary for ensuring that the 4-, 3-, and 2-chamber prescriptions conform to the long-axis of the heart. To overcome this, human MRI operators may optimally use multiple imaging planes for the planning of cardiac planes. Currently, it remains unclear how to integrate information from multiple imaging planes to accurately represent the 3-dimesnional nature of the heart. Prior authors have proposed various multiview analyses to improve cardiac segmentation performance [11], [47]. However, these implementations have relied on fusing multiple DCNN inputs, which may not allow for the independent optimization of each DCNN.

1.2 Outline of the Dissertation

In my first chapter, I explore if DCNNs can automatically plan the long and short-axis imaging planes. To improve the interpretability of my approach, I break this

problem down into two separate subproblems: 1) whether heatmap localization is a feasible method to localize the apex and valve positions on images routinely collected during a cardiac MRI exam and 2) whether these localizations can then be used to accurately plan the cardiac long and short-axis imaging planes.

In my second chapter, I extend this initial proof of concept work of landmark localization for planning an entire cardiac MRI scan. I also demonstrate the feasibility of using segmentation to identify the cardiac boundaries for prescribing a dedicated axial stack of images. Using this system, I further explore whether incorporating a diversity of field strength data enhances the performance of this system across multiple clinical environments. I then further evaluate the importance of integrating imaging information from multiple views of the heart to improve prescription of cardiac MR imaging planes.

In my third chapter, I explore the potential of adapting DCNNs to new clinical environments by combining uncertainty sampling and transfer learning. I first create a model system based on the visual differences between cardiac MR images acquired at either 1.5T or 3T field strength. I then develop two novel uncertainty metrics, based either on strength of DCNN prediction or the spatial variance of test-time rotational augmentation, and assess their potential to reduce the number of training cases required to generalize a DCNN to a new clinical environment.

This introduction was adapted in part from "Deep Learning–based Prescription of Cardiac MRI Planes" in Radiology Artificial Intelligence, 2019 by Kevin Blansit, Tara Retson, Evan Masutani, Naeim Bahrami, and Albert Hsaio. The dissertation author was the primary investigator and author of this paper.

Chapter 2: Deep Learning-based Prescription of MR Cardiac Imaging Planes

2.1 Abstract

Purpose:

Develop and evaluate a system to prescribe imaging planes for cardiac MRI, based on deep learning (DL) based localization of key anatomic landmarks.

Methods:

With HIPAA-compliance and IRB waiver of informed consent, we retrospectively collected and annotated landmarks on 892 long-axis (LAX) and 493 short-axis (SAX) cine SSFP series from cardiac MRIs at our institution between February 2012 to June 2017. U-Net-based heatmap regression was used for localization of cardiac landmarks, which were used to compute cardiac MR imaging planes.

Performance was evaluated by comparing localization distances and plane angle differences between DL predictions and ground truth. We further compared the plane angulations from DL to those prescribed by the technologist at the original time of acquisition. Data was split into 80% for training and 20% for testing, and results confirmed with 5x cross-validation.

Results:

On LAX images, DL localized the apex within 12.56±19.11 mm and the mitral valve within 7.68±6.91 mm. On SAX images, DL localized the aortic valve within 5.78±5.68 mm, mitral valve within 5.90±5.24 mm, pulmonary valve within 6.55±6.39 mm, and tricuspid valve within 6.39±5.89 mm. Based on these localizations, average angle bias and mean error of DL predicted imaging planes relative to ground truth annotations were -1.27±6.81°, 4.93±4.86° for SAX; 0.38±6.45°, 5.16±3.80 for 4-

chamber; 0.13±12.70°, 9.02±8.83° for 3-chamber; and 0.25±9.08°, 6.53±6.28° for 2-chamber prescriptions.

Conclusions:

DL-based anatomic localization is a feasible strategy for planning cardiac MRI planes. We show that this approach can produce imaging planes comparable to those defined by ground truth landmarks.

2.2 Introduction

Cardiac MRI is the gold standard for quantification of cardiac volumetry, function, and blood flow [1]. Cardiac MRI can be performed as a series of sequential image acquisitions, where earlier images inform the prescription of subsequent planes. This approach is typically composed of multiple acquisitions, including a short-axis (SAX) stack and multiple long-axis (LAX) planes, requiring multiple breath-holds by the patient. A key component of acquiring these images is the identification of specific cardiac structural landmarks by a physician or trained technologist. Proficient acquisition of high quality scans therefore requires extensive anatomical and technical expertise [48], [49]. Cardiac MRI has in large part been predominantly limited to major academic institutions and subspecialty centers, which to some extent may be related to the availability of specialized expertise [50]–[52].

Prior attempts to help automate the prescription of cardiac imaging planes predominantly have used traditional machine-learning based approaches for image analysis. Previously, Jackson et al proposed a semi-automated approach using expectation-maximization for planning cardiac MRI planes [4]. Part of this approach however required manual technologist labeling of blood pools. This method was

developed and tested in 50 healthy subjects, but not validated in a clinical population which may have greater variation in anatomical morphology. Another approach has used a mesh-based segmentation model based on canonical geometry or an anatomical atlas [5], [6]. However, these approaches used an additional volumetric scan not typically acquired during cardiac MRI.

Deep learning (DL) techniques have recently gained popularity for a variety of computer vision tasks in medical imaging ranging from disease risk stratification, segmentation of anatomic structures, and quantification of imaging features [53]–[56]. DL has also been used for anatomic localization [23], [57]. For example, this method has been previously used to localize bony structures in hand x-ray and MR images [14]. We hypothesize that this technique may be advantageously applied to localize key anatomic landmarks that define the cardiac imaging planes. Furthermore, we hypothesize that DL-based localizations of these landmarks may be sufficiently accurate to prescribe each of the cardiac imaging planes.

2.3 Methods

2.3.1 Patients and Data Description

With HIPAA compliance and IRB approval, we retrospectively collected 482 cardiac MRI studies performed on a 1.5T MRI scanner between February 2012 to June 2017. Within these studies there were 892 long-axis (LAX) cine SSFP series (including 257 4-chamber, 207 3-chamber, 197 2-chamber, and 231 other LAX views) and 493 short-axis (SAX) cine SSFP series. Of these studies, 303/482 (62.86%) were male and 179/482 (37.14%) were female patients, with an age range from 12 to 90. A typical imaging protocol from our institution is shown in Supplemental Figure 1.

Radiology resident Tara Retson was trained by a board-certified cardiac radiologist Albert Hsiao with ten years of experience to identify and annotate cardiac landmarks on each MRI series. Albert Hsiao had final approval of all ground truth annotations. LAX images were annotated for mitral valve (MV) and apex, while SAX stacks were annotated for aortic valve (AV), MV, pulmonic valve (PV), and tricuspid valve (TV). Cine SSFP images were each acquired on a 1.5T MRI scanner with mean flip angle of 54 (range: 45 to 60), Matrix = 256x200, mean field of view of 351 mm (range: 290 mm to 440 mm), slice thickness of 8 mm (range: 5 mm to 10 mm), mean repetition time of 3.96 ms (range: 3.14 ms to 4.45 ms), and echo time of 1 ms. SAX stack images were acquired in intervals of 10 mm (range: 5 mm to 13 mm).

2.3.2 Long-Axis Landmark Localization

To localize the LAX landmarks (apex and MV), we implemented a 2D U-net modified for heatmap regression, trained on a variety of LAX images (Figure 2) [14], [20]. The final convolutional layer was replaced with a linear activation and a kernel size of 1 using L_2 loss. Channel-by-channel isotropic gaussian pseudoprobability heatmaps centered at each localization (representing probability densities of landmark localization) were created for the apex and MV [57]. Deep learning (DL) predicted localizations were defined by the maximal index of the predicted pseudoprobability heatmaps.
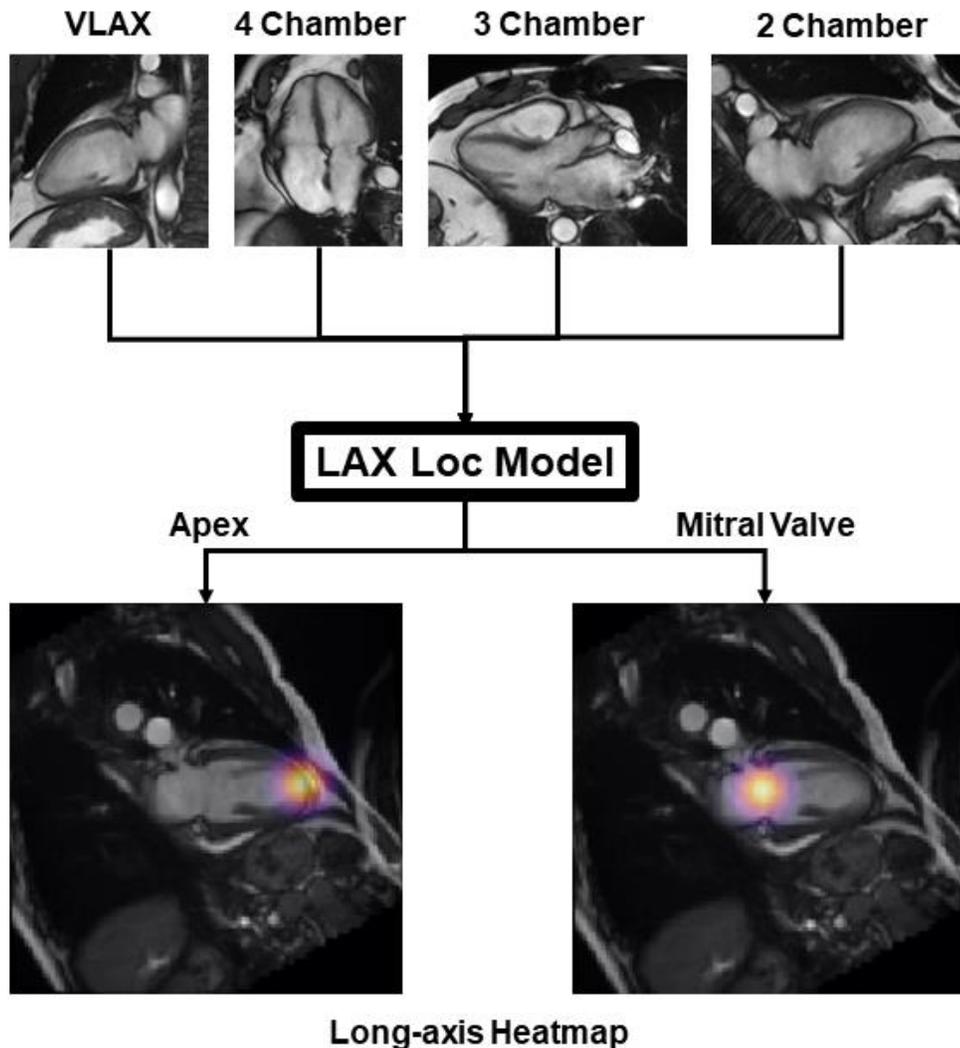
Figure 2: *Input data used to train the long-axis localization model*

Vertical long-axis (VLAX), 4-chamber, 3-chamber, and 2-chamber views were aggregated to train the long-axis model for localization of the mitral valve (MV) and apex.

2.3.3 Short-Axis Landmark Localization

SAX stack landmark localization was decomposed into multiple 2D problems (Figure 3). First, we identified the mitral valve slice (MVS) with a MVS localization model. Second, we created a bounding box around the heart to reduce the search space for the anatomic landmarks. Third, we created a final localization network for fine-grain anatomic localization of the AV, MV, PV, and TV.

To identify the MVS, we utilized a previously described method using a 2.5D VGG-19/LSTM ensemble network [58]–[60]. This redefined the MVS localization as a classification task to reduce the difficulties that arise from imbalanced data (Supplemental Figure 2) [22], [61]. The MVS model was trained to classify proposal slices as either atrial or ventricular to the ground truth labeled MVS. Spatial context was provided by adding two slices atrial and two slices ventricular to each target slice for a total of 5 channels. Within a given SAX stack of images, the first ventricular slice in the sorted stack was marked as the predicted MVS.

To reduce the localization search space, we implemented a 2.5D U-net to perform a rough in-plane bounding box around the heart, using all slices of the SAX stack. Bounding box labels were defined by identifying the minimum rectangles that surround a 20-pixel in-plane border that encompassed the AV, MV, PV, and TV landmarks.

The output of the MVS model was combined with the output of the in-plane bounding box to serve as inputs for a final localization model (Figure 3). This final localization model was implemented as a 2.5D heatmap regression model.

2.3.3 Plane Prescription

Based off the vertical LAX view, the SAX plane angle was defined as the plane orthogonal to the line between the apex and the MV landmarks (Supplemental Figure 1). A SAX stack may be therefore be prescribed at regularly spaced intervals along that line. Based off the SAX stack, the 4-chamber plane was defined as the plane intersecting the TV and MV, the 3-chamber plane was defined as the plane intersecting
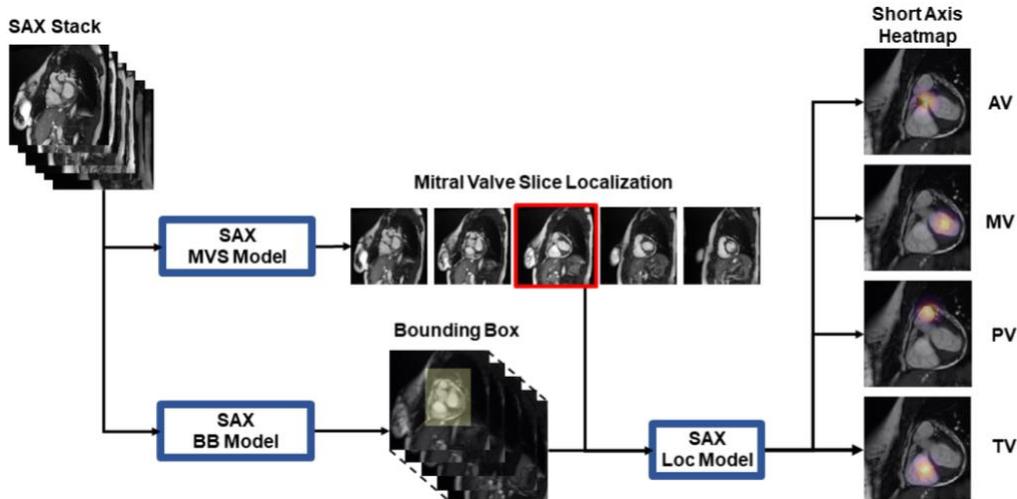
Figure 3: *Input data used to train short-axis localization models*
The short-axis stack images are used as inputs to train the mitral valve slice (MVS) and the bounding box (BB) models. The slice localization and bounding box outputs are then used to train the short-axis model for localization of aortic valve (AV), mitral valve (MV), pulmonary valve (PV), and tricuspid valve (TV).

the AV and MV, and the 2-chamber plane was defined as the plane bisecting the 4- and 3-chamber planes.

2.3.4 Model Training

DL models were each independently trained by Kevin Blansit on a GPU workstation running Ubuntu 16.04, equipped with four Titan X graphics card (NVIDIA; Mountain View, CA). We performed all DL experiments using Keras with TensorFlow backend. Hyperparameters for the final models are described in Supplemental Table 1.1.3.5 Model Analysis and Statistics

To assess localization accuracy, we compared ground truth expert annotation localizations to those predicted by DL, with results confirmed using 5-fold cross validation. For each cross-validation step, we reinitialized a blank model and independently trained with 80% of exams, using the remaining 20% of exams as an

13

independent test data. Cross-validation is a commonly used strategy to ensure that proposed machine learning algorithms are not dependent on the subset of training data used.

To assess plane angulation accuracy, we calculated the angle difference between deep learning (DL planes) and ground truth (GT planes). As a secondary assessment of performance, we also calculated the difference between DL planes and the retrospectively identified plane prescribed by the MRI technologist at the time of image acquisition. Finally, we compared these calculated angle differences against previously reported strategies for automated plane prescription. This last comparative analysis was performed using summary statistics reported in the previous studies. Differences in localization error and angulation error were compared using t-test in R (R Foundation for Statistical Computing, Vienna Austria) with a type I error threshold (α) of 0.05.

2.4 Results

2.4.1 Landmark Localization Performance

To assess localization accuracy on LAX images, we measured the distance between ground truth annotation and DL predicted localizations. For LAX images, the mean DL localization was within 12.56±19.11 mm for apex and 7.69±6.91 mm for the MV (Figure 4A). Given the importance of VLAX accuracy for subsequent derivation of the SAX stack, we examined the localizations for the VLAX. For VLAX images, mean predicted localizations were within 10.20±13.58 mm for the apex and 8.21±10.71 mm for MV.
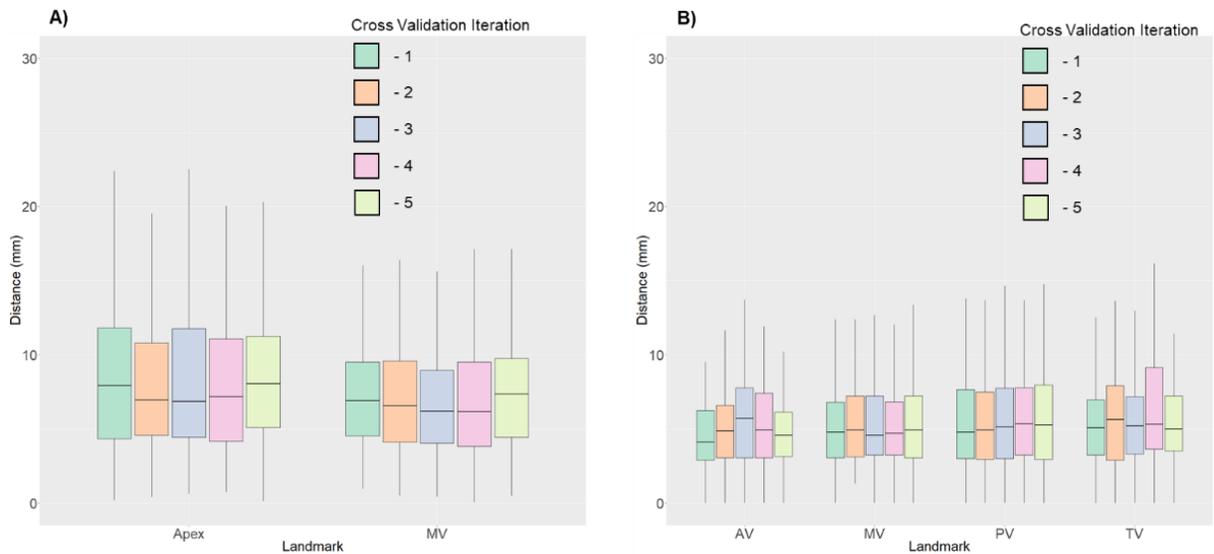
Figure 4: *Distances between predicted and ground truth for the A) long-axis and B) short-axis localization models.*
Each independent iteration of cross-validation is shown in a corresponding color

As a first step towards SAX localization, we first identified the MVS. The average

distance between ground truth and predicted MVS localization was within 4.87±8.35

mm, on average within the 10 mm spacing between slices typically used for planar

cardiac MRI at our institution. The majority of predicted MVS localizations (465/493,

94.32%) were within 1 slice of the labeled MVS. The second neural network was used

to identify an in-plane bounding box around the heart on SAX images. These predicted

segmentations had an average Dice score of 0.91±0.05, relative to ground truth

bounding boxes. Of these predicted bounding boxes, 493/493 (100%) contained the AV,

492/493 (99.8%) contained the MV, 490/493 (99.39%) contained the PV, and 491/493

(99.59%) contained the TV. After standardizing bounding boxes to native resolution,

there was only one case where the PV and TV localization were not contained within

the input image for SAX-localization. The results of the MVS localization and bounding

box were then combined to create the SAX-localization model. The average localization

for SAX landmarks was 5.78±5.68 mm for AV, 5.90±5.24 mm for MV, 6.55±6.39 mm for

PV, and 6.39±5.89 mm for TV (Figure 4B).

To further assess SAX landmark localizations, we compared slices that were

within a single slice error of the ground truth label (465/493, 94.32%), and slices that

were two or more slices away (28/493, 5.78%). For SAX series within a single slice,

average localization of the AV was within 5.24±3.33 mm compared to 14.60±17.66 mm

(p<0.01) for two or more slice error, MV was within 5.01±3.79 mm for within one slice

compared to 12.13 ±14.46 mm (p=0.02) for two slices, PV within 6.08±4.99 mm

compared to 14.03±15.73 mm (p=0.01), and TV within 5.81±3.69 mm compared to

15.96±17.19 (p<0.01).

2.4.2 Plane Prescription Performance

To assess the performance of landmark localizations obtained from DL, we computed plane angulations defined by these landmarks. SAX imaging planes were prescribed using DL predictions on VLAX images. These plane predictions were compared against expert ground truth. For SAX plane prescription, the mean angle bias was -1.27±6.81° and mean absolute angle difference was 4.93±4.86° (Figure 5A, Table 1). Similarly, LAX planes were computed from DL predictions on SAX images. For LAX plane prescription, the mean angle bias and mean absolute angle difference for 4-chamber were 0.38±6.45° and 5.16±3.80°, for 3-chamber were 0.13±12.70° and 9.02±8.83°, and for 2-chamber were 0.25±9.08° and 6.53±6.28°. Representative images of plane prescriptions and their frequency are shown in Supplemental Figure 3.
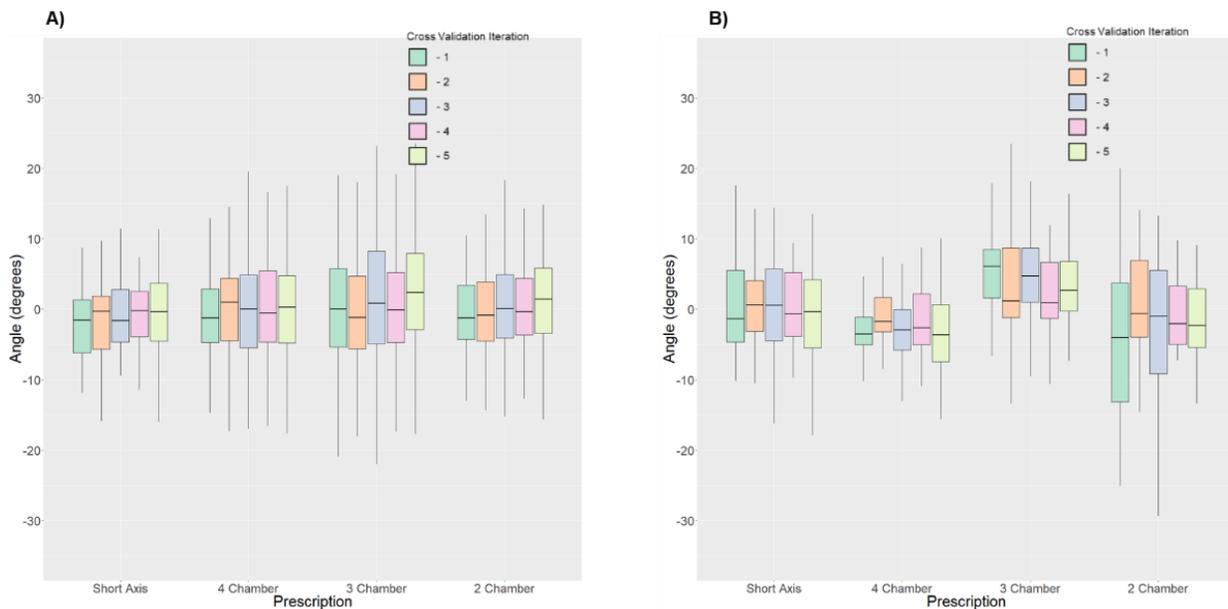
Figure 5: *Comparison of plane prescriptions against A) expert ground truth planes and B) retrospectively matched technologist planes used during examination.* Each independent iteration of cross-validation is shown in a corresponding color.

Table 1*: Accuracy of imaging planes prescribed by deep learning compared to A) expert ground truth and B) retrospectively matched planes prescribed by the MRI technologist.*

Comparison	Value	Prescription Plane			
		SAX	4-Chamber	3-Chamber	2-Chamber
A) Ground Truth	Angle Bias	-1.27±6.81°	0.38±6.45°	0.13±12.70°	0.25±9.08°
	Angle Difference	4.93±4.86°	5.16±3.80°	9.02±8.83°	6.53±6.28°
B) Technologist	Angle Bias	0.40±7.20°	-2.67±7.01°	4.29±7.68°	-2.36±9.83°
	Angle Difference	5.56±4.60°	5.49±5.06°	7.19±4.97°	8.00±6.03°

We report absolute angle differences. Means and standard deviations are reported. Statistical significance between the performance of DL and previously described strategies are denoted with asterisks.

To further validate our approach of planning cardiac MRI planes using DL predicted landmarks, the DL planes were compared to those prescribed by a technologist at the time of acquisition (Figure 4B, Table 1). For SAX plane prescription, the mean angle bias and mean absolute angle difference were 0.40±7.20° and 5.56±4.60°. For LAX plane prescriptions, the mean angle bias and mean angle difference for 4-chamber were -2.67±7.01° and 5.49±5.06°, for 3-chamber were 4.29±7.68° and 7.19±4.97°, and for 2-chamber were -2.36±9.83° and 8.00±6.03°. There was greater agreement and consistency between the DL prediction and technologist for the 4-chamber than the 2-chamber plane. The performance of 3-chamber plane prescription was in-between. Example plane prescriptions are in Figure 6.

Table 2*: Accuracy of imaging planes prescribed by deep learning compared to two recently described strategies.*

Comparison	Prescription Plane			
	SAX	4-Chamber	3-Chamber	2-Chamber
Current method	**4.93±4.86°**	**5.16±3.80°**	**9.02±8.83°**	**6.53±6.28°**
Frick et al	**6.7±3.6° ***	**7.7±6.1° ***	**9.1±6.3°**	**7.1±3.6°**
Lu et al	**8.6±9.7° ****	**19.2±8.5° *****	**12.3±11.0° *****	**18.9±2.1° *****

*   p<0.05
**   p<0.01
*** p<0.001

We report absolute angle differences. Means and standard deviations are reported. Statistical significance between the performance of DL and previously described strategies are denoted with asterisks.

We also compared our method against recently described strategies by *Frick et al* and *Lu et al* [5], [6]. Results are highlighted in Table 2. Though our studies don't share common reference data sets, we found statistically improved SAX and 4-chamber angulation relative to ground truths than prior methods. Further, 3-chamber and 2-chamber angulations also appear improved, relative to *Lu et al*.

2.5 Discussion

In this study, we demonstrated the feasibility of using DL to localize cardiac landmarks for prescription of SAX, 4-chamber, 3-chamber, and 2 chamber view planes. For LAX images, this was readily accomplished with a single 2D U-net modified for in-

Plane Prescriptions
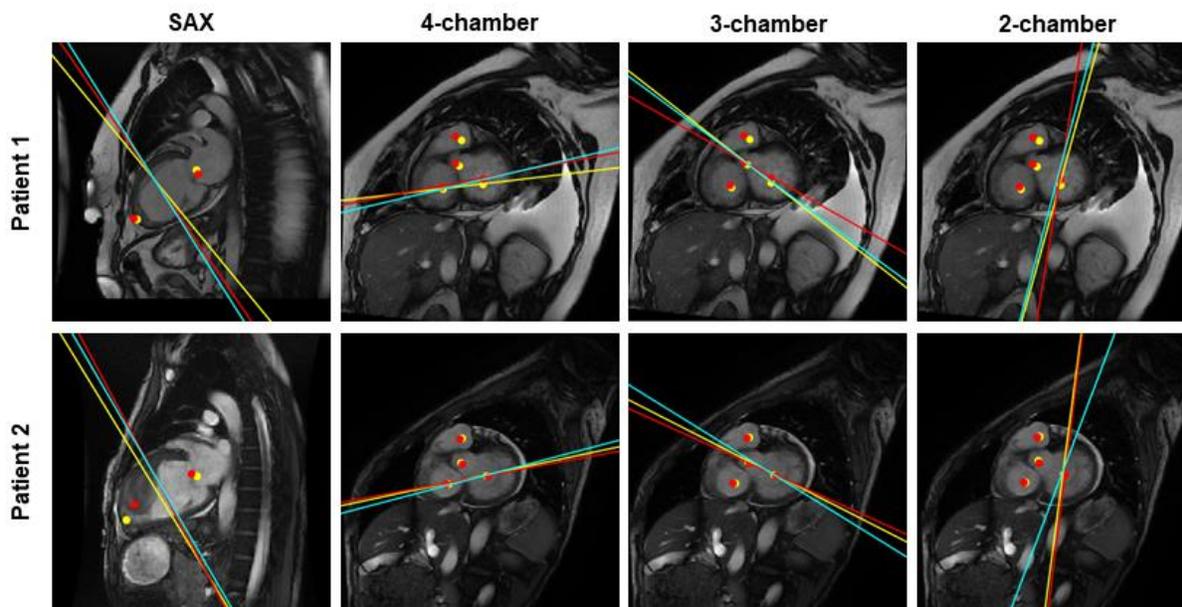
Figure 6: *Example images of short-axis (SAX), 4-chamber, 3-chamber, and 2-chamber plane prescription from two cases.*

Solid lines denote the imaging planes, while solid dots demarcate the localizations used to plan them. Ground truth is shown in yellow and the guidance system in red. The actual plane prescribed by the technologist is shown in teal. Note the large mass within the left ventricle in patient 2, which did not dramatically interfere with short-axis prescription.

plane heatmap regression. For localization of a SAX stack, we applied a cascaded system of neural networks to localize key anatomic landmarks by first identifying the basal slice at the plane of the mitral valve. We found that the MVS localization model correctly localized the MVS within a single slice in nearly all cases (465/493, 94.32% of SAX inputs). More importantly, these localizations yielded imaging planes very similar to those marked by a radiologist or those prescribed by a technologist at the time of image acquisition.

Previous view planning systems have been proposed for cardiac MRI. Lelieveldt et al proposed planning a SAX plane using MRI scans of the entire thorax [62]. Utilizing

a deformable atlas, these authors identified the organ landmarks (including lungs, ventricles, and heart) to prescribe the SAX plane. This approach was validated by showing that clinical measurements including ventricular mass and ejection volume were not significantly different between acquisition approaches. However, the study's authors did not identify the essential 4-chamber, 3-chamber, and 2-chamber LAX imaging planes that are necessary for the assessment of wall motion and valve function [63], [64]. More recently, other studies have utilized mesh segmentation-based approaches to plan sequences of view planes from a single 3D cardiac MRI acquisition [5], [6]. While promising, these approaches were developed using a more limited test population with the use of an additional acquisition that is not typically used in many cardiac MRI workflows. Recent works have shown that DL-based approaches may MRI benefit from the wider generalizability to other modalities and image contrasts than typically seen with traditional methods of machine learning [65]. Addy et al recently presented the use of a DL-based method to plan cardiac MRI views citing a similar strategy [66], [67]. Unlike prior efforts, we chose in our study to assess whether we could develop a system of convolutional neural networks which could be seamlessly integrated into a typical workflow of cardiac MRI, covering the heterogeneity of heart morphologies and disease states seen in clinical practice.

This study must be considered with the following limitations in mind. The current study is a proof of concept feasibility study showing the performance of a system of neural networks for landmark localization on planar MRI. There are many potential approaches for applying deep neural networks to solve this plane prescription problem, and it appears that DL-based landmark localization may be a feasible approach. While

21

the performance of this prototype system is promising, it is still relatively early in development. There is room for further optimization and refinement of this overall strategy. There is variability in the performance of different imaging planes, with greater consistency between DL and technologists for the 4-chamber view than the 2-chamber view. This may be related to exact angle bisections done for the DL prescription, which are only visually approximated by the technologist. However, DL appears to produce planes with greater agreement with ground truth than previous methods. Future work may be required to assess the performance of such a system prospectively within a clinical workflow and determine an acceptable error range for plane prescription. To implement this strategy in practice, there are multiple additional steps that need to be implemented. For example, a SAX stack may be optimally prescribed not just from one LAX view, but perhaps two. Achieving the first LAX prescription may require additional localization on an axial or sagittal stack. These may require further development before a fully-comprehensive workflow could be tested.

An important factor for overall generalizability of machine learning algorithms is the scope of the problem to be solved and the data that is ultimately used to train and test algorithm performance. We demonstrate that in a contiguous retrospective cohort on MRIs from a single vendor and field strength, that this strategy is generally effective with few exceptions and outliers in performance. It is not yet clear whether the system will achieve similar performance on images from other MRI vendors or at 3T, where blood-pool to myocardial contrast may not be relatively weaker or off-resonance banding artifacts may be present. Moreover, further work may be needed to evaluate performance of the algorithm on non-gated or single-shot images, rather than the Cine

SSFP which were included in this study. Incorporation of these data into further training may be necessary to further generalize this methodology.

Nevertheless, in this study, we demonstrated that a DL-based localization approach is adequate for cardiac MRI plane prescription. It is possible that the SAX and LAX localizations may be used to sequentially optimize imaging planes, as can be performed by skilled technologists or physicians. Future work may help determine whether such a system may fully automate prescription of cardiac imaging planes. As specialty technologist and physician training is currently a limiting factor for the availability of cardiac MRI, we believe that further development of this approach may alleviate this current barrier to access of this essential imaging modality.

2.6 Acknowledgements

This chapter was adapted from in part from "Deep Learning–based Prescription of Cardiac MRI Planes" in Radiology Artificial Intelligence, 2019 by Kevin Blansit, Tara

Retson, Evan Masutani, Naeim Bahrami, and Albert Hsiao. The dissertation author was

the primary investigator and author of this paper.

Chapter 3: Generalizability and Robustness of an Automated Deep Learning System for Multiplanar Cardiac MRI Plane Prescription

3.1 Abstract

Purpose:

Cardiac MRI is the reference standard for non-invasive assessment of cardiac morphology, function, and myocardial scar. Accessibility of cardiac MRI is limited by the need for specially trained operators to identify landmarks that define the imaging planes. With this clinical need in mind, we developed a multi-component deep convolutional neural network (DCNN) system to automatically identify and prescribe the cardiac imaging planes.

Methods:

To assess performance and generalizability, we collected 363 HIPAA compliant cardiac MRI studies from three tertiary academic hospitals including 1.5T and 3T field strengths. Images were annotated for ground truth position of plane prescription landmarks by a cardiovascular radiologist. We compared the position and angulation of imaging planes collected by technologists at the time of exam against those inferred by the DCNN. The quality of long-axis images was also assessed by two radiologists on a 5-point Likert scale to relate plane angulation error to image quality.

Results:

We found that images with higher angulation error ($>15°$) were more likely to be scored as lower (score: 1-3) vs higher plane quality (score: 4-5) (25% vs. 11%; $p<0.001$). DCNN inferred short-axis imaging planes had a median angulation error of $2.24°\pm6.09$, lower than technologist error $2.96°\pm3.73$, ($p<0.001$). Inferred 4, 3, and 2-

chamber imaging planes were either comparable or had lower median angulation errors than technologists (4-chamber: 5.61°±6.47 DCNN vs. 7.51°±5.28 technologist, p<0.001; 3-chamber: 6.23°±7.40 vs 6.63°±5.62, p=0.243; 2-chamber: 5.61°±6.47 vs. 8.68°±6.33, p<0.001).

Discussion

An automated deep learning system is capable of prescribing cardiac imaging planes with reduced angulation error compared to cardiac MRI technologists. The DCNN system shows robust performance across multiple institutions and field strengths and may help improve accessibility of cardiac MRI.

3.2 Introduction

Cardiac MRI is the reference standard for non-invasive quantification of cardiac volumetry, ventricular function, blood flow, and assessment of myocardial scar [1]. Acquisition of high-quality images requires detailed anatomic knowledge by the MRI operator. These double-oblique imaging planes are defined by patient-specific anatomic landmarks, including the left ventricular apex, mitral valve (MV), tricuspid valve (TV), and aortic valve (AV). Specially trained physicians and technologists are required to plan cardiac imaging planes, which limits the accessibility of cardiac MRI to primarily major academic and subspeciality centers.

Deep Convolutional Neural Networks (DCNNs) have shown promise for automating visual tasks in cardiac MRI, including heart segmentation, tissue characterization, and localization of anatomic structures [7]–[9], [11], [68]. While many DCNN algorithms have been described, very few have yet been implemented into clinical practice [24]. One concern for application of DCNNs in the clinical environment

is algorithm generalizability, where an algorithm's performance in the laboratory may not be maintained in clinical populations [26], [69]. For example, prior studies have shown poorer DCNN performance when tested on external patient populations [32]. Another factor for confident application of algorithms in the clinical environment is providing transparency into their mechanism of operation [43], [44]. Specifically for acquisition of cardiac MR images, heatmap localization has been shown as a feasible strategy for image plane planning [70]. This prior work illustrated that DCNNs can reliably recognize the anatomic landmarks needed for plane prescription on either long-axis (LAX) or short-axis (SAX) stack cine SSFP image series. This approach to planning the cardiac MR imaging planes using landmark localization parallels the approach used by human operators to scan cardiac MRI, and could be used to provide algorithm explainability. However, a human operator may use landmarks on multiple planes to obtain the optimal cardiac images, and the prior work only explored the feasibility of DCNNs to operate on a single view.

To address these limitations to DCNN application in the real world we developed a multi-stage system of DCNNs to comprehensively prescribe cardiac imaging planes, expanding upon the heatmap localization strategy. We hypothesized that a strategy of incorporating images from multiple field strengths would help provide algorithm generalizability across multiple institutions. We also explored the feasibility of incorporating DCNN inferences across multiple imaging planes to better capture the 3-dimensional nature of the heart and optimize plane prescription.

3.3 Methods

3.3.1 Multi-Stage Cardiac MRI System

We developed a multi-stage system for prescribing cardiac imaging planes consisting of 4 deep convolutional neural network (DCNN) modules (Figure 7): 1) CorSegNet for segmenting the lungs and mediastinum from coronal 3-plane localizer series to prescribe an axial stack, 2) AxLocNet to localize the MV and apex from the axial stack to prescribe a vertical long-axis (VLAX), 3) LaxLocNet to localize the MV and apex from LAX views to prescribe a SAX stack, 4) SaxLocNet to localize the MV, TV, and AV, which is combined with the apex localized by LaxLocNet to prescribe the 4-, 3-, and 2-chamber views. Details regarding training image parameters are shown in Supplemental Table 2.

Training data were labeled by one of two board certified cardiovascular radiologists (either Albert Hsiao or Seth Kligerman) each with over 10 years of experience in cardiac MRI or a senior radiology resident with radiologist review. For training of CorSegNet, coronal images from 3-plane localizers were annotated with lung and mediastinal masks. For AxLocNet, an axial stack of images was annotated with the center of the MV and left ventricular apex. Annotations for training of LaxLocNet and SaxLocNet were as previously described [70]. The details of the annotations used for algorithm evaluation are detailed in each of the relevant sections below.

The multi-stage algorithm was tested on an independent set of clinical exams from three geographically distinct tertiary academic medical centers in the United States. With HIPAA compliance and IRB waiver of informed consent, we collected a convenience sample of 363 examinations that included (174 from site 1, 90 from site 2,

CorSegNet Training/Validation: 118

Prescription	All	Site 1	Site 2	Site 3
Axial	279	184	-	95

AxLocNet Training/Validation: 700

Prescription	All	Site 1	Site 2	Site 3
Vertical Long-Axis	209	129	-	80

LaxLocNet Training/Validation: 1,216

Prescription	All	Site 1	Site 2	Site 3
Short-Axis	270	118	59	93
4Ch	259	109	65	85
3Ch	194	80	64	50
2Ch	186	75	63	48

SaxLocNet Training/Validation: 697

Prescription	All	Site 1	Site 2	Site 3
4Ch	263	110	66	87
3Ch	205	86	67	52
2Ch	194	80	66	48

Figure 7: *Left) Schematic of an automated, multi-stage system for prescribing cardiac imaging planes comprised of multiple DCNN modules and Right) number of image series used for either training/validation or testing.*

A) CorSegNet is used to segment the lungs andc mediastinum from a coronal localizer, which is used to define the superior and inferior coverage of an axial stack, B) AxLocNet localizes the mitral valve (MV) and apex on either a localizer or dedicated axial stack to prescribe a vertical long-axis image, C) LaxLocNet localizes the MV and apex from a vertical long-axis image to prescribe a short-axis stack, D) SaxLocNet first localizes the MV, aortic valve, and tricuspid valves on a short-axis stack, which is then combined with the apex localization of the apex from a cross referenced long-axis image using LaxLocNet to prescribe the 4-, 3-, and 2-chamber long-axis images.

and 99 from site 3). Of these studies, 197 (54%) were male, with an age range from 15

to 87. MRI examinations were performed for a variety of clinical indications

representative of each clinical practice, and further detailed in Table 3. These studies

Table 3*: Evaluation Patient Demographics*

	All	Site 1	Site 2	Site 3
n	363	174	90	99
Age	Median: 47 (Range: 15 - 87)	Median: 47 (Range: 18 - 87)	Median: 57 (Range: 15 - 86)	Median: 60 (Range: 15 - 86)
Field Strength	54% 1.5T (n=197)	30% 1.5T (n=53)	50% 1.5T (n=45)	100% 1.5T (n=99)
Sex	53% Male (191)	51% Male (n=89)	58% Male (n=52)	51% Male (N=50)
Date Range	June 30, 2018 - December 29, 2020	November 21, 2019 - December 29, 2020	June 30, 2018 - December 16, 2018	October 09, 2019 - December 06, 2019
Clinical Indications				
Cardiomyopathy	86	32	-	54
Structural evaluation	54	54	-	0
Other	31	20	-	11
Viability/scar	18	9	-	9
Pericardial Disease	17	0	-	17
Arrythmia / Syncope	17	17	-	0
Myocarditis / Sarcoid	15	10	-	5
Heart Failure	15	15	-	0
Mass	12	10	-	2
Research study	6	6	-	0
Amyloid	1	0	-	1
Scanner Model				
DISCOVERY MR750	111	110	1	0
Optima MR450w	124	0	25	99
Signa HDxt	56	53	3	0
SIGNA Premier	34	0	34	0
SIGNA Artist	17	0	17	0
DISCOVERY MR750w	14	11	3	0
SIGNA Architect	6	0	6	0
SIGNA PET/MR	1	0	1	0

Note that clinical indications were not available for site 2.

included 324 three-plane coronal localizer, 124 axial, 270 VLAX, 233 SAX, 263 4-chamber, 205 3-chamber, and 194 2-chamber series. Imaging parameters are listed in Supplemental Table 3. All ground truth annotations for evaluation were performed by one of the board-certified cardiovascular radiologists. The details of the annotations used for algorithm evaluation are detailed in each of the relevant sections below.

3.3.2 Chest Coverage (CorSegNet)

CorSegNet generated lung and mediastinal masks based on a 3-plane localizer. This DCNN architecture was based on a U-Net architecture [20]. The superior and inferior mask margins were used to infer the chest coverage. Ground truth was defined as the superior margin of the lung apices to the inferior most extent of the lungs and mediastinum. Labels were performed by a cardiovascular radiologist (Albert Hsiao) separate from the lung and mediastinal masks used for training. The superior and inferior margins were also recorded from those determined by a technologist during the clinical exam.

To evaluate performance, we calculated distances between ground truth margins and either 1) CorSegNet inferred margins, or 2) technologist acquired margins, and reported difference and standard deviations. To assess generalizability, a subset analysis was performed on sites 1 and 3 (site 2 was not included as all data from that site was used for training this module). Differences in margin error variations were compared using Levine's Test (R Foundation for Statistical Computing, Vienna Austria). P-Values < 0.05 were considered statistically significant.

3.3.3 Prescription of Cardiac Imaging Planes

3.3.3.1 Prescription of a Vertical Long-Axis Imaging Plane (AxLocNet)

The vertical long axis imaging plane was defined as the plane that intersects the mitral valve and apex, and we developed AxLocNet as a cascaded DCNN to localize these landmarks from either a 3-plane localizer or axial of images. A cardiac bounding box was first implemented to reduce the search space. Then, we implemented a 2.5D slice selection network to identify the axial slice containing the apex and mitral valve, similar to efforts by previous authors [70]. The mitral valve and apex locations were then determined with two 3-dimensional heatmap regression networks and computed as the point position with the maximal pseudoprobability from each network.

3.3.3.2 Prescription of a Short-Axis Imaging Plane (LaxLocNet)

The SAX imaging plane was defined as the plane orthogonal to the line intersecting the MV and apex. To prescribe the SAX image, based on previously described methods we implemented LaxLocNet as a 2-dimensional cascaded system to localize the apex and MV on a long-axis image [70]. Compared to the previously described DCNN, LaxLocNet is a two-stage network, where the initial localization is

31

used for a bounding box for a secondary fine-grain localization network, Moreover, this networked performed at both 1.5 and 3T field strength.

3.3.3.3 Multi-Planar Prescription of Cardiac Imaging Planes (Combining LaxLocNet and SaxLocNet)

To prescribe LAX imaging planes, we iterated on a previously described DCNN to develop SaxLocNet as a 3-dimensional cascaded system trained on 1.5T and 3T scout and cine images [70]. The 4-chamber imaging plane was defined as intersecting the mitral and tricuspid valves, the 3-chamber as intersecting the mitral and aortic valves, and the 2-chamber as bisecting the 4- and 3-chamber planes. To refine prescription, we used LaxLocNet to cross reference and localize the apex on a VLAX for 4-chamber prescription, and on the 4-chamber prescription for the 3- and 2-chamber prescriptions.

3.3.4 Evaluation of Cardiac Imaging Planes

3.3.4.1 Defining Ground Truth Landmarks and Angulation Error

For this fully independent test set, ground truth imaging planes were defined by landmark locations labeled by a cardiovascular radiologist (A.H.). Imaging planes obtained by the technologist during the clinical exam were also compared. To assess similarity, we calculated the angulation error between the ground truth planes, and either 1) the DCNN inferred planes, or 2) the planes acquired by the technologist at time of clinical exam. Given the multifactorial inputs necessary for optimally prescribing the 4-, 3-, and 2-chamber images as detailed above, these analyses are decomposed into the angulation errors on the basis of either LAX or SAX input images.

3.3.4.2 Plane Quality Score Evaluation

To develop a surrogate indicator of image plane quality, we assessed the relationship between qualitative image plane quality and the angulation error. Two board-certified cardiac radiologists (Michael Horowitz and Lewis Hahn) with 3 years of experience graded the 4-, 3-, and 2- chamber images by subjectively rating plane quality on a 5-point Likert scale (1 signifying poor quality, and 5 signifying ideal quality). Inter-reader variability was assessed for a subset of images with a linearly weighted Kappa statistic. If scores differed, the lower of the two was used. The angle error between the ground truth and LAX image acquired at time of exam was calculated, grouped by Likert score, and plotted against frequency of angulation error (binned in 5° increments). Differences were assessed by Chi-square.

3.3.4.3 Long-Axis and Short-Axis DCNN Module Evaluation

To evaluate AxLocNet, LaxLocNet, and SaxLocNet, we computed angulation errors between ground truth imaging planes and either 1) DCNN inferred planes or, 2) technologist acquired imaging planes. A subset analysis was also performed by site. Site 2 was excluded from our testing set for AxLocNet as it was used for training this module. We report mean errors compared with paired Wilcox test.

3.4 Results

3.4.1 Relationship between angulation error and plane quality

To establish the clinical relevance of angulation error, we performed a reader study to assess its relationship to image plane quality. Of the LAX images, 345 (53%) were scored a 5 (high quality), 212 (34%), scored 4, 52 (8%), scored 3, 24 (4%) scored 2, and 5 (1%) scored 1 (low quality). Of the 131 images that were scored by both

reviewers, the interrater variability Kappa statistic was fair ($\kappa: 0.34$). Frequency of

binned angulation errors was then plotted against their image plane quality scores.

Intuitively, higher angulation errors were more likely to be associated with lower image

plane quality (Figure 8) (*Chi-squared test: p<0.001*). Exemplar images that correspond

with different quality scores are shown in Figure 8B.
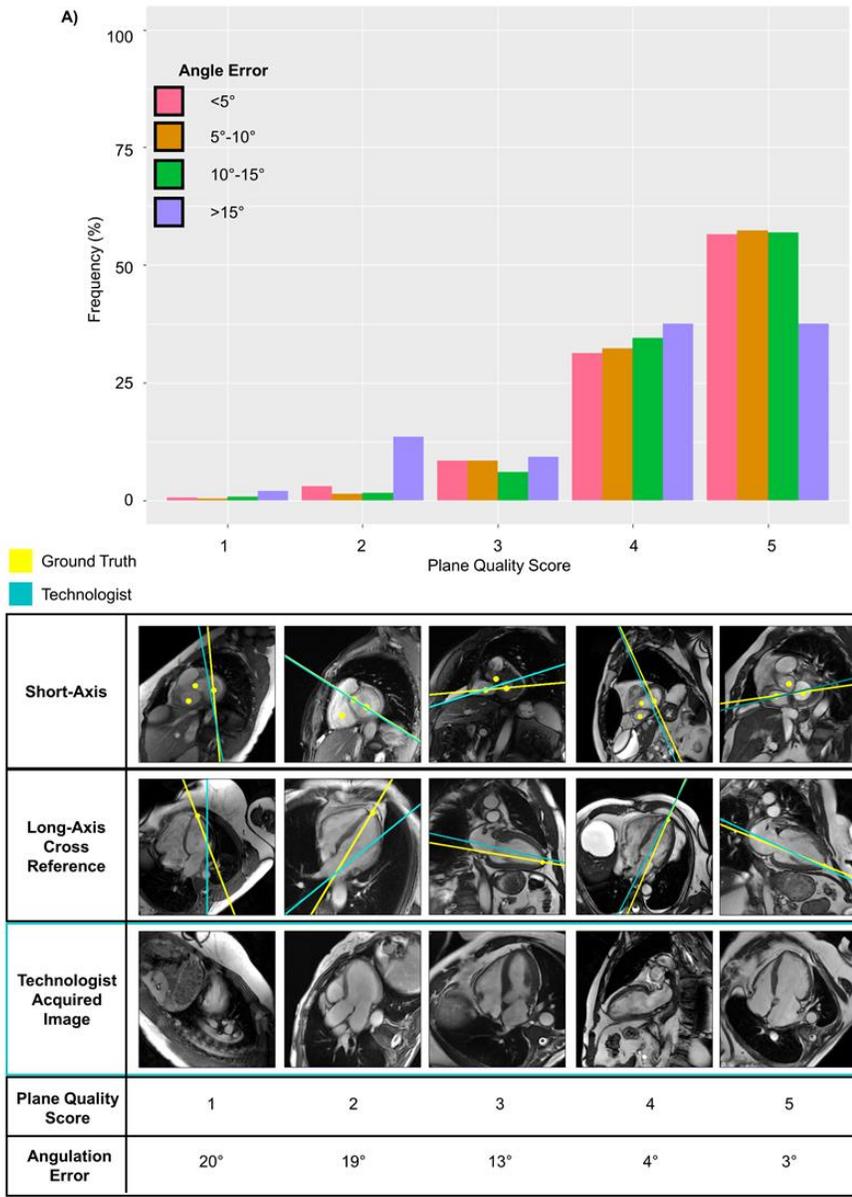
Figure 8*: Angulation error and its the relationship to plane quality score.*
A) Lower Likert plane quality scores had a greater representation of higher angulation errors. B) The top row represents short-axis images used for valve localization. The middle row represents cross referenced long-axis images used for apex localization. Combing the valve and apex localizations defines the plane of long-axis images shown in the bottom row. Cardiovascular radiologist defined ground truth valve and apex localization are shown as yellow dots and corresponding planes. Technologist acquired imaging planes are show in teal.

3.4.2 Axial Coverage Performance (CorSegNet)

We first assessed the ability of CorSegNet to infer and plan the superior and inferior extent of an axial stack of images. Variances were compared between ground truth margins and either those inferred by CorSegNet or those acquired by a technologist (Figure 9). For the superior margin, CorSegNet showed markedly lower variance than technologists (9.94 mm vs. 60.40 mm, *Levene's test:* p<0.001), both overall, and at the individual sites. Similarly, for the inferior margin, CorSegNet showed lower variance than technologists (39.0 mm vs. 35.4 mm, *Levene's test: p=0.288*).
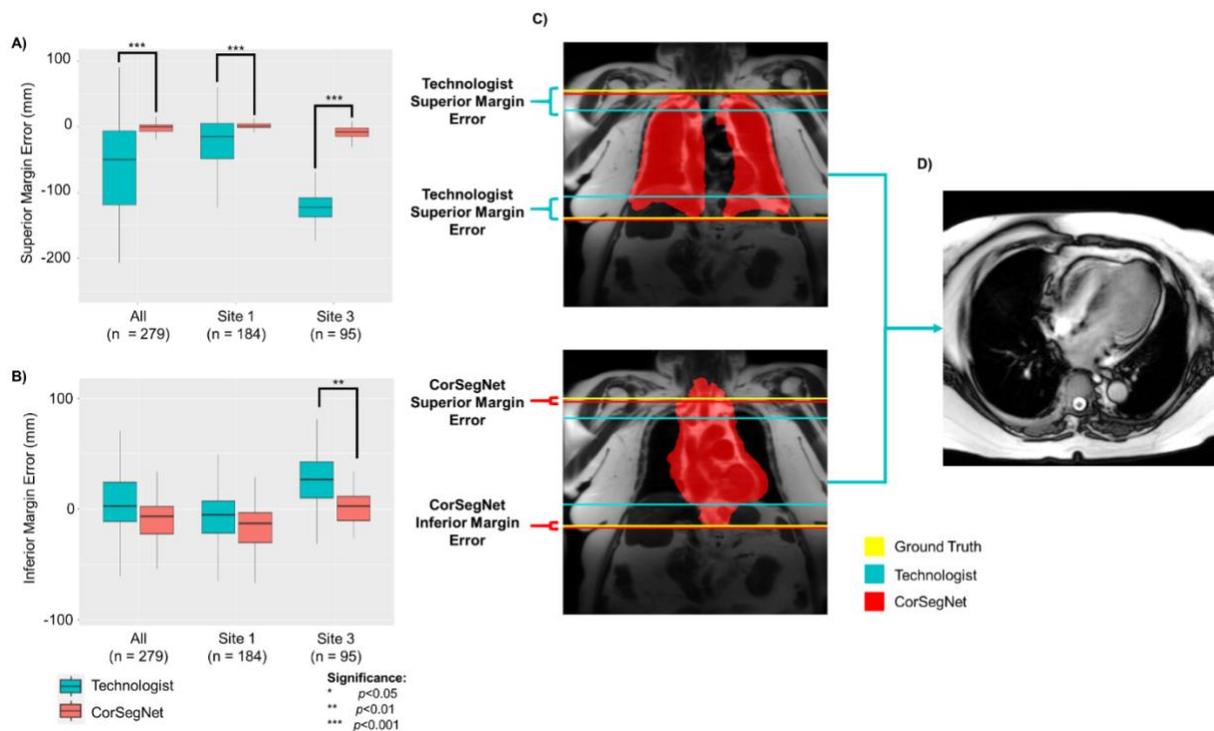
Figure 9*: Relationship between axial coverage by a technologist compared to the field of view inferred by CorSegNet.*

On the left, comparison graphs of the A) superior or B) inferior margin distances between ground truth and either technologist (teal) or CorSegNet (red) axial coverage. In the top right, exemplar coronal image with ground truth (yellow), technologist prescribed (red), or CorSegNet predicted superior and inferior margins. Predicted CorSegNet lung segmentation is shown in red on top, and

36

There was some variation in performance by site, with no statistical significance

difference at site 1, site 3 had greater technologist variance in inferior prescription.

3.4.3 Vertical Long-Axis Prescription Performance (AxLocNet)

AxLocNet was assessed by comparing the angulation errors between algorithm

inferred and technologist acquired VLAX imaging planes (Figure 10). AxLocNet inferred

errors were greater than those acquired by a technologist (5.21°±12.56 vs. 3.60°±6.22,

Wilcoxon signed-rank test: p>0.001) overall, with some variation by site. AxLocNet

inferred VLAX imaging planes had greater angulation error at site 1 compared to
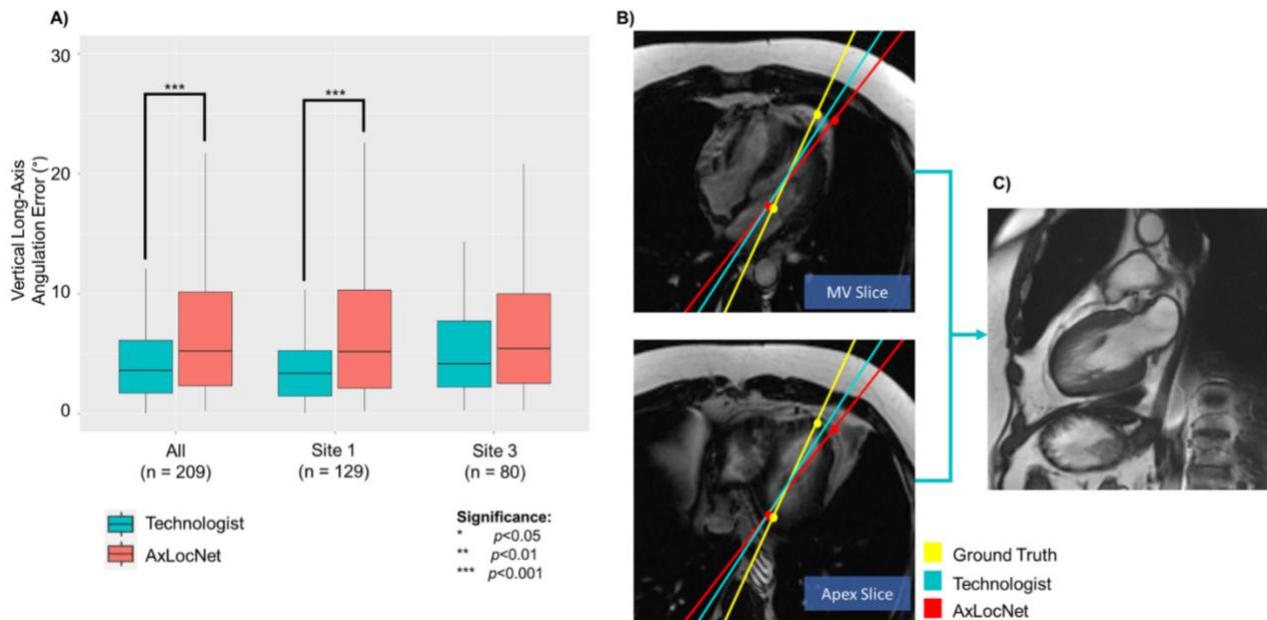
Figure 10*: Relationship between vertical long-axis angulation error by a technologist compared to AxLocNet.*

A) Comparison graph of angulation differences between ground truth and either technologist acquired (teal) or AxLocNet predicted (red) vertical long-axis planes (VLAX). B) Exemplar images from an axial stack shown at the mitral valve (top) and apex slice (bottom) levels with ground truth (yellow), technologist acquired (teal), and AxLocNet predicted (red) VLAX imaging plane. Ground truth and AxLocNet predicted mitral valve and apex localizations are shown as yellow and red dots, respectively. C) Technologist acquired VLAX image is shown.

technologist images, while angulation error was higher but not statistically significant at site 3.

3.4.4 Short-Axis Prescription Performance (LaxLocNet)

LaxLocNet was assessed by comparing angulation errors between the algorithm inferred and technologist acquired SAX imaging planes (Figure 11). LaxLocNet inferred SAX errors were smaller than those acquired by technologist (2.22°±6.17 vs. 2.95°±3.73, *Wilcoxon signed-rank test: p<0.001*) overall, with some variation by site. LaxLocNet inferred SAX imaging planes had lower angulation error compared to technologists at sites 1 and 3 and were equal at site 2.

3.4.5 Dedicated Long-Axis Prescription Performance (SaxLocNet and LaxLocNet)

Finally, we assessed the angulation errors between the combined SaxLocNet and LaxLocNet inferred 4-, 3-, and 2-chamber imaging planes. Overall, imaging planes
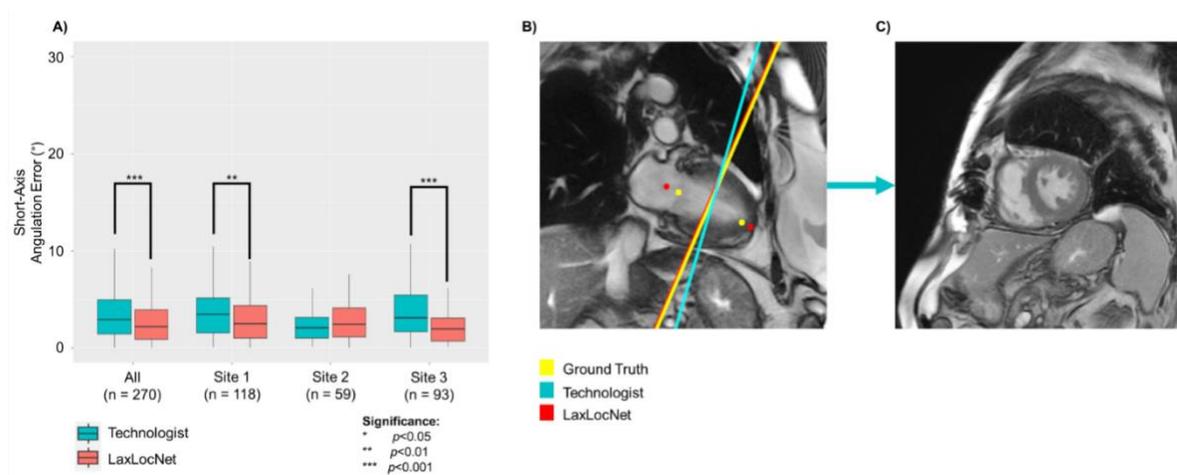
Figure 11*: Relationship between short-axis angulation error by a technologist compared to LaxLocNet.*

A) Comparison graph of angulation differences between ground truth and either technologist acquired (teal) or SaxLocNet predicted (red) short-axis planes (SAX). B) Exemplar vertical long-axis image is shown with ground truth (yellow), technologist acquired (teal), and LaXLocNet predicted (red) SAX imaging plane. Ground truth and SaxLocNet predicted mitral valve and apex localizations are shown as yellow and red dots, respectively. C) A midventricular technologist acquired SAX clinical image is shown.

had either similar or lower angulation error compared to technologist acquired images. For 4-chamber images, the error was lower compared to technologist acquired images (5.60°±6.48 vs. 7.51°±5.12, *Wilcoxon signed-rank test: p<0.001*), while the 3-chamber error was similar (6.12°±7.31 vs. 6.54°±5.51, *Wilcoxon signed-rank test: p=0.215*), and the 2-chamber was lower (4.99°±7.80 vs. 8.77°±6.33, *Wilcoxon signed-rank test: p<0.001*).

Given the multicomponent nature of the dedicated LAX imaging, we sub-analyzed the 4-, 3-, and 2-chamber planes, first on the basis of SAX image stacks, and then on the basis of cross reference SAX images. Assessing on the basis of the SAX mage stack, SaxLocNet inferred 4-chamber errors were lower than those acquired by technologist (4.31°±4.76 vs. 6.68°±5.40, *Wilcoxon signed-rank test: p<0.001*) (Figure 12A, left column). By site, inferred 4-chamber errors were equal at site 1, and they were lower at sites 2 and 3. SaxLocNet inferred 3-chamber errors were equal to those acquired by technologist (5.24°±5.43 vs. 5.45°±5.28, *Wilcoxon signed-rank test: p=0.789*) (Figure 12B, left column), and errors were equal at all sites. SaxLocNet inferred 2-chamber errors were lower than those acquired by technologist (4.16°±4.19 vs. 6.69°±6.38, *Wilcoxon signed-rank test: p<0.001*) (Figure 12C, left column). Errors were lower at sites 1 and 3, and equal at site 2.

Performance of LaxLocNet was sub analyzed to determine the importance of further refining the dedicated long-axis imaging planes based on LV apex localizations.
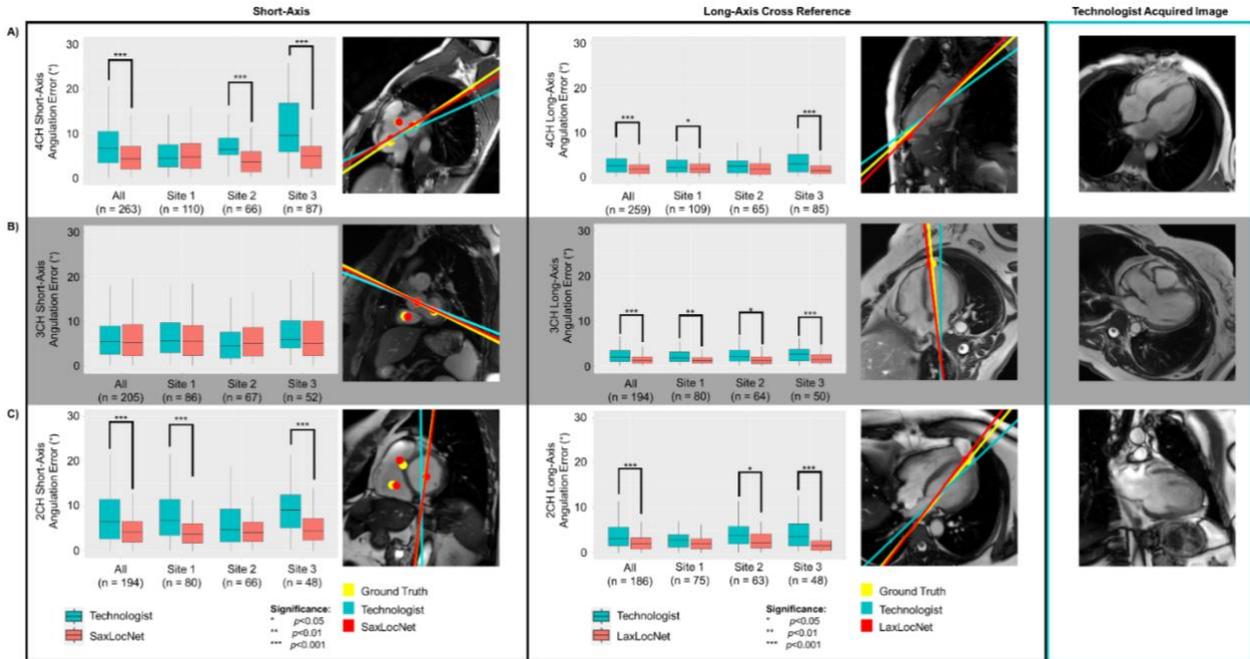
Figure 12*: Relationship between dedicated long-axis angulation errors by a technologist compared to SaxLocNet and LaxLocNet.*

In the left column, comparison graphs of angulation differences between ground truth and either technologist acquired (teal) or SaxLocNet predicted (red) for either A) 4-Chamber, B) 3-Chamber, and C) 2-Chamber planes. In the middle column, comparison graphs of angulation differences between ground truth and either technologist acquired (teal) or LaxLocNet predicted (red) long-axis imaging planes. Exemplar images of SAX and cross referenced LAX are each shown with ground truth (yellow), technologist acquired (teal), and AI (combining SaxLocNet predicted valve and LaxLocNet predicted apex) imaging planes. Ground truth and AI predicted valve and apex localizations are shown as yellow and red dots, respectively. In the right column, technologist acquired long-axis clinical images are shown.

On the basis of a VLAX image, LaxLocNet inferred 4-chamber errors were lower than those acquired by technologist (1.78°±5.32 vs. 2.56°±2.34, *Wilcoxon signed-rank test: p<0.001*) (Figure 12A, middle column). While cross referenced LaxLocNet inferred 4-chamber errors were lower at sites 1 and 3, they were equal at site 2. LaxLocNet inferred 3-chamber errors were lower than those acquired by technologist (1.19°±4.38 vs. 1.92°±2.88, *Wilcoxon signed-rank test: p<0.001*) (Figure 12B, middle column), with errors lower at all sites. Inferred 2-chamber errors were lower than those acquired by

technologist (2.10°±5.3 vs. 3.19°±3.17, *Wilcoxon signed-rank test: p<0.001*) (Figure 12C, middle column), and while errors were equal at site 1, they were lower at sites 2 and 3.

3.5 Discussion

In this paper, we assessed the ability of a multi-stage DCNN for the automation of cardiac MRI plane prescription on routine cardiac MRIs performed at three tertiary academic medical centers. We demonstrate that the performance of the system was comparable to or exceeded performance of cardiac MRI technologists for the majority of cardiac imaging plane prescriptions.

Many DCNN systems are trained and evaluated within a single institution or research laboratory [69]. However, prior studies have highlighted the importance of carefully evaluating DCNNs in local clinical environments to ensure translation of performance from laboratory to the real world [32], [71]. This is particularly relevant for cardiac MRI, where imaging parameters such as manufacturer and field strength may produce visually different images [72]. For example, Yan et al. showed that a left ventricular segmentation DCNN lost performance when tested on images from different MRI scanners [25]. In this study, despite two DCNN modules being trained on data from a single site, performance across all test sites and scanner models was preserved. Previous studies have shown that inclusion of heterogeneous training data performance on outside datasets [25], [32], [73]. Our study supports this, showing that incorporation of a large and diverse set of training data (spanning 1.5T and 3T field strengths) provided sufficient diversity for generalizable performance of our DCNN algorithms.

Another obstacle to the integration of DCNNs into clinical practice is their limited explainability [39]–[42]. Tonekaboni et al. discussed the importance of providing a rationale for algorithm decision making when integrating machine learning systems within existing clinical workflows [74]. We developed this system to localize the landmarks that define the cardiac MR imaging planes on standard cardiac acquisitions, thereby providing clinical end users a tangible rationale for plane prescription. This has perhaps greatest potential for MRI technologists inexperienced in cardiac MRI. At each of our respective centers, cardiac technologists require months to years of training to become proficient. An automated DCNN system for plane prescription has the potential to reduce the barrier to basic proficiency in cardiac MRI. Further, like any system, DCNNs may make errors. By having the DCNN provide localizations for the standardized landmarks used by technologists, the rationale for each plane becomes clear, and allows an expert technologist or radiologist the option to change or correct prescriptions.

Prior approaches to planning cardiac MRI imaging planes have focused primarily on planning all planes from a single acquisition [4]–[6]. However, stacks of 2-dimensional MR images are often anisotropic, and may have poor through-plane slice resolution [46]. For these reasons, previous authors have highlighted the importance of integrating multiple cardiac views for MRI analysis to fully capture the 3-dimensional nature of the heart [11], [47]. Thus, we developed a multiplanar approach to image plane prescription that integrates information from multiple views to precisely localize landmarks in 3-dimensional space. For example, prescription of a 3-chamber view

benefits from not only localization of valve landmarks on a short-axis stack of images, but also from the apex visualized on a 4-chamber view.

In this study, we developed angulation error as a surrogate metric for image plane quality, and saw that images scored as lower plane quality were associated with greater angulation errors. However, we note that this was based on a retrospectively acquired dataset, and images of insufficient quality due to factors such as obvious malalignment or patient motion may have not been saved at time of exam. Prospective studies are needed to further assess the relationship between angulation errors and subjective plane quality. Moreover, this study did not seek to make an authoritative judgment in the inter-site variability of technologist image acquisitions. Factors underlying any heterogeneity in technologist prescription may be related to differences in training and local institutional preference [75]–[77]. Another benefit of DCNN-assisted workflow could therefore be increased standardization of cardiac image prescription and acquisition.

Using angulation error, we demonstrated that each component of our multistage system performed with the range of clinical variability seen during a clinical exam. On further analysis, we found that SAX and dedicated LAX prescription performed more favorably than prescription of the VLAX. This may be explained by wider heterogeneity of cardiac morphologies and cardiac orientations within the thorax on axial imaging, compared to the more standardized presentation of VLAX and SAX images. Prior authors for example have highlighted the importance of standardizing cardiac image orientation for optimal performance of DCNNs [78]. Future work may be directed at

enhancing the performance of VLAX images with more diverse training data or DCNN training strategies.

Here we demonstrate that a multi-component system of DCNNs was able to prescribe cardiac imaging planes similar to gold standard planes defined by a cardiovascular radiologist. Moreover, this study showed generalizability of the DCNN system across multiple sites, including multiple scanners and field strengths.

3.6 Acknowledgements

This chapter was adapted from in part from "Deep Learning–based Prescription of Cardiac MRI Planes" in Radiology, 2021 by Kevin Blansit, Tara Retson, Samira Masoudi, Naeim Bahrami, Phillip Young, Christopher Francois, Lewis Hahn, Michael

Horowitz, Seth Kligerman, Albert Hsiao. The dissertation author was the primary investigator and author of this paper.

Chapter 4: Uncertainty Sampling Enables Data-Efficient DCNN Active Learning – Feasibility for Cardiac MRI Landmark Localization

4.1 Abstract

Purpose:

Deep Convolutional Neural Network (DCNN) algorithms show promise automating aspects of cardiac MRI. However, algorithms developed on one platform may not generalize to another. The amount and type of data required to make DCNNs extensible beyond the scope of their training environments is an area of open investigation. Here, we investigated the effectiveness of uncertainty sampling to improve performance of a DCNN designed for cardiac landmark localization.

Methods:

A DCNN was previously trained on 1.5T images using heatmap localization. We retrospectively curated 1,028 1.5T and 389 3T long-axis cardiac MR image series and manually-annotated the location of apex and mitral valve. We propose two metrics of 'confidence' of the initial DCNN: pseudoprobability maximum ($\hat{\phi}_{max}$) and the spatial variance of rotational entropy (S_l) for uncertainty sampling. Using transfer learning, the DCNN was trained using variable subsets of 1.5T and 3T images based on $\hat{\phi}_{max}$ and S_l.

Results:

Both uncertainty metrics stratified 3T images, which correlated with localization error. When 3T cases were sorted by $\hat{\phi}_{max}$ or S_l, only 60 were required to achieve apex localization within 16mm. Reversed and random sorting required greater number of 3T cases. Inclusion of 1.5T image data was necessary to preserve performance on 1.5T images.

Conclusions:

Our proposed uncertainty sampling metrics are effective at stratifying cases with higher localization error. Use of these metrics may reduce the number of training cases required to generalize performance of DCNNs across variations in MRI equipment or acquisition parameters, while minimizing expense of curating training data.

4.2 Introduction

Cardiac magnetic resonance imaging (MRI) is an essential imaging technique for accurate assessment of cardiac morphology, size, and function [1]. However, these exams are challenging to perform, requiring expertise from dedicated cardiac technologists and physicians to accurately place imaging planes [48], [49]. Cardiac imaging planes have previously been automatically prescribed by employing deep convolutional neural networks (DCNN) with heatmap regression to localize cardiac landmarks at 1.5T field strength [70]. While many centers perform cardiac MRI with equipment at 1.5T field strength, cardiac MRI at 3T field strength is becoming increasingly available at clinical imaging centers [79]–[81]. Images acquired at the higher 3T field strength appear visually distinct from those acquired at 1.5T; for example, 3T acquisitions generally have poorer myocardium to blood pool contrast due to specific absorption rate limits [82]. Moreover, acquiring cardiac MRIs at the higher 3T field strength may introduce off-resonance, chemical shift, and flow artifacts, impacting the resulting image and potentially obscuring cardiac anatomy. It therefore remains unclear if the initial performance of DCNNs at 1.5T field strength may generalize to 3T.

The uniform performance of DCNNs across all MRI acquisition parameters is not guaranteed. Prior work by Knoll et al. highlighted the importance of including

47

heterogeneous training data to generalize DCNN performance across MRI acquisition parameters [31]. However, curating a diverse biomedical imaging dataset may be challenging due to legal, ethical, or logistical obstacles [34]–[37]. Given these constraints, there has been prominent attention given to improving the data-efficiency of DCNN training – where one is able to achieve a similar level of performance with a reduced amount of data [38]. Comparatively less attention, however, has been paid to improving DCNN training data-efficiency within the context of different MRI parameters.

Many strategies have been described for improving the generalizability of DCNNs, but not all of them may be suitable for medical imaging due to the time, labor, and monetary costs associated with collecting high quality ground truth annotations [34]–[37]. Uncertainty sampling, a form of active learning, aims to improve the data-efficiency of machine learning training by identifying the data with the greatest uncertainty, resulting in fewer cases needed for training [83], [84]. However, quantifying uncertainty for imaging data is not a straightforward task. One simple formulation for uncertainty quantification uses the output layer as a surrogate for DCNN confidence in predictions [85], [86]. These formulations were used with an active learner to improve the data-efficiency across multiple non-medical benchmark datasets. Despite initial promise, these formulations were specific to multiclass classification, and may not generalize to regression DCNNs. While prior strategies for uncertainty quantification focused on DCNN output layer confidence in prediction, testing by image transformations (e.g. rotation, translation, and warping) was proposed by Wang et al. to estimate the uncertainty related to the image [87], [88]. Here, instead of focusing on

model robustness, we propose using image rotations of test data for uncertainty sampling.

Our paper makes the following contributions: After we observed that localization performance of an initial DCNN trained on 1.5T long-axis images (referred to as the initial LaxLocNet) performed poorly at 3T field strength, we sought to leverage active learning to teach it to perform well at both field strengths. Specifically, we define two novel uncertainty measures; maximum pseudoprobability metric based on DCNN activation, and spatial variance of rotational entropy based on test-time image augmentations. We demonstrate that these metrics can be used for uncertainty sampling to increase the data-efficiency of transfer learning by reducing the number of
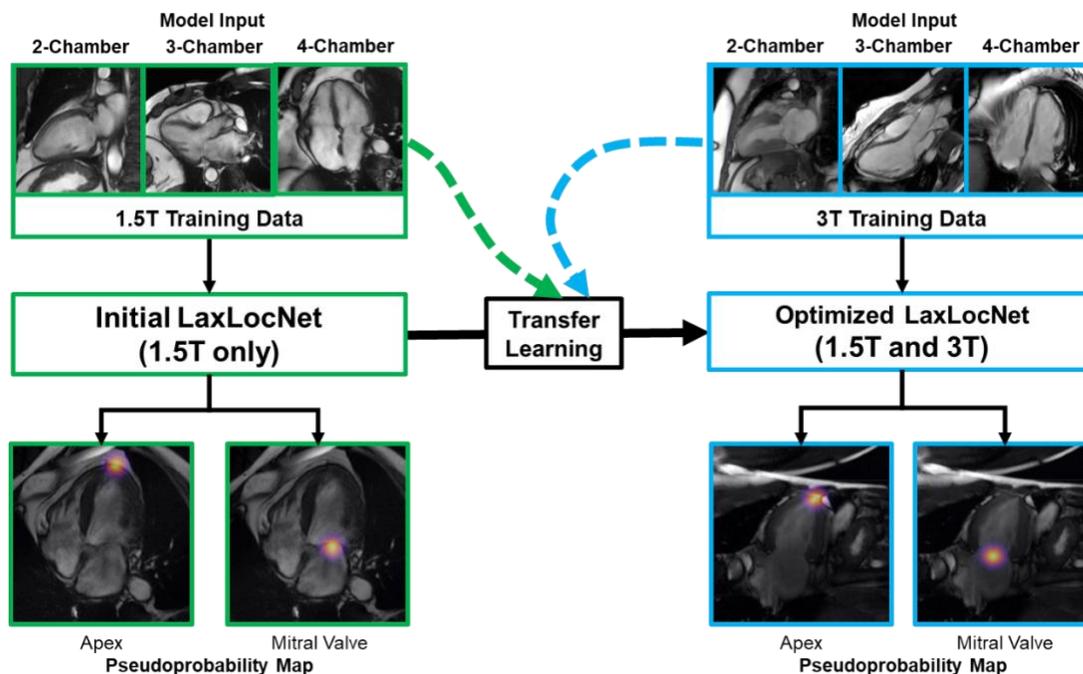
Figure 13*: Schematic of transfer learning strategy for optimizing performance of a deep convolutional neural network.*

We took an initial implementation of LaxLocNet, trained with only on 1.5T long-axis images to localize the mitral valve and apex. Controlling the number of 1.5T and 3T images, transfer learning was then used to improve performance on 3T images, while preserving performance on 1.5T images.

training cases required to generalize performance to both 1.5 and 3T field strengths. We anticipate that this strategy can be extended to improve the performance of DCNNs across a wide variety of MRI equipment and acquisition parameters, while reducing the data overhead required.

4.3 Methods

4.4.1 Heatmap Localization

We implemented a variant of heatmap localization as proposed by Payer et al. to localize the cardiac apex and the mitral valve (MV) (Figure 6) [13], [14]. We define target pseudoprobabilities for a set of i landmarks; $L = \{l_1, l_2, \ldots l_i\}$. For each of the l_i landmarks, ground truth pseudoprobabilities (P_l) are defined as a Gaussian function centered at the expert-defined ground truth coordinates ($x^*{}_l$) and spread (σ_l) according to (1). Our DCNN (2) learns to map the training set of images (I) to predicted pseudoprobabilities ($\hat{P}_l$) by identifying the weights **w** and **b** that minimize the L$_2$ loss as defined in (3). Predicted anatomic localizations ($\hat{x}_l$) can then be determined by identifying the coordinates with maximal pseudoprobability according to (4).

$$P_l = f_l(x; x^*{}_l, \sigma_l) = e^{-\frac{1}{2\sigma_l^2}(x - x^*{}_l)^2} \tag{1}$$

$$\hat{P}_l = g_l(I; \boldsymbol{w}, \boldsymbol{b}) \tag{2}$$

$$\min_{w,b} \sum_l^L \sum_x \|f(x; x^*{}_l, \sigma_l) - g_l(I; w, b)\|^2 \tag{3}$$

$$\hat{x}_l = \underset{x}{argmax}\, g_l(I; \boldsymbol{w}, \boldsymbol{b}) \tag{4}$$

4.4.2 Pseudoprobability Maximum

In previous descriptions of the heatmap localization method, only the maximal pseudoprobability coordinate is used for localization, without an explicit interpretation of

the confidence of prediction. We therefore extend the work of Payer et al. by applying a confidence interpretation of the predicted pseudoprobabilities [13], [14]. A wide variety of uncertainty metrics are possible, but for sake of simplicity, we propose the pseudoprobability maximum ($\hat{\phi}_{max}$) as a metric of neural network confidence for each landmark, described in (5) and Figure 14, top. Confident DCNN predictions are proposed to have higher predicted pseudoprobability maximum values, whereas less confident predictions are proposed to have lower values, reflecting greater uncertainty.

$$\hat{\phi}_{max} = max \; g_l(I; \boldsymbol{w}, \boldsymbol{b}) \tag{5}$$
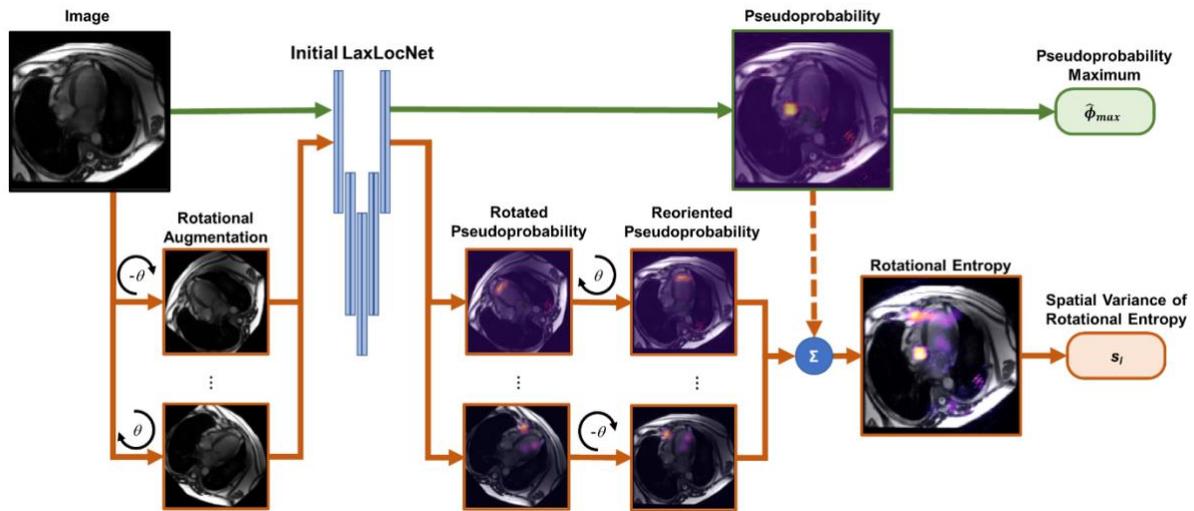
Figure 14*: Schematic of transfer learning strategy for optimizing performance of a deep convolutional neural network.*

The initial localization network is provided test images from which the image-wide pseudoprobability maximum is calculated. The initial localization network is also presented rotated test images. Rotated pseudoprobabilities are then reoriented into the original image orientation to generate a rotational entropy map. The spatial variance of the rotational entropy map is then calculated as a second uncertainty metric.

4.4.3 Spatial Variance of Rotational Entropy

Given the variety of possible DCNN formulations and training strategies, we sought to develop an additional uncertainty metric to capitalize on a concept of rotational invariance. For a DCNN with high confidence and broad generalizability, its predictions should be invariant to simple perturbations, such as image rotation. A visual representation of DCNN uncertainty can thus be defined by the composite of pseudoprobabilities generated with these perturbations. We developed a metric to quantify these perturbations, which we refer to as spatial variance of rotational entropy (s_l) for each landmark as follows and illustrated in Figure 14, bottom.

Given the rotational range set $\Theta = \{\theta_1, \theta_2, ... \theta_m\}$, we define $I_{l,\theta}$ as the rotation of image I_l about its center by angle θ_j. This can be implemented using test time augmentation to generate a set of rotated predicted pseudoprobabilities ($\hat{P}_{l,\theta}$) (6). These pseudoprobabilities may be reoriented into the original orientation ($\hat{P}_{l,\theta}{}^*$) and the mean absolute error calculated to determine their deviations from the non-rotated prediction ($\hat{P}_l$). This cumulative entropy map (E_l), thus provides a composite visual representation of the uncertainty and instability of the DCNN inference to rotation, described in (7).

$$\hat{P}_{l,\theta} = g_l(I_{l,\theta}; \boldsymbol{w}, \boldsymbol{b}) \tag{6}$$

$$E_l = \frac{1}{m}\Sigma_\theta |\hat{P}_l - \hat{P}_{l,\theta}{}^*| \tag{7}$$

We calculated the spatial variation of the rotational entropy maps to quantify uncertainty into a singular value. We first calculated Hu's image moments of the rotational entropy map (8) [89]. Horizontal and vertical variances were calculated as the second order moments in (9) and (10). We then defined the spatial variance of

rotational entropy (s_l) as the quadratic mean of the horizontal and vertical variances
(11).

$$M_{i,j} = \Sigma_x^i \Sigma_y^j x^i y^j E(x,y) \tag{8}$$

$$\mu'_{2,0} = M_{2,0} / M_{0,0} - \left(M_{1,0} / M_{0,0}\right)^2 \tag{9}$$

$$\mu'_{0,2} = M_{0,2} / M_{0,0} - \left(M_{0,1} / M_{0,0}\right)^2 \tag{10}$$

$$s_l = \sqrt{\mu_{2,0}^2 + \mu_{0,2}^2} \tag{11}$$

4.4.5 Data

With HIPAA compliance and IRB waiver of informed consent, we retrospectively

collected 405 cardiac MRI examinations at our institution between February 2012 to

July 2019. Patients were 60% male and 40% female with a median age of 53 (range of

18-111). Among these examinations, 285 (70%) were collected with a 1.5T GE Signa

HDxt MRI, (GE Healthcare, Chicago, IL) and 120 (30%) were collected with a 3T GE

750 MRI, (GE Healthcare, Chicago, IL). These examinations included 259 2-chamber

series, 311 3-chamber series, 403 4-chamber series, and 444 other long-axis series.

For patients scanned at 1.5T, cine steady-state free precession (SSFP) images were

performed with the following parameters: median flip angle 50.08° (range: 24.71° to

78.08°), 256x200 matrix, median field of view of 340 mm (range: 310 mm to 440 mm),

median slice thickness of 8 mm (range: 6 mm to 10 mm), median repetition time of 4.11

ms (range: 3.19 to 4.45 ms) and echo time of 1 ms. For patients scanned at 3T, cine

steady-state free precession (SSFP) images were performed with the following

parameters: median flip angle, 43.46° (range: 41.98° to 84.92°), 192x192 matrix,

median field of view of 360 mm (range: 320 mm to 400 mm), median slice thickness of 8

mm (range: 8 mm to 10 mm), median repetition time of 3.61 (range: 3.27 to 4.60 ms and echo time of 1 ms). All images were annotated by radiologists Albert Hsiao or Seth Kligerman who marked the location of the apex and MV using custom localization software written with python and Matplotlib.

4.4.6 Initial DCNN

We first created an initial DCNN to localize the apex and MV on 1.5T long-axis images, which we refer to as LaxLocNet, implemented as a U-net modified for heatmap regression [13], [14], [20]. The final convolutional layer was replaced with linear activation and kernel size of 1 to predict pseudoprobability values. We performed all deep learning experiments using Keras with TensorFlow backend. The initial LaxLocNet was trained with 80% of the 1.5T images. We applied rotational augmentations by rotating these images and their paired pseudoprobability maps from -60° to +60°. Initial training was performed with stochastic gradient descent, a learning rate of 10^{-4}, and momentum of 0.99. We then confirmed the performance of initial LaxLocNet by measuring the distance between ground truth and predicted localizations in a held-out validation set of the remaining 20% 1.5T data. Differences in localization error means were compared using t-test in R (R Foundation for Statistical Computing, Vienna Austria). P-Values < 0.05 were considered statistically significant.

4.4.7 DCNN Transfer Learning

We used transfer learn ing to generalize our initial LaxLocNet from 1.5T to 3T field strength (Figure 6). Transfer learning is a technique to generalize an initial DCNN to a greater range of tasks using a smaller amount of data [90]. From our 3T dataset, we separated 60% of cases for training data to be used for transfer learning and

reserved the remaining 40% for model evaluation. Our transfer learning DCNN was initialized with the weights from the initial LaxLocNet model and was further trained with varying amounts of the 1.5T training and 3T training data. We applied rotational augmentations by rotating images and paired pseudoprobability maps from -60° to +60°. Transfer learning was performed using stochastic gradient descent with a learning rate of 10^{-5} and momentum of 0.99. We evaluated our generalized models by measuring the distance between predicted and ground truth localizations from the reserved evaluation 1.5T and 3T datasets. Transfer learning experiments were validated for robustness using 5-fold cross validation.

4.4.8 Uncertainty Metric Assessment

To assess the ability of our uncertainty metrics (pseudoprobability maximum and spatial variance of rotational entropy) for identifying 3T cases with high vs. low initial LaxLocNet prediction confidence, we first divided our uncertainty metrics into quartiles. Quartiles were then plotted against initial LaxLocNet 3T localization error for each landmark. Differences in inter-quartile localization error and localization variances were compared using Kruskal–Wallis and Levene's test in R (R Foundation for Statistical Computing, Vienna Austria) respectively. P-Values < 0.05 were considered statistically significant.

4.4.9 Uncertainty Sampling

Uncertainty sampling is a machine learning methodology for selecting cases by greatest uncertainty, thereby potentially reducing the total volume of data required for DCNN performance [83], [84], [91]. To assess if we could improve the data-efficiency of improving 3T localization by incorporating uncertainty sampling, we developed a series

55

of transfer learning experiments. The data-efficiency of uncertainty sampling (using either pseudoprobability maximum or spatial variance of rotational entropy) was compared to reversed sorting (using our proposed uncertainty metrics) and to random order. We assessed the data-efficiency for each transfer learning experiment by plotting the averaged localization error for each of the 5-fold cross validation experiments against the number of cases used for transfer learning.

To gauge how well our 3T optimized LaxLocNet preserved performance at 1.5T field strength, we conducted a series of transfer learning experiments using varying amounts of the initial 1.5T and 3T training data. We assessed how well our 3T optimized LaxLocNet preserved initial 1.5T performance by plotting each landmark's 1.5T test data localization error against increasing amounts of the initial 1.5T and 3T training for each of the 5-fold cross validation experiments.
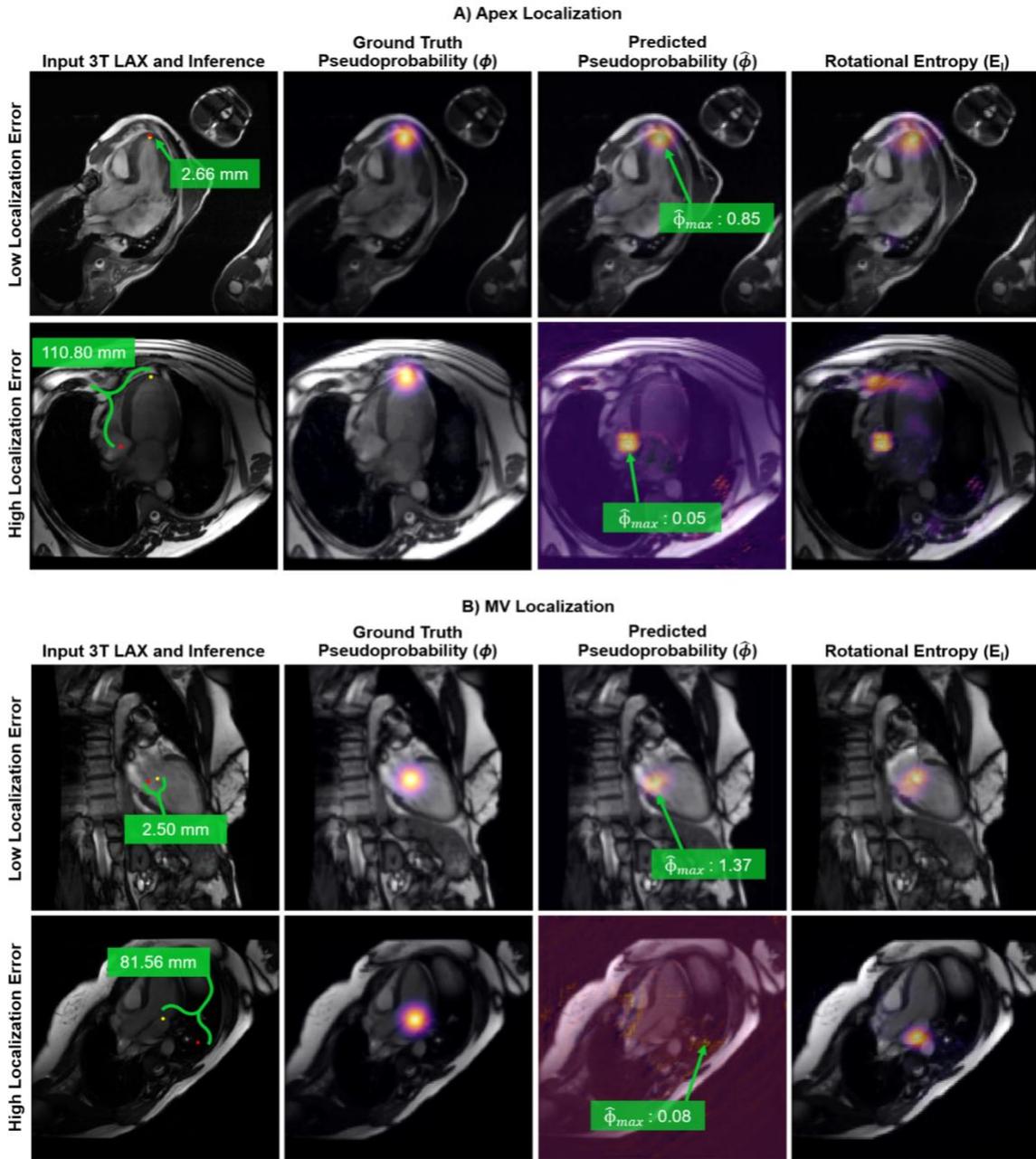
Figure 15*: Representative images of initial LaxLocNet predictions and uncertainty map*

Cases with low and high localization error for a) apex and b) mitral valve (MV) are shown. Input images are on the left, with a yellow dot indicating the ground truth and a red dot indicating the maximum inference. When localization error is low, the predicted pseudoprobability map often shows high peak intensity and narrow width and is invariant to rotation. When localization error is high, predicted pseudoprobability is diffuse with lower peak intensity and wide rotational variation.

4.5 Results

4.5.1 Initial LaxLocNet Performance

We first assessed the performance of our initial LaxLocNet by measuring the localization error between predicted and ground truth localization on an independent test set of 1.5T long-axis data. We found that the apex was localized within 9.64 mm, and the MV within 7.18 mm. To assess how well the initial LaxLocNet generalized to 3T, we compared the localization results of images acquired at 1.5T to images acquired at 3T. We found that the initial LaxLocNet had greater localization error on 3T images, with the apex localized within 29.79 mm vs. MV within 15.5 mm ($p < 0.01$).

4.5.2 Qualitative Assessment of Initial LaxLocNet Predictions

We reviewed the initial LaxLocNet pseudoprobability predictions on 3T long-axis images with high vs. low localization error to help guide the development of our proposed uncertainty metrics (Figure 15). First, images with low localization error had higher predicted pseudoprobabilities, which we quantified as the proposed maximum pseudoprobability uncertainty metric. Moreover, these images also had greater spread in their rotational entropy map, which we quantified as spatial variance of rotational entropy. We found that the 3T long-axis images with higher localization error had off-resonance artifacts and a greater range of image contrast.
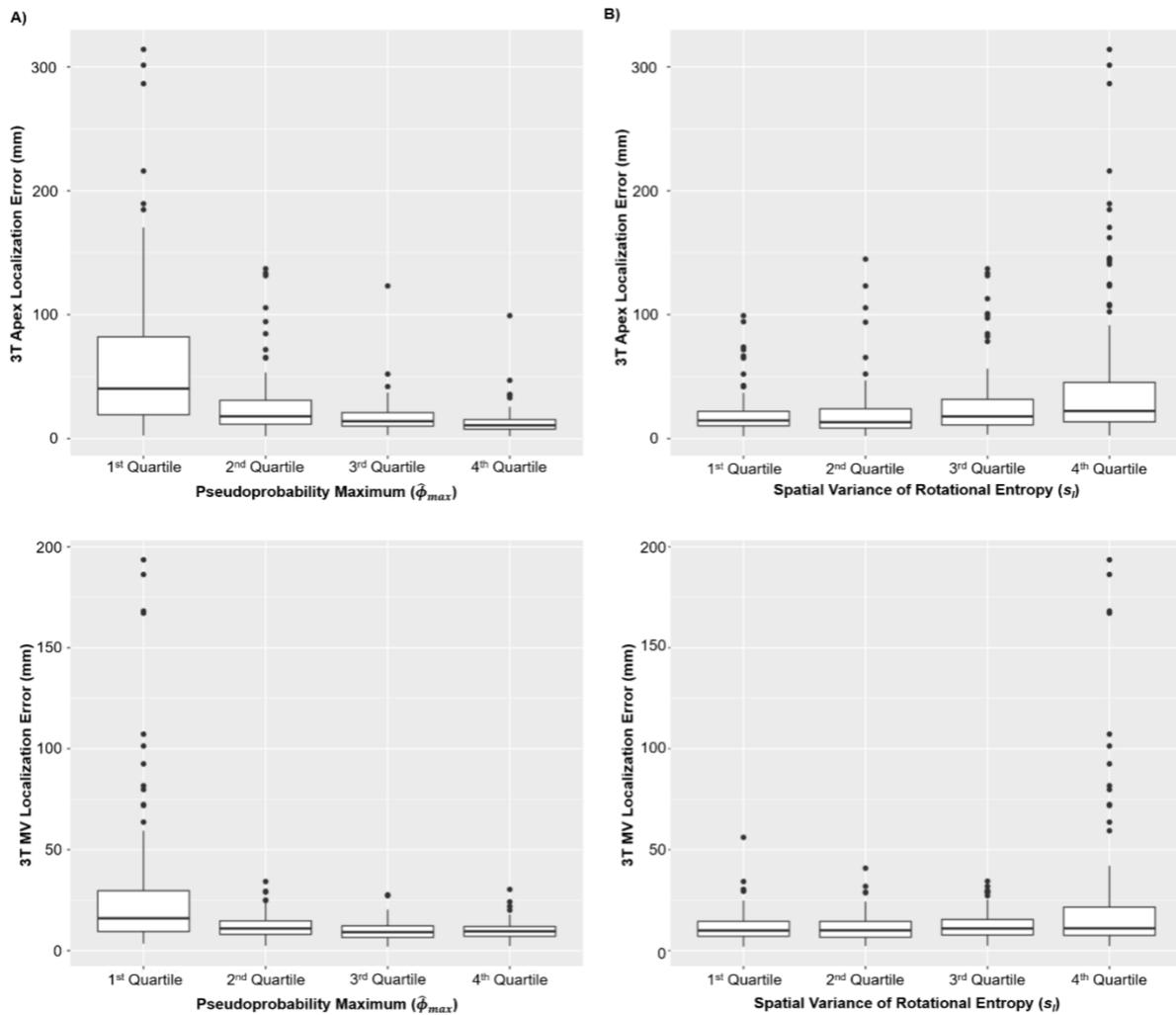
Figure 16*: Relationship between localization error and metrics of model uncertainty.*

To gauge the relationship between our uncertainty metrics a), pseudoprobability maximum ($\hat{\phi}_{max}$) and b) spatial variance in spatial variation in rotational entropy (s_l), we binned our uncertainty metrics into quartiles. We then plotted these quartiles against initial LaxLocNet 3T localization error for the apex and mitral valve (MV). Higher uncertainty (low pseudoprobability maximum values and higher spatial variation in rotational entropy) were associated with greater 3T localization error and variance.

4.5.3 Uncertainty Metric Performance

To assess the potential of our uncertainty metrics to gauge initial LaxLocNet

confidence in 3T predictions, we first divided our uncertainty metrics into quartiles, and

plotted each quartile's 3T localization error for each landmark (Figure 16, Table 4). 3T

Table 4*: Error in localization of the apex and mitral valve prior to transfer learning*

Landmark	Uncertainty Metric	Q1	Q2	Q3	Q4	Kruskal Test	Levene's Test
Apex	$\widehat{\phi}_{max}$	40.32±4,084.93 mm	17.93±722.02 mm	13.99±191.68 mm	10.70±134.38 mm	p<0.01	p<0.01
	S_l	14.59±327.79 mm	13.19±552.40 mm	17.85±878.44 mm	22.26±4359.92 mm	p<0.01	p<0.01
MV	$\widehat{\phi}_{max}$	15.99±1,428.23 mm	10.95±35.98 mm	9.15±24.15 mm	9.54±21.45 mm	p<0.01	p<0.01
	S_l	9.95±59.58 mm	10.04±43.16 mm	10.96±47.08 mm	11.04±1,503.55 mm	p=0.09	p<0.01

Cases with the lowest quartile of pseudoprobability maximum ($\widehat{\phi}_{max}$), showed the greatest localization error. Median standard deviation are shown displayed. Cases with the highest quartile of spatial variance of rotational entropy (s_l) showed the highest localization error.

images with low pseudoprobability maximum (i.e. higher uncertainty) had both greater

localization error and variance for the apex and MV. Similarly, 3T images with higher

spatial variance of rotational entropy (i.e. higher uncertainty) had higher localization

error and variance for the apex. However, while greater values of spatial variance of

rotational entropy had greater MV localization variance, there was not a statistical

difference in MV localization error across quartiles (p=0.09).

4.5.4 Uncertainty Sampling Improves Data-Efficiency

We aimed to assess the potential of uncertainty sampling to improve the data-
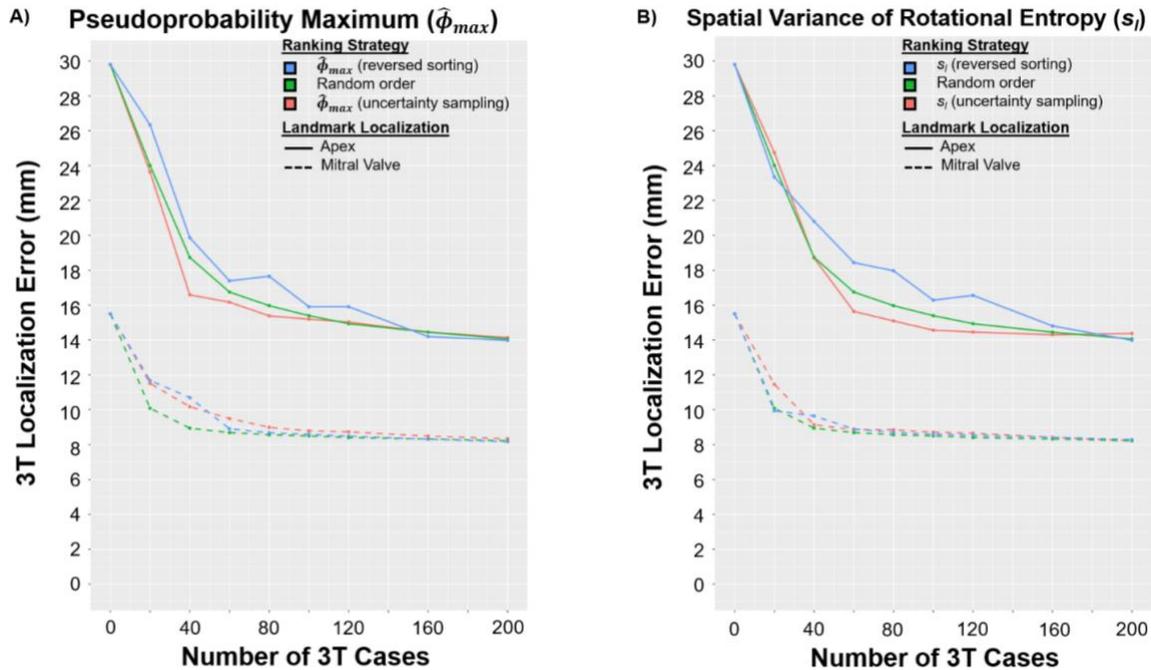
Figure 17: *Uncertainty sampling increases data-efficiency for transfer learning when selecting data by either a) pseudoprobability maximum ($\hat{\phi}_{max}$) or b) spatial variance of rotational entropy (s_l)*

Each line represents an average of five sets of trainings, while varying the number of input 3T training images. When sorting properly in order of decreasing uncertainty (red line), apex localization error improved more rapidly with fewer number of cases than random sorting (green line) or reversed ordering (blue line). We did not observe the same effect for the mitral valve, which started from a small baseline localization error.

efficiency of transfer learning by conducting a series of transfer learning experiments. In

each transfer learning experiment, we compared the average of each 5-fold cross

validation, varying the number of 3T images. When selecting 3T images for transfer

learning, we selected top ranking cases by our uncertainty metrics, random order, and

reversed uncertainty order. Performance of each selection method was assessed by

determining the number of 3T images required to improve 3T localization performance

(Figure 17).

We first assessed the ability to use uncertainty sampling to improve initial LaxLocNet localization of the apex on 3T images. This required substantially fewer cases when using uncertainty sampling with either pseudoprobability maximum or spatial variance of rotational entropy for selecting transfer learning cases than random selection and reversed order. Regardless of the uncertainty metric used, only 60 cases were required to localize the apex within 16 mm, compared to 80 cases for random selection. Backward sorting by reversing pseudoprobability maximum required 100 cases to localize the apex within 16 mm and required 160 cases for spatial variance of rotational entropy.

We also assessed the effectiveness of uncertainty sampling to improve initial LaxLocNet 3T localization of the MV on 3T images. Here, we found that the MV localization was already within 15.5 mm prior to additional training. Selection of 3T cases by either of our proposed uncertainty metrics did not improve the data-efficiency of 3T MV localization compared to random and reversed order.
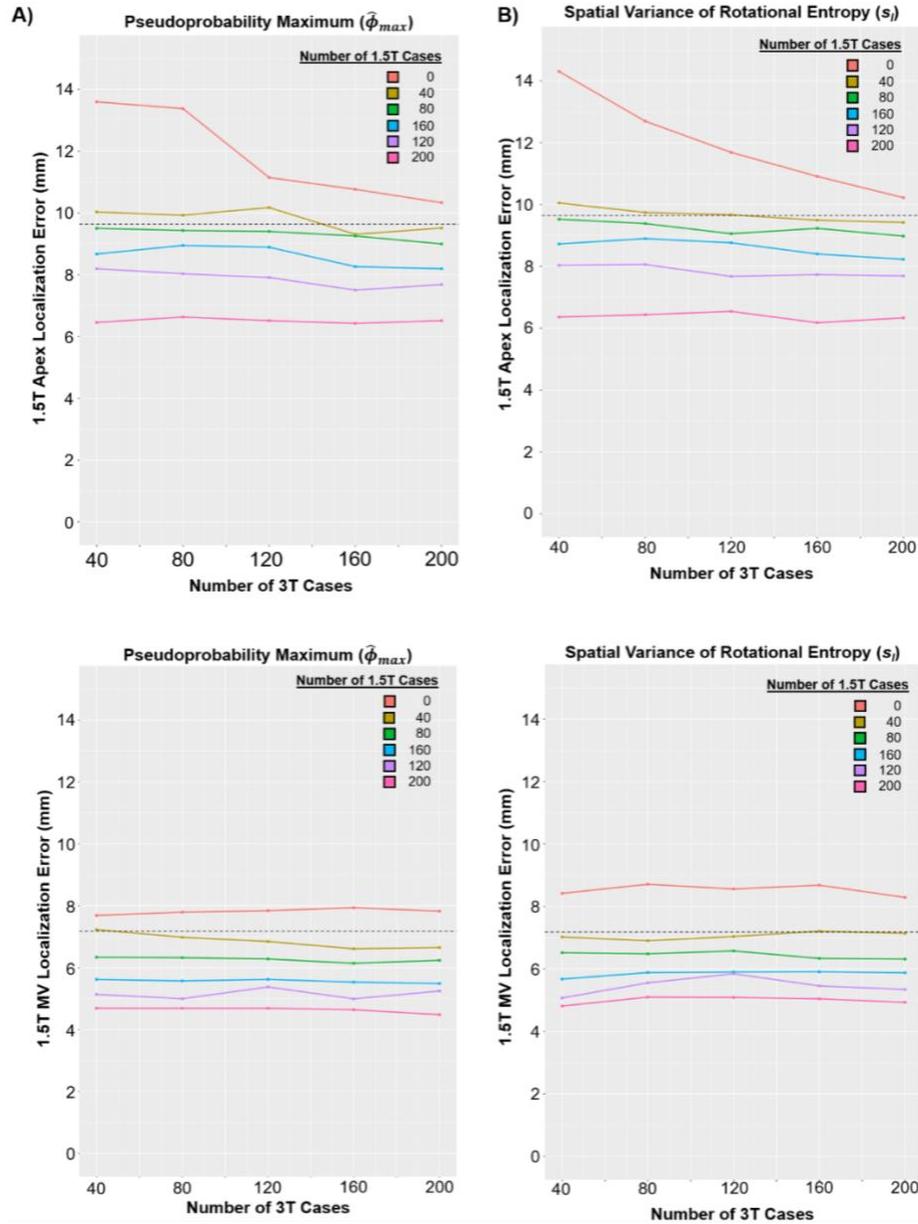
Figure 18*: Preservation of Performance for 1.5T Requires Inclusion of 1.5T Training Data.*

To gauge the importance of including 1.5T training data for transfer learning, we assessed the effect of training with varying numbers of 1.5T and 3T images. Regardless whether either a) pseudoprobability maximum ($\hat{\phi}_{max}$) or b) spatial variance in rotational entropy (s_l) was used, 1.5T localization performance improved with greater numbers of 1.5T images. When very few 1.5T images were used, 3T images surprisingly also improved performance on 1.5T images.

4.5.5 Preservation of 1.5T Localization Performance

To determine how well this active learning strategy preserved the localization of the apex and MV for images obtained 1.5T field strength, we assessed the localization error while varying the amount of 1.5T and 3T data used in transfer learning (Figure 18). Increased number of 1.5T images improved localization performance for both the apex and MV. This was true when using either uncertainty metric. Surprisingly, increasing the number of 3T images also improved localization of the apex, especially when few 1.5T images (less than 80) were applied to transfer learning. The choice of uncertainty metric used to select 3T images for transfer learning did not affect the localization error on 1.5T images.

4.6 Discussion

In this work, we proposed a methodology to efficiently generalize a DCNN to multiple field strengths by combining uncertainty sampling with transfer learning. Starting with an initial LaxLocNet that had poor performance on 3T long-axis images, we proposed two uncertainty metrics (the maximum pseudoprobability metric and the spatial variance of rotational entropy metric) to identify images with low confidence in initial LaxLocNet prediction. Moreover, we demonstrated that these uncertainty metrics could be used to reduce the number of cases to improve apex localization on images from 1.5T to 3T magnets. We further show that inclusion of 1.5T images in transfer learning helps to preserve performance on 1.5T images.

There has been great promise of DCNNs to automate and augment labor intensive aspects of cardiac MRI [70], [92]–[94]. However, it is not guaranteed that these DCNNs are extensible beyond the scope of their initial training environments.

Differing MRI acquisition protocols may have substantial differences in imaging

characteristics, due to factors such as field strength, pulse sequence, or underlying

hardware [79]–[81], [95]. To expand the utility of DCNNs outside of their initial scope of

training environment, it is necessary to generalize these DCNNs across a wide range of

imaging parameters [31]. In our work, we found that our initial LaxLocNet performed

poorly on 3T long-axis images, especially for those images with substantial off-

resonance banding artifacts and greater myocardial-to-blood pool contrast (Figure 15A).

Since these imaging characteristics are uncommon in the initial training set of images

acquired at 1.5T field strength, the initial LaxLocNet never had adequate exposure to

these features. To overcome the lack of these imaging characteristics in our initial 1.5T

training data, we used transfer learning to generalize our initial LaxLocNet efficiently by

identifying the 3T images with the greatest uncertainty.

In our uncertainty sampling experiments, we demonstrated that combining

transfer learning with uncertainty sampling increased the data-efficiency for 3T apex

localization. These data-efficiencies were followed by random selection and reversed

sorting, suggesting that pseudoprobability maximum and spatial variance of rotational

entropy were able to routinely identify 3T images with visual information not contained in

the initial 1.5T training data important for apex localization. However, this increased

data-efficiency for 3T generalization was not replicated for localizing the MV on 3T

images. Selecting 3T transfer learning cases by uncertainty sampling failed to improve

data-efficacy compared to random selection and reversed ordering. Since the MV

localization was substantially better than apex localization on 3T images to begin with,

there was not as much opportunity to improve localization performance to be gained from learning from highly uncertain cases.

The qualitative differences between 1.5 and 3T imaging are representative of a larger suite of differences between MRI formats and parameters [95]. Our current study highlights the potential of MRI acquisition parameters to confound DCNN predictions and led to a proposed methodology to efficiently generalize the initial LaxLocNet from 1.5T field strength to 3T. Further evaluation of how well this technique improves the data-efficiency for generalizing DCNNs across other MRI acquisition parameters is therefore warranted. Nevertheless, the kinds of methodological strategies introduced here can be broadly implemented to DCNN active learning, enabling the clinical implementation of DCNN-based tasks.

4.7 Conclusions

In this study, we demonstrated that the combination of uncertainty sampling with transfer learning improved can be used to efficiently generalize DCNNs to perform well beyond their initial training environment. This proposed strategy and uncertainty metrics may be applied to ensure robust performance of these algorithms in the clinical environment. Uncertainty sampling can reduce the amount of data required, to minimize the cost and labor of curating ground truth annotations that require radiological expertise and enable the broad clinical use of DCNNs.

4.8 Acknowledgements

I wish acknowledge Drs. Albert Hsiao, Tara Retson and Seth Kligerman for assistance in collecting and annotating the cardiac MR data used in this manuscript, Drs. Albert Haiso and Naeim Bahrami for assistance in the experimental design, and

Chapter 5. Conclusions

With this dissertation, I have demonstrated the feasibility of planning cardiac MR imaging planes by localizing the apex and valve image positions on routinely collected cardiac images. I then developed this method into a platform to interrogate important deep convolutional neural networks (DCCNs) concepts, including generalizability, the fusion of multi-image information, and adapting DCNNs to a changing clinical environment.

In the first chapter, I hypothesized that heatmap regression-based DCNNs could identify the apex and valve landmarks for the purposes of prescribing the cardiac MR imaging planes. I then explored the potential of prescribing these landmarks on cardiac MR images that are routinely collected during an exam, providing a rationalization of the basis for prescription. This additionally provided expert MR operators the opportunity to seamlessly intervene in the rare case of an incorrect localization.

In the second chapter, I expanded the previous DCNN-based system for prescribing the cardiac MR imaging planes to include prescription of the dedicated axial and vertical long-axis imaging planes. This system prescribed cardiac MR imaging planes that were either comparable or exceeded the performance of cardiac MR technologists across multiple clinical environments and field strengths. To accomplish this, I hypothesized that training DCNNs with a complement of both 1.5T and 3T data would provide a wide range of distinct visual information. Performance of the DCNNs were found to be preserved across each clinical site, indicating that curation of diverse datasets may be required to ensure adequate DCNN performance across different clinical environments.

I additionally assessed how to integrate multiple DCNNs to improve the analysis of cardiac MR images. In this subaim, I hypothesized that an initial long-axis prescription could be refined from a short-axis image by cross-referencing previously acquired long-axis image for the apex image position. By integrating imaging information from multiple cardiac views, I was able to refine the automatic prescription of the dedicated long-axis images. In the current implementation of the system for prescribing the cardiac MR imaging planes, the intermediate step of acquiring a VLAX image is required for prescription of the SAX image stack. However, further refinement of the AxLocNet to accurately localize the MV and apex from an axial stack of images may be further investigated to directly prescribe the SAX prescription. This may allow for either reduction of scan time or site-specific modification of the system due to fit institutional preference.

In the third chapter, I use the long-axis localization model to explore the potential of adapting existing models to work in a new clinical environment. I hypothesized that combining transfer learning and uncertainty sampling could efficiently adapt a model trained using only 1.5T data to work at both 1.5T and 3T field strengths. To test this, two novel uncertainty metrics were developed—pseudoprobability maximum based on a DCNN's inferred pseudoprobability map, and spatial variance of rotational entropy based on test-time rotational augmentations. I demonstrated that these uncertainty metrics could be used to reduce the number of images required to generalize performance of DCNNs while retaining initial performance.

Despite the wide availability of and inexpensive nature of echocardiography equipment, this modality is often limited in the amount of diagnostic information that can

be gleaned [96]. For example, echocardiography is often inaccurate with respect to volumetric and functional metrics compared to the gold standard of cardiac MRI [1], [97]. Cardiac MRI on the other hand has higher through-slide resolution, allowing for detailed structural evaluation and tissue characterization [96], [98]. Additionally, single photon emission computed tomography (SPECT) and cardiac computed topography (CT) are both additional complimentary imaging modalities to cardiac MRI which do not require technologists to prescribe the cardiac views at the time of image acquisition. While these modalities are commonly used to evaluate cardiac perfusion, they require the use of ionizing radiation. Ordering physicians must therefore balance the clinical need of the exam with the effects of additional doses of radiation [99], [100]. Despite the fact that newer CT scanners and scanning techniques have dramatically lowered the amount of radiation exposure, the cumulative effects of repeated radiation exposure associated with perfusion CT exams have remained a concerns among public health researchers [101]. Additionally, recurring safety incidents involving technologist errors in radiation dosing highlight the potential safety aspects of any techniques that utilize radiation [102].

Since the supply of specialty-trained physicians and technologists is the primary limiting factor of the availability of cardiac MRI, this system of DCNNs for prescribing the cardiac MR imaging planes can potentially alleviate some of the current barriers to accessing this essential cardiac imaging modality. For clinical institutions that have already have an established cardiac MR program, integration of this system may enhance acquisition workflows. For example, full-implementation of this system into a clinical environment may only require minimal technologist supervision of the system,

allowing the technologist to focus on tasks such as IV contrast administration and patient monitoring. This could result in significant time savings for cardiac MRI exams, which studies have highlighted as a major bottleneck for cardiac MRI utilization [103], [104].

The financial and fiscal reality for most hospitals is that they are largely dependent on patient volume to cover the considerable fixed and variable costs of operating a radiology department [105]. The volume of expected cardiac MRIs based on average turnaround times may determine the operational and financial viability of offering these specialized services [106]. Adopting the proposed DCNN based system for cardiac MRI planning may therefore decrease the marginal costs associated with cardiac MRI examination, thereby enabling greater patient access to this essential modality.

From a patient perspective, reduction in scan durations may increase patient satisfaction, especially with those who experience anxiety when undergoing MRI imaging. Currently, up to 37% of patients report moderate to high levels of anxiety during MRI exams, and 3-5% are not able to complete the exam [107]–[109]. Reducing the time required to fully scan a cardiac MRI may therefore reduce potential psychological burden and increase the willingness of patients to undergo and complete cardiac MRI studies.

While there is great promise and potential for the ability of this system of DCNNs to help standardize and improve accessibility of cardiac MRI, one must be proactive to minimize potential unintended consequences. One major concern in deploying DCNN-based systems into clinical practice is automation complacency, where clinicians

availed to such conveniences may become overly reliant on technology substituting for clinical judgement [110], [111]. In the case of a system malfunction or DCNN error, the human operator responsible for correcting any faults may lack the proper training to operate without the technological assistance. Failure of prescribing the cardiac imaging planes accurately could result in either suboptimal diagnostic scans or a repeat scanning procedure, increasing both the physical and financial cost to the patient as well as the hospital. In safety-critical environments, these automated systems must provide a graceful method to provide end-user feedback and the opportunity for corrective input [112]–[114]. In the proposed system of using DCNN to prescribe the cardiac MR image planes, we localized the image positions of the valve and apex landmarks which define these imaging planes. Further research within the human-computer interface domain that leverages human psychology may be necessary to explore the optimal way to provide clinical end-users with this feedback. For example, it is unclear if the image position alone or the underlying inferred pseudoprobability heatmaps may optimally aid in providing a robust human-machine system. Incorporating uncertainty metrics such as pseudoprobability maximum or spatial variance of rotational entropy may additional provide understanding of the underlying confidence the DCNN based system has in prescription [115]. Future work investigating how to integrate these systems to provide a robust prescription platform will be necessary to further optimize the implementation of this system into clinical practice.

In conclusion, while there is future work that is required to investigate the most optimal integration of this system for planning cardiac MR imaging planes, there is great promise of this system to positively augment acquisition of cardiac MRI. Increasing

access to cardiac MRI may have substantial benefits to patient care, allowing clinicians

to better monitor and treat cardiac disease.

WORKS CITED:

[1] A. L. Gerche, G. Claessen, A. Van de Bruaene, N. Pattyn, J. Van Cleemput, M. Gewillig, J. Bogaert, S. Dymarkowski, P. Claus, and H. Heidbuchel, "Cardiac MRI: A New Gold Standard for Ventricular Volume Quantification During High-Intensity Exercise," *Circ. Cardiovasc. Imaging*, vol. 6, no. 2, pp. 329–338, Mar. 2013.

[2] D. T. Ginat, M. W. Fong, D. J. Tuttle, S. K. Hobbs, R. C. Vyas, G. Dt, F. Mw, T. Dj, H. Sk, and V. Rc, "Cardiac Imaging: Part 1, MR Pulse Sequences, Imaging Planes, and Basic Anatomy," *AJR*, p. 197, 2011.

[3] P. K. Woodard, D. A. Bluemke, P. N. Cascade, J. P. Finn, A. E. Stillman, C. B. Higgins, R. D. White, and E. K. Yucel, "ACR Practice Guideline for the Performance and Interpretation of Cardiac Magnetic Resonance Imaging (MRI)1 1 The American College of Radiology, with more than 30,000 members, is the principal organization of radiologists, radiation oncologists, and clinical medical physicists in the United States.," *J. Am. Coll. Radiol.*, vol. 3, no. 9, pp. 665–676, Sep. 2006.

[4] C. E. Jackson, M. D. Robson, J. M. Francis, and J. A. Noble, "Computerised planning of the acquisition of cardiac MR images," *Comput. Med. Imaging Graph.*, vol. 28, no. 7, pp. 411–418, Oct. 2004.

[5] M. Frick, I. Paetsch, C. den Harder, M. Kouwenhoven, H. Heese, S. Dries, B. Schnackenburg, W. de Kok, R. Gebker, E. Fleck, R. Manka, and C. Jahnke, "Fully automatic geometry planning for cardiac MR imaging and reproducibility of functional cardiac parameters," *J. Magn. Reson. Imaging*, vol. 34, no. 2, pp. 457–467, Aug. 2011.

[6] X. Lu, M.-P. Jolly, B. Georgescu, C. Hayes, P. Speier, M. Schmidt, X. Bi, R. Kroeker, D. Comaniciu, P. Kellman, E. Mueller, and J. Guehring, "Automatic View Planning for Cardiac MRI Acquisition," Springer, Berlin, Heidelberg, 2011, pp. 479–486.

[7] Q. Tao, B. P. F. Lelieveldt, and R. J. Van Der Geest, "Deep learning for quantitative cardiac MRI," *American Journal of Roentgenology*, vol. 214, no. 3. American Roentgen Ray Society, pp. 529–535, 31-Oct-2020.

[8] A. S. Fahmy, J. Rausch, U. Neisius, R. H. Chan, M. S. Maron, E. Appelbaum, B. Menze, and R. Nezafat, "Automated Cardiac MR Scar Quantification in Hypertrophic Cardiomyopathy Using Deep Convolutional Neural Networks," *JACC: Cardiovascular Imaging*, vol. 11, no. 12. Elsevier Inc., pp. 1917–1918, 01-Dec-2018.

[9] N. Zhang, G. Yang, Z. Gao, C. Xu, Y. Zhang, R. Shi, J. Keegan, L. Xu, H. Zhang, Z. Fan, and D. Firmin, "Deep Learning for Diagnosis of Chronic Myocardial Infarction on Nonenhanced Cardiac Cine MRI," *Radiology*, vol. 291, no. 3, pp. 606–617, Jun. 2019.

[10] M. van Assen, G. Muscogiuri, D. Caruso, S. J. Lee, A. Laghi, and C. N. De Cecco, "Artificial intelligence in cardiac radiology," *Radiol. Medica*, vol. 125, no. 11, pp. 1186–1199, Nov. 2020.

[11] J. Duan, G. Bello, J. Schlemper, W. Bai, T. J. W. Dawes, C. Biffi, A. de Marvao, G. Doumoud, D. P. O'Regan, and D. Rueckert, "Automatic 3D Bi-Ventricular Segmentation of Cardiac Images by a Shape-Refined Multi- Task Deep Learning Approach," *IEEE Trans. Med. Imaging*, vol. 38, no. 9, pp. 2151–2164, Sep. 2019.

[12] X. Wang, L. Bo, and L. Fuxin, "Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression."

[13] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Integrating spatial configuration into heatmap regression based CNNs for landmark localization," *Med. Image Anal.*, vol. 54, pp. 207–219, May 2019.

[14] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Regressing Heatmaps for Multiple Landmark Localization Using CNNs," Springer, Cham, 2016, pp. 230–238.

[15] C. Payer, D. Darkoˇdarkoštern, H. Bischof, and M. Urschler, "Coarse to Fine Vertebrae Localization and Segmentation with SpatialConfiguration-Net and U-Net."

[16] Z. Zhong, J. Li, Z. Zhang, Z. Jiao, and X. Gao, "An Attention-Guided Deep Regression Model for Landmark Detection in Cephalograms," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11769 LNCS, pp. 540–548.

[17] H. Zhang, Q. Li, and Z. Sun, "Joint Voxel and Coordinate Regression for Accurate 3D Facial Landmark Localization," in *Proceedings - International Conference on Pattern Recognition*, 2018, vol. 2018-August, pp. 2202–2208.

[18] H. Lee, M. Park, and J. Kim, "Cephalometric landmark detection in dental x-ray images using convolutional neural networks," in *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017, vol. 10134, p. 101341W.

[19] X. Yang, W. T. Tang, G. Tjio, S. Y. Yeo, and Y. Su, "Automatic detection of anatomical landmarks in brain MR scanning using multi-task deep neural networks," *Neurocomputing*, vol. 396, pp. 514–521, Jul. 2020.

[20] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, vol. 9351. Springer Verlag, 2015, pp. 234–241.

[21] J. Long, E. Shelhamer, T. Darrell, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651.

[22]   M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, Oct. 2017.

[23]   T. Pfister, J. Charles, and A. Zisserman, "Flowing ConvNets for Human Pose Estimation in Videos," Jun. 2015.

[24]   C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, no. 1. BioMed Central Ltd., 29-Oct-2019.

[25]   W. Yan, L. Huang, L. Xia, S. Gu, F. Yan, Y. Wang, and Q. Tao, "MRI Manufacturer Shift and Adaptation: Increasing the Generalizability of Deep Learning Segmentation for MR Images Acquired with Different Scanners," *Radiol. Artif. Intell.*, vol. 2, no. 4, p. e190195, Jul. 2020.

[26]   D. A. Bluemke, L. Moy, M. A. Bredella, B. B. Ertl-Wagner, K. J. Fowler, V. J. Goh, E. F. Halpern, C. P. Hess, M. L. Schiebler, and C. R. Weiss, "Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers-from the Radiology Editorial Board," *Radiology*, vol. 294, no. 2. Radiological Society of North America Inc., pp. 487–489, 31-Dec-2020.

[27]   J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," *PLOS Med.*, vol. 15, no. 11, p. e1002683, Nov. 2018.

[28]   J. Krois, A. Garcia Cantu, A. Chaurasia, R. Patil, P. K. Chaudhari, R. Gaudin, S. Gehrung, and F. Schwendicke, "Generalizability of deep learning models for dental image analysis," *Sci. Rep.*, vol. 11, no. 1, p. 6102, Dec. 2021.

[29]   L. S. Hesse, G. Kuling, M. Veta, and A. L. Martel, "Intensity Augmentation to Improve Generalizability of Breast Segmentation across Different MRI Scan Protocols," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 3, pp. 759–770, Mar. 2021.

[30]   M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. G. Guttmann, F. E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and W. M. Wells, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10435 LNCS, pp. 516–524.

[31]   F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson, "Assessment of the generalization of learned image reconstruction and the potential for transfer learning," *Magn. Reson. Med.*, vol. 81, no. 1, pp. 116–128, Jan. 2019.

[32] Q. Tao, W. Yan, Y. Wang, E. H. M. Paiman, D. P. Shamonin, P. Garg, S. Plein, L. Huang, L. Xia, M. Sramko, J. Tintera, A. de Roos, H. J. Lamb, and R. J. van der Geest, "Deep Learning–based Method for Fully Automatic Quantification of Left Ventricle Function from Cine MR Images: A Multivendor, Multicenter Study," *Radiology*, vol. 290, no. 1, pp. 81–88, Jan. 2019.

[33] A. N. Bhuva, W. Bai, C. Lau, R. H. Davies, Y. Ye, H. Bulluck, E. McAlindon, V. Culotta, P. P. Swoboda, G. Captur, T. A. Treibel, J. B. Augusto, K. D. Knott, A. Seraphim, G. D. Cole, S. E. Petersen, N. C. Edwards, J. P. Greenwood, C. Bucciarelli-Ducci, A. D. Hughes, D. Rueckert, J. C. Moon, and C. H. Manisty, "A Multicenter, Scan-Rescan, Human and Machine Learning CMR Study to Test Generalizability and Precision in Imaging Biomarker Analysis," *Circ. Cardiovasc. Imaging*, vol. 12, no. 10, Oct. 2019.

[34] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation."

[35] H. J. W. L. Aerts, "Data Science in Radiology: A Path Forward," *Clin. Cancer Res.*, vol. 24, no. 3, pp. 532–534, Feb. 2018.

[36] M. A. Levy, J. B. Freymann, J. S. Kirby, A. Fedorov, F. M. Fennessy, S. A. Eschrich, A. E. Berglund, D. A. Fenstermacher, Y. Tan, X. Guo, T. L. Casavant, B. J. Brown, T. A. Braun, A. Dekker, E. Roelofs, J. M. Mountz, F. Boada, C. Laymon, M. Oborski, and D. L. Rubin, "Informatics methods to enable sharing of quantitative imaging research data," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1249–1256, Nov. 2012.

[37] J. B. Freymann, J. S. Kirby, J. H. Perry, D. A. Clunie, and C. C. Jaffe, "Image Data Sharing for Biomedical Research—Meeting HIPAA Requirements for De-identification," *J. Digit. Imaging*, vol. 25, no. 1, pp. 14–24, Feb. 2012.

[38] H. Davíð Hlynsson, A. N. Escalante-B, and L. Wiskott, "Measuring the Data Efficiency of Deep Learning Methods."

[39] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5. Nature Research, pp. 206–215, 01-May-2019.

[40] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays," *Comput. Methods Programs Biomed.*, vol. 196, p. 105608, Nov. 2020.

[41] H. Lee, S. Yune, M. Mansouri, M. Kim, S. H. Tajmir, C. E. Guerrier, S. A. Ebert, S. R. Pomerantz, J. M. Romero, S. Kamalian, R. G. Gonzalez, M. H. Lev, and S. Do, "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nat. Biomed. Eng.*, vol. 3, no. 3, pp. 173–182, Mar. 2019.

[42] M. Sendak, M. C. Elish, M. Gao, J. Futoma, W. Ratliff, M. Nichols, A. Bedoya, S. Balu, and C. O'Brien, "'The human body is a black box': Supporting clinical decision-making with deep learning," in *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 99–109.

[43] S. Gaube, H. Suresh, M. Raue, A. Merritt, S. J. Berkowitz, E. Lermer, J. F. Coughlin, J. V. Guttag, E. Colak, and M. Ghassemi, "Do as AI say: susceptibility in deployment of clinical decision-aids," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–8, Dec. 2021.

[44] C. M. Cutillo, K. R. Sharma, L. Foschini, S. Kundu, M. Mackintosh, K. D. Mandl, T. Beck, E. Collier, C. Colvis, K. Gersing, V. Gordon, R. Jensen, B. Shabestari, and N. Southall, "Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency," *npj Digital Medicine*, vol. 3, no. 1. Nature Research, pp. 1–5, 01-Dec-2020.

[45] R. Challen, J. Denny, M. Pitt, L. Gompels, T. Edwards, and K. Tsaneva-Atanasova, "Artificial intelligence, bias and clinical safety," *BMJ Quality and Safety*, vol. 28, no. 3. BMJ Publishing Group, pp. 231–237, 01-Mar-2019.

[46] K. K. Bhatia, A. N. Price, W. Shi, J. V. Hajnal, and D. Rueckert, "Super-resolution reconstruction of cardiac MRI using coupled dictionary learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014*, 2014, pp. 947–950.

[47] A. Mortazi, R. Karim, K. Rhode, J. Burt, and U. Bagci, "CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10434 LNCS, pp. 377–385.

[48] A. Suinesiaputra, D. A. Bluemke, B. R. Cowan, M. G. Friedrich, C. M. Kramer, R. Kwong, S. Plein, J. Schulz-Menger, J. J. M. Westenberg, A. A. Young, and E. Nagel, "Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours," *J. Cardiovasc. Magn. Reson.*, vol. 17, no. 1, p. 63, Dec. 2015.

[49] M. Ferguson and R. Otto, "Cardiac MRI Prescription Planes," *MedEdPORTAL Publ.*, no. 10, 2014.

[50] M. B. Stokes and R. Roberts-Thomson, "The role of cardiac imaging in clinical practice.," *Aust. Prescr.*, vol. 40, no. 4, pp. 151–155, Aug. 2017.

[51] A. Crean, "Cardiovascular MR and CT in congenital heart disease.," *Heart*, vol. 93, no. 12, pp. 1637–47, Dec. 2007.

[52] J. W. Goldfarb, "Abstract 13092: Cardiac MRI Practice: An Analysis of 2012 and 2013 Medicare Provider Utilization and Payment Data," *Circulation*, vol. 132, no.

Suppl 3, 2015.

[53]   K. J. Choi, J. K. Jang, S. S. Lee, Y. S. Sung, W. H. Shim, H. S. Kim, J. Yun, J.-Y. Choi, Y. Lee, B.-K. Kang, J. H. Kim, S. Y. Kim, and E. S. Yu, "Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent–enhanced CT Images in the Liver," *Radiology*, p. 180763, Sep. 2018.

[54]   G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, "Deep Learning: A Primer for Radiologists," *RadioGraphics*, vol. 37, no. 7, pp. 2113–2131, Nov. 2017.

[55]   J. Lieman-Sifry, M. Le, F. Lau, S. Sall, and D. Golden, "FastVentricle: Cardiac Segmentation with ENet," Springer, Cham, 2017, pp. 127–138.

[56]   J. G. Nam, S. Park, E. J. Hwang, J. H. Lee, K.-N. Jin, K. Y. Lim, T. H. Vu, J. H. Sohn, S. Hwang, J. M. Goo, and C. M. Park, "Development and Validation of Deep Learning–based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs," *Radiology*, p. 180237, Sep. 2018.

[57]   Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," Nov. 2016.

[58]   K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. .

[59]   S. Hochreiter and J. J. Urgen Schmidhuber, "Long Short-Term Memory," 1997.

[60]   N. Bahrami, T. Retson, K. Blansit, K. Wang, and A. Hsiao, "Automated selection of myocardial inversion time with a convolutional neural network: Spatial temporal ensemble myocardium inversion network (STEMI-NET)," *Magn. Reson. Med.*, vol. 81, no. 5, pp. 3283–3291, May 2019.

[61]   Haibo He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[62]   B. P. F. Lelieveldt, R. J. Van Der Geest, H. J. Lamb, H. W. M. Kayser, J. H. C. Reiber, and H. J. C. R.) Re, "Automated Observer-independent Acquisition of Cardiac Short-Axis MR Images: A Pilot Study," *Radiology*, vol. 221, pp. 537–542, 2001.

[63]   J. C. Lopez-Mattei and D. J. Shah, "The role of cardiac magnetic resonance in valvular heart disease.," *Methodist Debakey Cardiovasc. J.*, vol. 9, no. 3, pp. 142–8, 2013.

[64]   T. N. Bloomer, S. Plein, A. Radjenovic, D. M. Higgins, T. R. Jones, J. P. Ridgway, and M. U. Sivananthan, "Cine MRI using steady state free precession in the radial long axis orientation is a fast accurate method for obtaining volumetric data of the

left ventricle," *J. Magn. Reson. Imaging*, vol. 14, no. 6, pp. 685–692, Dec. 2001.

[65]   K. Wang, A. Mamidipalli, A. Hsiao, and Claude Sirlin, "A Convolutional Neural Network-Based Automated Hepatic-Fat Quantification Method in Non-Alcoholic Fatty Liver Disease," in *Int. Soc. Magn. Reson. Med. Workshop Mach. Learn*, 2018.

[66]   M. Le, J. Lieman-Sifry, F. Lau, S. Sall, A. Hsiao, and D. Golden, "Computationally Efficient Cardiac Views Projection Using 3D Convolutional Neural Networks BT - Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support," 2017, pp. 109–116.

[67]   O. Addy, W. Jiang, W. Overall, J. Santos, and B. Hu, "Autonomous CMR: Prescription to Ejection Fraction in Less Than 3 Minutes," 2018.

[68]   C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *arXiv*, vol. 7. arXiv, p. 25, 09-Nov-2019.

[69]   B. Allen, S. E. Seltzer, C. P. Langlotz, K. P. Dreyer, R. M. Summers, N. Petrick, D. Marinac-Dabic, M. Cruz, T. K. Alkasab, R. J. Hanisch, W. J. Nilsen, J. Burleson, K. Lyman, and K. Kandarpa, "A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop," *J. Am. Coll. Radiol.*, vol. 16, no. 9, pp. 1179–1189, Sep. 2019.

[70]   K. Blansit, T. Retson, E. Masutani, N. Bahrami, and A. Hsiao, "Deep Learning–based Prescription of Cardiac MRI Planes," *Radiol. Artif. Intell.*, vol. 1, no. 6, p. e180069, Nov. 2019.

[71]   L. K. Tan, R. A. McLaughlin, E. Lim, Y. F. Abdul Aziz, and Y. M. Liew, "Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression," *J. Magn. Reson. Imaging*, vol. 48, no. 1, pp. 140–152, Jul. 2018.

[72]   R. A. Pooley, "Fundamental physics of MR imaging," *Radiographics*, vol. 25, no. 4, pp. 1087–1099, 2005.

[73]   E. Gibson, Y. Hu, N. Ghavami, H. U. Ahmed, C. Moore, M. Emberton, H. J. Huisman, and D. C. Barratt, "Inter-site Variability in Prostate Segmentation Accuracy Using Deep Learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11073 LNCS, pp. 506–514.

[74]   S. Tonekaboni, S. Joshi, M. D. Mccradden, A. Goldenberg, and A. G. Ai, "What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use," PMLR, Oct. 2019.

[75]   C. M. Kramer, J. Barkhausen, C. Bucciarelli-Ducci, S. D. Flamm, R. J. Kim, and E. Nagel, "Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update," *J. Cardiovasc. Magn. Reson.*, vol. 22, no. 1, p. 17, Feb. 2020.

[76]   R. M. Menchón-Lara, F. Simmross-Wattenberg, P. Casaseca-de-la-Higuera, M. Martín-Fernández, and C. Alberola-López, "Reconstruction techniques for cardiac cine MRI," *Insights into Imaging*, vol. 10, no. 1. Springer Verlag, pp. 1–16, 01-Dec-2019.

[77]   Y. Jo, J. Kim, C. H. Park, J. W. Lee, J. H. Hur, D. H. Yang, B. Y. Lee, D. J. Im, S. J. Hong, E. Y. Kim, E. A. Park, P. K. Kim, and H. S. Yong, "Guideline for cardiovascular magnetic resonance imaging from the korean society of cardiovascular imaging—part 1: Standardized protocol," *Korean J. Radiol.*, vol. 20, no. 9, pp. 1313–1333, Sep. 2019.

[78]   D. M. Vigneault, W. Xie, C. Y. Ho, D. A. Bluemke, and J. A. Noble, "Ω-Net (Omega-Net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks," *Med. Image Anal.*, vol. 48, pp. 95–106, Aug. 2018.

[79]   O. Auti, K. Bandekar, N. Kamat, and V. Raj, "Cardiac magnetic resonance techniques: Our experience on wide bore 3 tesla magnetic resonance system," *Indian J. Radiol. Imaging*, vol. 27, no. 4, p. 404, 2017.

[80]   P. Rajiah and M. A. Bolen, "Cardiovascular MR Imaging at 3 T: Opportunities, Challenges, and Solutions 1 Content Codes: 1 From the Cardiothoracic Imaging Section SA-CME LEARNING OBJECTIVES," *RadioGraphics*, vol. 34, pp. 1612–1635, 2014.

[81]   S. R. Saleh, G. D. Lohan, K. Nael, M. Grover-McKay, and P. Finn, "Article - Cardiovascular MRI at 3T," *Appl. Radiol.*, 2007.

[82]   M. A. Bernstein, J. Huston, and H. A. Ward, "Imaging artifacts at 3.0T," *J. Magn. Reson. Imaging*, vol. 24, no. 4, pp. 735–746, Oct. 2006.

[83]   D. D. Lewis and J. Catlett, "Heterogeneous Uncertainty Sampling for Supervised Learning," in *Machine Learning Proceedings 1994*, Elsevier, 1994, pp. 148–156.

[84]   R. J. Brachman, W. W. Cohen, T. G. Dietterich, and B. Settles, *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan Claypool Publishers.

[85]   D. Wang and Y. Shang, "A new active labeling method for deep learning," in *Proceedings of the International Joint Conference on Neural Networks*, 2014, pp. 112–119.

[86]   T. He, X. Jin, G. Ding, L. Yi, and C. Yan, "Towards better uncertainty sampling:

Active learning with multiple views for deep convolutional neural network," in *Proceedings - IEEE International Conference on Multimedia and Expo*, 2019, vol. 2019-July, pp. 1360–1365.

[87] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34–45, Apr. 2019.

[88] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation," *Front. Comput. Neurosci.*, vol. 13, p. 56, Aug. 2019.

[89] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[90] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015.

[91] L. Yang, Y. Zhang, J. Chen, S. Zhang, and D. Z. Chen, "Suggestive annotation: A deep active learning framework for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10435 LNCS, pp. 399–407.

[92] T. Retson, E. Masutani, D. Golden, and A. Hsiao, "Clinical Performance and Role of Expert Supervision of Deep Learning for Cardiac Ventricular Volumetry: A Validation Study," *Radiol. Artif. Intell.*, vol. 2, no. 4, p. e190064, Jul. 2020.

[93] N. Bahrami, T. Retson, K. Blansit, K. Wang, and A. Hsiao, "Automated Selection of Myocardial Inversion Time with a Convolutional Neural Network," in *North American Society for Cardiovascular Imaging*, 2018.

[94] T. Leiner, D. Rueckert, A. Suinesiaputra, B. Baeßler, R. Nezafat, I. Išgum, and A. A. Young, "Machine learning in cardiovascular magnetic resonance: Basic concepts and applications," *Journal of Cardiovascular Magnetic Resonance*, vol. 21, no. 1. BioMed Central Ltd., p. 61, 07-Oct-2019.

[95] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson, and B. Fischl, "Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer," *Neuroimage*, vol. 32, no. 1, pp. 180–194, Aug. 2006.

[96] G. Captur, C. Manisty, and J. C. Moon, "Cardiac MRI evaluation of myocardial disease," *Heart*, vol. 102, no. 18. BMJ Publishing Group, pp. 1429–1435, 01-Sep-2016.

[97] N. H. J. Prakken, A. J. Teske, M. J. Cramer, A. Mosterd, A. C. Bosker, W. P. Mali, P. A. Doevendans, and B. K. Velthuis, "Head-to-head com parison between echocardiography and cardiac MRI in the evaluation of the athlete's heart," *Br. J. Sports Med.*, vol. 46, no. 5, pp. 348–354, Apr. 2012.

[98] P. Pattanayak and D. A. Bleumke, "Tissue characterization of the myocardium: State of the art characterization by magnetic resonance and computed tomography imaging," *Radiologic Clinics of North America*, vol. 53, no. 2. W.B. Saunders, pp. 413–423, 01-Mar-2015.

[99] P. Sharma, S. Sharma, S. Ballal, C. Bal, A. Malhotra, and R. Kumar, "SPECT-CT in routine clinical practice: Increase in patient radiation dose compared with SPECT alone," *Nucl. Med. Commun.*, vol. 33, no. 9, pp. 926–932, Sep. 2012.

[100] A. M. Larkin, Y. Serulle, S. Wagner, M. E. Noz, and K. Friedman, "Quantifying the Increase in Radiation Exposure Associated with SPECT/CT Compared to SPECT Alone for Routine Nuclear Medicine Examinations," *Int. J. Mol. Imaging*, vol. 2011, pp. 1–5, Jul. 2011.

[101] A. A. Qayyum and J. Kastrup, "Measuring myocardial perfusion: The role of PET, MRI and CT," *Clinical Radiology*, vol. 70, no. 6. W.B. Saunders Ltd, pp. 576–584, 01-Jun-2015.

[102] C. for D. and R. Health, "Safety Communications - Safety Investigation of CT Brain Perfusion Scans: Initial Notification."

[103] H. Xue, P. Kellman, G. Larocca, A. E. Arai, and M. S. Hansen, "High spatial and temporal resolution retrospective cine cardiovascular magnetic resonance from shortened free breathing real-time acquisitions," *J. Cardiovasc. Magn. Reson.*, vol. 15, no. 1, p. 102, Nov. 2013.

[104] E. M. Masutani, N. Bahrami, and A. Hsiao, "Deep learning single-frame and multiframe super-resolution for cardiac MRI," *Radiology*, vol. 295, no. 3, pp. 552–561, Jun. 2020.

[105] J. J. Cavallo and H. P. Forman, "The economic impact of the COVID-19 pandemic on radiology practices," *Radiology*, vol. 296, no. 3, pp. E141–E144, Sep. 2020.

[106] X. V. Nguyen, S. Tahir, B. W. Bresnahan, J. B. Andre, E. V. Lang, M. Mossa-Basha, N. A. Mayr, and E. C. Bourekas, "Prevalence and Financial Impact of Claustrophobia, Anxiety, Patient Motion, and Other Patient Events in Magnetic Resonance Imaging," *Top. Magn. Reson. Imaging*, vol. 29, no. 3, pp. 125–130, Jun. 2020.

[107] I. Eshed, C. E. Althoff, B. Hamm, and K. G. A. Hermann, "Claustrophobia and premature termination of magnetic resonance imaging examinations," *J. Magn. Reson. Imaging*, vol. 26, no. 2, pp. 401–404, Aug. 2007.

[108] E. Törnqvist, A. Mnsson, E. M. Larsson, and I. Hallström, "Impact of extended written information on patient anxiety and image motion artifacts during magnetic resonance imaging," *Acta radiol.*, vol. 47, no. 5, pp. 474–480, Jun. 2006.

[109] R. Katznelson, G. N. Djaiani, L. Minkovich, L. Fedorko, J. Carroll, M. A. Borger, R. J. Cusimano, and J. Karski, "Prevalence of claustrophobia and magnetic resonance imaging after coronary artery bypass graft surgery," *Neuropsychiatr. Dis. Treat.*, vol. 4, no. 2, pp. 487–493, 2008.

[110] R. B. Parikh, S. Teeple, and A. S. Navathe, "Addressing Bias in Artificial Intelligence in Health Care," *JAMA - Journal of the American Medical Association*, vol. 322, no. 24. American Medical Association, pp. 2377–2378, 24-Dec-2019.

[111] M. Grissinger, "Understanding human over-reliance on technology," *P and T*, vol. 44, no. 6. Medi Media USA Inc, pp. 320–321, 01-Jun-2019.

[112] A. Holzinger, "Interactive machine learning for health informatics: when do we need the human-in-the-loop?," *Brain Informatics*, vol. 3, no. 2, pp. 119–131, Jun. 2016.

[113] M. Sujan, D. Furniss, R. Hawkins, and I. Habli, "Human Factors of Using Artificial Intelligence in Healthcare: Challenges That Stretch Across Industries," 2020.

[114] O. Asan, A. E. Bayrak, and A. Choudhury, "Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians," *Journal of Medical Internet Research*, vol. 22, no. 6. JMIR Publications Inc., 01-Jun-2020.

[115] S. Daftry, S. Zeng, J. A. Bagnell, and M. Hebert, "Introspective perception: Learning to predict failures in vision systems," in *IEEE International Conference on Intelligent Robots and Systems*, 2016, vol. 2016-November, pp. 1743–1750.
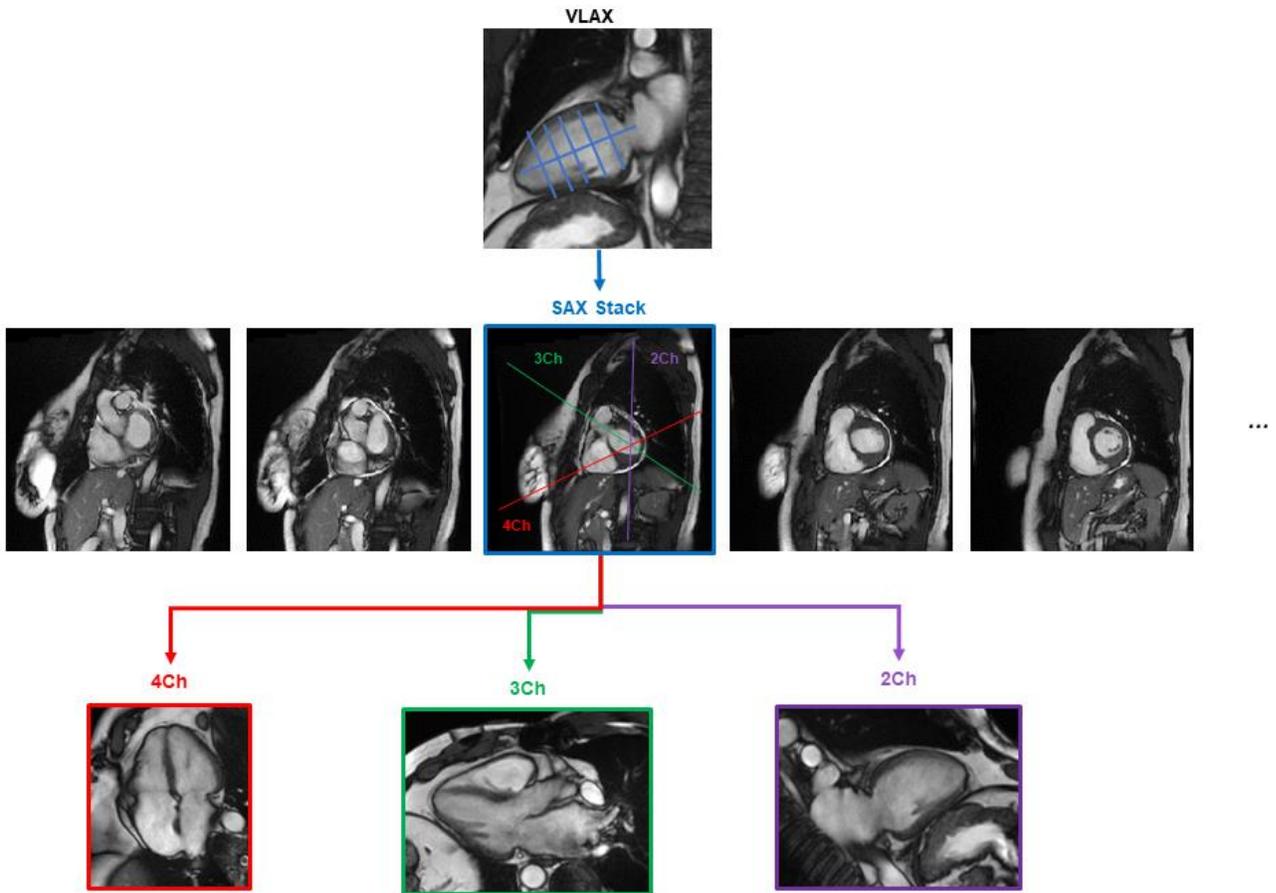
Supplemental Figure 1: *Typical sequence of imaging planes obtained for cardiac MRI at our institution*

Images are sequentially acquired. From an initial vertical long axis (VLAX), a short axis stack (SAX) is prescribed. From the SAX stack, the 4-chamber, 3-chamber, and 2-chamber long axis views are prescribed.
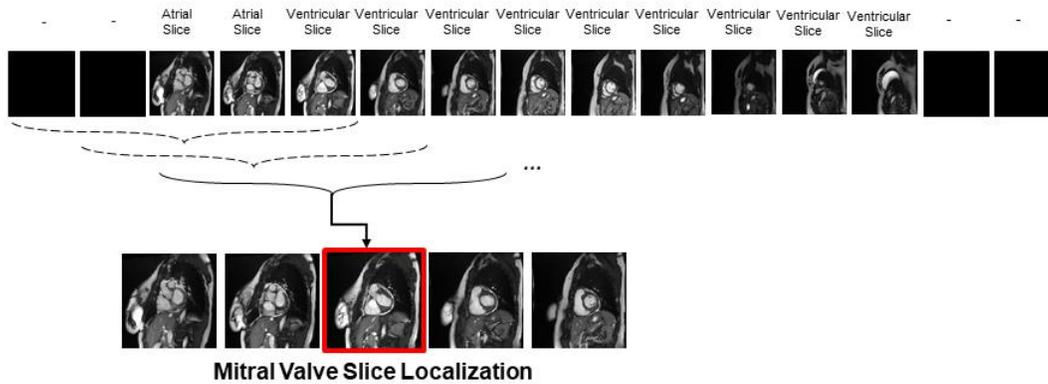
Supplemental Figure 2: *Sliding window localization of mitral valve slice*

Each slice in the short axis stack is classified as either "atrial" or "ventricular" using a VGG-19/LSTM ensemble classifier. The adjacent slices are used to provide context for the neural network. Zero-padding was added to the beginning and end of the stack of slices.
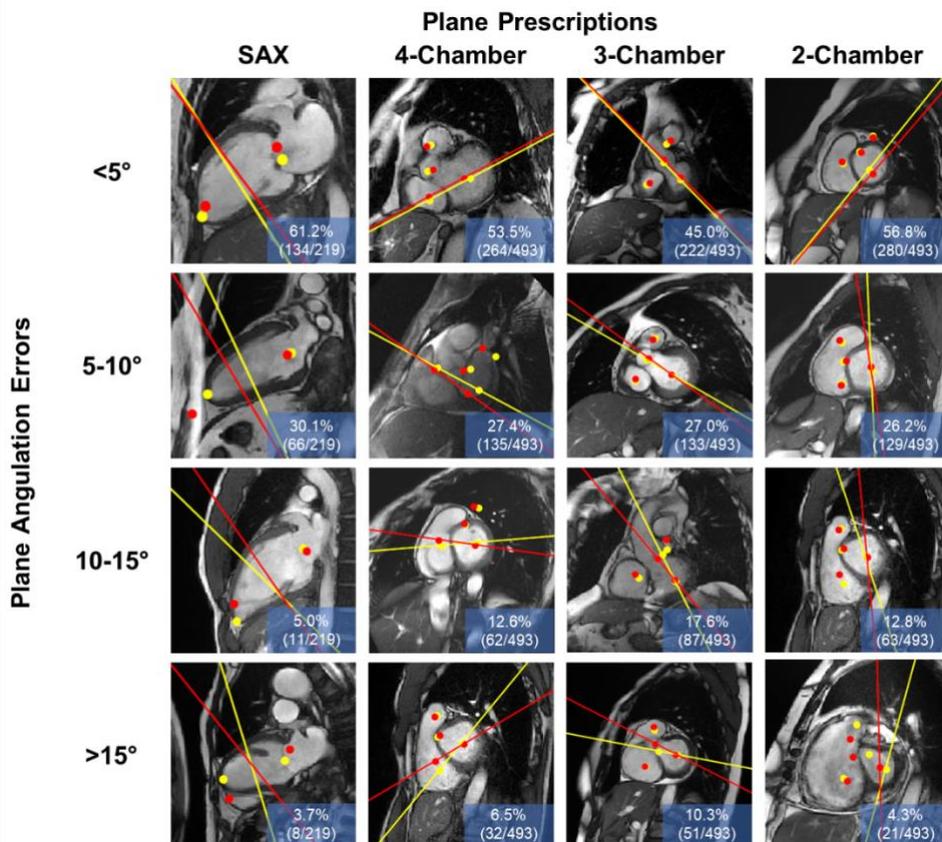
Supplemental Figure 3: *Example images of short-axis (SAX), 4-chamber, 3-chamber, and 2-chamber plane prescription angulation errors.*

Differences between GT planes and DL predicted planes are shown for the ranges of: <5°, 5-10°, 10-15°, and >15°. Frequency of each are shown in the bottom-right corner of each panel.

SUPPLEMENTAL TABLES

Supplemental Table 1: *Hyperparameters and details of each of the deep neural networks used.*

Attributes	LAX Localization	MV Slice Localization	Bounding Box	SAX Localization
Module Input	VLAX	SAX stack		
Module Output	MV, Apex Localization	AV, MV, PV, TV Localization		
Preprocessing	Down sampling to 128x128	Rolling windows using focal slice and 2 basal and 2 apical slices, rotated into standard orientation, field of view cropped to 280°	Rotated into standard orientation, field of view cropped to 280°	Rotated into standard orientation, cropped bounding box to 128x128 using predicted segmentation, use slices from predicted minimum window
Training Data	4 Ch, 3Ch, 2 Ch, HLAX, VLAX Planes	5 Slice Windows From SAX Stacks	Interpolated SAX Stacks	Bounding Box Cropped SAX Windows
Model Type	Heatmap Regression	Classification	Segmentation	Heatmap Regression
Base Model	2D U-net	Multi-channel VGG-19	2.5D U-net	2.5D U-net
Purpose	To localize MV and Apex	To identify the first window where the focal slice is ventricular	To create a heart bounding box	To in-plane localize AV, MV, PV, TV
Sigma	60	-	-	60 for 65 epochs, then 40 for 65 epochs
Epochs	70	35	40	130
Kernel size	32 for top two conv layers, then 9 for each additional conv layer	3	3	36 for top two conv layers, then 9 for each additional conv layer
Learning Rate	10e-4	10e-4	10e-4	10e-3

Supplemental Table 2: *Deep convolutional neural network module and training information.*

Model Name	CorSegNet	AxLocNet	LaxLocNet	SaxLocNet
Model Input	Coronal 3-Plane	Axial 3-Plane or Dedicated Axial	Vertical Long-Axis or 4-chamber	Short-Axis
Training Data Source	Sites 1 + 2	Sites 1 + 2	Site 1	Site 1
Pulse Sequence	SSFSE + SSFP	SSFP	SSFP	SSFP
Dimentionality	2D	3D	2D	3D
DCNN Type	Segmentation	Heatmap Regression	Heatmap Regression	Heatmap Regression
Number of Images Used For Training	1,390	4,647	24,320	11,109
Output	Lung and Mediastinum Mask	Mitral Valve and Apex	Mitral Valve and Apex	Mitral, Tricpuspid, and Aorotic Valves

Supplemental Table 3*: Evaluation Imaging Acquisition Parameters*

	3-Plane Localizer - SSFP	3-Plane Localizer - SSFSE	AX, LAX, SAX - SSFP
Echo Time (ms)	1.61 (range: 0.75 – 2.14)	58.15 (range: 0.79 – 83.14)	1.42 (range: 0.90 – 2.50)
Repition Time (ms)	3.86 (range: 2.48 – 4.88)	599.27 (range: 2.48 – 1398.66)	3.51 (range: 2.48 – 6.21)
Field of View (mm^2)	410.70 (range: 339.99 – 480.00)	458.44 (range: 380.01 – 480.00)	365.41 (range: 259.99 – 440.01)
Thickness (mm)	8.72 (range: 8.00 – 10.00)	7.78 (range: 5.00 – 10.00)	7.89 (range: 6.00 – 12.00)
Flip Angle (°)	0.00 (range: 0.00 – 0.00)	0.00 (range: 0.00 – 0.00)	40.80 (range: 0.00 – 101.86)