Automatic Plane Pose Estimation for Cardiac Left Ventricle Coverage Estimation via Deep Adversarial Regression Network

Le Zhang, Kevin Bronik, Stefan K. Piechnik, Joao A. C. Lima, Stefan Neubauer, Steffen E. Petersen and Alejandro F. Frangi, *Fellow, IEEE*

***Abstract*—Accurate segmentation of the ventricles plays a crucial role in determining cardiac functional parameters such as ventricular volume, ventricular mass, and ejection fraction. However, poor image quality, such as inadequate coverage of the left ventricle (LV) and right ventricle (RV) in cardiac magnetic resonance (CMR) image sequences, can significantly affect the assessment of cardiac function. This study investigates issues related to missing or corrupted imaging planes, which often lead to incomplete ventricle coverage. To address the challenge of estimating ventricle coverage in CMR images regardless of variations in imaging parameters such as device type, magnetic field strength, and protocol execution, we introduce a novel convolutional neural network (CNN) based on adversarial learning. Additionally, we integrate supplementary information (e.g., cross-view image data) as privileged information to enhance the interpretability of our model's predictions and identify potential biases or inaccuracies. This research represents the first attempt to automatically estimate ventricular coverage by identifying missing slices and plane orientations in CMR images using a dataset-agnostic approach. The effectiveness of the proposed model is demonstrated through the evaluation of datasets from three diverse and sizable image acquisition cohorts, demonstrating superior performance compared to existing methods.**

***Impact Statement*—Cardiac functional parameters, such as the ejection fraction of both ventricles and cardiac output, are crucial clinical indicators of cardiac function, providing insights into whether it is within normal or abnormal ranges. Accurate calculation of these parameters is based on precise measurements of ventricular volumes at the end of diastole and systole. The accuracy of volume measurements depends on correctly determining the heart's position, including its location, orientation, and size in CMR image sequences, thus establishing the full extent of the LV and RV. This study introduces a fully automatic detection method to identify missing slices and estimate heart pose parameters in CMR volumes, which is robust across different datasets. Unlike the previous research that focused solely on identifying missing base or apex slices to assess ventricular coverage, this study goes further by detecting incorrect slice plane orientations for a more precise estimation of ventricle coverage. This innovation has the potential to be integrated not only into postprocessing of images but also directly within the scanning protocol to promptly identify suboptimal acquisition. This advancement is poised to provide valuable guidance to radiographers, urging them to review faulty scans, streamlining image quality control, and facilitating the prompt correction of missing or misaligned image slices while the patient is still undergoing the scan.**

***Index Terms*—Adversarial learning (AL), deep learning (DL), privileged information (PI), regression network, ventricle pose estimation.**

I. Introduction

CARDIAC magnetic resonance imaging (CMR) is the standard reference imaging technique used to evaluate the morphology and functionality of the heart. After image acquisition, automatic techniques can extract volumetric information and derive clinical indices that place the subject within predetermined population ranges of normality. Acquisition of CMR images is the most automatic part, except for initial localization and enlargement of the heart performed by a trained radiologist or image technician [1]. Since the heart is a moving organ and the duration of the procedure requires the patient to hold his breath multiple times during the exam, the resulting images may suffer from artifacts due to variability in the position of breath hold adopted by the patient during each breath hold [2]. If the initial framing of the heart does not allow a sufficient margin around the organ, these differences in breath hold may cause the heart to move out of the frame, resulting in incomplete coverage in the basal or apical region of the organ [3], [4]. A related source of variability in organ coverage is determining what constitutes a sufficient margin around the heart. Although

Manuscript received 12 October 2023; revised 29 January 2024 and 2 April 2024; accepted 24 April 2024. Date of publication 8 May 2024; date of current version 10 September 2024. This article was recommended for publication by Associate Editor Monjoy Saha upon evaluation of the reviewers' comments. *(Corresponding author: Alejandro F. Frangi.)*

Le Zhang is with the School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, B15 2TT Birmingham, U.K.

Kevin Bronik is with the Department of Engineering Science, University of Oxford, OX1 3PJ Oxford, U.K.

Stefan K. Piechnik and Stefan Neubauer are with Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, OX3 9DU Oxford, U.K.

Joao A. C. Lima is with the Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD 21287 USA.

Steffen E. Petersen is with the Centre for Cardiovascular Medicine, The William Harvey Research Institute, Queen Mary University of London, E1 1BB London, U.K., and also with the Barts Heart Center, Barts Health NHS Trust, E1 1BB London, U.K.

Alejandro F. Frangi is with the Division of Informatics, Imaging and Data Science, School of Health Sciences and the Department of Computer Science, School of Engineering, The University of Manchester, M13 9PL Manchester, U.K. (e-mail: alejandro.frangi@manchester.ac.uk).

Digital Object Identifier 10.1109/TAI.2024.3394798

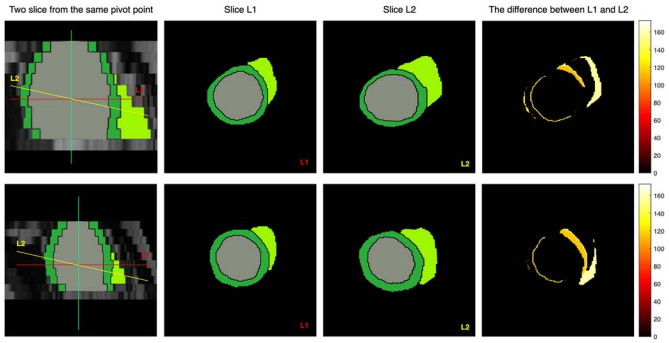

Fig. 1. Potential issues affecting CMR image acquisitions. Short-axis slices (SAXs) were acquired with incorrect orientations in *slice L2* compared to *slice L1* in the top and bottom volumes. The last column is the difference between the two planes. Best viewed in color.

anatomical characteristics allow for the precise location of the base and apex of the heart, "a sufficient margin above and below the base and apex" may not be precisely defined. This means that slightly different practices may be in place at different imaging facilities or by different experts, resulting in image volumes that, while providing full coverage, may present with one image slice above/below the cardiac base/apex. On the contrary, others with two image slices above/below the cardiac base/apex [5]. These variations in the procedure may present problems for subsequent image-analysis algorithms trained under the assumption of consistent object coverage.

Another parameter related to object coverage is the consistent orientation of slice planes regarding the cardiac ventricles [6] (see the top and bottom volumes in Fig. 1). If the slice orientation deviates significantly from expected values, the local image structure may change enough to cause subsequent algorithms based on image features to fail in localizing critical features required for further morphological and functional analysis. These sources of variability can affect the subsequent application of automatic methods for the computation of tissue volumes, cardiovascular indices, and statistics derived from them [7], [8]. Typically, volume computations are performed on the output of image segmentation algorithms. Whether these algorithms are generative or discriminative, having incomplete or incorrect organ coverage can cause inaccurate segmentation, leading to biased estimation of volume parameters [9]. For example, 2-D CNN-based segmentation methods will grossly under/overestimate blood volumes with incomplete/incorrect image stacks. In Fig. 1, the L2 slice with an incorrect plane orientation shows a larger left ventricle (LV) volume than the L1 slice in standard orientation. Also, 3-D generative-based models, such as active shape models (ASMs), can handle incomplete/incorrect volumes as shape priors constrain them [10]. However, if the initial shape estimate is poorly positioned regarding the target structure, the performance of ASMs can be affected by incorrect model initialization and thus may cause the algorithm not to converge to the best solution.

To address the above CMR ventricle coverage problems, several automated and learning-based image quality assessment (QA) algorithms have been proposed for base/apex slice detection [3], [4], [11]. However, some works focused on automatic view planning for CMR image acquisition [6], [8], [12], which involves manual annotation of anatomical landmarks, a tedious and time-consuming task. Therefore, algorithms that can automatically detect the base/apex slices and determine the pose parameters of each slice in CMR images are needed to intervene and correct problems in the datasets. Such corrections may include imputing missing data, choosing the appropriate segmentation method, and removing faulty image volumes from their use to compute aggregated statistics in large cohorts of patients [13]. This paves the way for "quality-aware" image analysis [14].

In this article, we adopt the adversarial learning (AL) strategy to train the convolutional neural network (CNN) and innovatively refer to the cross-view image (long-axis slice in this work) as the privileged information (PI) in the training phase, to detect the base/apex slice and estimate the slice pose parameters across multiple CMR imaging cohorts. Our model is the first work for robust CMR image QA that estimates the ventricle coverage-related parameters, e.g., base/apex slice location and orientation of slice plane. The main contributions of this work are summarized as follows.

1) This article first uses CNNs for automatic image QA by estimating the parameters related to ventricle coverage in multiple CMR imaging cohorts.
2) To solve the domain shift problem, we present a dataset invariance adversarial learning (DIAL) to process the data from any cohort, which allows transferring the task knowledge from the annotated source dataset to the unannotated target dataset.
3) We propose an end-to-end multilabel multitask (MLMT) regression network to jointly optimize ventricle coverage-related tasks (slice distance and plane orientation parameter estimation). Our proposed MLMT model has great generalizability and works well on different tasks.
4) We formulate a new problem, which combines DIAL and MLMT models with novel PI loss, namely DARN*. To the best of our knowledge, this is the first work that exploits the multiview information as PI in CMR images, and it is a much more practical approach in real-world applications.
5) We comprehensively evaluated the performance of our model in three large-scale 3-D CMR image datasets and achieved promising results compared to the related baseline methods.

II. Related Work

A. Base/Apex CMR Slice Detection

There are several studies published on automating LV base/apex detection. Paknezhad et al. [11] proposed an automatic tool that uses the horizontal long-axis (HLA) view to find the basal slice. The temporal binary profiles, created from the segmented HLA slice for each SAX, were used to detect the base slice. The drawbacks of this technique are its dependence on correct segmentation and the existence of the HLA slice. Mahapatra [15] proposed a learning-based method to detect

the basal slice. Intensity, texture, and contextual characteristics were extracted by training a random forest classifier from a bounding box around the annotated points on both sides of the mitral valve (MV). The use of such methods comes from the assumption that the basal slice is the first SAX slice below the line that connects the MV points. However, this is not always true across different imaging cohorts. According to the guidelines for cardiac image analysis [16] published by the Society for Cardiovascular Magnetic Resonance (SCMR) that describes precisely how to select the basal slice, the basal slice is the top view slice on the short axis that has more than 50% myocardium around the blood cavity [11].

Data-driven approaches using CNNs have significantly improved image classification tasks for medical image analysis, such as identifying the presence or absence of a disease, identifying the stage of a disease, and identifying different anatomical structures in the image. With this inspiration, Zhang et al. [17] determined the LV coverage from CMR images by using Fisher discriminative 3-D CNNs (FD3D). This approach utilizes spatial contextual information in CMR volumes and extracts more representative high-level features, thus achieving superior detection accuracy. Furthermore, Xia et al. [3] proposed a novel deep conditional GAN architecture to generate missing SAX slices for CMR images at different positions. A regression network learns features relevant to identifying the position of the missing CMR slice and predicts its corresponding position. This method can be easily extended to particular slice scenarios, such as detecting missing slices at the apical or basal positions. Duan et al. [18] and Joyce et al. [19] introduced anatomical shape prior knowledge to the network segmentation, which is a refinement step that is carried out using atlas propagation. The LV coverage issues may also be addressed by these altas-to-subject registration approaches. However, these methods involve segmenting the ventricles and detecting landmarks as an intermediate step before estimating ventricle coverage.

B. Standard CMR Slice Plane Detection

In recent years, significant research in automated plane detection using machine learning algorithms has led to the development of several standard approaches. Blansit et al. [20] proposed to develop and evaluate a system to prescribe imaging planes for CMR slices based on deep learning (DL)-based localization of key anatomic landmarks. Wei et al. [8] proposed a clinic-compatible automatic view planning system for CMR. This work is inspired by keypoint-based object detection [21], [22] and then proposes regressing the intersecting lines between the views, which can be easily computed using the spatial relationship. The authors [6], [12] developed an automatic view planning algorithm for CMR image acquisition. They use techniques based on learning anatomy segmentation and the detection of anchor points to calculate typical cardiac views. These techniques require prior knowledge of the entire 3-D image to recognize planes. This involves the complex and time-consuming task of manually annotating anatomical landmarks. Therefore, using the acquired images for training and avoiding manual labeling is necessary to develop an automatic standard plane detection method. Furthermore, Alansary et al. [23] proposed a fully automatic method to find standardized view planes in 3-D image acquisitions. The application of this method to cardiac MRI data shows target plane detection in real-time with an accuracy of around 5 mm. Although these *base/apex slice detection* and *standard slice plane detection* methods can help to achieve whole-heart coverage in CMR imaging. However, these methods work separately on the specific task, while both factors usually cause incomplete ventricle coverage. Combining base/apex slice detection and standard plane detection makes it desirable and possible to achieve complete ventricle coverage in CMR imaging.

C. AL

It has emerged as a prominent approach in machine learning, particularly in the context of deep neural networks, to enhance model robustness and generalization. This technique draws inspiration from game theory, introducing an adversarial component that involves training the model against a set of deliberately crafted adversarial examples [24]. Alansary et al. [25] seminally introduced the concept of adversarial training and demonstrated the vulnerability of deep neural networks to adversarial examples. The authors proposed adversarial training to enhance model robustness. Madry et al. [26] focused on creating robust models through adversarial training by formulating a minimax optimization problem. The authors developed adversarial training techniques that showed improved resistance against various types of adversarial attacks.

III. Methodology

A. Problem Formulation

We formulate our problem as two tasks as follows.

1) Dataset Invariance Learning: Given a set of 3-D images $\mathcal{X}^s = [\mathbf{X}_1^s, ..., \mathbf{X}_N^s] \in \mathbb{R}^{m \times n \times z^s \times N^s}$ and corresponding labels $\mathcal{Y}^s = [\mathbf{Y}_1^s, ..., \mathbf{Y}_N^s]$ of the modality $\mathcal{M}_s$ in the source dataset and $\mathcal{X}^t = [\mathbf{X}_1^t, ..., \mathbf{X}_N^t] \in \mathbb{R}^{m \times n \times z^t \times N^t}$ of the modality $\mathcal{M}_t$ in the target dataset; m and n are the dimensions of the axial view of the image, and z^s and z^t denote the size of the image along the z-axis, while N^s and N^t are the numbers of elements in the source and target datasets, respectively. Let $\{\mathbf{X}^s, \mathbf{Y}\} = \{\mathbf{x}_i^s, \mathbf{y}_i^s\}_{i=1}^{Z^s}$ and $\{\mathbf{X}^{*s}, \mathbf{Y}\} = \{\mathbf{x}_i^{*s}, \mathbf{y}_i^s\}_{i=1}^{Z^s}$ be a labeled 3-D CMR volume of the source modality $\mathcal{M}_s$ on the short and long axes, respectively; let $\mathbf{X}^t = \{\mathbf{x}_i^t\}_{i=1}^{Z^t}$ represents an unlabeled sample from the target dataset on the short axis, and Z be the total of CMR slices. We aim to build a symmetric mapping between the source and target datasets in the shared feature space, reducing the difference between the distributions of the source and target dataset.

2) Slice Pose Estimation: In this task, slice pose estimation is enhanced by using MLMT learning, for example, distance regression task and orientation regression task, in a single deep regression neural network. In MLMT learning, assume that there is an image sequence denoted by $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{Z}$, where each $\mathbf{X}$ has Z slices and $\mathbf{y}$ contains the labels (distance and orientation) associated with $\mathbf{x}$. We represent $\mathbf{y}_i$ as a vector

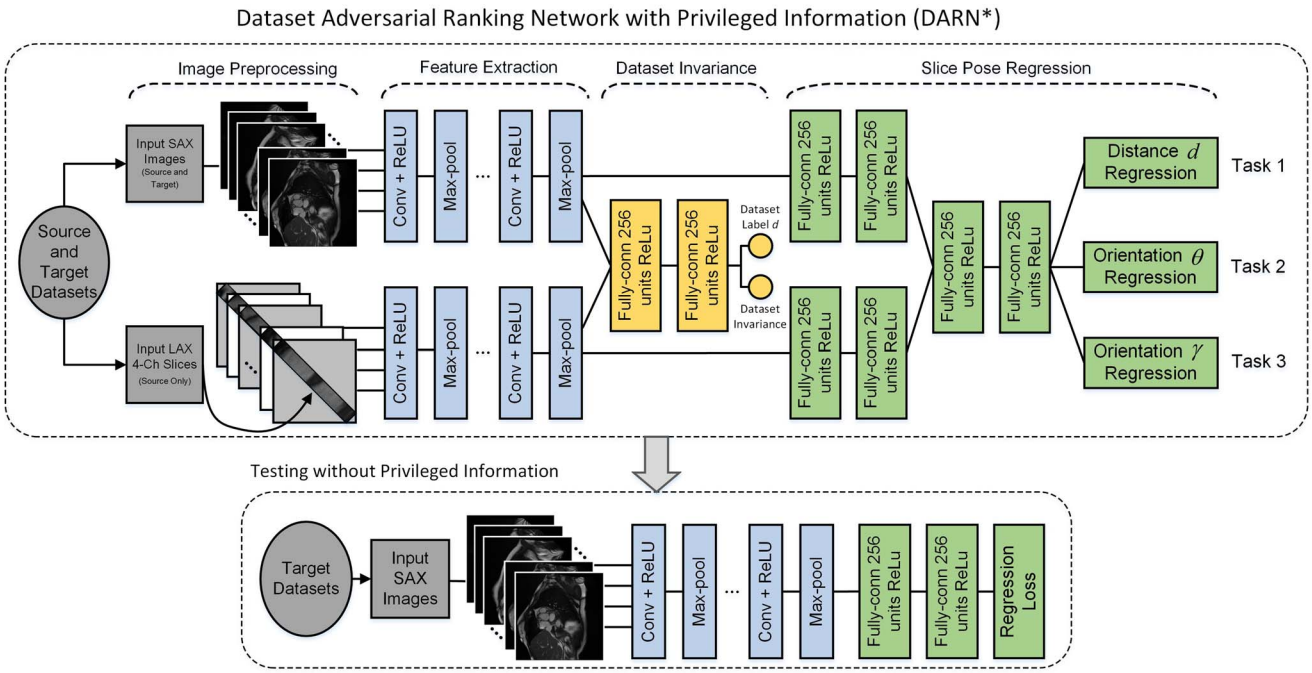

Fig. 2. Proposed framework of PI-based DARN. Our approach consists of three steps: 1) the CNN acts as a feature extractor to extract the spatial pattern of the cardiac image volume to facilitate the dataset invariance (DI) phase; 2) we use a DIAL model to fit the joint distribution over the images from different datasets with a minimax game; and 3) we extend the DIAL model to handle the MTRN model with learning using privileged information (LUPI) scenarios. The joint network can be trained to learn the complex spatial patterns of cardiac sequences in different cardiac MRI datasets and predict the pose without PI during the test. Best viewed in color.

of length C, where C is the number of labels. For example, the kth dimension $\mathbf{y}_i(k)$ denotes the distance between the basal slice and the ith slice. Our goal is to train a regression network processed with the MLMT procedure, mapping images to corresponding probabilities by the function $\eta(\cdot) : \mathcal{X} \to \mathcal{Y}$. We incorporate the long-axis patches $\mathbf{X}^{*s}$ as PI into the learning system at training time, and the testing stage continues to make use of only $\mathbf{X}^t$ without access to $\mathbf{X}^{*t}$. This training process produces a ConvNet with the parameters learnt $\mathbf{W}$ that is effectively a mapping between the input images $\mathcal{X}$, $\mathcal{X}^*$ and the estimated output vector $\mathcal{Y}$, represented by

$$\mathcal{Y} = \eta(\mathcal{X}, \mathcal{X}^*; \mathbf{W}). \tag{1}$$

We use the U.K. Biobank (UKBB) [27] cardiac MRI data cohort annotated with the multiethnic study of atherosclerosis (MESA) [28] and defibrillators to reduce risk using magnetic resonance imaging evaluation (DETERMINE) [29] datasets, and apply our method to cross-dataset slice position and orientation regression tasks. We describe deep AL approaches and configure an MLMT regression network incorporated with PI to perform automated slice pose estimation across datasets. During training, DIAL promotes the emergence of indiscriminate features with respect to the change between datasets but discriminative for the main learning task in the source dataset. Instead of using metric regression to identify the pose of the slice, the MLMT is trained to regress the distance and orientation of each slice simultaneously with the aim of mutual benefit. Our goal is to learn the discriminative features of $\mathbf{x}_i^s$ and utilize $\mathbf{x}_i^{*s}$ as PI to train more effective models and estimate the slice poses in the short axis for CMR volumes in the target dataset.

B. Dataset-Invariant Deep AL

In this section, we propose a DIAL model, which incorporates AL with dataset adaptation (DA) and implements them in a unified framework. Furthermore, we extend the DIAL model to a multiview learning strategy by creating multiple input channels (MC) from the image, which are resampled to the same spatial grid and visualize the same anatomical structure. Fig. 2 shows the framework of our method. Given a set of slices $\{\mathbf{x}_i^s\}_{i=1}^Z$ with corresponding labels $\{\mathbf{y}_i^s\}_{i=1}^Z$ for training, we optimize our objective function in several stages to achieve a model that can generalize well from one dataset to another, while maintaining the regression ability for estimating the pose of the slice. In the first stage, we optimize the label regression loss

$$\arg\min_{\mathbf{w}_f, \mathbf{w}_y^t} \left\{ \frac{1}{N^s} \sum_{i=1}^{N^s} \mathcal{L}_y^i(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i^s; \mathbf{w}_f); \mathbf{w}_y^t), y_i^t) \right\} \tag{2}$$

where $\mathbf{w}_f$ is the representation parameter in the feature extraction layers (blue in Fig. 2). $\mathbf{w}_y^t$ is the parameter in the regression layers (green in Fig. 2). y_i^t denotes the ith slice position label. $\mathbf{w}_f$ and $\mathbf{w}_y^t$ are trained for the image ith using labeled source data $\{\mathbf{x}_i^s, \mathbf{y}_i^s\}_{i=1}^{Z^s}$.

To achieve the dataset adaptation and minimize source and target representation distances, we choose the AL mechanism

by alternating *minimax* between two loss functions: one is the dataset discriminator loss

$$\arg\min_{\mathbf{w}_d}\left\{-\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\mathcal{L}_d^i(G_{\text{soft}}(G_{\text{conv}}(\mathbf{x}_i^s,\mathbf{x}_i^t;\mathbf{w}_f);\mathbf{w}_d),d_i)\right\} \tag{3}$$

which aims to discriminate whether an image is drawn from the source or target dataset. $\mathbf{w}_d$ is the parameter in the DI network and is used to predict the output of the dataset classification. d_i denotes the dataset label for the example slice ith. The other is a mapping invariant loss for the source and target dataset

$$\arg\min_{\mathbf{w}_f}\left\{-\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\mathcal{L}_f^i(G_{\text{soft}}(G_{\text{conv}}(\mathbf{x}_i^s,\mathbf{x}_i^t;\mathbf{w}_d);\mathbf{w}_f),d_i)\right\} \tag{4}$$

where the cross entropy is calculated between the predicted dataset labels and the uniform distribution over the dataset. The mapping invariant loss is optimized with a constrained adversarial objective with $\mathcal{N}=N^s+N^t$ being the total of samples. Then, our full method optimizes the joint loss function

$$\begin{aligned}E(\mathbf{w}_f,\mathbf{w}_d,\mathbf{w}_y^t)=&\mathcal{L}_y(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}^s;\mathbf{w}_f);\mathbf{w}_y^t),y^t)\\&+\lambda\mathcal{L}_f(G_{\text{soft}}(G_{\text{conv}}(\mathbf{x}^s,\mathbf{x}^t;\mathbf{w}_f);\mathbf{w}_d),d)\end{aligned} \tag{5}$$

where λ is the hyperparameter that determines how strongly the DI influences optimization; $G_{\text{conv}}(\cdot)$ is a convolution layer function (feature extraction) that maps an example to a new representation; $G_{\text{sigm}}(\cdot)$ is a label prediction layer function (sigmoid); and $G_{\text{soft}}(\cdot)$ is a dataset prediction layer function (softmax).

We chose the stochastic gradient procedure to optimize the DIAL model, similar to the classical CNN learning methods. The gradient of (5) is updated in the opposite direction to minimize the loss for parameters $\mathbf{w}_d,\mathbf{w}_y$ and in the direction along the gradient to maximize the loss for parameters $\mathbf{w}_f$ [30].

C. MLMT Learning With PI

In this section, we provide the learning procedure of our MLMT* learning network for estimating the position and orientation of the slice, instead of using the metric regression network. Then, we show that our framework can be trained end-to-end by optimizing the regression and spatial structured constraints.

1) Multilabel Multitask Learning (MLMT): To fully capture the spatial information relevant to the LV in every slice from one cardiac volume, we employ CNN as a feature extractor to efficiently encode spatial information. We choose DIAL to extract the features of each slice across datasets and predict the corresponding position and orientation using multiple regression tasks.

Traditional multitask learning (MTL) seeks to improve the generalization of multiple related tasks by learning them jointly. Suppose that we have T tasks and the training data for the tth task are denoted as $(\mathbf{x}_i^t,y_i^t)$, where $t=\{1,...,T\}$, with $\mathbf{x}_i,y_i\in\{\{\mathbf{x}_i^n,y_i^n\}_{i=1}^{Z}\}_{n=1}^{N^s}$ being the input image and the label, respectively. MTL tries to minimize the following:

$$\arg\min_{\{\mathbf{w}^t\}_{t=1}^T}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathcal{L}(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i;\mathbf{w}_f^t);\mathbf{w}_y^t),y_i^t)+\Phi(\mathbf{w}^t) \tag{6}$$

where $\mathbf{w}^t=\{\mathbf{w}_f^t,\mathbf{w}_y^t\}$ is the weight vector for tth task and y_i^t is the label for ith image for the tth task. The loss function is denoted by $\mathcal{L}(\cdot)$. A typical choice is the mean square loss for regression and the cross-entropy loss for classification. $\Phi(\mathbf{w}^t)$ is the regularization term that penalizes weight complexity.

In this article, we divided all the tasks into two groups: regression tasks t_d, for distance parameter d, and regression tasks t_o, for orientation parameters θ and γ (the definition will be discussed in experimental section). Inspired by the work of Niu et al. [31], we adopt a *data specific scheme* for each task in the two groups, which obtains the distribution of the sample number over their distance and orientation. We set the important parameters according to this distribution. For our MLMT with multiple outputs, each corresponding to the regression task for the ith image. Let α_y denotes the important coefficient of the label y ($y\in\{y_d,y_o\}$) in regression tasks. In our approach, the important parameters are set according to the reliability of different regression parameters. We set $\alpha_y^{t_d}=\sqrt{N_d}/(\sum_{d=1}^{D}\sqrt{N_d})$ for the distance regression task and $\alpha_y^{t_o}=\sqrt{N_o}/(\sum_{o=1}^{O}\sqrt{N_o})$ for the orientation regression task, where N_d is the number of samples with distance label d, and N_o is the number of samples with orientation label o. t_d corresponds to a distance regression task, which is trained to regress the slice distance d in a sequence. Therefore, for task t_d, the number of samples with distances nearby d, for example, samples with distance $\{(d-\delta d),d,(d+\delta d)\}$, ($\delta d\in\{0,1/Z\}$) is more important than other samples for the training of the output of task t_d. If more samples have a distance close to d, we could better train the corresponding position features, and hence, it is better to give it relatively greater importance. The loss function of our regression network can be formulated as

$$\begin{aligned}&\arg\min_{\{\mathbf{w}^t\}_{t=1}^T}\sum_{t=1}^{T}\sum_{i=1}^{N}\alpha_y\beta_t\mathcal{L}(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i;\mathbf{w}_f^t);\mathbf{w}_y^t),y_i^t)=\\&\arg\min_{\{\mathbf{w}^{t_d}\}_{t_d=1}^{T_d},}\sum_{t_d=1}^{T_d}\sum_{i=1}^{N}\alpha_y^{t_d}\beta_{t_d}\mathcal{L}(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i;\mathbf{w}_f^{t_d});\mathbf{w}_y^{t_d}),y_i^{t_d})+\\&\arg\min_{\{\mathbf{w}^{t_o}\}_{t_o=1}^{T_o}}\sum_{t_o=1}^{T_o}\sum_{i=1}^{N}\alpha_y^{t_o}\beta_{t_o}\mathcal{L}(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i;\mathbf{w}_f^{t_o});\mathbf{w}_y^{t_o}),y_i^{t_o})\end{aligned} \tag{7}$$

where T_d and T_o indicate the total of regression tasks for distance and orientation, respectively, and $T=T_d+T_o$; β_t denotes the important coefficient of the tth task's error. The regularization terms are omitted here for simplification.

2) MLMT With PI (MLMT):* In many image processing tasks, additional information can often help us to learn a better model in the training stage. In an image recognition system, we call this information PI, such as image captions. In other

Algorithm 1: DIAL Algorithm.

Input: source data $\{\mathbf{x}_i^s\}_{i=1}^{Z^s}$, $\{\mathbf{x}_i^{*s}\}_{i=1}^{Z^s}$;
target data $\{\mathbf{x}_i^t\}_{i=1}^{Z^t}$, $\{\mathbf{x}_i^{*t}\}_{i=1}^{Z^t}$;
ground-truth dataset label y_i^d.
Initialize: $\mathbf{w}_f, \mathbf{w}_d, \mathbf{w}_y^d \leftarrow$ random_init; $d \leftarrow 0$;
while *stopping criterion has not been met* **do**
  for *i from 1 to* N^s **do**
    1) Calculate $\mathbf{w}_f$, $\mathbf{w}_y^d$ using Eq. (2);
    2) Calculate $\mathbf{w}_d$ using Eq. (3) with fixed $\mathbf{w}_f$;
    3) Calculate $\mathbf{w}_f$ using Eq. (4) with fixed $\mathbf{w}_d$;
    4) Update parameters using gradient descent [30].
  end
end
Output: Neural network parameters $\{\mathbf{w}_f, \mathbf{w}_d, \mathbf{w}_y^d\}$;

Algorithm 2: MLMT* Algorithm.

Input: training data $\{\mathbf{x}_i^s\}_{i=1}^{Z^s}$, $\{\mathbf{x}_i^{*s}\}_{i=1}^{Z^s}$;
testing data $\{\mathbf{x}_i^t\}_{i=1}^{Z^t}$;
ground-truth distance label $y_i^{t_d}$;
ground-truth orientation label $y_i^{t_\theta}$ and $y_i^{t_\gamma}$.
Initialize: $\mathbf{w}_f, \mathbf{w}_y^t \leftarrow$ random_init;
while *stopping criterion has not been met* **do**
  for *i from 1 to* $\mathcal{N}$ **do**
    1) Calculate $\mathbf{w}_f$ and $\mathbf{w}_y^t$ using Eq. (7);
    2) Calculate $\mathbf{w}^t$ and $\mathbf{w}_f^{*,t}$ with PI using Eq. (8);
    3) Update parameters using gradient descent [30].
  end
end
Output: Neural network parameters and predicted regressor for test images $\{\eta(\mathbf{x}_i; \mathbf{w}_f, \mathbf{w}_y^t)\}_{i=1}^{Z}$;

words, the PI provides additional correct information during training, but in the test stage, the model operates without the supervision of the PI. This paradigm is called LUPI and was introduced by Vapnik and Vashist [32]. Our model constructs a two-stream framework: we train the first stream model for SAX images, and the second stream model with the PI (LAX patches). Our framework could also be helpful if, instead of LAX images, we utilize other PI like other image metadata (e.g., patient demographics and scanner type).

We need a PI loss to replace the original MLMT loss in the training phase to use the PI as "teacher" to train a more effective model. We propose to use PI to model the loss of training data, penalize the difference of PI modeled loss and true loss, and add the difference as a regularization term to (7). In particular, we assume that we have a privileged LAX patch for each SAX training image $\mathbf{x}_i$, we have a privileged LAX patch $\mathbf{x}_i^*$. We used the second network stream (MLMT-PI) to model PI. Compared to the first network stream, which models the training SAX images, this second stream aims not to learn a regression model but to model the loss of the first stream. Denote the output of the second stream for an input privileged patch $\mathbf{x}_i^*$ as $f^*(\mathbf{x}_i^*)$. The two streams share the same loss layer defined by

$$\begin{aligned}&\underset{\{\mathbf{w}^t, \mathbf{w}_f^{*,t}\}_{t=1}^T}{\arg\min} \sum_{t=1}^{T}\sum_{i=1}^{N} \alpha_y\beta_t\mathcal{L}(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i; \mathbf{w}_f^t); \mathbf{w}_y^t), y_i^t) + \\ &\gamma\left\|\alpha_y\beta_t\mathcal{L}(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_i; \mathbf{w}_f^t); \mathbf{w}_y^t), y_i^t) - G_{\text{conv}}^*(\mathbf{x}_i^*; \mathbf{w}_f^{*,t})\right\|_2^2 \end{aligned} \tag{8}$$

where $\mathbf{x}_i^{*,t}$ is the ith privileged patch and parameterized by the weight vector $\mathbf{w}_f^{*,t}$, $\|\cdot\|_2^2$ is the L2 norm. Our main hyperparameter is the tradeoff parameter γ, tuned by cross validation in a small subset of the training data.

The proposed MLMT with PI can be optimized in an alternating fashion. Specifically, we update the mainstream while fixing the parameters of the privileged stream until convergence and then update the privileged stream while fixing the parameters of the mainstream. This process is repeated several times until the whole system converges.

We formulate our complete DARN* model based on (5) and (8) in the following way. Suppose we have a set of feature vectors in a shared feature space across tasks $\{\mathbf{x}_i\}_{i=1}^Z$ and their corresponding labels $\{y_i^d, y_i^{t_d}, y_i^{t_\theta}, y_i^{t_\gamma}\}_{i=1}^Z$, where y_i^d is the target of DI and the remaining are the targets of slice pose regression, including inferences of "distance" and "orientation." In particular, $y_i^d \in \{0, 1\}$ is a binary dataset, $y_i^{t_d}$, $y_i^{t_\theta}$, and $y_i^{t_\gamma}$ are multiple values that represent the distances and orientations in the 3-D space. We employ least squares and cross entropy as loss functions for the main task (regression) and dataset-invariant learning task (classification). Therefore, the objective function can be rewritten as

$$\begin{aligned}&\underset{\mathbf{w}^d, \{\mathbf{w}^t, \mathbf{w}^{*,t}\}_{t=1}^T}{\arg\min} \sum_{t=1}^{T}\sum_{i=1}^{N} \frac{1}{2}\alpha_y\beta_t\left\|y_i^t - f(\mathbf{x}_i; \mathbf{w}^t)\right\|_2^2 \\ &+ \gamma\left\|\frac{1}{2}\alpha_y\beta_t\left\|y_i^t - f(\mathbf{x}_i; \mathbf{w}^t)\right\|^2 - f(\mathbf{x}_i^*; \mathbf{w}^{*,t})\right\|_2^2 \\ &- \sum_{i=1}^{N}\lambda y_i^d \log(p(y_i^d \,|\mathbf{x}_i; \mathbf{w}^d)) + \sum_{t=1}^{T}\left(\left\|\mathbf{w}^d\right\|_2^2 + \left\|\mathbf{w}^t\right\|_2^2\right) \end{aligned} \tag{9}$$

where $f(\mathbf{x}_i; \mathbf{w}^t) = (\mathbf{w}^t)^\top \mathbf{x}_i$ is a linear function. $p(y_i^d = m \,|\mathbf{x}_i) = \exp\{(\mathbf{w}_m^d)^\top \mathbf{x}_i\} \big/ \sum_j \exp\{(\mathbf{w}_j^d)^\top \mathbf{x}_i\}$ is a softmax function, which models the class posterior probability ($\mathbf{w}_j^d$ denotes the jth column of the matrix). Since the unique structure of CNN allows for multitasking and shared representation, in this study, we adopt the CNN to jointly learn the shared feature space $\mathbf{x}$.

D. Model Implementation

1) Network Structure: As shown in Fig. 2, our network consists of three parts: feature extraction, DIAL, and pose regression (MLMT*). Feature extraction network is adopted from [30] including three 5×5 convolutional layers (C1, C2, and C3), each followed by a 2×2 max-pooling layers (P1, P2, and P3) with stride 2. Followed by P3, there are two branches: one

TABLE I
NOTATIONS WE USED IN OUR MODEL AND THEIR REPRESENTATIONS

Notations	Representations
$\mathcal{X}$	3-D image sets
$\mathbf{X}$	3-D SAX images
$\mathcal{Y}$	Label sets
$\mathbf{x}$	2-D slice
$\mathbf{y}$	Label vectors (e.g., distance and orientation angles)
Z	Number of slices in a 3-D CMR volume
C	Number of labels in $\mathbf{y}$
k	kth value in $\mathbf{y}$
i	ith slice in $\mathbf{X}$
$\mathbf{X}*$	3-D HLA images
N^s	Number of source images
N^t	Number of target images
$\mathcal{N}$	Number of total images
$\mathbf{w}_f$	Parameter in the feature extraction layers
$\mathbf{w}_y$	Parameter in the regression layers
$\mathbf{w}_d$	Parameter for the dataset classification output
λ	Hyperparameter for optimization
$G_{\text{conv}}(\cdot)$	Convolution layer function (feature extraction)
$G_{\text{sigm}}(\cdot)$	Label prediction (sigmoid) layer function
$G_{\text{soft}}(\cdot)$	Dataset prediction (softmax) layer function
T	Number of tasks
$\mathcal{L}(\cdot)$	loss function
$\Phi(\mathbf{w})$	Regularization term that penalizes weights complexity
t_d	Regression tasks for distance parameter d
t_o	Regression tasks for orientation parameters θ and γ
α_y	Important coefficient of the label y
β_t	Important coefficient of the tth task's error
γ	Tradeoff parameter

Note: Matrices and 3-D images are written in bold uppercase (e.g., image $\mathbf{X}, \mathbf{Y}$), vectors and vectorized 2-D images in bold lowercase (e.g., slice $\mathbf{x}, \mathbf{y}$), and scalars are denoted in lowercase (e.g., slice position label r).

is DIAL, which consists of two fully connected layers (F1 and F2), each with 256 rectified linear unit activation neurones; the other one is MLMT*, which includes two fully connected layers (F3 and F4). The fully connected layers (F5 and F6) following two streams produce a feature vector shared by multiple tasks in the estimation stage.

2) Implementation Details and Parameters: Our method is implemented in PyTorch. All models were trained on an NVIDIA RTX 2080 Ti GPU at least three times with different random initializations to compute the mean performance and its standard deviation (run three times of the experiments with the same initialization). The same regularization strength is used across all datasets. Adam optimizer was used in all experiments with default hyperparameter settings (learning rate = 0.01, momentum = 0.9, and dropout rate = 0.1). We also kept the training details the same between our model and the compared methods. In addition, the notation used in our model is presented in Table I.

IV. EXPERIMENTS

A. Annotated Datasets

1) UKBB: CMR image data with gold standard image quality annotations are available for about 5000 volunteers of the UKBB imaging resource [34]. Based on the visual inspection of experts, a simple three-grade quality score [35] is used for manual annotation: 1) optimal quality for diagnosis (4361 sequences); 2) suboptimal quality, yet analyzable (527 sequences); and 3) poor quality and diagnostically unusable (177 sequences). Since these data have full coverage of the heart from base to apex, all data with optimal quality were used to construct the ground-truth classes in our experiments. We manually checked one slice above and below the detected basal slice to ensure it is the right one and generated the annotations; ditto for apical slice checking. It is worth noting that the data with full heart coverage here do not mean that the SAX plane orientation is correct for all these optimal samples. Therefore, we use the manual contours of the left ventricle of [36] to calculate the ground truth of the cut distance and orientation labels (see Section IV-B). The UKBB data are the source dataset for training our model.

2) MESA and DETERMINE: To evaluate the performance of DI learning, we used 598 subjects of CMR images obtained from the Cardiac Atlas Project (CAP) [33] (see Table II). CAP is a web-accessible resource (http://www.cardiacatlas.org/), which provides a resource for cardiac image data sharing and atlas-based shape analysis for population studies. The datasets used in this study are part of two cohorts: MESA [37] and DETERMINE [38]. The imaging protocols included cine images acquired in SAX planes from the heart base to the apex and three LAX planes. We manually checked one slice above and below the detected basal slice to ensure that it was right and generated the annotations; ditto for apical slice checking. For the MESA dataset, contours were drawn manually as a series of points by the MESA CMR core lab on SAX slices for all cases at end diastole (ED) and end systole (ES) [39] using-Q-MASS software. As the DETERMINE dataset does not provide manual contours, we manually generated the LV contours for each sample. We used the same method to create the distance and orientation labels.

To evaluate models with input from the PI (long-axis patches), the HLA image information was extracted by collecting pixel values along the intersecting line between the four-chamber view plane and the corresponding short-axis plane during the cardiac cycle. We extracted four pixels above and below the two-plane intersection. We embedded the constructed profile within a square image with zeros everywhere except at the profile diagonal [see Fig. 2(a), bottom channel].

B. Data Augmentation and Label Generation

1) Data Augmentation: To prevent overfitting because of insufficient target data (DETERMINE and MESA) for training, we artificially enlarge the two datasets by data augmentation techniques, such as realistic rotations, scaling factors, and corresponding mirror images. In particular, we chose $-45°$ and

TABLE II
CARDIOVASCULAR MAGNETIC RESONANCE PROTOCOLS FOR UKBB, MESA, AND DETERMINE DATASETS

Dataset	View	Number of Sequences	Cardiac Phases	Matrix Size	Slice Thickness	Slice Gap	Slice Spacing	Slices per Volume
UKBB	SAX	4280	50	208 × 187	8 mm	2 mm	10 mm	ca. 10
	LAX	4280	50	208 × 187	6 mm	n.a	n.a	1
	Balanced steady state free precession (bSSFP) cine images in short-axis planes and three long-axis planes [27].							
MESA	SAX	298	20 ∼ 30	256 × 160	6 mm	4 mm	10 mm	ca. 10
	LAX	298	20 ∼ 30	256 × 160	6 mm	n.a	n.a	1
	Cine images (using the GRE pulse sequence) in short-axis planes and three long-axis planes [33].							
DETERMINE	SAX	300	25	128 × 256	⩽ 10 mm	⩽ 2 mm	10 mm	ca. 10
	LAX	300	25	128 × 256	6 mm	n.a	n.a	1
	Cine images (using the SSFP pulse sequence) acquired in short-axis planes and three long-axis planes [33].							

45° as the rotations and 0.75 and 1.25 as the scaling factors. After data augmentation, the two datasets increased by a factor of eight and had 2400 and 2384 sequences for the DETERMINE and MESA datasets, respectively.

2) Plane Pose Parameters Definition: We define the SAX slice pose using four parameters: $\{d_b, d_a, \theta, \gamma\}$. The ground truth of the position of the slice d_b (the distance to the base slice), d_a (the distance to the apical slice) and the orientation parameterized with deflection angles θ in the xoy plane and γ in the z direction can be calculated from the SAX slices in realistic cardiac volumes. According to Table II, the interslice spacing value δd is constant in each cardiac volume. Therefore, we can represent d_a and d_b by multiples δd. The orientation is defined based on θ and γ, which are calculated as the following process: 1) standard cardiac vector: steady-state free precession short-axis cine images are acquired from the MV plane through the apex [40]. Therefore, we choose the vector $\widehat{O}_L^S$ that is perpendicular to the MV plane as the standard cardiac vector; 2) normal vector: the normal vector $\widehat{O}_L^A$ is perpendicular to the given SAX image plane; and 3) deflection angles θ and γ: θ is the angle between the x-axis and the projection of the deflection vector at the xoy plane, γ is the angle between $\widehat{O}_L^S$ and $\widehat{O}_L^A$, which is calculated as $\gamma = \angle\left\langle \hat{O}_L^S, \hat{O}_L^A \right\rangle$.

3) Label Generation: We construct the multilabel $\{y^{t_d}, y^{t_\theta}, y^{t_\gamma}\}$ with a set of realistic distance and orientation values in the images of the three datasets for training our model. For each cardiac volume, we normalize the distance between the basal slice and the apical slice as unit 1, setting the distance label of the basal slices as 0. The distance label of the apical slice is 1, then the label of the rest of the slices can be synthesized using $y_i^{t_d} = (i - Z_b)/(Z_a - Z_b)$ (Z_a and Z_b represent the numbers of the basal and basal slices, respectively, in the sequence Z). Here, the normalized y^{t_d} can not only represent d_b but also represent d_a. Based on the analysis of the plane orientation distribution for the three datasets available for manual segmentation (and therefore y^{t_θ}, y^{t_γ} can be computed), we found that y^{t_θ} ranges at the median value of 132.8° with standard deviation 8.0° and y^{t_γ} ranges at the median value of 7.1° with standard deviation 3.9°. The orientation labels used for training were chosen from these realistic distributions, and all used labels were balanced by data augmentation.

C. Comparison Methods and Evaluation Metrics

To evaluate the effectiveness of the proposed method in different datasets, we conduct a comprehensive comparison of our approach with various state-of-the-art (related) approaches for cross-dataset slice pose estimation.

1) *MC+CNN*: Metric Classification with CNN [4];
2) *MC+CNN**: MC+CNN with PI [41];
3) *MCDA+CNN*: MC-CNN with DI [30];
4) *MCDA+CNN**: MCDA+CNN with PI;
5) *MR+CNN*: Metric Regression with CNN [31];
6) *DARN-DA* (MLMT): Metric Regression with MLMT;
7) *DARN*-DA*: DARN-DA with PI;
8) *DARN*: DARN* without PI;
9) *DARN*-MLMT*: DARN* without MLMT [42]; and
10) *DARN**: Fully fledged dataset adversarial regression network with PI method.

MC+CNN can be cast as a fundamental baseline only for considering the deep classification neural network, and MR+CNN can be cast as a fundamental baseline only for considering the deep regression neural network. MCDA+CNN is the most relevant and state-of-the-art cross-dataset image classification approach. For clarity, to validate the effectiveness of our DARN* method by removing parts of the fully functional model, an ablation study [43] is adopted. We consider three special cases of the proposed method by excluding DI (DARN*-DA) or excluding PI (DARN * -PI), or excluding the MLMT regression neural network (DARN * -MLMT) [42] to show that each of the added terms is useful for a more accurate pose estimation.

To evaluate the classification tasks, we use Accuracy $= (\text{TP+TN})/N$, where TP and TN are the numbers of true positive and true negative samples, respectively, N represents the number of subjects in the test set.

To evaluate the regression error between our predicted pose and the ground truth, we adopt a widely used evaluation method, mean absolute error (MAE) [44], in our experiments for different models. MAE computes the absolute costs between the real and the predicted slice position or orientation: $\text{MAE} = \sum_{i=1}^{M} e_i/M$, where $e_i = \left|\hat{l}_i - l_i\right|$ is the absolute cost of misclassifying the true label l_i with $\hat{l}_i$, and M is the total of the testing samples. We expect that the lower the MAE, the better. Also, we use the paired t-test to demonstrate the statistical

TABLE III
COMPARISON OF BASAL/APICAL SLICE DETECTION ACCURACY (MEAN ± STANDARD DEVIATION) BETWEEN ADAPTATION AND NONADAPTATION METHODS, EACH WITH SINGLE (SAX)- AND MULTIVIEW INPUTS (BS/AS INDICATE BASAL/APICAL SLICE DETECTION ACCURACY, RESPECTIVELY)

Dataset	No Dataset Adaptation (BS/AS)		With Dataset Adaptation (BS/AS)	
	MC+CNN [4]	MC+CNN* [41]	MCDA+CNN [30]	MCDN+CNN* (Ours)
UKBB	79.0±0.2/76.2±0.3	**89.2**±0.1/**92.4**±0.2	–	–
MESA	31.6±0.3/35.1±0.1	61.5±0.2/68.3±0.4	74.2±0.2/72.9±0.4	**87.1**±0.3/**90.2**±0.2
DETERMINE	48.3±0.2/51.1±0.3	75.6±0.3/78.4±0.3	77.2±0.3/76.5±0.2	**89.0**±0.2/**91.2**±0.2

Note: Best results are highlighted in bold. All experiments were trained with UKBB data.

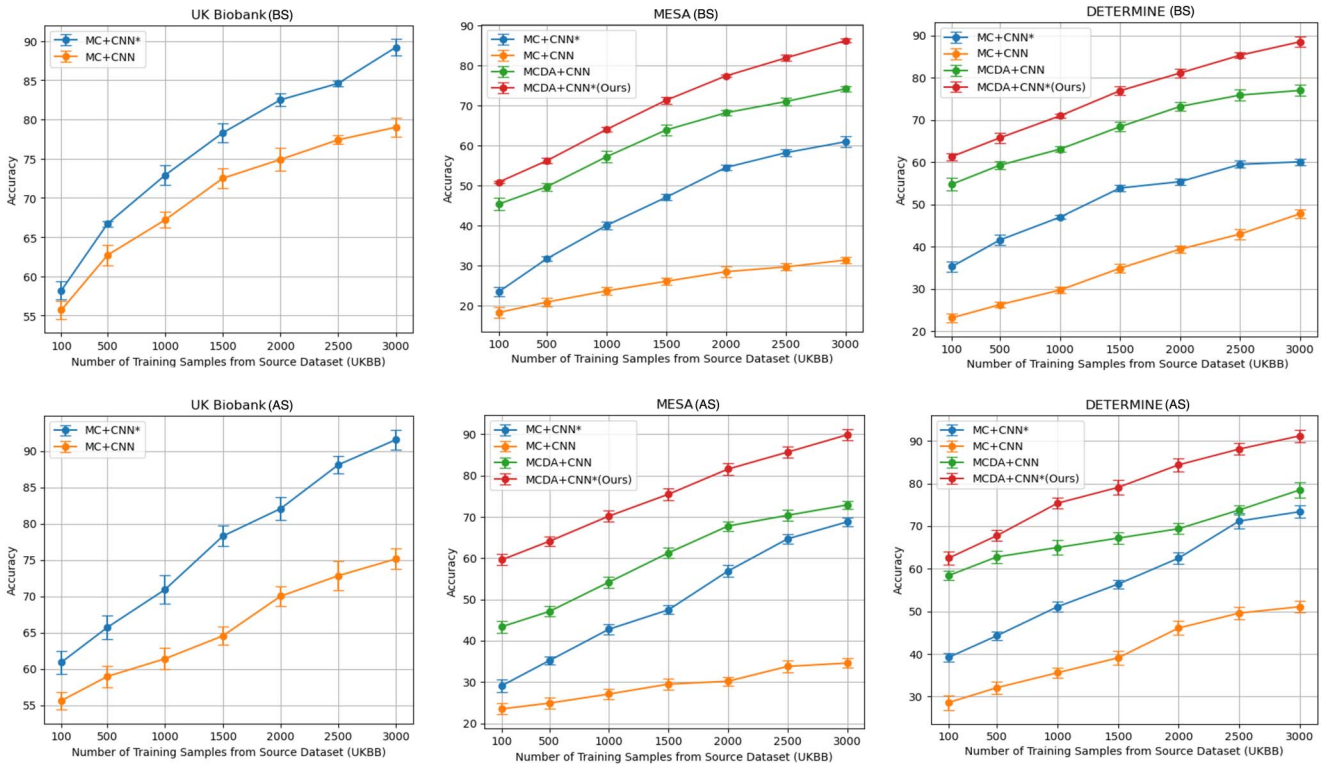

Fig. 3. Basal/apical slice detection accuracy (mean ± standard deviation) when the number of training samples from the source dataset (UKBB) increases. BS/AS indicates the detection accuracy of the basal/apical slice, respectively.

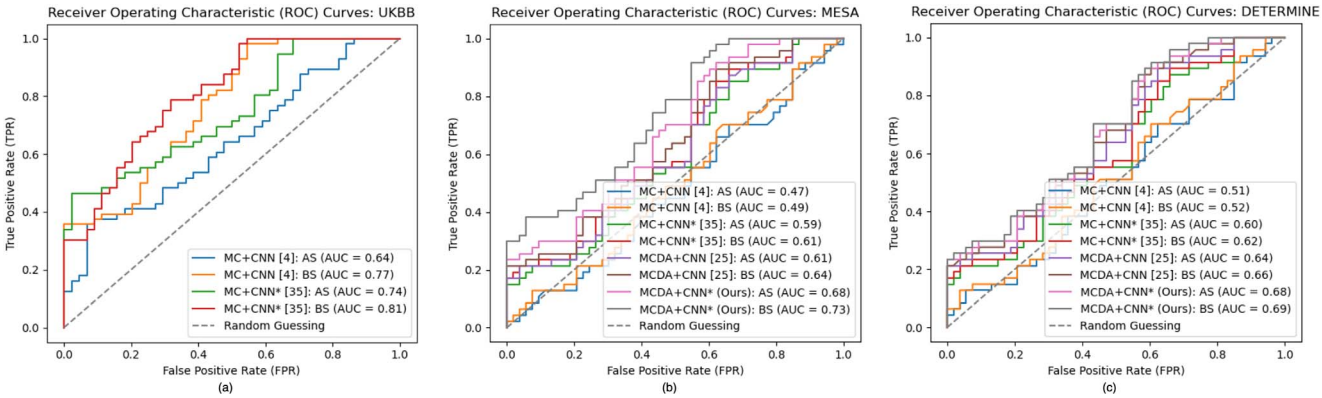

Fig. 4. ROC curve and the AUC for basal/apical slice detection using different methods: (a) UKBB; (b) MESA; and (c) DETERMINE. BS/AS indicates the basal/apical slice detection, respectively.

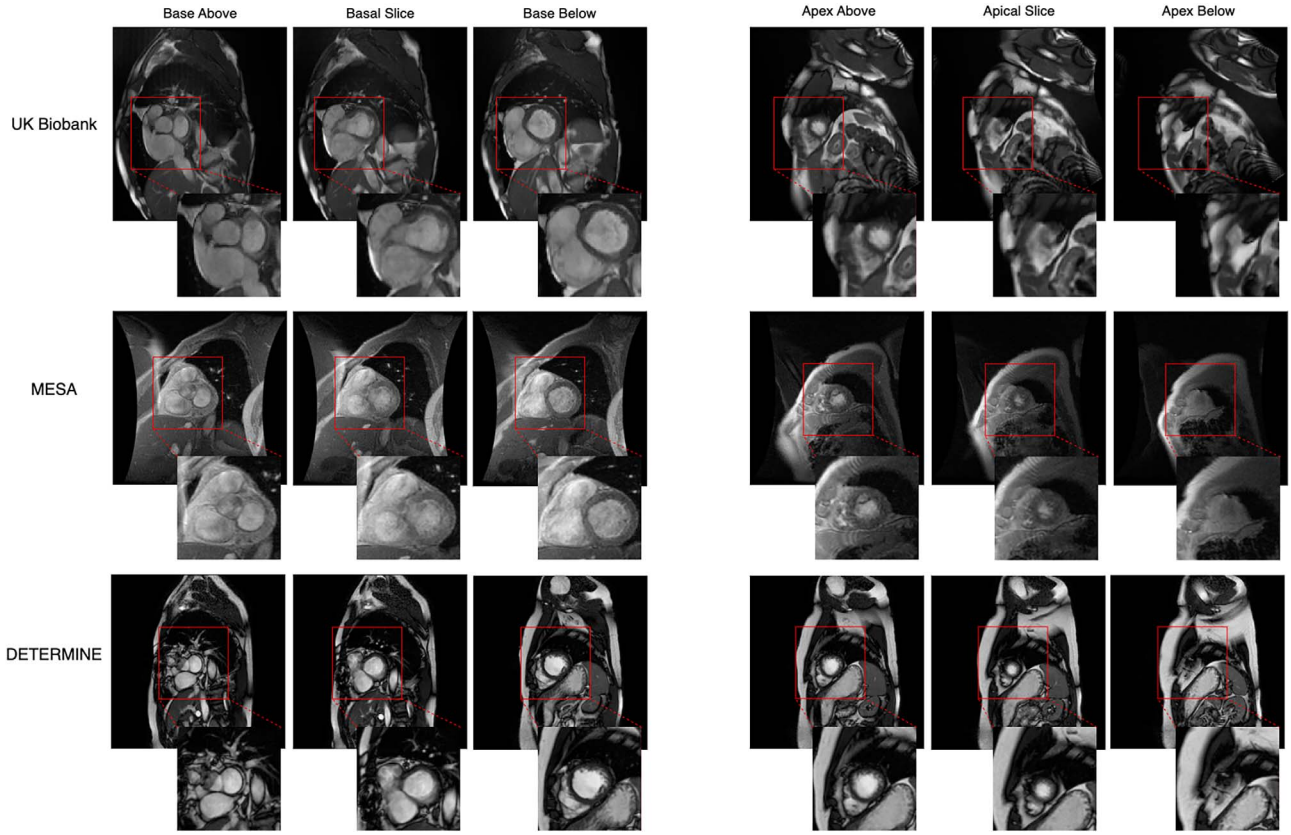

Fig. 5. Examples show the detected basal slice, apical slice, and the corresponding above and below slices. The red square area in each image shows the region-of-interest for model learning.

TABLE IV
SAX SLICE POSITION REGRESSION ERROR COMPARISON BETWEEN ADAPTATION AND NONADAPTATION METHODS, IN TERMS OF MAE (MEAN ± STANDARD DEVIATION) (MM)

Dataset	No Dataset Adaptation			With Dataset Adaptation		
	MR+CNN [31]	DARN-DA	DARN*-DA	DARN	DARN*-MLMT	DARN* (Ours)
UKBB	5.43 ± 1.4	4.11±m 1.6	**3.12± 1.1**	–	–	–
MESA	8.21± 1.6	7.94± 2.0	6.53 ± 1.7	4.97 ± 1.0	3.91 ± 1.1	**3.68 ± 1.1**
DETERMINE	7.42 ± 1.3	6.47 ± 1.9	5.96 ± 1.3	4.77 ± 1.6	4.27 ± 1.0	**4.05 ± 1.0**

Note: Best results are highlighted in bold. The unit of the distance regression error is *mm*.

significance of our empirical comparison if our DARN* outperforms other methods.

We extract every slice from top to bottom for each volume during testing and apply them to our DARN* model. Our model output gives the slice position in millimeters and two angles (θ and γ) in degrees. There are no LAX patches (PI) in this phase.

D. Results of Left Ventricle Coverage Estimation

1) Results Analysis for Basal and Apical Slice Detection: To evaluate the performance of DI learning and learning with PI, we propose a baseline method that deals with object detection, which only keeps the end-to-end CNN learning part and drops the part of transforming framework, i.e., it casts the basal/apical slice detection problem as a metric classification problem. It addresses it with/without DI (adaptation versus nonadaptation) and PI, by transferring object classifiers from the UKBB to MESA and DETERMINE. For clarity, we compared the metric classification with CNN (MC+CNN) in [4], MC+CNN* in [41], and the GTSRB architecture in [30] (MCDA+CNN) with our MCDA+CNN*. Table III shows the accuracy of adaptation and nonadaptation detection for traditional CNN and CNN with PI. The best improvement results from combining these features (DI plus PI) for both target datasets. For MESA, the detection accuracy increased by 64%; for DETERMINE, the improvement reaches 44% (rightmost column). All experiments are significantly different at $p < 0.05$. In Fig. 3, we analyzed the detection accuracy when the number of training samples

TABLE V
COMPARISON BETWEEN ADAPTATION AND NONADAPTATION METHODS, EACH WITH A SINGLE VIEW (SAX) AND PI INPUTS FOR CARDIAC SAX SLICE ORIENTATION ESTIMATION IN TERMS OF MAE (MEAN ± STANDARD DEVIATION)

Dataset	No Dataset Adaptation ($\Delta\theta/\Delta\gamma$)			With Dataset Adaptation ($\Delta\theta/\Delta\gamma$)		
	MR+CNN [31]	DARN-DA	DARN*-DA	DARN	DARN*-MLMT	DARN* (Ours)
UKBB	5.94±1.4/3.68±1.6	5.42±1.6/3.37±1.9	**5.25**±1.1/**3.24**±1.7	–	–	–
MESA	7.32±2.6/5.02±1.9	6.78±2.0/4.83±2.4	6.47±1.7/4.54±1.4	6.31±2.0/3.92±2.4	6.12±1.6/3.86±1.9	**5.96**±1.7/**3.74**±1.4
DETERMINE	7.14±2.1/5.11±2.3	6.63±1.9/4.81±2.0	6.32±1.3/4.17±1.4	6.27±1.6/3.81±1.3	5.82±1.6/3.69±1.9	**5.24**±1.0/**3.45**±1.3

Note: $\Delta\theta$ and $\Delta\gamma$ indicate the MAE of the deflection angles in degree (°). Best results are highlighted in bold.

from the source dataset (UKBB) increased. Given more labeled source data, DI learning still works reliably and contributes to the detection tasks for the target datasets. Fig. 4 provides the ROC curve and the AUC for basal/apical slice detection with different methods among the three datasets. This analysis includes a comparison of the true positive rate (sensitivity) and false positive rate (one specificity) for each method, as well as the AUC values representing the overall discriminative ability of the methods in detecting basal and apical slices. For instance, the PI branch boosts the performance on BS/AS detection by looking at the AUC results on the UKBB dataset. From the AUC results in the MESA and DETERMINE datasets, the adversarial training approach generates superior performance for the BS/AS detection. Meanwhile, we present image samples from the three datasets in Fig. 5 to show the detected basal, apical slices and the corresponding above and below slices.

2) Results Analysis for Slice Pose Estimation: We propose another model, which only keeps the end-to-end CNN learning part with DI and PI, and drops the part of transforming framework, i.e., it casts the metric classification module instead of a regression module, which transforms slice pose estimation, including distance and orientation, into the MLMT regression problem. To discover the factor that contributes the most to the final improvement in performance and to confirm that our regularization terms are beneficial, we also compare our proposed DARN * with DARN-DA, DARN *-DA, and DARN *-PI and show that these results in Tables IV and V represent position estimation and orientation regression, respectively.

Table IV shows the average estimation errors of the cutting distance using the MAE metric in millimeters (mm). Even without the PI input channel, our DI framework can reduce the slice distance estimation error to less than half the average slice spacing in our test datasets, that is, < 5 mm. With PI in the training process, we reduced the MAE to 4.27 and 4.05 mm on average to estimate the cut distance on the DETERMINE dataset. We also visualized the performance when the number of training samples from the source dataset (UKBB) increases. Regression errors are decreasing, and our proposed model achieved the best results in each group of experiments. Fig. 6 shows that the slice distance regression error on the three datasets provides a visual representation of the performance of different methods in estimating the distance to the basal and apical slices in CMR imaging. The results indicate that our proposed method demonstrates the best performance in terms of the slice distance regression error in three large datasets.

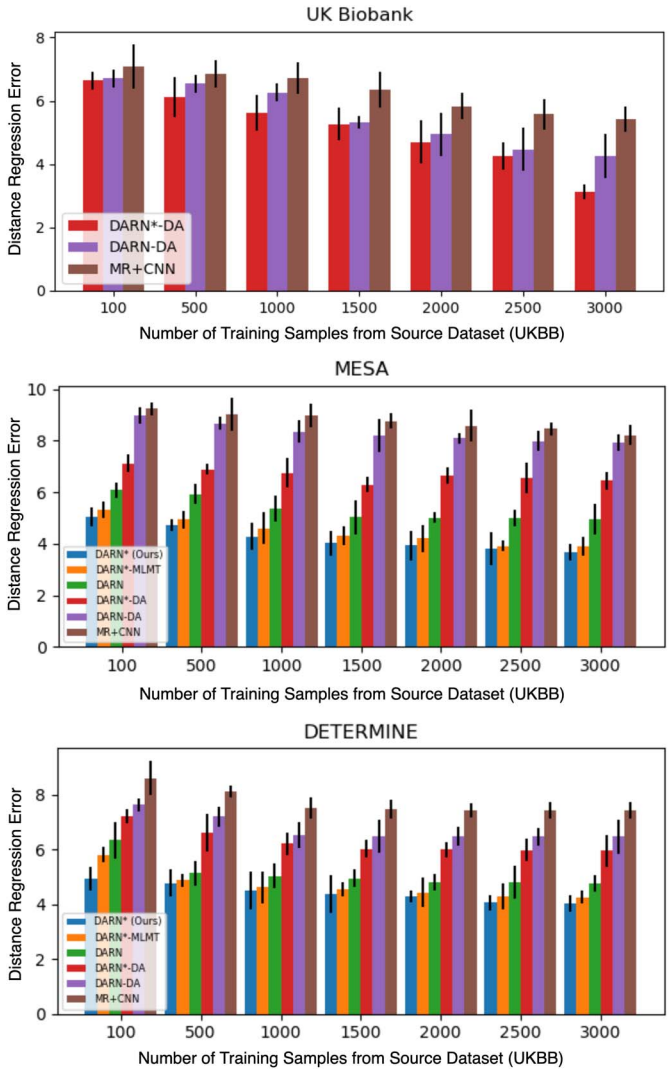

Fig. 6. Performance of slice distance regression (mean error ± standard deviation) when the number of training samples from the source dataset (UKBB) is increasing.

Table V shows the MAE of the estimation of the orientation of the slice by regression θ and γ in degree (°). Even without using the PI input channels, our DI framework can obtain smaller estimation errors, that is, $\Delta\theta < 7°$ and $\Delta\gamma < 4°$. With PI input, we reduced the estimation errors of $\Delta\theta$ and $\Delta\gamma$ to 5.24° and 3.45° on average for each volume on the DETERMINE dataset.

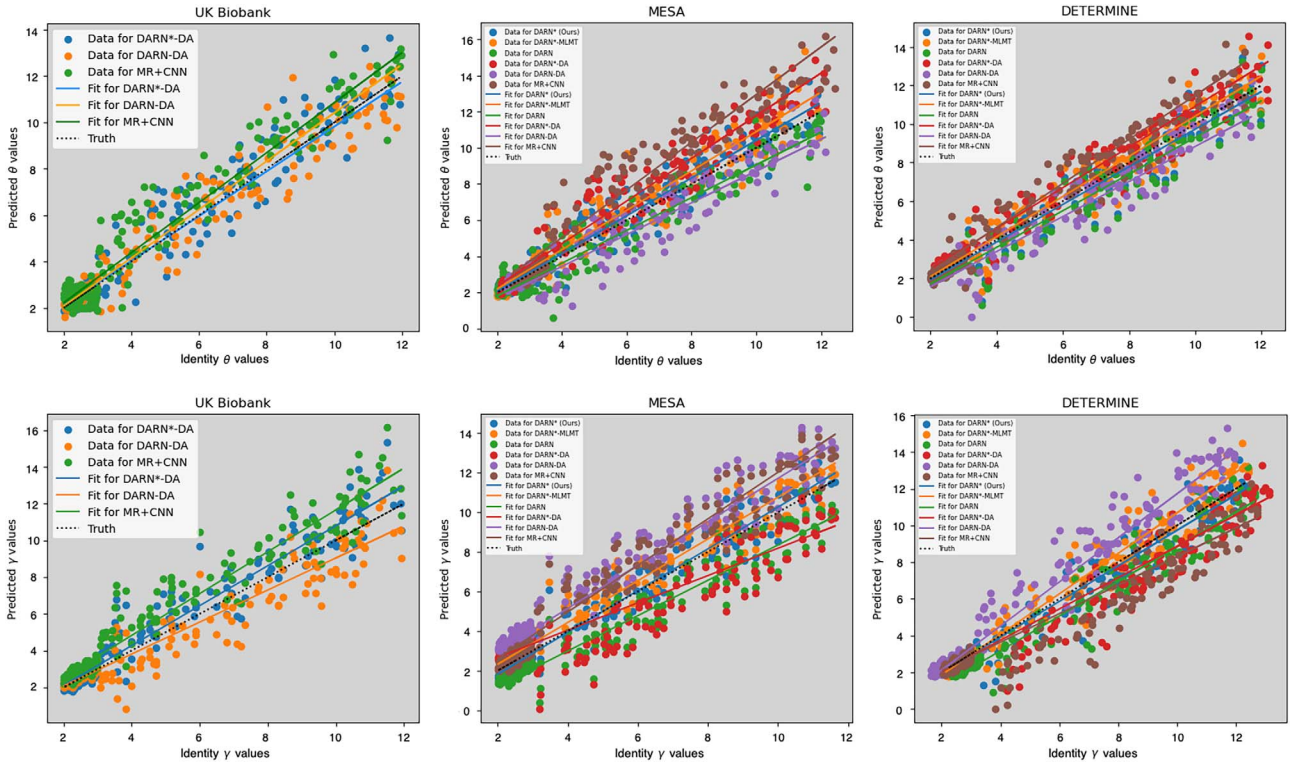

Fig. 7. Regression results of plane orientation parameters: θ, γ, and their corresponding true results. We only provide results without dataset adaptation for the UKBB dataset.

Meanwhile, we plot the regression results for θ and γ in Fig. 7. Looking at the fitting lines of each method, our model shows a minimal regression error with significant margins, especially when testing on the MESA and DETERMINE datasets. All experiments are significantly different at $p < 0.05$.

The training and validation loss curves in Fig. 8 for the SAX slice regression error reflect the model's performance in minimizing the error between predicted and ground-truth distances to the basal and apical slices. A decreasing loss curve indicates effective learning and convergence, while fluctuations or a plateau may indicate challenges in model performance. Similarly, the loss curves for the slice plane orientation parameters $\Delta\theta$ and $\Delta\gamma$ provide information about the model's ability to minimize errors in estimating the orientation of the slice plane. These curves offer insight into the model's capacity to accurately predict deflection angles in the *xoy* plane and the z direction.

V. Discussion

Accuracy and *robustness* should be considered key criteria when designing the automatic image QA system for cardiac population imaging. False positives and negatives should be minimized when detecting basal/apical slices using these methods, and the MAE of the slice distance/orientation must cope with considerable image quality variation. Most deep learning methods can perform well by training and testing a single dataset. However, this may not be true when processing with different datasets or retraining is required as new data become available. This study used a large dataset that included 5000 subjects (each with 50 time points) from the same number of individually annotated CMR image scans in UKBB. However, when we transfer our well-trained model to other CMR image datasets, we observe that deep learning methods without DI cannot achieve good performance. We had to design an efficient network learning common representation across datasets. Since there is no label information in the target datasets, we also need to learn discriminative information from the source dataset and transfer it to our target datasets. To reduce the difference between the training and test domain distributions and improve generalization performance, AL has been among the most promising solutions. However, most AL works have been focused on image-generative tasks, and little effort has been devoted to minimizing an approximate domain discrepancy distance. We propose novel AL to detect and localize the basal/apical slices across datasets, incorporating the PI (cross-view slice) into the training phase. Then, an MLMT regression network is trained to estimate the slice distance and orientation. In particular, our proposed DIAL and PI learning strategy can achieve a high accuracy rate of almost 87%/90% for the detection of MBS/MAS by training in UKBB and testing in MESA, which is better than CNN methods without adaptation of the dataset. Meanwhile, with the MLMT network, DARN* can decrease MAE by around 6% compared to the DARN*-MLMT [42] approaches for the estimation of the basal cut distance in the MESA dataset.

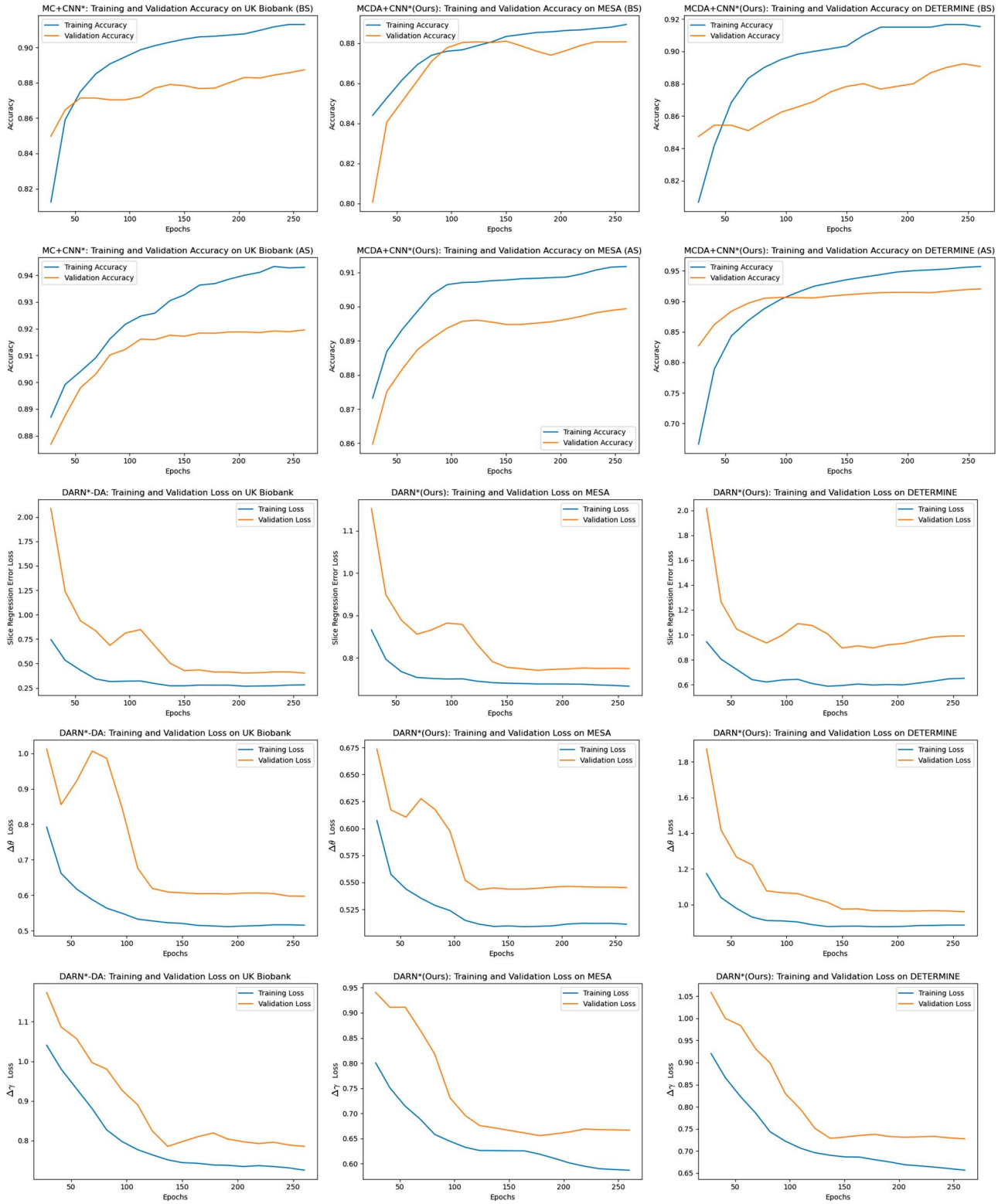

Fig. 8. Accuracy curves of BS and AS classification; the loss curves of SAX slice regression error and the slice plane orientation parameters: $\Delta\theta$, $\Delta\gamma$.

A. Addressing Incomplete Coverage With Atlas-Based Methods

Integrating anatomical shape priors into network segmentation, as introduced by Duan et al. [18] and Joyce et al. [19], represents a significant advancement in refining segmentation processes through atlas propagation. This technique, particularly beneficial for addressing segmentation challenges under incomplete left ventricular (LV) coverage, relies on the

preacquisition of a relatively dense set of imaging planes. By segmenting the ventricles or generating volumetric meshes as an intermediate step, these methods aim to estimate ventricle coverage accurately. Both [18] and [19] employ shape priors to mitigate the limitations posed by potentially incomplete image coverage, a strategy that inherently depends on the anatomical coverage provided by the atlases (e.g., LV+RV in [18] and LV in [19]). However, the reliance on atlas-based approaches introduces a notable vulnerability: the potential for bias. This bias primarily emerges when atlases, developed from specific imaging views or single-site data, are applied to multiview or multisite studies, or in scenarios involving significant pathology not represented in the atlas. Moreover, these methods, whilst innovative, are computationally intensive and have yet to be validated on extensive cross-center datasets. Consequently, their capacity to generalize to diverse real-world scenarios is still under scrutiny. Despite these challenges, the value of atlas-to-subject registration methods becomes particularly evident when dealing with already collected data that presents incomplete coverage. In such cases, these methods offer a viable solution for managing data limitations postacquisition. Nonetheless, the context shifts when considering population imaging studies or clinical trials, where the paramount concern is image quality. In these instances, the ability to promptly identify incomplete coverage immediately after MRI acquisition is invaluable. Such timely feedback allows for the possibility of reacquiring the image before the patient or volunteer departs from the imaging suite, ensuring the collection of high-quality data essential for robust analysis. This nuanced perspective underscores the complexity of employing atlas-based segmentation methods in medical imaging. While they offer substantial benefits in certain contexts, their limitations and the need for careful application must be acknowledged to leverage their full potential effectively.

In this study, we propose to solve the image QA problem in terms of SAX slice position and plane orientation, and generalize the model trained on UKBB cohorts to other image domains, e.g., MESA and DETERMINE cohorts. Therefore, in our experiments, we validated our image QA pipeline on MESA and DETERMINE, which can be recognized as real datasets with different image domains from UKBB. In all three datasets, our method has shown better accuracy in detecting basal and apical slices, as well as a decrease in MAE compared to other approaches. The study's results demonstrate the potential for robust population image analysis and the ability to adapt to image quality variations. Our proposed automatic image QA framework for CMR image plane orientation estimation has great potential for later robust population image analysis. One could imagine that the proposed image-analysis methods are adaptive to image quality and design, depending on whether the image under analysis has incomplete ventricle coverage. If such a QA platform or system can be performed online and provide immediate feedback to technical staff when acquiring new images, thousands of suboptimal CMR images could be avoided in future image analysis. More importantly, the proposed model has the potential for immediate integration within the scanner system, offering a highly advantageous ability to instruct radiographers effectively. This guide helps them promptly identify, verify, and address discrepancies in image slices acquisition, while the patient remains within the scanning apparatus. Although we introduced incomplete coverage of LV and RV that can have an important impact on calculating cardiac function, this work focuses primarily on estimating LV coverage, which is easier to evaluate our proposal model. However, our model can potentially be generalized to the RV as well by taking a look at the RV characters of each slice from base to apex. When analyzing RV coverage using our model, it may require addressing specific anatomical and imaging challenges unique to the RV, such as its complex geometry and differences in image characteristics compared to the LV. One of our future studies will investigate the possibility of quantifying ventricle coverage, which is not specific for estimating the position and orientation of the slice, so that we can predict the percentage of ventricle coverage directly. It is difficult to calculate the percentage of ventricle coverage. Therefore, training the volume classifier could be a nontrivial task due to the different shapes of contiguous ventricle slices. Another future study will expand the deep learning method to synthesize missed slices, that is, synthesis of the basal/apical slice if a cardiac sample without them and the missing slices acquired from the middle positions. One possible way to achieve this using the machine learning approach for image synthesis would be to apply a generative network (e.g., synthesize the missed slices using the diffusion model [45] in UKBB).

VI. Conclusion

In this article, we proposed an adversarial regression network in the dataset with the PI (DARN*) model to learn a common image representation and use it to estimate the CMR slice pose (SAX). We achieve this by first using the DIAL model with a *minmax* game to learn the common representation between images from different datasets. The DIAL model was then incorporated with the PI input (HLA) scenarios to achieve better performance for the estimation of the plane pose in the CMR image (SAX). Finally, we proposed an MLMT regression network to predict the slice pose. Our DARN* model is evaluated on three large datasets, UKBB, MESA, and DETERMINE. Extensive experimental results show that the capability of the proposed model outperforms state-of-the-art nondataset-adaptive and non-PI methods. Our DARN* approach lies in its integration of adversarial regression tasks, PI, and dataset-invariant strategies to offer a unique and effective solution to the challenges associated with incomplete ventricle coverage in CMR imaging. Our model provides valuable guidance to radiographers, urging them to review faulty scans, streamlining image quality control, and facilitating the prompt correction of missing or misaligned image slices while the patient is still undergoing the scan.

References

[1] B. Ruijsink et al., "Fully automated, quality-controlled cardiac analysis from CMR: Validation and large-scale application to characterize cardiac function," *Cardiovascular Imag.*, vol. 13, no. 3, pp. 684–695, 2020.

[2] Q. Meng et al., "Mulvimotion: Shape-aware 3D myocardial motion tracking from multi-view cardiac MRI," *IEEE Trans. Med. Imag.*, vol. 41, no. 8, pp. 1961–1974, Aug. 2022.

[3] Y. Xia et al., "Recovering from missing data in population imaging–Cardiac MR image imputation via conditional generative adversarial nets," *Med. Image Anal.*, vol. 67, 2021, Art. no. 101812.

[4] L. Zhang et al., "Automated quality assessment of cardiac MR images using convolutional neural networks," in *Proc. Int. Workshop Simul. Synthesis Med. Imag.,* Springer-Verlag, 2016, pp. 138–145.

[5] J. Schulz-Menger et al., "Standardized image interpretation and post-processing in cardiovascular magnetic resonance-2020 update," *J. Cardiovascular Magn. Reson.*, vol. 22, no. 1, pp. 1–22, 2020.

[6] X. Lu et al., "Automatic view planning for cardiac MRI acquisition," in *Proc. Med. Image Comput. Comput. Assist. Intervention,* Springer-Verlag, 2011, pp. 479–486.

[7] M. Edalati et al., "Implementation and prospective clinical validation of AI-based planning and shimming techniques in cardiac MRI," *Med. Phys.*, vol. 49, no. 1, pp. 129–143, 2022.

[8] D. Wei, K. Ma, and Y. Zheng, "Training automatic view planner for cardiac MR imaging via self-supervision by spatial relationship between views," in *Proc. Med. Image Comput. Comput. Assisted Intervention (MICCAI),* Strasbourg, France, Berlin, Germany: Springer-Verlag, 2021, pp. 526–536.

[9] I. Oksuz et al., "Automatic left ventricular outflow tract classification for accurate cardiac MR planning," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI),* Piscataway, NJ, USA: IEEE Press, 2018, pp. 462–465.

[10] P. Peng, K. Lekadir, A. Gooya, L. Shao, S. E. Petersen, and A. F. Frangi, "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 29, pp. 155–195, Apr. 2016.

[11] M. Paknezhad, S. Marchesseau, and M. S. Brown, "Automatic basal slice detection for cardiac analysis," *J. Med. Imag.*, vol. 3, no. 3, 2016, Art. no. 034004.

[12] M. Le, J. Lieman-Sifry, F. Lau, S. Sall, A. Hsiao, and D. Golden, "Computationally efficient cardiac views projection using 3D convolutional neural networks," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support: 3rd Int. Workshop (DLMIA), 7th Int. Workshop (ML-CDS),* Québec City, QC, Canada. Berlin, Germany: Springer-Verlag, 2017, pp. 109–116.

[13] P. J. García-Laencina, J.-L. Sancho-Gómez, and A. R. Figueiras-Vidal, "Pattern classification with missing data: A review," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 263–282, 2010.

[14] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[15] D. Mahapatra, "Landmark detection in cardiac MRI using learned local image statistics," in *Proc. Int. Workshop Statist. Atlases Comput. Models Heart*, Springer-Verlag, 2012, pp. 115–124.

[16] J. Schulz-Menger et al., "Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing," *J. Cardiovascular Magn. Reson.*, vol. 15, no. 1, 2013, Art. no. 35.

[17] L. Zhang et al., "Automatic assessment of full left ventricular coverage in cardiac cine magnetic resonance imaging with fisher discriminative 3-D CNN," *IEEE Trans. Biomed. Eng.*, vol. 22, no. 1, pp. 102–107, Nov. 2018.

[18] J. Duan et al., "Automatic 3D Bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2151–2164, Sep. 2019.

[19] T. Joyce, S. Buoso, C. T. Stoeck, and S. Kozerke, "Rapid inference of personalised left-ventricular meshes by deformation-based differentiable mesh voxelization," *Med. Image Anal.*, vol. 79, 2022, Art. no. 102445.

[20] 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, 2019, Art. no. e180069.

[21] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6569–6578.

[22] X. Zhou, J. Zhuo, and P. Krahenbuhl, "Bottom-up object detection by grouping extreme and center points," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 850–859.

[23] A. Alansary et al., "Automatic view planning with multi-scale deep reinforcement learning agents," in *Proc. Med. Image Comput. Comput. Assisted Intervention (MICCAI),* Granada, Spain, Berlin, Germany: Springer-Verlag, 2018, pp. 277–285.

[24] X. Ma et al., "Understanding adversarial attacks on deep learning based medical image analysis systems," *Pattern Recognit.*, vol. 110, 2021, Art. no. 107332.

[25] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, *arXiv:1412.6572*.

[26] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," 2017, *arXiv:1706.06083*.

[27] S. E. Petersen et al., "UK Biobank's cardiovascular magnetic resonance protocol," *J. Cardiovascular Magn. Reson.*, vol. 18, no. 1, pp. 8–14, 2015.

[28] K. Yoneyama et al., "Natural history of myocardial function in an adult human population: Serial longitudinal observations from mesa," *JACC Cardiovascular Imag.*, vol. 9, no. 10, pp. 1164–1173, 2016.

[29] A. H. Kadish et al., "Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (DETERMINE) trial," *J. Cardiovascular Electrophysiol.*, vol. 20, no. 9, pp. 982–987, 2009.

[30] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[31] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output CNN for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 4920–4928.

[32] V. Vapnik and A. Vashist, "A new learning paradigm: Learning using privileged information," *Neural Netw.*, vol. 22, nos. 5–6, pp. 544–557, 2009.

[33] C. G. Fonseca, "The Cardiac Atlas Project—An imaging database for computational modeling and statistical atlases of the heart," *Bioinformatics*, vol. 27, no. 16, pp. 2288–2295, 2011.

[34] S. E. Petersen et al., "Imaging in population science: Cardiovascular magnetic resonance in 100,000 participants of UK Biobank—Rationale, challenges and approaches," *J. Cardiovascular Magn. Reson.*, vol. 15, no. 1, pp. 1–10, 2013.

[35] V. Carapella et al., "Towards the semantic enrichment of free-text annotation of image quality assessment for UK Biobank cardiac cine MRI scans," in *Proc. Deep Learn. Data Labeling Med. Appl.,* Springer-Verlag, 2016, pp. 238–248.

[36] S. E. Petersen et al., "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort," *J. Cardiovascular Magn. Reson.*, vol. 19, no. 1, pp. 1–19, 2017.

[37] P. Medrano-Gracia et al., "Left ventricular shape variation in asymptomatic populations: The multi-ethnic study of atherosclerosis," *J. Cardiovascular Magn. Reson.*, vol. 16, pp. 1–10, Dec. 2014.

[38] P. Medrano-Gracia et al., "Continuous spatio-temporal atlases of the asymptomatic and infarcted hearts," in *Proc. 4th Int. Workshop Statist. Atlases Comput. Models Heart. Imag. Modell. Challenges (STACOM),* Nagoya, Japan, Berlin, Germany: Springer-Verlag, 2014, pp. 143–151.

[39] S. Natori et al., "Cardiovascular function in multi-ethnic study of atherosclerosis: Normal values by age, sex, and ethnicity," *Amer. J. Roentgenol.*, vol. 186, no. 6_supplement_2, pp. S357–S365, 2006.

[40] C. M. Kramer, J. Barkhausen, S. D. Flamm, R. J. Kim, and E. Nagel, "Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update," *J. Cardiovascular Magn. Reson.*, vol. 15, no. 1, 2013, Art. no. 91.

[41] J. Lambert, O. Sener, and S. Savarese, "Deep learning under privileged information using heteroscedastic dropout," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR), 2018, pp. 8886–8895.

[42] L. Zhang, M. Pereañez, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Multi-input and dataset-invariant adversarial learning (MDAL) for left and right-ventricular coverage estimation in cardiac MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.* Assisted Intervention, Berlin, Germany: Springer-Verlag, 2018, pp. 481–489.

[43] O. Ishaq, S. K. Sadanandan, and C. Wählby, "Deep fish: Deep learning–based classification of zebrafish deformation for high-throughput screening," *SLAS DISCOVERY Adv. Life Sci. R & D*, vol. 22, no. 1, pp. 102–107, 2017.

[44] N. M. R. Suri, M. N. Murty, and G. Athithan, *Outlier Detection: Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2019.

[45] K. Preechakul, N. Chatthee, S. Wizadwongsa, and S. Suwajanakorn, "Diffusion autoencoders: Toward a meaningful and decodable representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10619–10629.