



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/145286/>

Version: Accepted Version

---

**Proceedings Paper:**

Zhang, L, Pereañez, M, Piechnik, SK et al. (2018) Multi-input and dataset-invariant adversarial learning (MDAL) for left and right-ventricular coverage estimation in cardiac MRI. In: Lecture Notes in Computer Science. 21st International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2018, 16-20 Sep 2018, Granada, Spain,. Springer Verlag, pp. 481-489. ISBN: 9783030009335. ISSN: 0302-9743. EISSN: 1611-3349.

[https://doi.org/10.1007/978-3-030-00934-2\\_54](https://doi.org/10.1007/978-3-030-00934-2_54)

---

© Springer Nature Switzerland AG 2018. This is an author produced version of a conference paper published in Lecture Notes in Computer Science. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# Multi-Input and Dataset-Invariant Adversarial Learning (MDAL) for Left and Right-Ventricular Coverage Estimation in Cardiac MRI

Le Zhang<sup>1</sup>, Marco Pereañez<sup>1</sup>, Stefan K. Piechnik<sup>2</sup>,  
Stefan Neubauer<sup>2</sup>, Steffen E. Petersen<sup>3</sup>, Alejandro F. Frangi<sup>1</sup>

<sup>1</sup>Centre for Computational Imaging and Simulation Technologies in Biomedicine, Department of Electronic and Electrical Engineering, University of Sheffield, UK.

<sup>2</sup>Oxford Center for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, UK.

<sup>3</sup>Cardiovascular Medicine at the William Harvey Research Institute, Queen Mary University of London and Barts Heart Center, Barts Health NHS Trust, UK.

**Abstract.** Cardiac functional parameters, such as, the Ejection Fraction (EF) and Cardiac Output (CO) of both ventricles, are most immediate indicators of normal/abnormal cardiac function. To compute these parameters, accurate measurement of ventricular volumes at end-diastole (ED) and end-systole (ES) are required. Accurate volume measurements depend on the correct identification of basal and apical slices in cardiac magnetic resonance (CMR) sequences that provide full coverage of both left (LV) and right (RV) ventricles. This paper proposes a novel adversarial learning (AL) approach based on convolutional neural networks (CNN) that detects and localizes the basal/apical slices in an image volume independently of image-acquisition parameters, such as, imaging device, magnetic field strength, variations in protocol execution, etc. The proposed model is trained on multiple cohorts of different provenance, and learns image features from different MRI viewing planes to learn the appearance and predict the position of the basal and apical planes. To the best of our knowledge, this is the first work tackling the fully automatic detection and position regression of basal/apical slices in CMR volumes in a dataset-invariant manner. We achieve this by maximizing the ability of a CNN to regress the position of basal/apical slices within a single dataset, while minimizing the ability of a classifier to discriminate image features between different data sources. Our results show superior performance over state-of-the-art methods.

**Keywords:** Deep Learning · Dataset Invariance · Adversarial Learning · Ventricular Coverage Assessment · MRI.

## 1 Introduction

To obtain accurate and reliable volume and functional parameter measurements in CMR imaging studies, recognizing basal and apical slices for both ventricles is crucial. Unfortunately, current practice to detect basal/or apical slice positions is still carried out by visual inspection of experts on the image. This practice is costly, subjective, error prone, and time consuming [1]. Although significant

progress [14] has been made in automatic assessment of full LV coverage in cardiac MRI, to accurately measure volumes and functional parameters for both ventricles where the basal/apical slices are missing, methods to estimate the position of the missing slices are required [10]. Such methods would be critical to prompt the intervention of experts to correct problems in data measurements, or to trigger algorithms that can cope with missing data by, for instance, imputation [5] through image synthesis, or shape based extrapolation. This paves the way to “quality-aware image analysis” [13]. To the best of our knowledge, previous work regarding image quality control has focused solely on coverage detection of the LV, but not on missing slice position estimation.

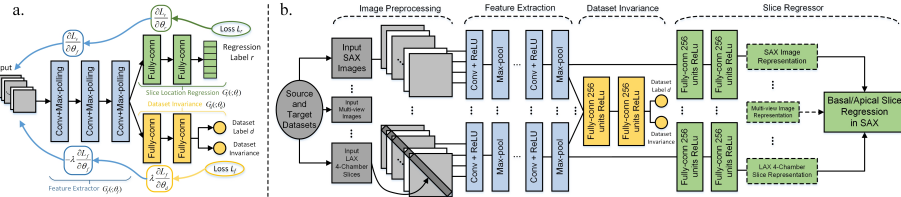
In medical image analysis, it is sometimes convenient or necessary to infer an image in one modality from another for image quality assessment purposes. One major challenge of basal/apical slice estimation for CMR comes from differences between data sources, which are tissue appearance and/or spatial resolution of images sourced from different physical acquisition principles or parameters. Such differences make it difficult to generalize algorithms trained on specific datasets to other data sources. This is problematic not only when the source and target datasets are different, but more so, when the target dataset contains no labels. In all such scenarios, it is highly desirable to learn a discriminative classifier or other predictor in the presence of a shift between training and test distributions, which is called *dataset invariance*. The general approach of achieving dataset adaptation has been explored under many facets. Among the existing cross-dataset learning works, dataset adaptation has been adopted for re-identification hoping labeled data from a source dataset can provide transferable identity-discriminative information for a target dataset. [7] explored the possibility of generating multimodal images from single-modality imagery. [8] [9] employed multi-task metric learning models to benefit the target task. However, these works are focused mainly on linear assumptions.

In this paper, we focus on the non-linear representations and analysis of short-axis (SA) and long-axis (LA) cine MRI for the detection and regression of the basal and apical slices of both ventricles in CMR volumes. To deal with the problem where there is no labeled data for a target dataset, and one hopes to transfer knowledge from a model trained on sufficient labeled data of a source dataset sharing the same feature space, but with a different marginal distribution we present these contributions: 1) We present a unified model (MDAL) for any cross-dataset basal/apical slice estimation problem in CMR volumes; 2) We integrate adversarial feature learning by building an end-to-end architecture of CNNs and transferring non-linear representations from a labeled source dataset to a target dataset where labels are non-existent. Our deep architecture effectively improves the adaptability of learning with data of different databases; 3) A multi-view image extension of the adversarial learning model is proposed and exploited. By making use of multi-view images acquired from short- and long-axis views, one can further improve and constrain the basal/apical slice position. We evaluate our method on three datasets and compare with state-of-the-art methods. Experimental results show the superior performance of our method compared to other approaches.

## 2 Methodology

### 2.1 Problem Formulation

The cross-dataset localization of basal or apical slices can be formulated as two tasks: (i) *Dataset Invariance*: given a set of 3D images  $\mathcal{X}^s = [\mathbf{X}_1^s, \dots, \mathbf{X}_N^s] \in \mathbb{R}^{m \times n \times z^s \times N^s}$  of modality  $\mathcal{M}_s$  in the source dataset, and  $\mathcal{X}^t = [\mathbf{X}_1^t, \dots, \mathbf{X}_N^t] \in \mathbb{R}^{m \times n \times z^t \times N^t}$  of modality  $\mathcal{M}_t$  in the target dataset.  $m, n$  are the dimensions of axial view of the image, and  $z^s$  and  $z^t$  denote the size of images along the  $z$ -axis, while  $N^s$  and  $N^t$  are the number of volumes in source and target datasets, respectively. Our goal is to build mappings between the source (training-time) and the target (test-time) datasets, that reduce the difference between the source and target data distributions; (ii) *Multi-view Slice Regression*: In this task, slice localization performance is enhanced by using multiple image stacks, *e.g.* SA and LA stacks, into a single regression task. Let  $\mathbf{X}^s = \{\mathbf{x}_i^s, r_i^s\}_{i=1}^{Z^s}$  and  $\mathbf{Y}^s = \{\mathbf{y}_i^s, r_i^s\}_{i=1}^{Z^s}$  be a labeled 3D CMR volume from source modality  $\mathcal{M}_s$  in short- and long-axis, respectively, and  $\mathbf{x}_b^s, \mathbf{x}_a^s$ , and  $\mathbf{y}_b^s, \mathbf{y}_a^s$  be the short-axis slices, and long-axis image patches of the basal and apical views; let  $\mathbf{X}^t = \{\mathbf{x}_i^t\}_{i=1}^{Z^t}$  and  $\mathbf{Y}^t = \{\mathbf{y}_i^t\}_{i=1}^{Z^t}$  represent an unlabeled sample from the target dataset in short- and long-axis,  $i$  represents the  $i^{\text{th}}$  slice and  $Z$  is the total number of CMR slices. Our goal is to learn the discriminative features from  $\mathbf{x}_b^s, \mathbf{x}_a^s$ , and  $\mathbf{y}_b^s, \mathbf{y}_a^s$  to localize the basal and apical slices in two axes for CMR volumes in the target dataset<sup>1</sup>. We use the labeled UK Biobank (UKBB) [11] cardiac MRI data cohort together with the MESA<sup>2</sup> and DETERMINE<sup>3</sup> datasets, and apply our method to cross-dataset basal and apical slice regression tasks.



**Fig. 1.** *a*: Schematic of our dataset-invariant adversarial network; *b*: System overview of our proposed dataset-invariant adversarial model with multi-view input channels for bi-ventricular coverage estimation in cardiac MRI. Each channel contains three conv layers, three max-pooling layers and two fully-connected layers. Additional dataset invariance net (yellow) includes two fully-connected layers. Kernel numbers in each conv layer are 16, 16 and 64 with sizes of  $7 \times 7$ ,  $13 \times 13$  and  $10 \times 10$ , respectively; filter sizes in each max-pooling layer are  $2 \times 2$ ,  $3 \times 3$  and  $2 \times 2$  with stride 2.

<sup>1</sup> **Notation:** Matrices and 3D images are written in bold uppercase (*e.g.*, image  $\mathbf{X}$ ,  $\mathbf{Y}$ ), vectors and vectorized 2D images in bold lowercase (*e.g.*, slice  $\mathbf{x}$ ,  $\mathbf{y}$ ) and scalars in lowercase (*e.g.*, slice position label  $r$ ).

<sup>2</sup> <http://www.cardiacatlas.org/studies/mesa/>

<sup>3</sup> <http://www.cardiacatlas.org/studies/determine/>

## 2.2 Multi-Input and Dataset-Invariant Adversarial Learning

Inspired by Adversarial Learning (AL)[6] and Dataset Adaptation (DA)[12] for cross-dataset transfer, we propose a Dataset-Invariant Adversarial Learning model, which extends the DA formulation into a AL strategy, and performs them jointly in a unified framework. We propose multi-view adversarial learning by creating multiple input channels (MC) from images which are re-sampled to the same spatial grid and visualize the same anatomy. An overview of our method is depicted in Fig. 1. Given two sets of slices  $\{\mathbf{x}_i^s\}_{i=1}^N$ ,  $\{\mathbf{y}_i^s\}_{i=1}^N$  with slice position labels  $\{r_i^s\}_{i=1}^N$  for training, to learn a model that can generalize well from one dataset to another, and is used both during training and test time to regress the basal/apical slice position, we optimize this objective in stages: 1) we optimize the label regression loss

$$\begin{aligned} \mathcal{L}_r^i &= \mathcal{L}_r(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s; \theta_f); \theta_r), r_i) \\ &= \sum_i \|r_i - G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s; \theta_f); \theta_r), r_i\|_2^2 + \frac{1}{2} \left( \|\theta_f\|_2^2 + \|\theta_r\|_2^2 \right), \end{aligned} \quad (1)$$

where  $\theta_f$  is the representation parameter of the neural network feature extractor, which corresponds to the feature extraction layers.  $\theta_r$  is the regression parameter of the slice regression net, which corresponds to the regression layers.  $r_i$  denotes the  $i^{\text{th}}$  slice position label.  $\theta_f$  and  $\theta_r$  are trained for the  $i^{\text{th}}$  image by using the labeled source data  $\{\mathbf{X}_i^s, \mathbf{r}_i^s\}_{i=1}^{N^s}$  and  $\{\mathbf{Y}_i^s, \mathbf{r}_i^s\}_{i=1}^{N^s}$ . 2) Since dataset adversarial learning satisfies a dataset adaptation mechanism, we minimize source and target representation distances through alternating *minimax* between two loss functions: one is the dataset discriminator loss

$$\begin{aligned} \mathcal{L}_d^i &= \mathcal{L}_d(G_{\text{disc}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d_i) \\ &= - \sum_i \mathbb{1}[o_d = d_i] \log(G_{\text{disc}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d_i), \end{aligned} \quad (2)$$

which classifies whether an image is drawn from the source or the target dataset.  $o_d$  indicates the output of the dataset classifier for the  $i^{\text{th}}$  image,  $\theta_d$  is the parameter used for the computation of the dataset prediction output of the network, which corresponds to the dataset invariance layers;  $d_i$  denotes the dataset that the example slice  $i$  is drawn from. The other is the source and target mapping invariant loss

$$\begin{aligned} \mathcal{L}_f^i &= \mathcal{L}_f(G_{\text{conf}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d_i) \\ &= - \sum_{d_i} \frac{1}{D} \log(G_{\text{conf}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d_i), \end{aligned} \quad (3)$$

which is optimized with a constrained adversarial objective by computing the cross entropy between the output predicted dataset labels, and a uniform distribution over dataset labels.  $D$  indicates the number of input channels. Our full method then optimizes the joint loss function

$$\begin{aligned} E(\theta_f, \theta_r, \theta_d) &= \mathcal{L}_r(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s; \theta_f); \theta_r), r) \\ &\quad + \lambda \mathcal{L}_f(G_{\text{conf}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d), \end{aligned} \quad (4)$$

where hyperparameter  $\lambda$  determines how strongly the dataset invariance influences the optimization;  $G_{\text{conv}}(\cdot)$  is a convolution layer function that maps an example into a new representation;  $G_{\text{sigm}}(\cdot)$  is a label prediction layer function;  $G_{\text{disc}}(\cdot)$  and  $G_{\text{conf}}(\cdot)$  are the dataset prediction and invariance layer functions.

### 2.3 Optimization

Similar to classical CNN learning methods, we propose to tackle the optimization problem with the stochastic gradient procedure, in which updates are made in the opposite direction of the gradient of Equation (4) to minimize parameters, and in the direction of the gradient to maximize other parameters [4]. We optimize the objective in the following stages.

**Optimizing the Label Regressor:** In adversarial adaptive methods, the main goal is to regularize the learning of the source and target mappings, so as to minimize the distance between the empirical source and target mapping distributions. If so then the source regression model can be directly applied to the target representations, eliminating the need to learn a separate target regressor. Training the neural network then leads to this optimization problem on the source dataset:

$$\arg \min_{\theta_f, \theta_r} \left\{ \frac{1}{N^s} \sum_{i=1}^{N^s} \mathcal{L}_r^i(G_{\text{sigm}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s; \theta_f); \theta_r), r_i) \right\}. \quad (5)$$

**Optimizing for Dataset Invariance:** This optimization corresponds to the true *minimax* objective ( $\mathcal{L}_d$  and  $\mathcal{L}_f$ ) for the dataset classifier parameters and the dataset invariant representation. The two losses stand in direct opposition to one another: learning a fully dataset invariant representation means the dataset classifier must do poorly, and learning an effective dataset classifier means that the representation is not dataset invariant. Rather than globally optimizing  $\theta_d$  and  $\theta_f$ , we instead perform iterative updates for these two objectives given the fixed parameters from the previous iteration:

$$\arg \min_{\theta_d} \left\{ -\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathcal{L}_d^i(G_{\text{disc}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d_i) \right\}, \quad (6)$$

$$\arg \max_{\theta_f} \left\{ -\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathcal{L}_f^i(G_{\text{conf}}(G_{\text{conv}}(\mathbf{x}_s, \mathbf{y}_s, \mathbf{x}_t, \mathbf{y}_t; \theta_f); \theta_d), d_i) \right\}, \quad (7)$$

where  $\mathcal{N} = N^s + N^t$  being the total number of samples. These losses are readily implemented in standard deep learning frameworks, and after setting learning rates properly so Equation (6) only updates  $\theta_d$  and (7) only updates  $\theta_f$ , the updates can be performed via standard backpropagation. Together, these updates ensure that we learn a representation that is dataset invariant.

### 2.4 Detection and Regression for Basal/Apical Slice Position

We denote  $\hat{\mathcal{H}}_t, \hat{\mathcal{G}}_t$  as extracted query features, and  $\hat{\mathcal{H}}_s, \hat{\mathcal{G}}_s$  as extracted basal/apical slice representations from SAX and LAX, respectively. In order to regress basal and apical slices according to query features, we compute the dissimilarity matrix  $\delta_{i,j}$  based on  $\hat{\mathcal{H}}_t, \hat{\mathcal{G}}_t$  and  $\hat{\mathcal{H}}_s, \hat{\mathcal{G}}_s$  using the volume’s inter-slice distance as:

$\delta_{i,j}(\hat{\mathcal{H}}_t, \hat{\mathcal{H}}_s, \hat{\mathcal{G}}_t, \hat{\mathcal{G}}_s) = \sqrt{(\hat{\mathcal{H}}_t^i - \hat{\mathcal{H}}_s^j)^2 + (\hat{\mathcal{G}}_t^i - \hat{\mathcal{G}}_s^j)^2}$ . Then, ranking can be carried out based on the ascending order of each row of the dissimilarity distance, *i.e.*, the lower the entry value  $\delta_{i,j}$  is, the closer the basal/apical slice and the query slice are.

### 3 Experiments and Analysis

**Data specifications:** Quality-scored CMR data is available for circa 5,000 volunteers of the UK Biobank imaging (UKBB) resource. Following visual inspection, manual annotation for SAX images was carried out with a simple 3-grade quality score [2]. 4,280 sequences correspond to quality score 1 for both ventricles, these had full coverage of the heart from base to apex and were the source datasets to construct the ground-truth classes for our experiments. Note that having full coverage should not be confused with having the top/bottom slices corresponding exactly to base/apex. Basal slices including the left ventricular outflow tract, pulmonary valve and right atrium, and apical slices with a visible ventricular cavity were labeled manually. The distance between the actual location of the basal/apical slice to other slices in the volume were used as training labels for the regression. We validated the proposed MDAL on three target datasets: UKBB, DETERMINE and MESA (protocols of the three datasets are shown in Table. 1). To prevent over-fitting due to insufficient target data, and to improve the detection rate of our algorithm, we employ data augmentation techniques to artificially enlarge the target datasets. For this purpose we chose a set of realistic rotations, scaling factors, and corresponding mirror images, and applied them to the MRI images. The set of rotations chosen were  $-45^\circ$  and  $45^\circ$ , and the scaling factors 0.75 and 1.25. This increased the number of training samples by a factor of eight. After data augmentation, we had 2400, and 2384 sequences for DETERMINE and MESA datasets, respectively. For evaluating of multi-view models, we defined two input channels, one for SAX images, and another for LAX (4-chamber) from the UKBB, MESA and DETERMINE. The LAX image information was extracted by collecting pixels values along the intersecting line between the 4-chamber view plane and corresponding short-axis plane over the cardiac cycle. We extracted 4 pixels above and below the two plane intersection. We embedded the constructed profile within a square image with zeros everywhere except the profile diagonal (see Fig. 1b bottom channel).

**Table 1.** 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
UKB	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
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
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

**Experimental set-up:** The architecture of our proposed method is shown in Fig. 1. To maximize the number of training samples from all datasets, while preventing biased learning of image features from a particular dataset and given that the number of samples from the UKBB is at least an order of magnitude larger than from MESA or DETERMINE, we augmented both the MESA and DETERMINE datasets, to match the resulting number of samples from the UKBB. This way our dataset classification task will not over-fit to anyone sam-

**Table 2.** The comparison of basal/apical slice detection accuracy (Mean  $\pm$  standard deviation) (%) between adaptation and non-adaptation methods, each with single (SAX)- and multi-view inputs (BS/AS indicate basal/apical slice detection accuracy). Best results are highlighted in bold.

Dataset	No dataset adaptation (BS/AS)		With dataset Adaptation (BS/AS)	
	Single-view [14]	Multi-view [14]	Single-view [4]	Multi-view (Ours)
UKBB	79.0 $\pm$ 0.2/76.2 $\pm$ 0.3	<b>89.2<math>\pm</math>0.1/92.4<math>\pm</math>0.2</b>	78.2 $\pm$ 0.2/75.4 $\pm$ 0.3	88.7 $\pm$ 0.1/91.4 $\pm$ 0.3
MESA	31.6 $\pm$ 0.3/35.1 $\pm$ 0.1	61.5 $\pm$ 0.2/68.3 $\pm$ 0.4	74.2 $\pm$ 0.2/72.9 $\pm$ 0.4	<b>87.1<math>\pm</math>0.3/90.2<math>\pm</math>0.2</b>
DETERMINE	48.3 $\pm$ 0.2/51.1 $\pm$ 0.3	75.6 $\pm$ 0.3/78.4 $\pm$ 0.3	77.2 $\pm$ 0.3/76.5 $\pm$ 0.2	<b>89.0<math>\pm</math>0.2/91.2<math>\pm</math>0.2</b>

**Table 3.** Regression error comparison between adaptation and non-adaptation methods, each with single (SAX)- and multi-view inputs for cardiac SAX slice position regression in terms of MAE (Mean  $\pm$  standard deviation)(mm)(BS/AS indicate basal/apical slice regression errors). Best results are highlighted in bold.

Dataset	No dataset adaptation (BS/AS)		With dataset adaptation (BS/AS)	
	Single-view [14]	Multi-view [14]	Single-view [4]	Multi-view (Ours)
UKBB	4.32 $\pm$ 1.6/5.73 $\pm$ 1.9	<b>3.42<math>\pm</math>1.1/3.98<math>\pm</math>1.7</b>	5.13 $\pm$ 2.1/6.33 $\pm$ 2.3	3.64 $\pm$ 1.9/4.02 $\pm$ 2.0
MESA	7.78 $\pm$ 2.0/8.34 $\pm$ 2.4	6.47 $\pm$ 1.7/6.83 $\pm$ 1.4	4.81 $\pm$ 1.0/5.73 $\pm$ 1.5	<b>3.98<math>\pm</math>1.1/4.07<math>\pm</math>1.3</b>
DETERMINE	6.43 $\pm$ 1.9/6.81 $\pm$ 2.0	6.01 $\pm$ 1.3/6.17 $\pm$ 1.4	4.73 $\pm$ 1.6/4.81 $\pm$ 1.3	<b>4.24<math>\pm</math>1.0/4.45<math>\pm</math>1.3</b>

ple. Our MDAL method processes images with small blocks ( $120 \times 120$ ), which are crop-centered on the images to extract specific regions of interest. The experiments here reported were conducted using the ConvNet library [3] on an Intel Xeon E5-1620 v3 @3.50GHz machine running Windows 10 with 32GB RAM and Nvidia Quadro K620 GPU. We optimize the network using a learning rate  $\mu$  of 0.001 and set the hyper-parameter parameter  $\lambda$  to be 0.01, respectively. To evaluate the detection process, we measure classification accuracy, and to evaluate the regression error between the predicted position and the ground truth, we use the Mean Absolute Error (MAE).

**Results:** We evaluate the performance of the multi-view basal/apical slice detection and regression tasks with and without dataset invariance (adaptation vs non-adaptation), by transferring object regressors from the UKBB to MESA and DETERMINE. To evaluate performance on MESA and DETERMINE, we manually generated annotations as follows: we checked one slice above and below the detected basal slice to confirm the slice is the basal and record true or false, ditto for apex. We chose the CNN architecture in [14] for single- and multi-view metrics with non-adaptation, and the GTSRB architecture in [4] for single-view adaption method. Table 2 shows the detection accuracy for basal/apical slice of the adaptation and non-adaptation from single and multi-view. For both test datasets, the best improvements are the result of combining both of these features. For MESA the detection accuracy was increased by 64%, and for DETERMINE best improvements are of 44% (right-most column). Table 3 shows the average regression errors of slice locations in millimeter (*mm*). Even without using the multi-input channels, our dataset invariance framework is able to reduce the slice localization error to less than half the average slice spacing found on our test datasets, *i.e.*,  $< 5mm$ . With multi-view we reduced the localization errors to 4.24 and 4.45mm on average for both basal/apical slices. All the experiences are significantly different at  $p < 0.05$ .

## 4 Conclusion

In this paper, we have proposed a Multi-Input and Dataset-Invariant Adversarial Learning (MDAL) framework capable of learning a common image representation, and using it to detect and localize basal and apical CMR slices, we achieve this by: first, using a Dataset-Invariant Adversarial Learning (DIAL) model to fit the joint distribution over the images from different datasets with a minimax game. Second, extending the DIAL model to handle multiple view input scenarios thereby obtaining better results for Left and Right-Ventricular coverage estimation in Cardiac MRI. And third, by introducing a regressor network able to predict the location of basal/apical slices. We evaluated our framework on two large datasets MESA and DETERMINE and found that our approach significantly outperforms state-of-the-art non-dataset-adaptive and single-input methods. Finally, Our MDAL framework can be easily generalized to any anatomical structure or image modality.

## References

1. Attili, A.K., Schuster, A., Nagel, E., Reiber, J.H., van der Geest, R.J.: Quantification in Cardiac MRI: Advances in Image Acquisition and Processing. *Int J Cardiovasc Imaging* 26(1), 27–40 (2010)
2. Carapella, V., et al.: Towards the semantic enrichment of free-text annotation of image quality assessment for UK Biobank cardiac Cine MRI scans. In: *Deep Learning and Data Labeling for Medical Applications*, pp. 238–248. Springer (2016)
3. Demyanov, S.: ConvNet library for Matlab [Online]. <https://github.com/sdemyanov/ConvNet>, accessed 15 Oct. 2017
4. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., et al.: Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* 17(1), 2096–2030 (2016)
5. García-Laencina, P.J., Sancho-Gómez, J.L., Figueiras-Vidal, A.R.: Pattern classification with missing data: a review. *Neural Comput Appl* 19(2), 263–282 (2010)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., et al.: Generative adversarial nets. In: *Adv Neural Inf Process Syst.* pp. 2672–2680 (2014)
7. Huang, Y., Shao, L., Frangi, A.F.: Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. *IEEE Conference on CVPR* pp. 6070–6079 (2017)
8. Lisanti, G., Masi, I., Bagdanov, A.D., Del Bimbo, A.: Person re-identification by iterative re-weighted sparse ranking. *IEEE TPAMI* 37(8), 1629–1642 (2015)
9. Ma, L., Yang, X., Tao, D.: Person re-identification over camera networks using multi-task distance metric learning. *IEEE TIP* 23(8), 3656–3670 (2014)
10. Paknezhad, M., Marchesseau, S., Brown, M.S.: Automatic basal slice detection for cardiac analysis. *J Med Imaging* 3(3), 034004–034004 (2016)
11. Petersen, S.E., Matthews, P.M., Francis, J.M., et al.: UK Biobanks Cardiovascular Magnetic Resonance Protocol. *J Cardiovasc Magn Reson* 18(1), 1 (2016)
12. Sharmanska, V., Quadrianto, N.: Learning from the mistakes of others: Matching errors in cross-dataset learning. *IEEE Conference on CVPR* pp. 3967–3975 (2016)
13. Wang, Z., Wu, G., Sheikh, H.R., Simoncelli, E.P., Yang, E.H., Bovik, A.C.: Quality-aware images. *IEEE TIP* 15(6), 1680–1689 (2006)
14. Zhang, L., Gooya, A., Petersen, S.E., Medrano-Gracia, P., Frangi, A.F., et al.: Automated quality assessment of cardiac MR images using convolutional neural networks. In: *MICCAI Workshop on SASHIMI*. pp. 138–145. Springer (2016)