

# Adaptive Hierarchical Dual Consistency for Semi-Supervised Left Atrium Segmentation on Cross-Domain Data

Jun Chen, Heye Zhang, Member, IEEE, Raad Mohiaddin, Tom Wong, David Firmin, Jennifer Keegan, and Guang Yang, Senior Member, IEEE

**Abstract**— Semi-supervised learning provides great significance in left atrium (LA) segmentation model learning with insufficient labelled data. Generalising semi-supervised learning to cross-domain data is of high importance to further improve model robustness. However, the widely existing distribution difference and sample mismatch between different data domains hinder the generalisation of semi-supervised learning. In this study, we alleviate these problems by proposing an *Adaptive Hierarchical Dual Consistency* (AHDC) for the semi-supervised LA segmentation on cross-domain data. The AHDC mainly consists of a Bidirectional Adversarial Inference module (BAI) and a Hierarchical Dual Consistency learning module (HDC). The BAI overcomes the difference of distributions and the sample mismatch between two different domains. It mainly learns two mapping networks adversarially to obtain two matched domains through mutual adaptation. The HDC investigates a hierarchical dual learning paradigm for cross-domain semi-supervised segmentation based on the obtained matched domains. It mainly builds two dual-modelling networks for mining the complementary information in both intra-domain and inter-domain. For the intra-domain learning, a consistency constraint is applied to the dual-modelling targets to exploit the complementary modelling information. For the inter-domain learning, a consistency constraint is applied to the LAs modelled by two dual-modelling networks to exploit the complementary knowledge among different data domains. We demonstrated the performance of our proposed AHDC on four 3D late gadolinium enhancement cardiac MR (LGE-CMR) datasets from

different centres and a 3D CT dataset. Compared to other state-of-the-art methods, our proposed AHDC achieved higher segmentation accuracy, which indicated its capability in the cross-domain semi-supervised LA segmentation.

**Index Terms**— Semi-supervised Learning; Cross-domain Study; Hierarchical Dual Consistency; Bidirectional Adversarial Inference.

## I. INTRODUCTION

SEMI-SUPERVISED learning provides great significance in left atrium (LA) segmentation model learning with insufficient labelled data. Automated and accurate LA segmentation is a crucial task to aid the diagnosis and treatment for the patients with atrial fibrillation (AF) [1]–[4]. Deep learning based approaches have great potential for the LA segmentation [5], [6]. However, it is expensive and laborious to annotate large amounts of data by experienced experts for training an accurate LA segmentation model based on deep learning [7]. Since semi-supervised learning can alleviate the need for the labelled data by effectively exploiting the unlabelled data to learn deep models [8]. Semi-supervised learning is able to overcome the insufficient labelled data for advancing the accurate LA segmentation, benefiting the subsequent diagnosis and treatment for the patients with AF.

Generalising semi-supervised learning to cross-domain data for the LA segmentation is of high importance to improve model robustness. Semi-supervised learning aims to mine effective hidden information from unlabelled data to support model learning [9]. Because of the noise interference and the limited collection capabilities of data sources, a single data domain can not always provide sufficient high-quality unlabelled data and abundant data characteristics for robust semi-supervised LA segmentation. For example, the single data domain is usually subject to the limited LA varieties of contrast, shape and texture for robust model learning. Compared to the single data domain, cross-domain data not only can provide more available high-quality data, but also can provide complementary domain information and more comprehensive data characteristics to describe the LA of interest [10]. Therefore, it is important to effectively ensemble cross-domain data for robust semi-supervised LA segmentation.

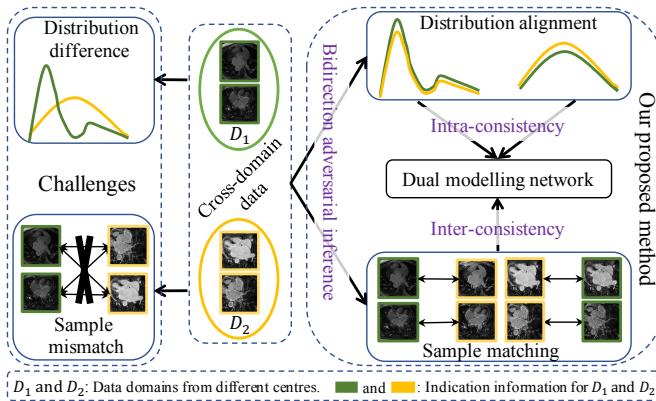
This study was supported in part by the Key-Area Research and Development Program of Guangdong Province (2019B010110001), the Key Program for International Cooperation Projects of Guangdong Province (2018A050506031), the National Youth Talent Support Program (RC2020-01), the Guangdong Natural Science Funds for Distinguished Young Scholar (2019B151502031), the Natural Science Foundation of Guangdong Province (2020B1515120061), and the National Natural Science Foundation of China (61771464, U1801265, U1908211); in part by the British Heart Foundation (Project Number: TG/18/5/34111, PG/16/78/32402), the European Research Council Innovative Medicines Initiative (DRAGON, H2020-JTI-IMI2 101005122), the AI for Health Imaging Award (CHAIMELEON, H2020-SC1-FA-DTS-2019-1 952172), and the UK Research and Innovation Future Leaders Fellowship (MR/V023799/1).

H. Zhang and G. Yang are corresponding authors (e-mail: zhangheye@mail.sysu.edu.cn; g.yang@imperial.ac.uk).

J. Chen and H. Zhang are with the School of Biomedical Engineering, Sun Yat-sen University, Shenzhen, Guangdong 518107, P.R. China.

R. Mohiaddin, T. Wong, D. Firmin, J. Keegan and G. Yang are with the Cardiovascular Research Centre, Royal Brompton Hospital, SW3 6NP, London, U.K and the National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, U.K.

However, generalising semi-supervised to cross-domain data is difficult due to the difference of distributions and the sample mismatch as shown in Fig. 1: (1) The difference of cross-domain data distributions. Semi-supervised learning with the generative model, low-density separation and graph-based method can work but relies on the consistent data distribution under certain model assumptions including smoothness assumption, cluster assumption or manifold assumption [9]. Performance degradation of the semi-supervised model may occur whenever the assumptions adopted for a particular task do not match the characteristics of the data distribution [9]. In the real world, cross-domain data collected from different sources exhibit heterogeneous properties [11], which can lead to the difference in distributions. For example, in medical image analysis, because of the different subject groups, scanners, or scanning protocols, the distributions of cross-domain data are different [12]. Therefore, generalising semi-supervised learning to cross-domain data directly is not trivial. (2) Sample mismatch of cross-domain data. Semi-supervised learning with the disagreement-based method requires matched samples from different domains, where the information of different domains is regarded as the different characteristics of matched samples [13]. Since the collection of cross-domain data is independent, the samples in different domains are not matched. This restricts the cross-domain generalisation of semi-supervised learning.



**Fig. 1:** Our proposed adaptive hierarchical dual consistency overcomes the difference of data distribution and sample mismatch in different domains for the cross-domain semi-supervised segmentation.

In order to overcome the issues mentioned above, we propose an *Adaptive Hierarchical Dual Consistency* framework called **AHDC** for semi-supervised LA segmentation on cross-domain data as shown in Fig. 1. The AHDC consists of two modules: (1) A Bidirectional Adversarial Inference module (BAI), which performs the mutual domain adaptation to align distributions and match samples for two different data domains. The adapted domains and two corresponding source domains are merged to obtain two matched domains. The obtained matched domains not only expand the number of data in a specific source domain, but also learns complementary representation for the samples in the specific source domain.

(2) A Hierarchical Dual Consistency learning module (HDC), which performs a hierarchical semi-supervised segmentation with dual consistency on the obtained matched domains. The HDC builds two dual-modelling networks applied to the matched domains for mining the complementary information in both intra-domain and inter-domain. Within a specific domain, the segmentation task is represented as global modelling and local modelling. Then we perform a consistency between the complementary modelling LAs for intra-domain semi-supervised learning. For the inter-domain, we build a consistency between the outputs of dual-modelling networks estimated from different domains to exploit the complementary domain information.

Our main contributions are summarised as follows:

- We propose a semi-supervised LA segmentation framework for generalising across domains. It provides a solution for generalising semi-supervised LA segmentation to cross-domain data with effectiveness on both different distributions and mismatched samples.
- We propose a paradigm of hierarchical dual consistency learning to mine the effective information in both inter-domain and intra-domain. It explicitly enforcing consistency under complementary information.
- We have conducted comprehensive experiments on four 3D MR datasets from different centres and one 3D CT dataset. The experiment results demonstrated the feasibility and the superiority of our proposed cross-domain semi-supervised segmentation framework.

## II. RELATED WORK

### A. Domain Adaptation

Domain adaptation, which aims to overcome the distribution difference of different domains, has drawn great attention in computer vision [14]. Because generative adversarial network (GAN) has great superiority in capturing data distribution, it has been widely used in domain adaptation for aligning distributions of different domains [15]–[19]. There are different GAN based structures for achieving domain adaptation. For the domain adaptation with a single direction, GAN usually leverages a generator and a discriminator to improve the distribution of the source domain to approximate it to the distribution of the target domain by adversarial learning. To focus on the high-resolution image with emphasis on pixel-level reconstruction, Pix2pixHD extends conditional GANs to leverage a decomposed generator and three multi-scale discriminators to achieve domain adaptation [20]. For the domain adaptation with bi-direction, CycleGAN [21], DualGAN [22] and DiscoGAN [23] concatenate two generators with two discriminators to ensure two cyclic consistency for the bidirectional domain adaptation of two different domains. ALI [24] and BiGAN [25] employ two generators and a discriminator to match joint distribution for different domains. However, ALI and BiGAN do not focus on pixel-level reconstruction, thus cannot effectively capture the position, colour, and style of targets. ALICE extends the ALI to exploit cycle-consistency to focus on pixel-level reconstruction for the target domain [26]. It also proposes to enforce cycle-consistency

using fully adversarial learning with an extra discriminator. Our used domain adaptation method is based on the ALICE framework. We extended it to focus on bidirectional pixel-level reconstruction for two domains simultaneously. In order to reduce computing resources and difficulty of training while using fully adversarial learning, we adopt the explicit cycle-consistency, thus exploiting two generators and a discriminator for bidirectional domain adaptation with pixel-level reconstruction.

### B. Semi-supervised Learning

Semi-supervised learning alleviates the problem of the lack of labelled data. Here we only discuss related consistency-based and disagreement-based semi-supervised learning. More information about semi-supervised learning can be found in [9]. The consistency-based methods constrain the prediction consistency under different perturbations and ensembles. For example, the  $\Pi$  model enforces the prediction consistency under the input perturbations with different Gaussian noise and the model perturbation with dropout operation [27]. Unsupervised data augmentation (UDA) replaces the traditional noise perturbations with high-quality data augmentations (e.g., RandAugment, Back-translation and TF-IDF) to improve consistency learning [28]. FixMatch uses a separate weak augmentation and a strong augmentation on input data for consistency regularisation [29]. In contrast to these methods, Temporal Ensembling (TE) penalises the inconsistency between the current prediction and the integration of previous predictions based on an exponential moving average (EMA) [27]. Compared to the TE, the Mean Teacher proposes to average the weights of a base model [30]. However, they need multiple reasoning processes to provide predictions for consistency learning, thus being subject to the computational cost.

The disagreement-based semi-supervised learning exploits the disagreement of predictions from multiple task learners during the learning process [13] including co-training and co-regularisation. Co-training leverages two sufficient and redundant views of data to train two task models for annotating the unlabelled data. Then the unlabelled data with high prediction confidence is added to the training set for further improving the model [31], [32]. Co-regularisation tries to directly minimise the prediction disagreement of unlabelled samples on different views [33].

## III. METHOD

### A. Overview

The overview of our proposed AHDC framework is illustrated in Fig. 2. The notations are summarised in TABLE I. The AHDC framework consists of two modules: a BAI module and a HDC module. Given two different data domains denoted by  $D_1$  and  $D_2$ .  $D_1$  contains both labelled data  $D_1^l$  and unlabelled data  $D_1^u$ , where  $D_1^l = \{(x_1^l)^i, y^i\}_{i=1}^{n_1}$  with  $n_1$  labelled samples and  $D_1^u = \{(x_1^u)^i\}_{i=n_1+1}^{n_1+n_2}$  with  $n_2$  unlabelled samples, respectively. The  $D_2$  only contains unlabelled data denoted as  $D_2^u = \{(x_2^u)^i\}_{i=1}^{m_1}$  with  $m_1$  unlabelled samples. The BAI module employs two mapping networks of  $G_1$  and  $G_2$  to generate complementary domains by adapting  $D_1$  and

TABLE I  
SUMMARY OF NOTATIONS

| Notion   | Definition  | Notion  | Definition  |
|--|---|---|---|
| $D_1$  | Domain from source1   | $D_2$   | Domain from source2   |
| $D_{1t2}$  | Domain adapted from $D_1$ to $D_2$  | $D_{2t1}$   | Domain adapted from $D_2$ to $D_1$  |
| $D_{p1}$   | $D_1 \cup D_{2t1}$  | $D_{p2}$  | $D_2 \cup D_{1t2}$  |
| $D^l, D^u$   | Labelled domain, Unlabelled domain  | $G_1, G_2$  | Mutual mapping nets of $D_1$ and $D_2$                                      |
| $S_1 = \{S_{l1}, S_{g1}\}$ ,<br>$S_2 = \{S_{l2}, S_{g2}\}$ | Dual-modelling nets<br>{local net, global net}                              | $T$   | Discriminator   |
| $x_1, x_2, x_{1t2},$<br>$x_{2t1}, x_{p1}, x_{p2}$          | Images from $D_1, D_2,$<br>$D_{1t2}, D_{2t1}, D_{p1}, D_{p2}$               | $\tilde{y}_{l1}, \tilde{y}_{g1},$<br>$\tilde{y}_{l2}, \tilde{y}_{g2}$ | Estimated LAs from<br>$S_{l1}, S_{g1}, S_{l2}, S_{g2}$                      |
| $\hat{x}_1, \hat{x}_2$                                     | Reconstructions of $x_1, x_2$   | $x^l, x^u$  | Labelled data, unlabelled data  |
| $j(x_{p1}, x_{p2})$  | Joint distribution of $D_{p1}, D_{p2}$                                      | $y$   | Ground truth  |
| $p(x_1), q(x_2),$<br>$p(x_{p1}), q(x_{p2})$                | Marginal distributions of $D_1, D_2, D_{p1}, D_{p2}$                        | $p_{\varphi_1}(x_2 x_1)$<br>$q_{\varphi_2}(x_1 x_2)$                  | Parameterised conditional distributions                                     |
| $\varphi_1, \varphi_2$                                     | Params of $G_1, G_2$  | $\psi_1$  | Param of $T$  |
| $\theta_1 = \{\theta_1^f, \theta_1^l, \theta_1^g\}$        | Param of $S_1$ in the modules of feature, local-modelling, global-modelling | $\theta_2 = \{\theta_2^f, \theta_2^l, \theta_2^g\}$                   | Param of $S_2$ in the modules of feature, local-modelling, global-modelling |
| $L(\cdot)$   | Loss function   | $\lambda$   | Weight Param  |

$D_2$  to each other, where the domain adapted from  $D_1$  to  $D_2$  is denoted as  $D_{1t2}$  while the domain adapted from  $D_2$  to  $D_1$  is denoted as  $D_{2t1}$ . Then the targeted domains ( $D_1$  and  $D_2$ ) and the corresponding adapted domains ( $D_{2t1}$  and  $D_{1t2}$ ) merge to form two matched domains of  $D_{p1}$  and  $D_{p2}$ . Finally, two dual-modelling networks of  $S_1 = \{S_{l1}, S_{g1}\}$  and  $S_2 = \{S_{l2}, S_{g2}\}$  are fed with matched samples sampled from  $D_{p1}$  and  $D_{p2}$  to predict LAs, where the LAs predicted by the local modelling  $S_{l1}$  and the global modelling  $S_{g1}$  are denoted as  $\tilde{y}_{l1}$  and  $\tilde{y}_{g1}$  while the LAs predicted by the local modelling  $S_{l2}$  and the global modelling  $S_{g2}$  are denoted as  $\tilde{y}_{l2}$  and  $\tilde{y}_{g2}$ , respectively.

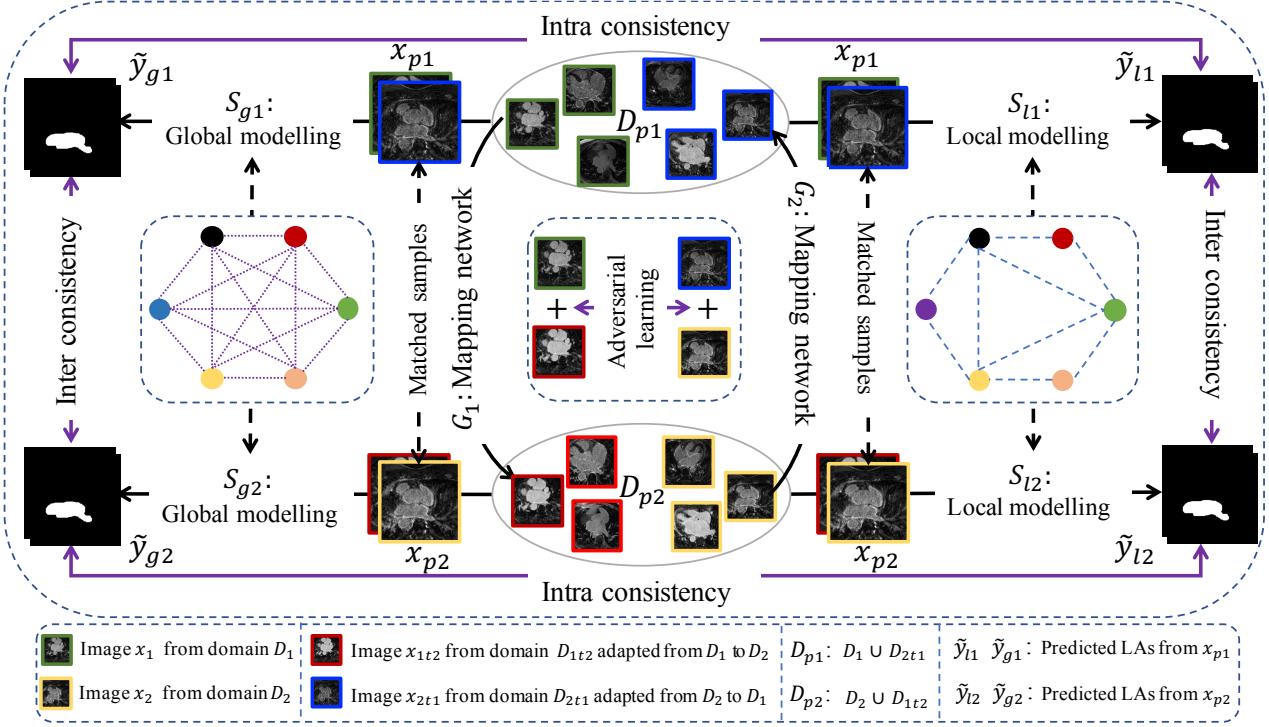
### B. Bidirectional Adversarial Inference for Distribution Alignment and Sample Matching.

Consider a  $D_1$  to  $D_2$  domain mapping network  $G_1 : x_1 \rightarrow x_2$ . Meanwhile, consider a  $D_2$  to  $D_1$  domain mapping network  $G_2 : x_2 \rightarrow x_1$ . We denote two domain marginal distributions of  $D_1$  and  $D_2$  as  $p(x_1)$  and  $q(x_2)$ . One domain can be inferred based on the other using parameterised conditional distributions,  $p_{\varphi_1}(x_2|x_1)$  and  $q_{\varphi_2}(x_1|x_2)$ , where  $\varphi_1$  and  $\varphi_2$  denote the parameters of two distributions. Then, we have the joint distributions of  $p_{\varphi_1}(x_1, x_2) = p_{\varphi_1}(x_2|x_1)p(x_1)$  and  $q_{\varphi_2}(x_1, x_2) = q_{\varphi_2}(x_1|x_2)q(x_2)$ . We aims to match  $p_{\varphi_1}(x_2) = \int p_{\varphi_1}(x_2, x_1)dx_1$  to  $q(x_2)$  and match  $q_{\varphi_2}(x_1) = \int q_{\varphi_2}(x_1, x_2)dx_2$  to  $p(x_1)$  by matching  $p_{\varphi_1}(x_1, x_2)$  and  $q_{\varphi_2}(x_1, x_2)$ . Then we use a discriminator network  $T_{\psi_1}(x_1, x_2)$  parameterised using  $\psi_1$  to penalise mismatches in the joint distributions of  $p_{\varphi_1}(x_1, x_2)$  and  $q_{\varphi_2}(x_1, x_2)$ . Specifically, we consider following objectives:

$$\begin{aligned} & \min_{\varphi_1, \varphi_2} \max_{\psi_1} O^d(\varphi_1, \varphi_2, \psi_1) \\ &= E_{(x_1, x_2) \sim p_{\varphi_1}(x_1, x_2)} [\log \sigma(T_{\psi_1}(x_1, x_2))] \\ &+ E_{(x_1, x_2) \sim q_{\varphi_2}(x_1, x_2)} [1 - \log \sigma(T_{\psi_1}(x_1, x_2))] \end{aligned} \quad (1)$$

where the  $\sigma(\cdot)$  denotes the sigmoid function.

Intuitively, if equation (1) is achieved,  $p_{\varphi_1}(x_1, x_2)$  and  $q_{\varphi_2}(x_1, x_2)$  match each other, which not only implies that  $p_{\varphi_1}(x_2)$  and  $q(x_2)$  match each other, but also implies that  $p_{\varphi_2}(x_1)$  and  $p(x_1)$  match each other. However, the relationship



**Fig. 2:** Overview of our proposed AHDC framework for cross-domain semi-supervised segmentation. The framework consists of a bidirectional adversarial inference (BAI) module and a hierarchical dual consistency learning (HDC) module. The BAI module employs two mapping networks to perform a mutual adaptation of two different domains of  $D_1$  and  $D_2$  to obtain matched domains of  $D_{p1}$  and  $D_{p2}$ . The HDC module applies two dual-modelling networks to the matched domains for performing semi-supervised segmentation tasks. Each dual-modelling network contains a global-modelling branch ( $S_{g1}/S_{g2}$ ) used to capture the global correlation of feature maps to estimate LA ( $\tilde{y}_{g1}/\tilde{y}_{g2}$ ), and a local-modelling branch ( $S_{l1}/S_{l2}$ ) used to capture the local correlation of feature maps to estimate LA ( $\tilde{y}_{l1}/\tilde{y}_{l2}$ ). In intra-domain, a consistency is performed between  $\tilde{y}_{l1}/\tilde{y}_{l2}$  and  $\tilde{y}_{g1}/\tilde{y}_{g2}$  estimated by complementary modellings, respectively. In inter-domain, a consistency is performed between  $\tilde{y}_{l1}/\tilde{y}_{g1}$  and  $\tilde{y}_{l2}/\tilde{y}_{g2}$  estimated by complementary domain networks, respectively.

253 between random variables  $x_1$  and  $x_2$  is not specified or  
 254 constrained by equation (1). In order to obtain paired samples,  
 255 according to [26], we extend the conditional entropies from  
 256 single constraint to bi-direction constraints ( $H(x_1|x_2)$  and  
 257  $H(x_2|x_1)$ ), which imposes constraints on the conditionals  
 258  $p_{\varphi_1}(x_2|x_1)$  and  $q_{\varphi_2}(x_1|x_2)$ , simultaneously. Because there is  
 259 no explicit distributions to compute the conditional entropies.  
 260 According to [26], we bound the conditional entropies using  
 261 the cycle-consistency ( $L^{x_1 \rightarrow \hat{x}_1}(\varphi_1, \varphi_2)$  and  $L^{x_2 \rightarrow \hat{x}_2}(\varphi_1, \varphi_2)$ ):

$$\begin{aligned}
 & H(x_1|x_2) \\
 &= -E_{x_1 \sim p(x_1), x_2 \sim p_{\varphi_1}(x_2|x_1)}[\log p_{\varphi_1}(x_1|x_2)] \\
 &= -E_{x_1 \sim p(x_1), x_2 \sim p_{\varphi_1}(x_2|x_1)}[\log q_{\varphi_2}(x_1|x_2)] \\
 &\quad - E_{x_1 \sim p(x_1), x_2 \sim p_{\varphi_1}(x_2|x_1)}[\log p_{\varphi_1}(x_1|x_2) - \log q_{\varphi_2}(x_1|x_2)] \\
 &= -E_{x_1 \sim p(x_1), x_2 \sim p_{\varphi_1}(x_2|x_1)}[\log q_{\varphi_2}(x_1|x_2)] \\
 &\quad - E_{q_{\varphi_2}(x_2)}[KL(p_{\varphi_1}(x_1|x_2) || q_{\varphi_2}(x_1|x_2))] \\
 &\leq -E_{x_1 \sim p(x_1), x_2 \sim p_{\varphi_1}(x_2|x_1)}[\log q_{\varphi_2}(x_1|x_2)] = L^{x_1 \rightarrow \hat{x}_1}(\varphi_1, \varphi_2)
 \end{aligned} \tag{2}$$

Similarly,

$$\begin{aligned}
 & H(x_2|x_1) \\
 &= -E_{x_2 \sim q(x_2), x_1 \sim q_{\varphi_2}(x_1|x_2)}[\log q_{\varphi_2}(x_2|x_1)] \\
 &\leq -E_{x_2 \sim q(x_2), x_1 \sim q_{\varphi_2}(x_1|x_2)}[\log p_{\varphi_1}(x_2|x_1)] = L^{x_2 \rightarrow \hat{x}_2}(\varphi_1, \varphi_2)
 \end{aligned} \tag{3}$$

where the  $\hat{x}_1$  and  $\hat{x}_2$  are denoted as the reconstructions of  $x_1$  and  $x_2$ . KL denotes the Kullback-Leibler divergence. According to the equations of (2) and (3), on the one hand, we have a function  $G_3 : x_1 \rightarrow \hat{x}_1$  defined by  $G_3 = G_1 \circ G_2$ , which first generates  $x_2$  from  $x_1$  based on  $G_1$ , then  $G_2$  produces  $\hat{x}_1$  from generated  $x_2$ . On the other hand, we also have a function  $G_4 : x_2 \rightarrow \hat{x}_2$  defined by  $G_4 = G_2 \circ G_1$ , which first generates  $x_1$  from  $x_2$  based on  $G_2$ , then  $G_1$  produces  $\hat{x}_2$  from generated  $x_1$ . In contrast to the fully adversarial training for solving  $L^{x_1 \rightarrow \hat{x}_1}(\varphi_1, \varphi_2)$  and  $L^{x_2 \rightarrow \hat{x}_2}(\varphi_1, \varphi_2)$ , we employ the reconstruction loss to reduce the difficulty of model training. Specifically, we consider following object:

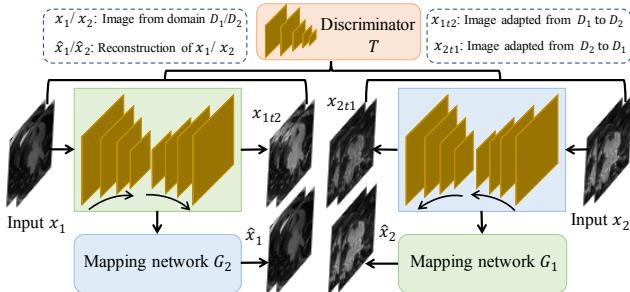
$$\begin{aligned}
 & \min_{\varphi_1, \varphi_2} O^{x_1 \rightarrow \hat{x}_1}(\varphi_1, \varphi_2) \\
 &= E_{\hat{x}_1 \sim q_{\varphi_2}(\hat{x}_1|x_2), x_1 \sim p_{\varphi_1}(x_1|x_2)} L_{mae}(x_1, \hat{x}_1)
 \end{aligned} \tag{4}$$

$$\begin{aligned} & \min_{\varphi_1, \varphi_2} O^{x_2 \rightarrow \hat{x}_2}(\varphi_1, \varphi_2) \\ & = E_{\hat{x}_2 \sim p_{\varphi_1}(\hat{x}_2|x_1), x_1 \sim q_{\varphi_2}(x_1|x_2)} L_{mae}(x_2, \hat{x}_2) \end{aligned} \quad (5)$$

where the  $L_{mae}(\cdot)$  denotes the mean absolute error. Finally, we have the following object for BAI:

$$\begin{aligned} & \min_{\varphi_1, \varphi_2} \max_{\psi_1} \lambda_d O^d(\varphi_1, \varphi_2, \psi_1) \\ & + \lambda_r O^{x_1 \rightarrow \hat{x}_1}(\varphi_1, \varphi_2) \\ & + \lambda_r O^{x_2 \rightarrow \hat{x}_2}(\varphi_1, \varphi_2) \end{aligned} \quad (6)$$

where  $\lambda_d$  and  $\lambda_r$  are hyperparameters to balance the adversarial loss and the reconstruction loss.



**Fig. 3:** Structure of bidirectional adversarial inference network. The mapping network  $G_1$  and the mapping network  $G_2$  have the same structure.

### C. Hierarchical Dual Consistency for Semi-supervised Segmentation

The BAI makes the cross-domain data adapt to each other to produce matched domains. In detail, the domain  $D_1$  adapted to source  $D_2$  is denoted as  $D_{1t2} = D_{1t2}^l \cup D_{1t2}^u$ , where  $D_{1t2}^l = \{(x_{1t2}^l)^i, y^i\}_{i=1}^{n_1}$  with  $n_1$  labelled samples and  $D_{1t2}^u = \{(x_{1t2}^u)^i\}_{i=n_1+1}^{n_1+n_2}$  with  $n_2$  unlabelled samples. The domain  $D_2$  adapted to source  $D_1$  is denoted as  $D_{2t1} = D_{2t1}^u = \{(x_{2t1}^u)^i\}_{i=1}^m$  with  $m$  unlabelled samples. Then we merge two source domains and two adapted domains to obtain the matched domains of  $D_{p1}$  and  $D_{p2}$ . The  $D_{p1} = D_1 \cup D_{2t1} = D_1^l \cup D_1^u \cup D_{2t1}^u = D_{p1}^l \cup D_{p1}^u$ , where  $D_{p1}^l = \{(x_{p1}^l)^i, y^i\}_{i=1}^{n_1}$  with  $n_1$  labelled samples and  $D_{p1}^u = \{(x_{p1}^u)^i\}_{i=n_1+1}^{n_1+n_2+m}$  with  $n_2 + m$  unlabelled samples. The  $D_{p2} = D_2 \cup D_{1t2} = D_{1t2}^l \cup D_{1t2}^u \cup D_{p2}^u = D_{p2}^l \cup D_{p2}^u$ , where  $D_{p2}^l = \{(x_{p2}^l)^i, y^i\}_{i=1}^{n_1}$  with  $n_1$  labelled samples and  $D_{p2}^u = \{(x_{p2}^u)^i\}_{i=n_1+1}^{n_1+n_2+m}$  with  $n_2 + m$  unlabelled samples. We denote two domain marginal distributions of  $D_{p1}$  and  $D_{p2}$  as  $p(x_{p1})$  and  $q(x_{p2})$ , respectively. The joint distribution of  $D_{p1}$  and  $D_{p2}$  is denoted as  $j(x_{p1}, x_{p2})$ .

Based on the matched domains, we investigate complementary LA modelling and complementary domain knowledge learning to provide inherent prediction perturbation for the consistency based cross-domain semi-supervised learning. Therefore, a hierarchical dual consistency is investigated. Specifically, for the intra-domain, we consider two dual-modelling networks  $S_1 : x_{p1} \rightarrow (\tilde{y}_{l1}, \tilde{y}_{g1})$  parameterised by  $\theta_1 = \{\theta_1^f, \theta_1^l, \theta_1^g\}$  and  $S_2 : x_{p2} \rightarrow (\tilde{y}_{l2}, \tilde{y}_{g2})$  parameterised by

$\theta_2 = \{\theta_2^f, \theta_2^l, \theta_2^g\}$  applied to the matched domains of  $D_{p1}$  and  $D_{p2}$ , respectively. Each dual-modelling network estimates two targets by considering local information and global information of image, where  $S_1$  simultaneously performs the global modelling of  $S_{g1} : x_{p1} \rightarrow \tilde{y}_{g1}$  parameterised by  $\{\theta_1^f, \theta_1^g\}$  and the local modelling of  $S_{l1} : x_{p1} \rightarrow \tilde{y}_{l1}$  parameterised by  $\{\theta_1^f, \theta_1^l\}$ . Similarly, the  $S_2$  simultaneously performs the global modelling of  $S_{g2} : x_{p2} \rightarrow \tilde{y}_{g2}$  parameterised by  $\{\theta_2^f, \theta_2^g\}$  and the local modelling of  $S_{l2} : x_{p2} \rightarrow \tilde{y}_{l2}$  parameterised by  $\{\theta_2^f, \theta_2^l\}$ . Then we encourage the global modelling and the local modelling of each dual-modelling network to predict consistent targets via the consistency loss:

$$\min_{\theta_1} O^{intra1}(\theta_1) = E_{x_{p1}^u \sim p(x_{p1})} L_d(S_{l1}(x_{p1}^u), S_{g1}(x_{p1}^u)) \quad (7)$$

$$\min_{\theta_2} O^{intra2}(\theta_2) = E_{x_{p2}^u \sim q(x_{p2})} L_d(S_{l2}(x_{p2}^u), S_{g2}(x_{p2}^u)) \quad (8)$$

where  $L_d(\cdot)$  denotes the dice loss function. For the dual consistency in inter-domain, we maximise the agreement on two matched domains. Therefore, we encourage  $S_1$  and  $S_2$  to predict similar outputs by:

$$\begin{aligned} & \min_{\theta_1, \theta_2} O^{inter}(\theta_1, \theta_2) \\ & = E_{(x_{p1}^u, x_{p2}^u) \sim j(x_{p1}, x_{p2})} L_c(S_1(x_{p1}^u), S_2(x_{p2}^u)) \\ & = E_{(x_{p1}^u, x_{p2}^u) \sim j(x_{p1}, x_{p2})} (L_c(S_{l1}(x_{p1}^u), S_{l2}(x_{p2}^u)) \\ & + L_c(S_{g1}(x_{p1}^u), S_{g2}(x_{p2}^u))) \end{aligned} \quad (9)$$

where  $L_c(\cdot)$  denotes the cross-entropy loss function. To avoid that  $S_1$  and  $S_2$  gradually resemble each other, we encourage the  $S_1$  and  $S_2$  to produce conditional independent features by orthogonalising the weights of feature layers:

$$\min_{\theta_1, \theta_2} O^{ow}(\theta_1, \theta_2) = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{K_i^2} \sum_{j=1}^{K_i^2} \left| \frac{(\theta_{1i}^f)^T \theta_{2i}^f}{||\theta_{1i}^f|| ||\theta_{2i}^f||} \right| \right) \quad (10)$$

where the  $N$  denotes the number of layers in  $S_1$  and  $S_2$ .  $K_i$  represents the number of features in  $i$ th layer.  $\theta_{1i}^f$  and  $\theta_{2i}^f$  denote the parameters of  $i$ th feature layer in  $S_1$  and  $S_2$ , respectively.

Beyond the consistency learning above,  $S_1$  and  $S_2$  can explicitly learn from  $D_{p1}^l$  and  $D_{p2}^l$  with the supervision of the labels:

$$\begin{aligned} \min_{\theta_1} O^{super1}(\theta_1) & = E_{x_{p1}^l \sim p(x_{p1})} L_s(S_1(x_{p1}^l), y) \\ & = E_{x_{p1}^l \sim p(x_{p1})} (L_s(S_{l1}(x_{p1}^l), y) \\ & + L_s(S_{g1}(x_{p1}^l), y)) \end{aligned} \quad (11)$$

$$\begin{aligned} \min_{\theta_2} O^{super2}(\theta_2) & = E_{x_{p2}^l \sim q(x_{p2})} L_s(S_2(x_{p2}^l), y) \\ & = E_{x_{p2}^l \sim q(x_{p2})} (L_s(S_{l2}(x_{p2}^l), y) \\ & + L_s(S_{g2}(x_{p2}^l), y)) \end{aligned} \quad (12)$$

where the  $y$  denotes the LA label.  $L_s(\cdot)$  denotes the supervised loss functions (cross-entropy loss function and dice loss function). Then the final training objective for the learning of  $S_1$  and  $S_2$  is denoted as:

$$\begin{aligned} \min_{\theta_1, \theta_2} O^{total}(\theta_1, \theta_2) & = \lambda_{super}(O^{super1} + O^{super2}) \\ & + \lambda_{intra}(O^{intra1} + O^{intra2}) \\ & + \lambda_{inter} O^{inter} + \lambda_{ow} O^{ow} \end{aligned} \quad (13)$$

341 where the  $\lambda_{super}$ ,  $\lambda_{intra}$ ,  $\lambda_{inter}$  and  $\lambda_{ow}$  are hyperparameters  
 342 to balance the loss terms.

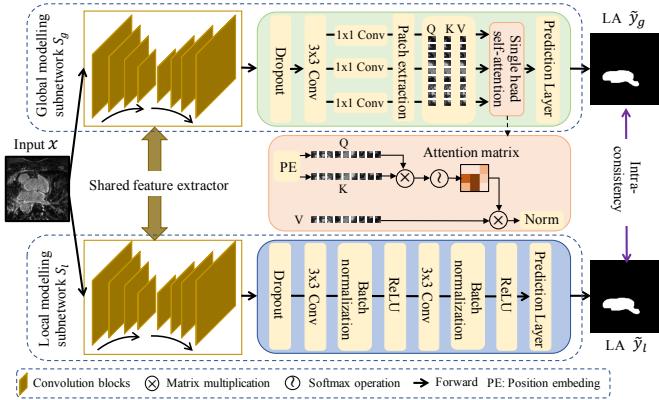


Fig. 4: Dual-modelling network for intra-consistency learning. The local-modelling branch and global-modelling branch share a feature extractor. For the global-modelling branch, the extracted feature maps from input images are split into  $8 \times 8$  patches. These  $8 \times 8$  patches are taken as a sequence of vectors to be fed to a self-attention based global-modelling structure.

#### 343 D. Network Configuration

344 The BAI module contains three subnetworks: two domain  
 345 mapping networks ( $G_1$ ,  $G_2$ ) and a discriminative network  $T$ .  
 346 We use the 2D U-Net with bilinear upsampling as network  
 347 backbones of both  $G_1$  and  $G_2$ .  $T$  has six convolution layers  
 348 with the numbers of filters of 32, 64, 128, 256, 256, 1, respec-  
 349 tively. Each of the first five  $3 \times 3$  convolutional layers with a  
 350 stride of 2 is followed by a batch normalisation layer and a  
 351 ReLU layer. The final  $1 \times 1$  convolutional layer with a stride  
 352 of 1 is followed by a sigmoid layer.

353 Hierarchical dual-modelling network contains two dual-  
 354 modelling networks with the same structure. Each dual-  
 355 modelling network contains a 2D U-Net with bilinear upsam-  
 356 pling used to extract image features and two branch networks  
 357 used to estimate targets. The two branch networks are the  
 358 global modelling network and the local modelling network.  
 359 The global modelling network is based on the self-attention  
 360 [34]–[36] as shown in Fig. 4. In the global modelling network,  
 361 we use the sinusoidal position encoding to emphasise the  
 362 sequential relationship between input feature patches [37]. The  
 363 local modelling network consists of three convolution blocks.  
 364 The details are shown in Fig. 4.

## IV. EXPERIMENTS

### 366 A. Overview of Experiments

367 Comprehensive experiments were performed to validate our  
 368 proposed AHDC.

369 **(1) The feasibility of AHDC for generalising across**  
 370 **domains:** Our proposed AHDC was validated on four 3D late  
 371 gadolinium enhancement cardiac MR (LGE CMR) datasets  
 372 and a 3D CT dataset combined in pairs, which followed the

373 independent validation protocol. Furthermore, we also investi-  
 374 gated the impact of different ratios ( $r = \{5\%, 10\%, 20\%\}$ ) of  
 375 the labelled data for validating our proposed AHDC.

376 **(2) The superiority of AHDC for generalising across**  
**377 domains:** We compared to widely used and state-of-the-art

378 semi-supervised methods on cross-domain data for compari-  
 379 son, including mean teacher (MT) method [30], uncertainty-  
 380 aware self-ensembling model (UA-MT) [7], Dual-Task con-  
 381 sistency (DTC) [38] and Dual-Teacher [39]. It is of note that  
 382 MT, UA-MT and DTC were proposed for the single-domain  
 383 semi-supervised learning while the Dual-Teacher method was  
 384 proposed for the cross-domain learning. Besides, the Dual-  
 385 Teacher required the labelled data from both cross-domain  
 386 data for model learning. For a fair comparison, MT, UA-MT  
 387 and DTC were performed on one of the matched domains,  
 388 i.e.,  $D_{p1}$ . We also compared with the joint training method  
 389 that combining the cross-domain data directly for the LA  
 390 segmentation based on our proposed semi-supervised method.

391 **(3) The effectiveness of the components in AHDC:**  
 392 Firstly, we compared the performance between different ar-  
 393 chitectures of the BAI module. On the one hand, to validate  
 394 the effectiveness of bidirectional reconstruction for speci-  
 395 fying the relationship of matched samples, an experiment  
 396 was performed on bidirectional adversarial inference without  
 397 using bidirectional reconstruction (BAI<sub>wbr</sub>/ALI/BiGAN). On  
 398 the other hand, to validate the effectiveness of skip connec-  
 399 tion of domain mapping network for keeping target structure  
 400 consistent, an experiment was performed on bidirectional  
 401 adversarial inference without using skip connection in domain  
 402 mapping network (BAI<sub>eds</sub>). Then, we further validated the  
 403 performance of BAI by comparing it with the fully adversarial  
 404 ALICE [26] on the downstream semi-supervised tasks. Finally,  
 405 for validating the effectiveness of HDC, we decomposed the  
 406 HDC into independent intra-domain dual consistency learning  
 407 (HDC<sub>intra</sub>) by removing a dual-modelling network and inter-  
 408 domain dual consistency learning (HDC<sub>inter</sub>) by removing  
 409 global modelling branch but retaining local modelling branch.

410 **(4) The effectiveness of the BAI for matching domains:**  
 411 Firstly, we performed the principal components analysis to  
 412 show the data distributions of source domains ( $D_1$  and  $D_2$ ) and  
 413 the adapted domains ( $D_{1t2}$  and  $D_{2t1}$ ). The data distributions  
 414 of source domains and the adapted domains were compared to  
 415 validate the effectiveness of AHDC for aligning distributions.  
 416 Then, we made a qualitative visualisation of images before  
 417 and after the bidirectional adversarial inference to validate the  
 418 effectiveness of AHDC for matching samples.

419 **(5) The effectiveness of the HDC for the availability**  
 420 **of complementary information:** To validate the availability  
 421 of complementary modelling information in the intra-domain,  
 422 we compared the segmentation performance of dual modelling  
 423 network (local-global modelling structure) to the ones with-  
 424 out using dual-modelling structures. Specifically, we replaced  
 425 the local-modelling branch with the global-modelling branch  
 426 (global-global modelling structure) and replaced the global-  
 427 modelling branch with the local-modelling branch (local-local  
 428 modelling structure) in dual modelling network for exper-  
 429 iments. To validate the availability of complementary domain  
 430 information in inter-domain, we compared the segmentation

TABLE II

COMPARISON OF FOUR LGE-CMRI DATASETS FROM DIFFERENT CENTRES. ABBREVIATIONS: TE, ECHO TIME; TR, REPETITION TIME; CARMA, COMPREHENSIVE ARRHYTHMIA RESEARCH AND MANAGEMENT.

| Centres | Acquired Resolution                           | TE/TR      | Scanner                              | Source                                    | Amount of Data    |
|---------|---|------------|--------------------------------------|---|-------------------|
| C1      | (1.4 ~ 1.5) × (1.4 ~ 1.5) × 4 mm <sup>3</sup> | 2.2/5.2 ms | 1.5 Tesla Avanto                     | Royal Brompton Hospital                   | 165 LGE-CMR scans |
| C2      | 1.25 × 1.25 × 2.5 mm <sup>3</sup>             | 2.3/5.4 ms | 1.5 Tesla Avanto,<br>3.0 Tesla Vario | CARMA, University of Utah                 | 153 LGE-CMR scans |
| C3      | 1.4 × 1.4 × 1.4 mm <sup>3</sup>               | 2.1/5.3 ms | 1.5T Philips Achieva                 | Beth Israel Deaconess Medical Center      | 20 LGE-CMR scans  |
| C4      | 1.3 × 1.3 × 4 mm <sup>3</sup>                 | 2.1/5.3 ms | 1.5T Philips Achieva                 | Imaging Sciences at King's College London | 20 LGE-CMR scans  |

431 performance of HDC with/without using the orthogonal weight  
 432 constraint (WOW and WOOW).

433 **(6) The effects of parameter settings on model performance:** We explored two important parameter settings. (i) The  
 434 impact of different patch sizes (4 × 4, 8 × 8 and 16 × 16) for  
 435 global modelling. (ii) The impact of different values of  $\lambda_{ow}$   
 436 (0.0, 0.1, and 1.0 ) for inter-domain learning.

### B. Datasets

437 To evaluate the performance of our proposed AHDC, four  
 438 3D LGE-MRI datasets (C1, C2, C3 and C4) and a 3D CT  
 439 dataset (C5) were collected as a retrospective study. In our  
 440 experiments, the collected datasets of C1 and C2 included  
 441 segmentation of the LA epicardium and LA endocardium  
 442 while the collected datasets of C3, C4 and C5 included seg-  
 443 mentation of the LA endocardium. We have summarised the  
 444 characteristics of the four 3D LGE-MRI datasets to emphasise  
 445 their differences as shown in TABLE II.

446 LGE-MRI scanning sequence of centre 1 (C1): Cardiac  
 447 MR data were acquired in patients with longstanding persis-  
 448 tent atrial fibrillation (AF) on a Siemens Magnetom Avanto  
 449 1.5T scanner (Siemens Medical Systems, Erlangen, Germany).  
 450 Transverse navigator-gated 3D LGE-CMRI [40] was per-  
 451 formed using an inversion prepared segmented gradient echo  
 452 sequence (TE/TR 2.2ms/5.2ms) 15 minutes after gadolinium  
 453 administration (Gadovist-gadobutrol, 0.1mmol/kg body  
 454 weight, BayerSchering, Berlin, Germany) [41]. The inversion  
 455 time was set to null the signal from normal myocardium. The  
 456 acquired resolution parameter of LGE-CMRI data was (1.4 –  
 457 1.5) × (1.4 – 1.5) × 4 mm<sup>3</sup> (reconstructed to (0.7 – 0.75) × (0.7 –  
 458 0.75) × 2 mm<sup>3</sup>). LGE-CMRI data were acquired during free-  
 459 breathing using a crossed-pairs navigator positioned over the  
 460 dome of the right hemi-diaphragm with navigator acceptance  
 461 window size of 5mm and CLAWS respiratory motion control  
 462 [42], [43]. The LGE CMR data were collected from the Royal  
 463 Brompton Hospital. In total, 165 scans were used in this study.

464 LGE-MRI scanning sequence of centre 2 (C2): Cardiac  
 465 MR data were obtained on a 1.5 Tesla Avanto scanners  
 466 or a 3.0 Tesla Vario (Siemens Medical Solutions, Erlangen,  
 467 Germany). The scan is acquired 20–25 minutes after 0.1  
 468 mmol/kg gadolinium contrast (Multihance, Bracco Diagnostics  
 469 Inc., Princeton, NJ) using a 3D respiratory navigated, inversion  
 470 recovery prepared gradient echo pulse sequence. Typical ac-  
 471 quisition parameters are free breathing using navigator gating,  
 472 a transverse imaging volume with voxel size = 1.25 × 1.25 × 2.5  
 473 mm<sup>3</sup> (reconstructed to 0.625 × 0.625 × 2.5 mm<sup>3</sup>), TR/TE =

474 5.4/2.3 ms, inversion time (TI)=270–310 ms. The TI value for  
 475 the LGE-MRI scan is identified using a scout scan. Typical  
 476 scan times for the LGE-MRI study were between 8 and 15  
 477 min at 1.5 T and 6–11 min using the 3T scanner (for Siemens  
 478 sequences) depending on subject respiration and heart rates.  
 479 The LGE CMR data were collected from the Comprehensive  
 480 Arrhythmia Research and Management, University of Utah. In  
 481 total, 153 scans were used in this study.

482 LGE-MRI scanning sequence of center 3 (C3): C3 is from  
 483 the ISBI 2012 Left Atrium Fibrosis and Scar Segmentation  
 484 Challenge [44], [45]. The LGE CMR data were collected from  
 485 the Beth Israel Deaconess Medical Center. In total, 20 scans  
 486 were used in this study.

487 LGE-MRI scanning sequence of center 4 (C4): C4 is also  
 488 from the ISBI 2012 Left Atrium Fibrosis and Scar Segmen-  
 489 tation Challenge [44], [45]. The LGE CMR data were collected  
 490 from the Imaging Sciences at King's College. In total, 20 scans  
 491 were used in this study.

492 CT scanning sequence of centre 5 (C5): C5 is from the  
 493 Multi-modality Whole Heart Segmentation (MM-WHS) 2017  
 494 dataset [46]–[49]. In total, 60 CT scans were used in this study.

### C. Experimental Setup

495 **(1) Data partitioning:** For C1, the 3D LGE-MRI dataset  
 496 with 165 scans was randomly split into a training set with 99  
 497 scans and a testing set with 66 scans (33 pre-ablation scans and  
 498 33 post-ablation scans). The training set then was randomly  
 499 split into a labelled training set with 20 scans (20%) and an  
 500 unlabelled training set with 79 scans (80%). For C2, the 3D  
 501 LGE-MRI dataset with 153 scans was randomly split into a  
 502 training set with 91 scans and a testing set with 62 scans (31  
 503 pre-ablation scans and 31 post-ablation scans). The training  
 504 set then was randomly split into a labelled training set with  
 505 18 scans (20%) and an unlabelled training set with 73 scans  
 506 (80%). For C3 and C4, each 3D LGE-MRI dataset with 20  
 507 scans was randomly split into a training set with 12 scans  
 508 and a testing set with 8 scans (4 pre-ablation scans and 4  
 509 post-ablation scans). The training set then was randomly split  
 510 into a labelled training set with 4 scans and an unlabelled  
 511 training set with 8 scans. Because C5 only provides 60 CT  
 512 scans including 20 labelled scans and 40 unlabelled scans,  
 513 we randomly selected 15 scans from 20 labelled scans as a  
 514 testing set. The remaining 5 labelled scans (labelled training  
 515 set) and 40 unlabelled scans (unlabelled training set) together  
 516 as a training set. Since each patient may contain multiple 3D  
 517 LGE-MRI scans, the 3D LGE-MRI datasets were split under  
 518 519 520

521 the strategy that all scans from each unique patient were only  
 522 in one of the training or testing sets.

523 **(2) Implementation details:** Experiments were performed  
 524 on five datasets combined in pairs for cross-centre study (C1  
 525 and C2, C3 and C4) and cross-modality study (C2 and C5).  
 526 To reduce the dependence of models on annotated data and  
 527 to avoid the impact of label variations from different centres,  
 528 there were two kinds of experiment settings for each cross-  
 529 domain data. Take experiments on C1 and C2 as an example:  
 530 one used C1 to support C2 that the model was trained using the  
 531 labelled training set (18 labelled cases) of C2, the unlabelled  
 532 training set (73 unlabelled cases) of C2 and the whole training  
 533 set (99 unlabelled cases) of C1. The other one used C2 to  
 534 support C1 that the model was trained using the labelled  
 535 training set (20 labelled cases) of C1, the unlabelled training  
 536 set (79 unlabelled cases) of C1 and the whole training set (91  
 537 unlabelled cases) of C2. We denoted the results obtained by  
 538 the fully supervised model trained with the labelled training  
 539 set from C1 (20 cases), C2 (18 cases), C3 (4 cases), C4 (4  
 540 cases) and C5 (5 cases) as the baseline and the results obtained  
 541 by the fully supervised model trained with the whole training  
 542 set from C1 (99 cases), C2 (91 cases), C3 (12 cases) and C4  
 543 (12 cases) as the upper bound.

544 We pre-processed the data with the normalisation. Smaller  
 545 patches of  $256 \times 256$  centred on the LA region were cropped.  
 546 To avoid overfitting, we applied data augmentations with  
 547 random rotation. The training time of our model is about  
 548 17.17 hours while the testing time for one 3D case is about  
 549 0.259 seconds. For the learning of the BAI network, we  
 550 used the Adam method to perform the optimisation of two  
 551 mapping networks with an initial learning rate of 0.001 and a  
 552 decayed rate of 0.98. The optimiser used in the discriminative  
 553 network was Adam with a fixed learning rate of 0.0001.  
 554 For the learning of two dual-modelling networks, we also  
 555 used the Adam method with an initial learning rate of 0.001  
 556 and a decayed rate of 0.98. The current statistics of batch  
 557 normalisation were used for both training and testing. All  
 558 experiments were performed with an independent test. For  
 559 the dual consistency learning, in each iteration, we first per-  
 560 formed the intra-consistency with both labelled and unlabelled  
 561 data simultaneously, then performed the inter-consistency with  
 562 both labelled and unlabelled data simultaneously, performed  
 563 supervised learning with labelled data in the last. Our deep  
 564 learning model was implemented using Tensorflow 1.2.1 on  
 565 an Ubuntu 16.04 machine (The code will be released pub-  
 566 licly once the manuscript is accepted for publication via  
 567 <https://github.com/Heye-SYSU/AHDC>). It was trained and  
 568 tested using an Nvidia RTX 8000 GPU (48GB GPU memory).

569 The coefficients  $\lambda_d$  and  $\lambda_r$  used to balance the adversarial  
 570 loss and the reconstruction loss, were automatically learned  
 571 based on the strategy of uncertainty [50]. The coefficient  
 572  $\lambda_{intra}$  was dynamically changed over time with the function  
 573 of  $f(t) = e^{-5*(1 - \frac{t}{t_{max}})^2}$ . The coefficients  $\lambda_{inter}$ ,  $\lambda_{super}$  and  
 574  $\lambda_{ow}$  were set to the values of 1.0, 0.5 and 0.1, respectively.

575 **(3) Evaluation criteria:** To evaluate the segmentation  
 576 performance, we used region-based metrics [51], [52], e.g.,  
 577 the Dice Similarity Coefficient (DSC) and the Jaccard Index  
 578 (JI), to validate the predicted segmentation map against the

579 manually defined ground-truth. We also used a surface-based  
 580 metric called Average Surface Distance (ASD) to provide the  
 581 distance in mm to quantify the accuracy of the predicted mesh  
 582 ( $S$ ) compared to the ground-truth mesh ( $S'$ ) [52].

## V. RESULTS AND ANALYSIS

583 In this section, we demonstrate the results of the above  
 584 mentioned experiments to validate our proposed AHDC for  
 585 the cross-domain semi-supervised segmentation.

### A. The Feasibility Analysis of AHDC for Generalising 587 Across Domains:

588 TABLE III and TABLE IV summarises the quantitative  
 589 segmentation results of AHDC on multi-centre data and multi-  
 590 modality data. As we can see, our proposed AHDC obtains  
 591 consistent improvements in terms of the DSC, JI and ASD  
 592 against the baselines. Furthermore, as the experiment results  
 593 are summarised in TABLE V, one can see that our proposed  
 594 AHDC obtains consistent improvements against the fully  
 595 supervised learning under the 5%, 10%, 20% labelled data  
 596 setting. Fig. 5 and Fig. 6 provide the 2D and 3D qualitative  
 597 LAs estimated by AHDC compared to the ground truth. It is  
 598 observed that our proposed AHDC has the ability to segment  
 599 LA accurately. These quantitative and qualitative results in-  
 600 dicate the feasibility of our proposed AHDC for generalising  
 601 across domains.

### B. The Superiority Analysis of AHDC for Generalising 603 Across Domains:

604 TABLE III and TABLE IV summarises the experiment re-  
 605 sults on multi-centre data and multi-modality data combined in  
 606 pairs for comparison. It is observed that the widely used semi-  
 607 supervised method of MT improves the segmentation accuracy  
 608 of LA compared to the baseline. One can see that after adding  
 609 uncertainty information to the MT, the performance of the MT  
 610 is improved (UA-MT). The DTC method further improves  
 611 the segmentation accuracy, indicating the effectiveness of  
 612 dual task consistency for semi-supervised learning. Although  
 613 these methods have the ability to mine effective information  
 614 from unlabelled data to support task learning, they have no  
 615 proper mechanism to exploit the cross-domain information,  
 616 thus leading to limited segmentation results. Compared to  
 617 these methods, Dual-Teacher leverages two teacher models to  
 618 guide a student model for the learning of both intra-domain  
 619 and inter-domain knowledge, thus achieving big improvements  
 620 in terms of segmentation accuracy. Notably, our proposed  
 621 AHDC obtains the best segmentation accuracy over these  
 622 widely used and state-of-the-art semi-supervised methods,  
 623 which shows its superiority for generalising across domains.  
 624 Furthermore, it is observed that our proposed AHDC generally  
 625 improves the segmentation accuracy compared to the joint  
 626 training, which combines the cross-domain data directly for  
 627 the semi-supervised LA segmentation. This demonstrates that  
 628 our proposed AHDC can leverage cross-domain information to  
 629 improve the model performance. We also provide qualitative  
 630 comparison between different methods in Fig. 5. It is observed

TABLE III

QUANTITATIVE COMPARISON BETWEEN OUR PROPOSED AHDC AND OTHER METHODS ON MULTI-CENTRE DATA.

ABBREVIATIONS: DSC, DICE SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD, AVERAGE SURFACE DISTANCE.

(a) EXPERIMENTS ON C1 (MR) AND C2 (MR)

| Method               | C2 (MR) supports C1 (MR)       |                                |                             | C1 (MR) supports C2 (MR)       |                                |                             |
|----------------------|--------------------------------|--------------------------------|-----------------------------|--------------------------------|--------------------------------|-----------------------------|
|                      | DSC                            | JI                             | ASD (mm)                    | DSC                            | JI                             | ASD (mm)                    |
| Upper Bound Baseline | 0.932 ± 0.026<br>0.869 ± 0.078 | 0.874 ± 0.044<br>0.775 ± 0.111 | 1.28 ± 0.847<br>2.81 ± 2.08 | 0.926 ± 0.021<br>0.860 ± 0.103 | 0.863 ± 0.036<br>0.765 ± 0.122 | 0.867 ± 0.46<br>4.00 ± 5.08 |
| MT                   | 0.882 ± 0.059                  | 0.793 ± 0.090                  | 2.45 ± 1.67                 | 0.880 ± 0.071                  | 0.792 ± 0.096                  | 1.79 ± 1.96                 |
| UA-MT                | 0.885 ± 0.060                  | 0.799 ± 0.092                  | 2.19 ± 1.47                 | 0.884 ± 0.072                  | 0.799 ± 0.098                  | 2.79 ± 3.76                 |
| DTC                  | 0.887 ± 0.061                  | 0.803 ± 0.094                  | 2.25 ± 1.54                 | 0.888 ± 0.076                  | 0.806 ± 0.102                  | 2.40 ± 3.48                 |
| Dual-Teacher         | 0.899 ± 0.046                  | 0.820 ± 0.073                  | 1.83 ± 1.02                 | 0.896 ± 0.064                  | 0.816 ± 0.088                  | 1.97 ± 3.13                 |
| Joint-training       | 0.889 ± 0.059                  | 0.805 ± 0.091                  | 2.05 ± 1.22                 | 0.887 ± 0.056                  | 0.801 ± 0.081                  | 1.59 ± 1.56                 |
| AHDC                 | <b>0.916 ± 0.041</b>           | <b>0.848 ± 0.066</b>           | <b>1.47 ± 0.846</b>         | <b>0.917 ± 0.026</b>           | <b>0.848 ± 0.043</b>           | <b>1.17 ± 1.60</b>          |

(b) EXPERIMENTS ON C3 (MR) AND C4 (MR)

| Method               | C4 (MR) supports C3 (MR)       |                                |                             | C3 (MR) supports C4 (MR)       |                                |                             |
|----------------------|--------------------------------|--------------------------------|-----------------------------|--------------------------------|--------------------------------|-----------------------------|
|                      | DSC                            | JI                             | ASD (mm)                    | DSC                            | JI                             | ASD (mm)                    |
| Upper Bound Baseline | 0.808 ± 0.035<br>0.684 ± 0.098 | 0.679 ± 0.050<br>0.528 ± 0.109 | 2.42 ± 0.645<br>6.17 ± 3.63 | 0.841 ± 0.043<br>0.742 ± 0.081 | 0.727 ± 0.062<br>0.596 ± 0.095 | 2.07 ± 0.543<br>4.20 ± 1.08 |
| MT                   | 0.749 ± 0.073                  | 0.604 ± 0.092                  | 3.59 ± 1.77                 | 0.797 ± 0.101                  | 0.673 ± 0.122                  | 2.42 ± 1.10                 |
| UA-MT                | 0.760 ± 0.081                  | 0.619 ± 0.100                  | 3.96 ± 2.10                 | 0.811 ± 0.086                  | 0.690 ± 0.108                  | 2.21 ± 0.718                |
| DTC                  | 0.765 ± 0.066                  | 0.624 ± 0.086                  | 3.40 ± 1.72                 | 0.809 ± 0.088                  | 0.687 ± 0.111                  | 2.94 ± 0.935                |
| Dual-Teacher         | 0.773 ± 0.050                  | 0.633 ± 0.067                  | 3.03 ± 0.988                | 0.817 ± 0.089                  | 0.699 ± 0.112                  | 2.48 ± 0.937                |
| Joint-training       | 0.770 ± 0.056                  | 0.629 ± 0.074                  | 3.45 ± 1.51                 | 0.811 ± 0.094                  | 0.691 ± 0.115                  | 2.70 ± 0.889                |
| AHDC                 | <b>0.795 ± 0.044</b>           | <b>0.661 ± 0.061</b>           | <b>2.47 ± 0.681</b>         | <b>0.830 ± 0.057</b>           | <b>0.713 ± 0.077</b>           | <b>2.07 ± 0.659</b>         |

TABLE IV

QUANTITATIVE COMPARISON BETWEEN OUR PROPOSED AHDC AND OTHER METHODS ON MULTI-MODALITY DATA.

ABBREVIATIONS: DSC, DICE SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD, AVERAGE SURFACE DISTANCE.

| Method               | C5 (CT) supports C2 (MR)       |                                |                            | C2 (MR) supports C5 (CT) |                      |                     |
|----------------------|--------------------------------|--------------------------------|----------------------------|--------------------------|----------------------|---------------------|
|                      | DSC                            | JI                             | ASD (mm)                   | DSC                      | JI                   | ASD (mm)            |
| Upper Bound Baseline | 0.923 ± 0.025<br>0.858 ± 0.107 | 0.858 ± 0.042<br>0.763 ± 0.121 | 1.20 ± 1.96<br>2.72 ± 4.59 | -<br>0.828 ± 0.115       | -<br>0.722 ± 0.157   | -<br>4.88 ± 3.53    |
| MT                   | 0.874 ± 0.072                  | 0.782 ± 0.99                   | 2.42 ± 3.14                | 0.861 ± 0.086            | 0.765 ± 0.121        | 4.54 ± 3.35         |
| UA-MT                | 0.881 ± 0.056                  | 0.791 ± 0.084                  | 1.84 ± 2.66                | 0.878 ± 0.050            | 0.786 ± 0.077        | 2.39 ± 1.47         |
| DTC                  | 0.888 ± 0.059                  | 0.803 ± 0.084                  | 1.93 ± 3.29                | 0.880 ± 0.064            | 0.791 ± 0.096        | 2.96 ± 2.10         |
| Dual-Teacher         | 0.888 ± 0.041                  | 0.801 ± 0.062                  | 1.41 ± 1.47                | 0.891 ± 0.036            | 0.806 ± 0.057        | 2.38 ± 1.14         |
| Joint-training       | 0.869 ± 0.069                  | 0.774 ± 0.100                  | 2.18 ± 3.89                | 0.834 ± 0.117            | 0.731 ± 0.156        | 5.62 ± 5.10         |
| AHDC                 | <b>0.911 ± 0.028</b>           | <b>0.837 ± 0.047</b>           | <b>1.07 ± 0.872</b>        | <b>0.916 ± 0.031</b>     | <b>0.846 ± 0.052</b> | <b>1.30 ± 0.319</b> |

632 that the LAs estimated by other methods present fragmentary  
 633 parts and unsMOOTH boundaries. While the LAs estimated by  
 634 our proposed method are closer to the ground truth with  
 635 smoother boundaries.

### C. Ablation Studies

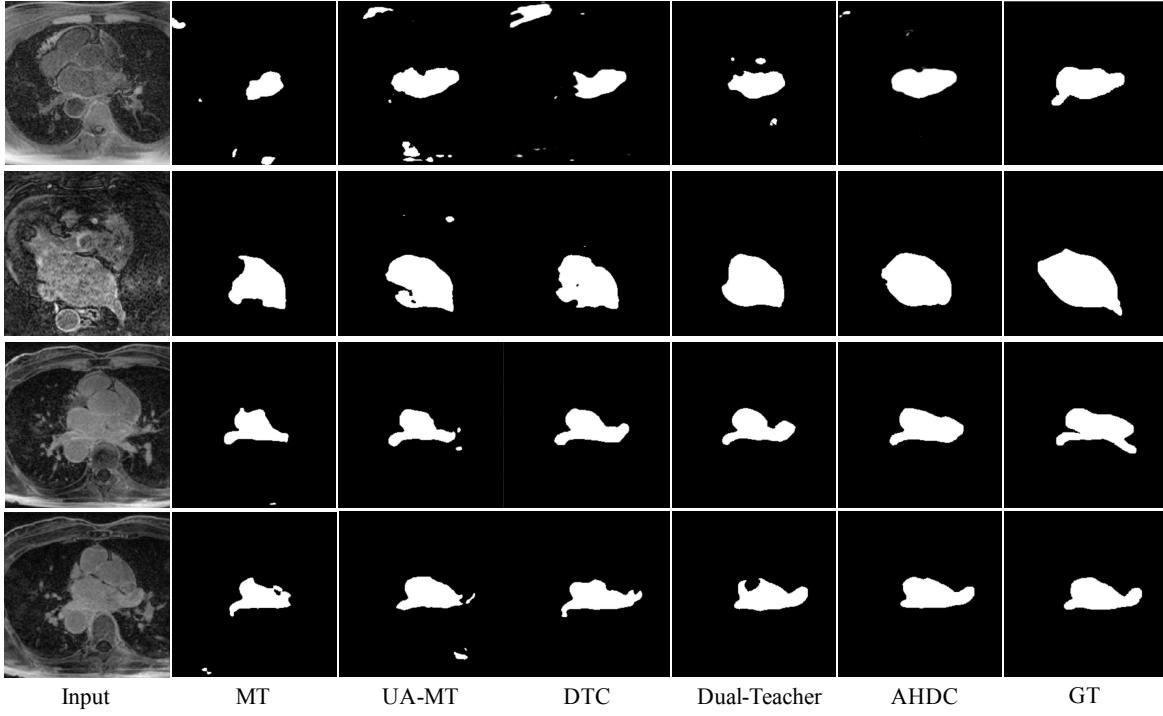
We performed ablation studies on *C1* and *C2* (*C1* supports *C2*) to validate the effectiveness of our proposed AHDC for the cross-domain semi-supervised segmentation.

**(1) Model variation study for bidirectional adversarial inference:** As the experimental results are summarised in TABLE VI, the bidirectional adversarial inference with bidirectional reconstruction improves the LA segmentation accuracy in terms of DSC, JI and ASD compared with the BAI<sub>wbr</sub>/ALI/BiGAN. The reason behind the improvements is that bidirectional reconstruction makes the relationship between matched samples specified and constrained. It guarantees that the matched samples are one-to-one correspondence

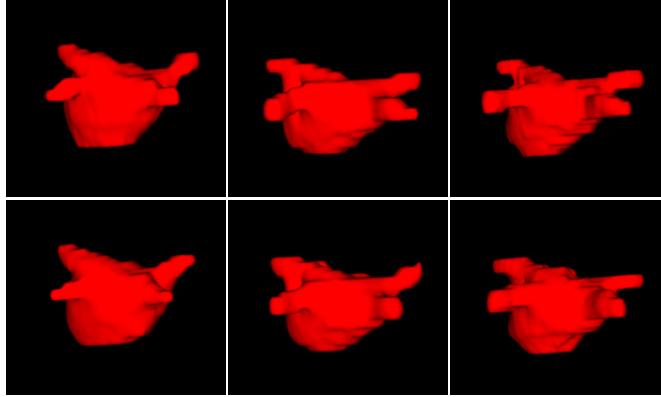
TABLE V  
 THE PERFORMANCE OF AHDC ON DIFFERENT PERCENTAGES OF LABELLED DATA. ABBREVIATIONS: LX (%) : LX (%) : THE RATIO OF LABELLED DATA IN THE TRAINING SET OF CENTRE X; UX (%) : THE RATIO OF UNLABELLED DATA IN THE TRAINING SET OF CENTRE X; DSC, DICE SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD, AVERAGE SURFACE DISTANCE.

| Method      | Rate      |           | Metrics       |               | ASD          |
|-------------|-----------|-----------|---------------|---------------|--------------|
|             | L2/U2 (%) | L1/U1 (%) | DSC           | JI            |              |
| Upper Bound | 100/0     | 0/0       | 0.926 ± 0.021 | 0.863 ± 0.036 | 0.867 ± 0.46 |
| Baseline    | 20/0      | 0/0       | 0.860 ± 0.103 | 0.765 ± 0.122 | 4.00 ± 5.08  |
| AHDC        | 20/80     | 0/100     | 0.917 ± 0.026 | 0.848 ± 0.043 | 1.17 ± 1.60  |
| Baseline    | 10/0      | 0/0       | 0.815 ± 0.142 | 0.706 ± 0.153 | 4.84 ± 6.09  |
| AHDC        | 10/90     | 0/100     | 0.891 ± 0.039 | 0.805 ± 0.060 | 1.98 ± 2.65  |
| Baseline    | 5/0       | 0/0       | 0.776 ± 0.134 | 0.650 ± 0.146 | 6.51 ± 5.52  |
| AHDC        | 5/95      | 0/100     | 0.871 ± 0.041 | 0.773 ± 0.060 | 1.95 ± 1.44  |

for subsequent effective hierarchical dual consistency learning on cross-domain data. It is also observed that the segmentation accuracy is dropped while removing the skip connection from



**Fig. 5:** 2D visual comparisons on LA segmentation results estimated by different methods. It is observed that our estimated LAs (AHDC) are more similar to the ground truth (GT) than others (DSC based segmentation accuracies of AHDC for the 2D slices from row 1 to row 4 are 0.859, 0.897, 0.907 and 0.949, respectively). Abbreviations: DSC, Dice Similarity Coefficient.



**Fig. 6:** 3D visualization of LA segmentation results estimated by AHDC. Each DSC score is calculated for the whole 3D LGE-MRI image (The DSC based segmentation accuracies of AHDC for the 3D slices from column 1 to column 3 are 0.936, 0.917, and 0.898, respectively). Abbreviations: DSC, Dice Similarity Coefficient.

the domain mapping network. The reason behind this is that the domain mapping network (UNet structure) employs the skip connection to deliver the low-level information. It allows the samples adapted to another domain to maintain the same LA structures, which makes subsequent dual consistency learning effective. Furthermore, one can see that our proposed BAI has better performance for the downstream semi-supervised LA segmentation task compared to the fully

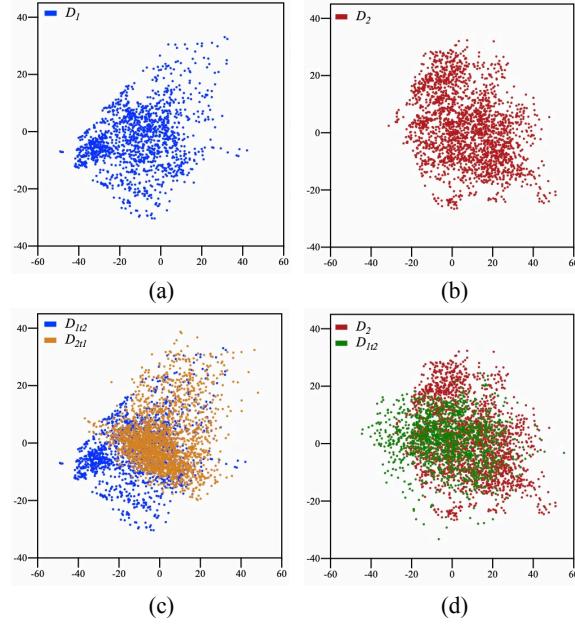
adversarial ALICE method, which indicates the superiority of our proposed BAI.

TABLE VI  
MODEL VARIATION STUDY ON C1 AND C2 (C1 SUPPORTS C2). ABBREVIATIONS: LX (%): THE RATIO OF LABELLED DATA IN THE TRAINING SET OF CENTRE X; UX (%): THE RATIO OF UNLABELLED DATA IN THE TRAINING SET OF CENTRE X; DSC, DICE SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD, AVERAGE SURFACE DISTANCE.

| Method                     | Metrics       |               |              |
|----------------------------|---------------|---------------|--------------|
|                            | DSC           | JI            | ASD          |
| Lower Bound                | 0.860 ± 0.103 | 0.765 ± 0.122 | 4.00 ± 5.08  |
| BAI <sub>eds</sub> + HDC   | 0.879 ± 0.039 | 0.786 ± 0.060 | 1.59 ± 0.875 |
| BAI <sub>wbr</sub> + HDC   | 0.885 ± 0.051 | 0.798 ± 0.076 | 1.37 ± 0.802 |
| ALICE + HDC                | 0.896 ± 0.033 | 0.814 ± 0.053 | 1.42 ± 1.09  |
| BAI + HDC <sub>intra</sub> | 0.893 ± 0.048 | 0.809 ± 0.073 | 1.90 ± 2.81  |
| BAI + HDC <sub>inter</sub> | 0.900 ± 0.045 | 0.822 ± 0.066 | 1.51 ± 1.24  |
| AHDC                       | 0.917 ± 0.026 | 0.848 ± 0.043 | 1.17 ± 1.60  |

(2) **Model variation study for hierarchical dual consistency:** As the experiment results are summarised in TABLE VI, the independent intra-domain and inter-domain dual consistency learning can both improve the LA segmentation accuracy compared to the lower-bound model. This indicates that the intra-domain dual consistency learning and the inter-domain dual consistency learning are effective to exploit the unlabelled data from cross-domain data. Furthermore, it is also observed that the intra-domain dual consistency and the inter-domain dual consistency can promote each other for the cross-domain semi-supervised segmentation. These results demon-

673 strate the effectiveness of the hierarchical dual consistency for  
674 semi-supervised segmentation on cross-domain data.



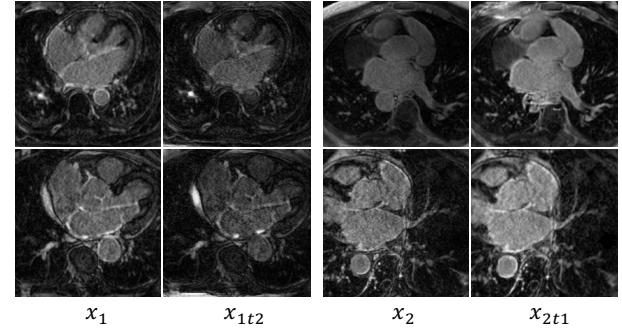
**Fig. 7:** Principal components analysis based visualisation for the data distribution of the testing tests of C1 and C2. (a) The data distribution of domain  $D_1$ . (b) The data distribution of domain  $D_2$ . (c) The data distribution of domain  $D_1$  and the domain  $D_{2t1}$  adapted from  $D_2$  to  $D_1$ . (d) The data distribution of domain  $D_2$  and the domain  $D_{1t2}$  adapted from  $D_1$  to  $D_2$ .

#### 675 *D. The Effectiveness Analysis of BAI for Matching 676 Domains*

677 The effectiveness of the bidirectional adversarial inference  
678 is further validated by the qualitative results on distribution  
679 alignment and sample matching in the testing set.

680 **(1) Distribution alignment:** In Fig. 7, we color samples  
681 from different domains and adapted domains to highlight their  
682 correspondence (brown and blue for the samples from domains  
683 of  $D_2$  and  $D_1$ , respectively. Peru and green for the samples from  
684 the domain  $D_{2t1}$  adapted from  $D_2$  to  $D_1$  and the domain  
685  $D_{1t2}$  adapted from  $D_1$  to  $D_2$ , respectively). It is observed  
686 that the domains of  $D_1$  and  $D_2$  have different distributions  
687 as shown in Fig. 7 (a) and Fig. 7 (b). Besides, as shown in  
688 Fig. 7 (c) and Fig. 7 (d), after the bidirectional adversarial  
689 inference, the distribution of the domain  $D_{2t1}$  adapted from  
690  $D_2$  to  $D_1$  is consistent with the distribution of  $D_1$ . Meanwhile,  
691 the distribution of the domain  $D_{1t2}$  adapted from  $D_1$  to  $D_2$   
692 is consistent with the distribution of  $D_2$ . One also can find  
693 the adapted domains of  $D_{1t2}$  and  $D_{2t1}$  make the distribution  
694 spaces of  $D_1$  and  $D_2$  more complete. These results indicate  
695 the effectiveness of BAI for the distribution alignment.

696 **(2) Sample matching:** Fig. 8 provides 2D visualisation  
697 of some examples before and after bidirectional adversarial  
698 inference. The images of the first two columns are from  
699 the domain  $D_1$  and the domain adapted from  $D_1$  to  $D_2$ .



**Fig. 8:** Qualitative visualisation of images and corresponding adapted images in the testing tests of C1 and C2. Abbreviations:  $x_1$ , image from domain  $D_1$ ;  $x_{1t2}$ , image adapted from domain  $D_1$  to domain  $D_2$ ;  $x_2$ , image from domain  $D_2$ ;  $x_{2t1}$ , image adapted from domain  $D_2$  to domain  $D_1$ .

The images of the last two columns are from the domain  
700  $D_2$  and the domain adapted from  $D_2$  to  $D_1$ . It is observed  
701 that the target shape and structure in corresponding images  
702 are consistent. However, the texture and the brightness in  
703 corresponding images are different. These results illustrate  
704 that the bidirectional adversarial inference is effective to produce  
705 the matched samples.

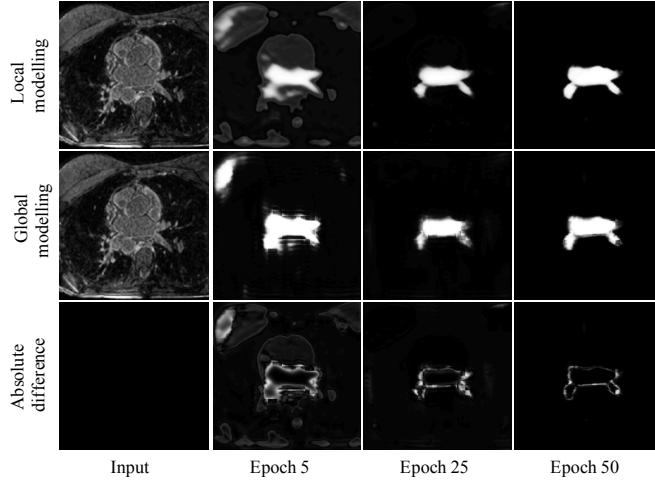
#### 707 *E. The Effectiveness Analysis of HDC for the Availability 708 of Complementary Information*

709 **(1) Availability of the complementary modelling information:** TABLE VII summarises the experiment results on  
710 different modelling structures (Local-Global, Local-Local and  
711 Global-Global) for the intra-domain semi-supervised learning.  
712 It is observed that the dual-modelling structure (Local-Global)  
713 achieved higher segmentation accuracy. The reason behind  
714 this is that the dual modelling can complement each other  
715 during the model training, thus can provide effective prediction  
716 perturbation for consistency-based learning. We also visualize  
717 the examples estimated by local modelling branch and global  
718 modelling branch in different training epochs as shown in Fig.  
719 9. One can see that the absolute difference between the local  
720 modelling and global modelling demonstrates that the local  
721 modelling branch and local modelling branch are modelled  
722 separately, which can provide effective prediction perturbation  
723 for consistency based learning.

724 **(2) Availability of complementary domain information:**  
725 Fig. 10 (a) provides the experiment results on hierarchical  
726 dual consistency learning with/without orthogonal weight con-  
727 straint. Fig. 10 (b) provides examples of feature correlations  
728 between corresponding layers of two dual-modelling networks  
729 with/without orthogonal weight constraint. It is observed that  
730 while removing the orthogonal weight constraint for inter-  
731 domain semi-supervised learning, the model segmentation  
732 performance is dropped. Meanwhile, the feature correlations  
733 between two dual-modelling networks become higher. The  
734 reason behind this is that the inter-domain semi-supervised  
735 learning with the orthogonal weight constraint can provide  
736 more effective prediction perturbation for consistency based

**TABLE VII**  
**PERFORMANCE COMPARISON BETWEEN DUAL STRUCTURE (LOCAL-GLOBAL) AND NON-DUAL STRUCTURES (LOCAL-LOCAL AND GLOBAL-GLOBAL) IN TERMS OF DCS, JI AND ASD. THE RESULTS ARE PRESENTED IN THE FORM OF THE MEAN (STANDARD DEVIATION).**  
**ABBREVIATIONS:** DSC, DICE SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD, AVERAGE SURFACE DISTANCE; LOCAL, LOCAL MODELLING NETWORK; GLOBAL, GLOBAL MODELLING NETWORK.

| Method        | Metrics           |                   |                 |
|---------------|-------------------|-------------------|-----------------|
|               | DSC               | JI                | ASD             |
| Local-Local   | $0.875 \pm 0.071$ | $0.785 \pm 0.102$ | $2.66 \pm 4.18$ |
| Global-Global | $0.879 \pm 0.069$ | $0.789 \pm 0.098$ | $3.01 \pm 4.80$ |
| Local-Global  | $0.893 \pm 0.048$ | $0.809 \pm 0.073$ | $1.90 \pm 2.81$ |



**Fig. 9:** Visualization for the evolution of dual-modelling results (first row and second row) and their absolute difference (third row). The second to fourth columns correspond to the estimated LAs from 5, 25 and 50 epochs during model learning.

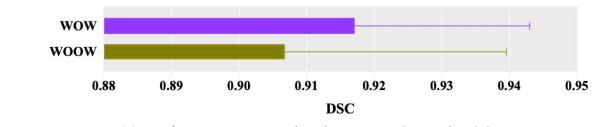
learning. It is also observed that while removing the orthogonal weight constraint for inter-domain semi-supervised learning, the feature correlations between two dual-modelling networks are not high ( $< 0.3$ ). In this case, two dual-modelling networks also can learn the complementary domain knowledge for providing effective prediction perturbation, thus achieving a high segmentation accuracy of 0.907 in terms of DSC.

#### F. The Effects of Parameter Settings on Model Performance

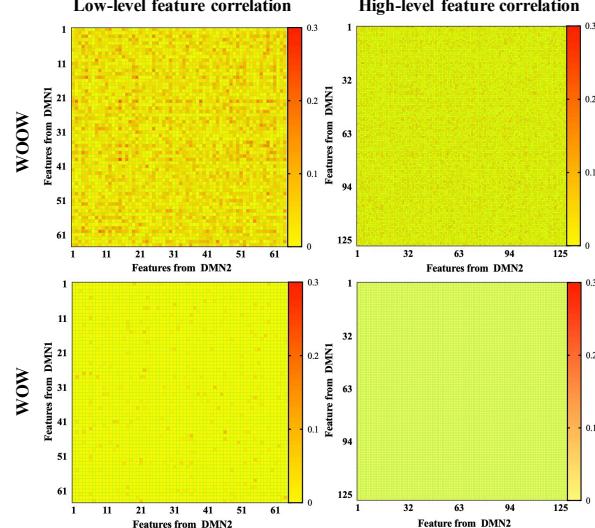
TABLE VIII presents the performances of our model for the LA segmentation using different parameter settings. It is observed that our model achieves the best performance when the patch size and the  $\lambda_{ow}$  are set as  $8 \times 8$  and 0.1, respectively.

## VI. DISCUSSION

In this study, we have developed a semi-supervised LA segmentation framework for generalising across domains. The



(a) Performance comparison between WOW and WOOW



(b) Feature correlation comparison between WOW and WOOW

**Fig. 10:** Segmentation performance comparison and feature correlation analysis between AHDC without orthogonal weights (WOOW) and AHDC with orthogonal weights (WOW). The experiments were performed on C1 and C2 (C1 supports C2). DMN1 and DMN2 represent two dual-modelling networks, respectively. The grape and asparagus bars denote the mean values with standard deviations.

**TABLE VIII**  
**PARAMETER VALIDATION FOR AHDC FRAMEWORK. THE RESULTS ARE PRESENTED IN THE FORM OF MEAN  $\pm$  STANDARD DEVIATION. ABBREVIATIONS: DSC, DICE SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD, AVERAGE SURFACE DISTANCE.**

| Parameter      | Value          | DSC               | JI                | ASD (mm)        |
|----------------|----------------|-------------------|-------------------|-----------------|
| Patch Size     | $4 \times 4$   | $0.874 \pm 0.062$ | $0.780 \pm 0.086$ | $2.12 \pm 2.22$ |
|                | $8 \times 8$   | $0.893 \pm 0.048$ | $0.809 \pm 0.073$ | $1.90 \pm 2.81$ |
|                | $16 \times 16$ | $0.881 \pm 0.069$ | $0.794 \pm 0.098$ | $2.40 \pm 3.85$ |
| $\lambda_{ow}$ | 0.0            | $0.907 \pm 0.033$ | $0.831 \pm 0.052$ | $1.34 \pm 1.58$ |
|                | 0.1            | $0.917 \pm 0.026$ | $0.848 \pm 0.043$ | $1.17 \pm 1.60$ |
|                | 1.0            | $0.913 \pm 0.032$ | $0.841 \pm 0.052$ | $1.34 \pm 1.68$ |

semi-supervised LA segmentation framework comprises a BAI module and a HDC module. The effectiveness of each module has been validated in our ablation study presented in TABLE VI. It is of note that self-attention based global modelling requires more computational resources, which are proportional to the dimensions of the image. In our proposed framework, we performed the self-attention based global modelling branch on the image feature maps for correlating  $8 \times 8$  patches instead of all pixels, which greatly reduces the requirements of computational resources during model training. Besides, during the testing phase or the practical applications, the self-

765 attention based global modelling branches will be removed  
 766 from our proposed framework. Then, the LA targets will  
 767 be only predicted by the local-modelling branch with low  
 768 computational resources.

TABLE IX  
 PERFORMANCE COMPARISON BETWEEN THE VANILLA  
 MODEL AND OUR PROPOSED AHDC USING ALL THE DATA  
 AVAILABLE FROM FOUR CENTRES. ABBREVIATIONS: LX  
 (%) : THE RATIO OF LABELLED DATA IN THE TRAINING SET  
 OF CENTRE X; UX (%) : THE RATIO OF UNLABELLED DATA  
 IN THE TRAINING SET OF CENTRE X; DSC, DICE  
 SIMILARITY COEFFICIENT; JI, JACCARD INDEX; ASD,  
 AVERAGE SURFACE DISTANCE.

| Method | Rate         |              |              |              | Metrics       |               |               |
|--------|--------------|--------------|--------------|--------------|---------------|---------------|---------------|
|        | L1/U1<br>(%) | L2/U2<br>(%) | L3/U3<br>(%) | L4/U4<br>(%) | DSC           | JI            | ASD           |
| U-Net  | 0/0          | 100/0        | 0/0          | 0/0          | 0.923 ± 0.025 | 0.858 ± 0.042 | 1.20 ± 1.96   |
| AHDC   | 100/0        | 100/0        | 100/0        | 100/0        | 0.927 ± 0.022 | 0.864 ± 0.037 | 0.728 ± 0.578 |
|        | 0/100        | 100/0        | 0/100        | 0/100        | 0.938 ± 0.015 | 0.883 ± 0.026 | 0.506 ± 0.164 |

769 The AHDC requires the complementary domain information  
 770 for inter-domain learning. For multi-centre studies, although  
 771 the domains from different sources exhibit heterogeneous  
 772 properties [11], they still share some specific information  
 773 because they come from the same image modality of LGE.  
 774 To make the model focus on the heterogeneous properties  
 775 of different domains for the inter-domain learning, we use  
 776 an orthogonal weight constraint to extract the conditional  
 777 independent features of different domains for subsequent target  
 778 modelling. We have explored the effectiveness of the orthogonal  
 779 weight constraint together with its weight coefficient  $\lambda_{ow}$   
 780 for the inter-domain learning. As the experiment results are  
 781 shown in TABLE VIII, one can see that the orthogonal weight  
 782 constraint generally improves the segmentation accuracy. Fur-  
 783 thermore, the performance of AHDC is not very sensitive to  
 784 the  $\lambda_{ow}$  values of 0.1 and 1.0 while using the orthogonal  
 785 weight constraint. Therefore, the orthogonal weight constraint  
 786 could exploit the heterogeneous properties among different  
 787 domains for inter-domain learning.

788 Considering the data annotation scarcity in medical image  
 789 analysis, our proposed method only requires the labelled  
 790 data from one of the multiple centres during cross-domain  
 791 learning, thus further reducing the dependence of the model  
 792 on annotated data. As the experiment results are shown in  
 793 TABLE III and TABLE IV, our proposed method is able  
 794 to generalise across two different domains simultaneously.  
 795 We further explore how the task model generalises across  
 796 multiple domains. Specifically, we have applied our proposed  
 797 method to the LGE CMRI data available from four centres.  
 798 We also trained a vanilla model (U-Net) with the LGE CMRI  
 799 data available from a single target centre and all the LGE  
 800 CMRI data available from four centres for comparison. As  
 801 the experiment results are shown in TABLE IX, compared  
 802 with the results obtained by using all annotated data from a  
 803 single domain, using all the data available from four centres  
 804 only makes small improvements in the segmentation accuracy  
 805 due to the domain shift and the label variations from different  
 806 centres. While our proposed AHDC generally improves the  
 807 segmentation accuracy, which indicates its ability for cross-

domain semi-supervised learning.

808

## VII. CONCLUSION

809

In this paper, we proposed an adaptive hierarchical dual consistency for the cross-domain semi-supervised LA segmentation. The adaptive hierarchical dual consistency firstly overcomes the distribution difference and sample mismatch of different domains by the bidirectional adversarial inference. Then, it explores the complementary modelling and domain information in intra-domain and inter-domain for semi-supervised LA segmentation based on the hierarchical dual consistency. Comprehensive experiments on four 3D LGE CMR datasets and one CT dataset demonstrated the feasibility and superiority of our proposed method for the cross-domain semi-supervised LA segmentation.

820

## REFERENCES

822

- [1] O. Razeghi, I. Sim, C. H. Roney, R. Karim, H. Chubb, J. Whitaker, L. O'Neill, R. Mukherjee, M. Wright, M. O'Neill, *et al.*, "Fully automatic atrial fibrosis assessment using a multilabel convolutional neural network," *Circulation: Cardiovascular Imaging*, vol. 13, no. 12, p. e011512, 2020.
- [2] Z. Xiong, V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 515–524, 2019.
- [3] G. Yang, J. Chen, Z. Gao, S. Li, H. Ni, E. Angelini, T. Wong, R. Mohiaddin, E. Nyktari, R. Wage, *et al.*, "Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention," *Future Generation Computer Systems*, vol. 107, pp. 215–228, 2020.
- [4] J. Chen, G. Yang, H. Khan, H. Zhang, Y. Zhang, S. Zhao, R. Mohiaddin, T. Wong, D. Firmin, and J. Keegan, "Jas-gan: Generative adversarial network based joint atrium and scar segmentation on unbalanced atrial targets," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [5] X. Zhang, M. Noga, D. G. Martin, and K. Punithakumar, "Fully automated left atrium segmentation from anatomical cine long-axis mri sequences using deep convolutional neural network with unscented kalman filter," *Medical Image Analysis*, vol. 68, p. 101916, 2021.
- [6] Z. Xiong, Q. Xia, Z. Hu, N. Huang, S. Vesal, N. Ravikumar, A. Maier, C. Li, Q. Tong, W. Si, *et al.*, "A global benchmark of algorithms for segmenting late gadolinium-enhanced cardiac magnetic resonance imaging," *Medical Image Analysis*, vol. 67, 2021.
- [7] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 605–613.
- [8] X. Cao, H. Chen, Y. Li, Y. Peng, S. Wang, and L. Cheng, "Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation," *IEEE transactions on medical imaging*, vol. 40, no. 1, pp. 431–443, 2021.
- [9] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.
- [10] Y. Yang, K.-T. Wang, D.-C. Zhan, H. Xiong, and Y. Jiang, "Comprehensive semi-supervised multi-modal learning," in *International Joint Conference on Artificial Intelligence*, 2019, pp. 4092–4098.
- [11] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martín-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma, *et al.*, "Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m&m challenge," *IEEE Transactions on Medical Imaging*, 2021.
- [12] V. Cheplygina, I. P. Pena, J. H. Pedersen, D. A. Lynch, L. Sørensen, and M. de Brujne, "Transfer learning for multicenter classification of chronic obstructive pulmonary disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1486–1496, 2018.
- [13] W. Dong-DongChen and Z.-H. WeiGao, "Tri-net for semi-supervised deep learning," in *International Joint Conferences on Artificial Intelligence*, 2018, pp. 2014–2020.

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

- 875 [14] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain  
876 adaptation: A survey of recent advances," *IEEE signal processing  
877 magazine*, vol. 32, no. 3, pp. 53–69, 2015.
- 878 [15] K. Li, L. Yu, S. Wang, and P.-A. Heng, "Towards cross-modality  
879 medical image segmentation with online mutual knowledge distillation,"  
880 in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.  
881 34, no. 1, 2020, pp. 775–783.
- 882 [16] Y. Zhang, Y. Wei, Q. Wu, P. Zhao, S. Niu, J. Huang, and M. Tan, "Col-  
883 laborative unsupervised domain adaptation for medical image diagnosis,"  
884 *IEEE Transactions on Image Processing*, vol. 29, pp. 7834–7844, 2020.
- 885 [17] J. Chen, H. Zhang, Y. Zhang, S. Zhao, R. Mohiaddin, T. Wong,  
886 D. Firmin, G. Yang, and J. Keegan, "Discriminative consistent domain  
887 generation for semi-supervised learning," in *International Conference  
888 on Medical Image Computing and Computer-Assisted Intervention*,  
889 Springer, 2019, pp. 595–604.
- 890 [18] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Unsupervised  
891 bidirectional cross-modality adaptation via deeply synergistic image and  
892 feature alignment for medical image segmentation," *IEEE transactions  
893 on medical imaging*, vol. 39, no. 7, pp. 2494–2505, 2020.
- 894 [19] C. Chen, Z. Chen, B. Jiang, and X. Jin, "Joint domain alignment and  
895 discriminative feature learning for unsupervised deep domain adapta-  
896 tion," in *Proceedings of the AAAI Conference on Artificial Intelligence*,  
897 vol. 33, 2019, pp. 3296–3303.
- 898 [20] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro,  
899 "High-resolution image synthesis and semantic manipulation with condi-  
900 tional gans," in *Proceedings of the IEEE conference on computer vision  
901 and pattern recognition*, 2018, pp. 8798–8807.
- 902 [21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-  
903 image translation using cycle-consistent adversarial networks," in *Pro-  
904 ceedings of the IEEE international conference on computer vision*, 2017,  
905 pp. 2223–2232.
- 906 [22] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual  
907 learning for image-to-image translation," in *Proceedings of the IEEE  
908 international conference on computer vision*, 2017, pp. 2849–2857.
- 909 [23] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to  
910 discover cross-domain relations with generative adversarial networks," in  
911 *International Conference on Machine Learning*, PMLR, 2017, pp. 1857–  
912 1865.
- 913 [24] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Ar-  
914 jovsky, and A. Courville. (2017). "Adversarially learned inference."  
915 [online]. Available: <https://arxiv.org/abs/1606.00704>.
- 916 [25] J. Donahue, P. Krähenbühl, and T. Darrell. (2017). "Adversarial feature  
917 learning." [online]. Available: <https://arxiv.org/abs/1605.09782>.
- 918 [26] C. Li, H. Liu, C. Chen, Y. Pu, L. Chen, R. Henao, and L. Carin,  
919 "Alice: towards understanding adversarial learning for joint distribution  
920 matching," in *Proceedings of the 31st International Conference on  
921 Neural Information Processing Systems*, 2017, pp. 5501–5509.
- 922 [27] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learn-  
923 ing," in *International Conference on Learning Representations*, 2017.
- 924 [28] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data aug-  
925mentation for consistency training," in *Advances in Neural Information  
926 Processing Systems*, 2020.
- 927 [29] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel,  
928 E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-  
929 supervised learning with consistency and confidence," in *Advances in  
930 Neural Information Processing Systems*, 2020.
- 931 [30] A. Tarvainen and H. Valpola, "Mean teachers are better role mod-  
932 els: Weight-averaged consistency targets improve semi-supervised deep  
933 learning results," in *Proceedings of the 31st International Conference  
934 on Neural Information Processing Systems*, 2017, pp. 1195–1204.
- 935 [31] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training  
936 for semi-supervised image recognition," in *Proceedings of the european  
937 conference on computer vision (eccv)*, 2018, pp. 135–152.
- 938 [32] Y. Xia, D. Yang, Z. Yu, F. Liu, J. Cai, L. Yu, Z. Zhu, D. Xu,  
939 A. Yuille, and H. Roth, "Uncertainty-aware multi-view co-training for  
940 semi-supervised medical image segmentation and domain adaptation,"  
941 *Medical Image Analysis*, p. 101766, 2020.
- 942 [33] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview:  
943 Recent progress and new challenges," *Information Fusion*, vol. 38,  
944 pp. 43–54, 2017.
- 945 [34] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai,  
946 T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al,  
947 "An image is worth 16x16 words: Transformers for image recognition at  
948 scale," in *International Conference on Learning Representations*, 2021.
- 949 [35] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and  
950 S. Zagoruyko, "End-to-end object detection with transformers," in  
951 *European Conference on Computer Vision*, Springer, 2020, pp. 213–229.
- [36] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural net-  
952 works," in *Proceedings of the IEEE conference on computer vision and  
953 pattern recognition*, 2018, pp. 7794–7803.  
954
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,  
955 L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances  
956 in neural information processing systems*, 2017, pp. 5998–6008.  
957
- [38] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image  
958 segmentation through dual-task consistency," in *Proceedings of the AAAI  
959 Conference on Artificial Intelligence*, vol. 35, 2021, pp. 8801–8809.  
960
- [39] K. Li, S. Wang, L. Yu, and P.-A. Heng, "Dual-teacher: Integrating intra-  
961 domain and inter-domain teachers for annotation-efficient cardiac seg-  
962 mentation," in *International Conference on Medical Image Computing  
963 and Computer-Assisted Intervention*, Springer, 2020, pp. 418–427.  
964
- [40] D. C. Peters, J. V. Wylie, T. H. Hauser, R. Nezafat, Y. Han, J. J. Woo,  
965 J. Taclas, K. V. Kissinger, B. Goddu, M. E. Josephson, et al, "Recur-  
966 rence of atrial fibrillation correlates with the extent of post-procedural  
967 late gadolinium enhancement: a pilot study," *JACC: Cardiovascular  
968 Imaging*, vol. 2, no. 3, pp. 308–316, 2009.  
969
- [41] M. Haissaguerre, P. Jaïs, D. C. Shah, A. Takahashi, M. Hocini, G. Quin-  
970 ion, S. Garrigue, A. Le Mouroux, P. Le Métayer, and J. Clémenty,  
971 "Spontaneous initiation of atrial fibrillation by ectopic beats originating  
972 in the pulmonary veins," *New England Journal of Medicine*, vol. 339,  
973 no. 10, pp. 659–666, 1998.  
974
- [42] J. Keegan, P. Jhooti, S. V. Babu-Narayan, P. Drivas, S. Ernst, and  
975 D. N. Firmin, "Improved respiratory efficiency of 3d late gadolinium  
976 enhancement imaging using the continuously adaptive windowing strat-  
977 egy (claws)," *Magnetic resonance in medicine*, vol. 71, no. 3, pp. 1064–  
978 1074, 2014.  
979
- [43] J. Keegan, P. Drivas, and D. N. Firmin, "Navigator artifact reduction in  
980 three-dimensional late gadolinium enhancement imaging of the atria,"  
981 *Magnetic resonance in medicine*, vol. 72, no. 3, pp. 779–785, 2014.  
982
- [44] R. Karim, R. J. Housden, M. Balasubramaniam, Z. Chen, D. Perry,  
983 A. Uddin, Y. Al-Beyatti, E. Palkhi, P. Acheampong, S. Obom, et al,  
984 "Evaluation of current algorithms for segmentation of scar tissue from  
985 late gadolinium enhancement cardiovascular magnetic resonance of the  
986 left atrium: an open-access grand challenge," *Journal of Cardiovascular  
987 Magnetic Resonance*, vol. 15, no. 1, pp. 1–17, 2013.  
988
- [45] L. Li, V. A. Zimmer, J. A. Schnabel, and X. Zhuang. (2021). "Atrialgen-  
989 eral: Domain generalization for left atrial segmentation of multi-center  
990 lge mris." [online]. Available: <https://arxiv.org/abs/2106.08727>.  
991
- [46] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases  
992 for whole heart segmentation of mri," *Medical image analysis*, vol. 31,  
993 pp. 77–87, 2016.  
994
- [47] X. Zhuang, "Challenges and methodologies of fully automatic whole  
995 heart segmentation: a review," *Journal of healthcare engineering*, vol. 4,  
996 no. 3, pp. 371–407, 2013.  
997
- [48] X. Zhuang, W. Bai, J. Song, S. Zhan, X. Qian, W. Shi, Y. Lian, and  
998 D. Rueckert, "Multiatlas whole heart segmentation of ct data using  
999 conditional entropy for atlas ranking and selection," *Medical physics*,  
1000 vol. 42, no. 7, pp. 3822–3833, 2015.  
1001
- [49] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin,  
1002 "A registration-based propagation framework for automatic whole heart  
1003 segmentation of cardiac mri," *IEEE transactions on medical imaging*,  
1004 vol. 29, no. 9, pp. 1612–1625, 2010.  
1005
- [50] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncer-  
1006 tainty to weigh losses for scene geometry and semantics," in *Proceedings  
1007 of the IEEE conference on computer vision and pattern recognition*,  
1008 2018, pp. 7482–7491.  
1009
- [51] L. R. Dice, "Measures of the amount of ecologic association between  
1010 species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.  
1011
- [52] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image  
1012 segmentation: analysis, selection, and tool," *BMC medical imaging*,  
1013 vol. 15, no. 1, pp. 1–28, 2015.  
1014