Multitask Learning for Estimating Multitype Cardiac Indices in MRI and CT Based on Adversarial Reverse Mapping

Chengjin Yu, Zhifan Gao, Member, IEEE, Weiwei Zhang, Guang Yang, Member, IEEE, Shu Zhao, Heye Zhang, Yanping Zhang, and Shuo Li

Abstract—The estimation of multitype cardiac indices from cardiac magnetic resonance imaging (MRI) and computed tomography (CT) images attracts great attention because of its clinical potential for comprehensive function assessment. However, the most exiting model can only work in one imaging modality (MRI or CT) without transferable capability. In this article, we propose the multitask learning method with the reverse inferring for estimating multitype cardiac indices in MRI and CT. Different from the existing forward inferring methods, our method builds a reverse graph networks that maps the multitype cardiac indices to cardiac images. The task dependencies are then learned and shared to multitask learning networks using an adversarial training approach. Finally, we transfer the parameters learned from MRI to CT. A series of experiments were conducted in which we first optimized the performance of our framework via ten-fold cross-validation of over 2900 cardiac MRI images. Then, the fine-tuned network was run on an independent data set with 2360 cardiac CT images. The results of all the experiments conducted on the proposed adversarial reverse mapping show excellent performance in estimating multitype cardiac indices.

Index Terms—Adversarial training, multitask learning, multitype cardiac indices, reverse mapping.

I. INTRODUCTION

MULTITYPE cardiac indices (see Fig. 1) estimation faces two challenges. Primarily, the complicated relationships between multitype cardiac indices are the main difficult (i.e., the mining and the representation of commonalities and differences among indices) for learning proper task dependencies. These indices have considerably different dimensions (i.e., 2-D cardiac segmentation, 1-D linear indices, and planar indices). Concerning specific indices, regional wall thicknesses (WT) also vary according to the orientation of the myocardium segments in different regions [1]–[4]. Furthermore, most of existing multitask methods can only work on one image modality because of the significantly different in appearance between magnetic resonance imaging (MRI) and computed tomography (CT). MRI can accurately define cardiovascular structures and characterize tissue composition [5], while short- and long-axis CT can be used for cardiac function analysis [6]. This diagnostic capability has attracted considerable attention because these MRI and CT techniques allow angiography to be performed noninvasively [1], [6]. However, MRI may not be used for patients with cardiac rhythm management devices for safety reasons [7]. In such circumstances, knowledge transfer from CT to MRI is necessary to obtain a better clinical evaluation.

Existing multitask learning methods may be unsuitable for estimating multitype cardiac indices in different modalities for two reasons. First, they are unable to learn the complex task dependencies among multitype cardiac indices.

1) For the assumption-based regularization methods, task dependencies are assumed to be known as the prior information, which includes a shared set of features [8], and shared cluster structures among tasks [9]. However, these assumptions are not always accurate or suited to all tasks.

2) Data-driven methods can learn task dependencies automatically from training data without any assumptions [10], [11]. However, most of these methods express task dependencies via a covariance matrix, which might be insufficient for 2-D and 1-D indices because the relationship between the indices cannot be represented directly owing to their different dimensions [12].
Second, existing multitask learning methods cannot efficiently transfer the task dependencies learned to other modalities. This is because most existing multitask learning methods only mine the relationships among tasks [12]–[14] but lack a mechanism for sharing the learned knowledge between different image modalities, especially for complex task dependencies.

To address the abovementioned challenges, a multitask learning method based on adversarial reverse mapping is proposed for estimating multitype cardiac indices in different imaging modalities (in this case, MRI and CT). The proposed method investigates multitask learning and the knowledge transfer problem from the perspective of a reverse generation that further learns the mapping from multitype cardiac indices to cardiac images via an adversarial training approach. This is loosely inspired by the recent progress of reverse mapping and adversarial training [15], [16], which shows the promising ability for modeling the data distribution, especially for assessing joint distributions of complex semantic variables [17]. As shown in Fig. 2, in addition to the multitask learning network that learns the mapping from cardiac image to multitype cardiac indices, we propose a reverse mapping network that further learns the reverse mapping to facilitate mining and the representation of task dependencies. Then, an adversarial training approach is applied to integrate the multitask learning network and the reverse mapping network. Finally, the parameters from the two networks learned from the source modality (MRI) are shared with the target modality (CT).

Specifically, our contributions also include the following.

1) We propose an efficient multitask learning framework based on adversarial reverse mapping that can obtain task dependencies and can also be applied to data in a different modality.

2) We propose a reverse mapping network that can facilitate the mining of task dependencies. This network reveals the role of each cardiac index and the dependencies among different cardiac indices from a generational perspective. Thus, the reverse mapping network can act as a regularization enabling the multitask network to learn task dependencies.

3) We propose a symmetric adversarial training for realizing the regularization of the reverse network to the multitask learning network. Learning occurs in this adversarial training, and the joint distribution from the reverse mapping network and the multitask network are matched. The convexity of the matching problem is then proven, which indicates that the reverse mapping network can effectively constrain the learning of the multitask network.

4) Furthermore, we propose a bidirectional parameter sharing scheme that shares the parameters learned from both the multitask learning network and the reverse mapping network to facilitate training on different modalities. Compared with parameters from one-way deep neural networks, our strategy is more powerful for representing complex task dependencies and is, therefore, more effective at sharing them.

5) Finally, our comprehensive experiments show promising results on estimating multitype cardiac indices from MRI and CT, which validates the feasibility of our adversarial reverse mapping-based multitask learning framework.

A. Estimation of Multitype Cardiac Indices

Existing cardiac indices estimation methods can be categorized into two classes: 1) segmentation methods and 2) regression methods.

1) Segmentation Methods: Existing segmentation-based method, Max Flow [18], initially performs cardiac segmentation and then computes other cardiac indices from the cardiac segmentation results. However, cardiac segmentation is still a challenge. Previously, cardiac segmentation methods were based on traditional techniques, such as region growing [19] and clustering [20], and energy minimization techniques, such as graph cuts [21]. Other model-based methods include the multivariate mixture model for cardiac segmentation from multisquence MRI [22] and the statistical shape model-based methods [23]. Recently, deep neural networks have demonstrated powerful capability for cardiac image segmentation. FCN [24] is a basic method for biomedical image segmentation. GAN [25] improved the FCN based on adversarial training, and V-net [26] improved the FCN based on 3-D convolutional neural network (CNN) that is often used in 3-D image. Patravali et al. [27] used a 2-D CNN to implement slice-by-slice segmentation and a 3-D CNN for 2D + T MRI. In contrast to most existing 3-D methods, Qiao et al. [28] proposed a more advanced approach to obtaining 3-D context information using spatial propagation and have achieved promising results. Poudel et al. [29] and Alayba et al. [30] combined a 2-D CNN and recurrent neural
Regression methods that calculate the cardiac indices directly from cardiac images instead of using cardiac segmentation have recently attracted a lot of attention. The benefit of regression methods is that they can estimate these indices directly from cardiac images more efficiently. However, for segmentation methods, additional computational cost and intermediate operations are required to calculate 1-D cardiac indices from segmentation results. In contrast, our method estimates 1-D cardiac indices and 2-D segmentation jointly. This is because the joint estimation directly models the relationship between image appearance and the cardiac indices, which allows the model to take further advantage of the dependencies between 1-D cardiac indices and 2-D segmentation. Therefore, our method provides more information more efficiently.

2) Regression Methods: Regression methods that calculate the cardiac indices directly from cardiac images instead of using cardiac segmentation have recently attracted a lot of attention. The benefit of regression methods is that they can estimate these indices directly from cardiac images more conveniently, efficiently, and accurately by directly modeling the relationship between the cardiac image appearance and cardiac indices. Therefore, regression methods reduce computational cost and avoid errors induced by intermediate operations that may be applicable in clinical applications. Primarily, regression methods can be divided into two categories: two-phase methods and multitask learning methods. Most two-phase methods usually extract a feature representation and then employ a regression model to achieve the final estimation. The feature representation can include statistical features, meshfree representation, CNN-based feature representation, multifeatures composed of pyramidal Gabor features, steerable features and a histogram of oriented gradients; the features extracted by an unsupervised multiscale convolutional deep belief network and a supervised descriptor. Regression models include artificial neural networks, random forests, K-cluster regression forests, and deep regression networks. Regarding multitask learning methods, recent efforts are primarily based on deep multitask learning. Xue et al. proposed a framework of joint representation and regression learning for extracting task-aware features. Several other studies have also modeled the dependencies among different tasks and obtained promising results.

However, most existing methods focus only on MRI but pay no attention to CT, which is also important for clinical diagnosis. In contrast, the proposed method can estimate multitype cardiac indices from MRI and can also be transferred to CT.

B. Reverse Mapping and Generative Adversarial Networks

Reverse mapping aims to improve the learned model via learning the reverse process of the main task. An important research area is dual learning used for translation tasks. It has two models: one model that translates English into French and a reverse model that translates French into English. During the training process, the two models would be complementary via the performance of dual tasks. Recently, reverse mapping has also been employed for multitask learning. A representative study is, which first predicts segmentation and tags via a multitask network and then regenerates original images from the predicted results via a reverse mapping network.

Generative adversarial networks (GANs) show impressive results for data synthesis. In a standard GAN, the generator would try to generate fake samples to fool the discriminator, and the discriminator would try to distinguish between fake and real samples. In this game, if the discriminator cannot distinguish between fake and real samples correctly, we suppose that the generator would model the
data distribution accurately. Recently, some studies also take advantage of both GAN and reverse mapping (e.g., in realizing image translation). Dual GAN [50] employs two GANs to learn dual tasks. Cycle GAN [51] and Disco GAN [52] concatenate the two generators to ensure cyclic consistency.

II. METHOD

A. Overview of Our Method

Our proposed method uses a multitask learning framework with a reverse mapping network that reconstructs the image from learned features and multitype cardiac indices (see Section II-B), an adversarial training approach to solve multitask learning (see Section II-C), and a bidirectional parameter sharing mechanism (see Section II-D). Notably, we first inference a shared feature representation to predict multitype cardiac indices. Then, we also reconstruct the original images from the learned feature representation and predicted multitype cardiac indices, as illustrated in Fig. 2. Furthermore, we propose a discriminator network to distinguish between the two joint distributions from the multitask learning network and the reverse mapping network. In this case, we treat the discriminator network as a binary classification subtask in the multitask learning framework. We then explicitly learn the multitask relation. Finally, we transfer the learned parameters and multitask relation to different imaging modalities, as illustrated in Fig. 3.

B. Multitask Learning With Reverse Mapping for Estimating Multitype Cardiac Indices

Consider multitype cardiac indices \( I = \{i_{\text{seg}}, i_{\text{linear}}, i_{\text{planar}}\} \), and \( i_{\text{seg}}, i_{\text{linear}}, \) and \( i_{\text{planar}} \) denote segmentation, linear indices, and planar indices, respectively. The prior distribution on the multitype cardiac indices is denoted by \( p(I) \). Considering an observed cardiac image \( x \), the prior distribution on the cardiac image is denoted by \( q(x) \). Most existing methods aim to establish multitask learning mapping \( G_1 : x \rightarrow I \) to accurately predict \( I \). In this study, we further learn a reverse mapping \( G_2 : I \rightarrow x \) that generates \( x \) from \( I \). This reverse mapping establishes the mapping relationship from a reverse perspective capable of revealing the complex relationships between \( x \) and \( I \).

A common practice to learn \( G_1 \) and \( G_2 \) is maximum likelihood estimation based on parameterized conditional distributions \( q_o(I|x) \) and \( p_o(x|I) \)

\[
G_1(x; \varphi) = \arg \max_I q_o(I|x) \\
G_2(I; \theta) = \arg \max_x p_o(x|I)
\]

where \( \varphi \) and \( \theta \) denote the learned parameters of the two distributions. Apparently, \( G_1 \) and \( G_2 \) are reverse processes of each other. We aim to learn \( G_1 \) and \( G_2 \) jointly and enable them to complement each other. As shown in Fig. 2, we consider two ways to learn \( G_1 \) and \( G_2 \).

1) We assume that \( X \) are the observable variables, thus concatenating \( G_1 \) and \( G_2 \). Then, we have a function \( f : \hat{X} \rightarrow \hat{X} \), defined by \( f(x) = G_2(G_1(x)) \) for all \( x \) in \( \hat{X} \). In this case, \( G_1 \) first predicts \( I \) from the observed cardiac image \( x \), and then, \( G_2 \) regenerates \( x \) from the predicted \( I \).

2) We assume that \( I \) are the observable variables, thus concatenating \( G_2 \) and \( G_1 \). Then, we have another function \( g : \hat{I} \rightarrow \hat{I} \), defined by \( g(I) = G_1(G_2(I)) \) for all \( I \) in \( \hat{I} \). In this case, \( G_2 \) first predicts \( x \) from the multitype cardiac index \( I \), and then, \( G_1 \) regenerates \( I \) from the predicted \( x \). A mathematical description of the learning of \( f \) and \( g \) is as follows:

\[
f(x; \theta, \varphi) = \arg \max_x p_o(x) \arg \max_I q_o(I|x) \\
g(I; \theta, \varphi) = \arg \max_I q_o(I) \arg \max_x p_o(x|I).
\]

In previous studies, \( f \) and \( g \) are solved using an \( L_1 \)-based reconstruction loss that focuses on image translation but pays no attention to multitask problems [53]. The reverse mapping network reveals the role of each index (task) and dependence among different indices (tasks) from a generational perspective. It can act as a regularization enabling the multitask network to learn task dependencies via checking the cardiac image reconstructed from predicted cardiac indices. Notably, the reverse mapping network is trained to generate...
Then, we can rewrite (5) and (6) as the information entropy of $q_\phi(x, I)$ learned by the multilabel training network and the joint distribution $p_\theta(x, I) = p_\theta(x) p(I)$ learned by the reverse mapping network would be matched. Then, we consider the Kullback–Leibler (KL) divergence [16] between $q_\phi(x, I)$ and $p_\theta(x, I)$

$$\mathcal{L}_f = \mathbb{E}_{q(I)} \log p_\theta(x) - KL(q_\phi(x, z, I) || p_\theta(x, I))$$

$$\mathcal{L}_g = \mathbb{E}_{p(I)} \log q_\phi(I) - KL(p_\theta(x, z, I) || q_\phi(x, z, I))$$

(7)

(8)

where $p_\theta(x) = \int p_\theta(x|I)p(I)dI$ such that it synthesizes samples that are well matched to those drawn from $q(x)$. $q_\phi(x, I)$ is $\int q_\phi(x|I)q(I)dx$ such that it synthesizes samples that are well matched to those drawn from $p(I)$. Therefore, (5) and (6) can be regarded as automatic encoders, which would be well learned if $q_\phi(x, I)$ and $p_\theta(x, I)$ were matched. However, it is difficult to directly learn (5) and (6) because the information entropy of $I$ is lossy compared with that of $x$. To address this problem, we introduce a feature map $z$ to increase the information entropy. On implementation, our reverse mapping reconstructs the cardiac image from $z$ and $I$. Then, we can rewrite (5) and (6) as

$$\mathcal{L}_f = \mathbb{E}_{q(I)} \log p_\theta(x) - KL(q_\phi(x, z, I) || p_\theta(x, z, I))$$

(9)

$$\mathcal{L}_g = \mathbb{E}_{p(I)} \log q_\phi(I) - KL(p_\theta(x, z, I) || q_\phi(x, z, I))$$

(10)

Based on (7) and (8), our goal becomes learning and matching $q_\phi(x, z, I)$ and $p_\theta(x, z, I)$. In Section II-C, we employ an adversarial training approach to learn and match the two distributions simultaneously.

C. Adversarial Training Approach for Joint Distribution Matching

To effectively learn and match $q_\phi(x, z, I)$ and $p_\theta(x, z, I)$, we consider the sampling method shown in Fig. 2: For the function $f$, we can first sample $(x, z, I)$, where $x$ is the real cardiac image, $z$ is the feature map predicted by $G_1$, and $I$ is estimated by $G_1$. Then, we can sample $(x, z, I)$, where $x$ is the image reconstructed from $z$ and the real $I$ by $G_2$. For the function $g$, we can first sample $(x, z, I)$, where $x$ is a cardiac image reconstructed by $G_2$, $z$ is a feature map predicted by $G_2$, and $I$ is a real multitype cardiac index. Then, we can sample $(x, z, I)$, where $x$ is the real image, and $z$ and $I$ are predicted by $G_1$. Because $z$ is a feature representation of $x$, we do not consider the local relationship between $x$ and $z$; and then, let $X = (x, z)$ for convenience. We use $q(X)$ to denote the prior distribution of $X$. Then, we have $q_\phi(X, I) = q_\phi(I|X)q(X)$ and $p_\theta(X, I) = p_\theta(X)p(I)$. Furthermore, we have $p_\theta(X) = \int p_\theta(X|I)p(I)dI$ and $q_\phi(I) = \int q_\phi(I|X)q(X)dX$.

Then, we use two real-valued discriminator networks, $T_{\psi_1}(X, I)$ and $T_{\psi_2}(X, I)$, parameterized using $\psi_1$ and $\psi_2$, respectively [54], to distinguish between joint distributions $(x, I)$. Precisely, we consider the following objectives:

$$\max_{T_{\psi_1}} \mathbb{E}_{X \sim q(X)} \mathbb{E}_{l \sim q_l(I|X)} \log \sigma(T_{\psi_1}(X, I)) + \mathbb{E}_{X \sim p_\theta(X,I), I \sim p(I)} \mathbb{E}_{l \sim p(l)} \log (1 - \sigma(T_{\psi_1}(X, I)))$$

(11)

$$\max_{T_{\psi_2}} \mathbb{E}_{l \sim p(l)} \mathbb{E}_{X \sim p_\theta(X,I), I \sim p(I)} \log \sigma(T_{\psi_2}(X, I)) + \mathbb{E}_{X \sim q(X), I \sim q(I)} \mathbb{E}_{l \sim q(l)} \log (1 - \sigma(T_{\psi_2}(X, I)))$$

(12)

In (9) and (10), $\sigma(t) = (1 + e^{-t})^{-1}$ denotes the sigmoid function. Intuitively, $T_{\psi_1}(X, I)$ tries to distinguish between joint distributions $(X, I)$ in $f$, while $T_{\psi_2}(X, I)$ tries to distinguish between joint distributions $(X, I)$ in $g$. If parameters $\theta$ and $\varphi$ are given, we can rewrite (9) and (10) as

$$T_{\psi_1}(X, I) = \log q_\phi(x, I) - \log p_\theta(X) - \log q_\phi(I)q(X)$$

(13)

$$T_{\psi_2}(X, I) = \log p_\theta(X) - \log q_\phi(x, I)$$

(14)

Therefore, the complete objective can be written as

$$\min_{\psi_1, \psi_2} \mathbb{E}_{q_\phi(x, I)} [\mathcal{L}_{\psi_1}(X, I) + \log p_\theta(X)]$$

(15)

$$\min_{\psi_1, \psi_2} \mathbb{E}_{p_\theta(x, I)} [\mathcal{L}_{\psi_2}(X, I) + \log q_\phi(I)|X)]$$

Algorithm 1 Multitask Learning With Adversarial Reverse Mapping

1. $\varphi, \theta, \psi_1, \psi_2 \leftarrow$ Initialize parameters
2. repeat
3. $[x^1 , \ldots, x^M]$, $[l^1 , \ldots, l^M] \leftarrow$ Random minibatch of $M$ images and corresponding $M$ multitype cardiac indices.
4. Compute $\varphi$ gradient:
   $$g_{\varphi} \leftarrow \frac{1}{M} \sum_{m=1}^M \Delta \{ -\log \sigma(T_{\psi_1}(x^m, I)) + \log p_\theta(x^m) \}$$
5. Compute $\theta$ gradient:
   $$g_{\theta} \leftarrow \frac{1}{M} \sum_{m=1}^M \Delta \{ -\log \sigma(T_{\psi_2}(x^m, I)) + \log q_\phi(I|x^m) \}$$
6. Compute $\psi_1$ gradient:
   $$g_{\psi_1} \leftarrow \frac{1}{M} \sum_{m=1}^M \Delta \{ \log \sigma(T_{\psi_1}(x^m, I)) + \log (1 - \sigma(T_{\psi_2}(x^m, I))) \}$$
7. Compute $\psi_2$ gradient:
   $$g_{\psi_2} \leftarrow \frac{1}{M} \sum_{m=1}^M \Delta \{ \log \sigma(T_{\psi_2}(x^m, I)) + \log (1 - \sigma(T_{\psi_1}(x^m, I))) \}$$
8. $\varphi, \theta, \psi_1, \psi_2 \leftarrow$ Update parameters using gradients $g_{\varphi}, g_{\theta}, g_{\psi_1}, g_{\psi_2}$ based on the SGD.
9. until convergence with parameters $(\varphi, \theta, \psi_1, \psi_2)$
10. return $\varphi, \theta, \psi_1, \psi_2$

Based on (12), we propose using Algorithm 1 to optimize the objects in (9) and (10) via the stochastic gradient descent (SGD). We regard (13) as a min–max game. As shown in Algorithm 1, we first update the parameters $\theta$ and $\varphi$ while keeping $\psi_1$ and $\psi_2$ fixed. Then, we update $\psi_1$ and $\psi_2$ using the adversarial objects in (9) and (10).

Finally, if the Nash-equilibrium of the min–max game in (13) is achieved via parameters $[\theta^*, \varphi^*, \psi_1^*, \psi_2^*]$, we have $p_{\theta^*}(X, I) \approx q_{\varphi^*}(X, I)$, which indicates that the knowledge
learned by the reverse mapping network is shared to the multitask learning network. Furthermore, due to \( X = (x, z) \), we can get \( p_{\theta}(x, z, l) \approx q_{\phi}(x, z, l) \). This indicates that our method can learn the joint distribution \( p_{\theta}(x, z, l) \) and \( q_{\phi}(x, z, l) \) via adversarial training. Furthermore, this also shows that the proposed method can establish a one-to-one relationship between \( l \) and \( x \) as well as a one-to-one relationship between \( l \) and \( z \). Furthermore, the information on \( x \) can be maintained in \( z \) to avoid steganography or the overfitting effect [55].

D. Bidirectional Parameter Sharing Scheme

Although the knowledge leveraged from MRI can reflect the common characteristics, such as cardiac structures and dependencies among multitype cardiac indices, shared in MRI and CT, additional effort is still required to represent and transfer such common characteristics.

In this section, we propose a new transfer learning scheme (as shown in Fig. 3), i.e., a bidirectional parameter sharing scheme that transfers the learned parameters from deep layers (colored blue) in both the multitask learning network and the reverse mapping network learned from MRI.

1) Structure of the Multitask Learning Network: We improved a dense convolutional network (DenseNet) [56] to fit our multitask learning network, which has one convolution layer \( (M_{\text{cnm}}) \) and three dense blocks with four, eight, and eight layers \( (M_{D1}, M_{D2}, \text{and } M_{D3}) \) (as shown in Fig. 3). Our model takes 2-D cardiac image slices across one cardiac cycle as input. Thus, a 3-D CNN is employed for all convolution layers because of its excellent capability in action modeling. For the first convolution layer, we use three different sizes with convolution kernel \( (3^3, 5^3, \text{and } 7^3) \) to generate hierarchical information for further learning. For the rest of the 3-D convolution layers in dense blocks, we use a \( (3^3) \) kernel. Then, we can obtain a feature map \( (z) \) output using the third dense block. Finally, one 3-D deconvolution-based pixel-level classifier \( (M_{\text{dcm}}) \) is employed to generate 2-D segmentation results from \( z \), and one fully connected network-based regression network \( M_{F} \) is employed to estimate 1-D cardiac indices from \( z \). Furthermore, the feature map \( z \) is also output by a branch fea. Notably, the outputs of \( M_{\text{dcm}} \) and \( M_{F} \) correspond to the 2-D segmentation and 1-D cardiac indices, respectively, in 2-D slices across one cardiac cycle.

2) Structure of the Reverse Mapping Network: Our reverse mapping network has a similar but reverse structure of the multitask learning network. It regenerates corresponding cardiac images from the given multitype cardiac indices. In particular, for \( f \), we first employ both convolution layers and fully connected layers as a joint learning network \( (R_{\text{cmn}}) \) to learn a joint representation from \( z \) and the real \( I \). For \( g \), the joint representation is directly generated from the real \( I \). Then, three dense blocks with eight, eight, and four layers \( (R_{D1}, R_{D2}, \text{and } R_{D3}) \) are employed for further learning. Finally, we reconstruct final cardiac images using deconvolution layers \( (R_{\text{dcm}}) \).

In contrast to existing methods that only share the parameters from one-way network, our bidirectional parameter sharing scheme shares the parameters of the multitask learning network and the reverse mapping network. The entire model is first trained on the MRI data using the adversarial training approach mentioned in Section II-B. Then, the parameters of layers \( M_{D1}, M_{D2}, M_{D3}, M_{\text{dcm}}, \text{and } M_{F} \) in the multitask learning network and layers \( R_{\text{cmn}}, R_{D1}, R_{D2}, \text{and } R_{D3} \) in the reverse mapping network are shared to CT images (as shown in Fig. 3, the shared layers are the blue parts of the networks). In this step, we concatenate new untrained layers \( M_{\text{cmn}} \) and \( R_{\text{dcm}} \) with the shared layers. Then, we fine-tune the new network using labeled CT images. For the multitask learning network, we employ the Dice-based [26] loss \( L_{\text{Dice}} \) for 2-D cardiac segmentation and \( L_{2} \) (ridge) [57] loss for 1-D cardiac indices. For the reverse mapping network, we employ the reconstruction loss \( L_{\text{recon}} = \|x - G_{2}(G_{1}(x))\|_{1} \) [50]. Then, the total loss can be written as

\[
L_{M&R} = \lambda_{1} L_{\text{Dice}} + \lambda_{2} L_{\text{recon}} + \lambda_{3} L_{2}.
\]

It is important to note that only \( L_{M&R} \) is used for model fine-tuning on the CT data. For the model training on the MRI data, we use a joint loss \( L_{\text{join}} = \lambda_{4} L_{\text{Adv}} + \lambda_{5} L_{M&R} \) that combines the adversarial training loss mentioned in Section II-C and \( L_{M&R} \) loss mentioned in Section II-D.

III. EXPERIMENTS

A. Data Acquisition

Our data include one MRI data set and two CT data sets. The MRI data are collected from a public data set (Left Ventricle Full Quantification Challenge of MICCAI 2018). There are 2900 2-D short-axis cine MR images of 145 subjects (20 frames for each subject) collected from three hospitals affiliated with two healthcare centers (London Healthcare Center and St. Josephs HealthCare). The subjects’ age is from 16 to 97 years, with an average of 58.9 years. The pixel spacings of the MR images range from 0.6836 to 2.0833 mm/pixel, with a mode of 1.5625 mm/pixel. The CT data sets are private, which includes one testing set with 360 images and one training set with 2000 images. All the images are 2-D short-axis CT images of 118 subjects (20 frames for each subject) collected from the Beijing Anzhen Hospital. The subjects’ age is from 38 to 85 years, with an average of 60.4 years. The pixel spacings of these CT images range from 0.5627 to 2.1232 mm/pixel, with a mode of 1.6435 mm/pixel.

All cardiac images undergo several preprocessing steps, including landmark labeling, rotation, ROI cropping, and resizing. The resulted images are approximately aligned with dimensions of \( 80 \times 80 \). Then, these cardiac images are manually contoured to obtain the epicardial and endocardial borders, which are double-checked by four experienced cardiac radiologists (two for MRI and two for CT). The interobserver error ranged between 0.33% and 4.11% (average: 1.21%) for MRI and between 0.42% and 5.57% (average: 1.63% for CT). The ground-truth values of the 1-D cardiac indices can be obtained from the two borders. The values of WTs and cavity dimensions are normalized by the image dimension, while the areas are normalized by the pixel number (6400). During the evaluation, the obtained results are converted to
physical thickness (in mm) and area (in mm²) by reversing the resizing procedure and multiplying by the pixel spacing for each subject.

B. Experimental Settings and Evaluation Criteria

In our experiments, in order to test the performance of our method on multitype cardiac indices estimation, ten-fold cross-validation is employed for performance evaluation for all comparison and ablation studies in MRI data set, and the reported results are the average of ten-fold cross-validation. To test the performance on transfer learning from MRI to CT, the model is trained with all MRI data, and the CT data set with 360 images are used for fine-tuning and testing. From CT to MRI, the other CT data set with 2000 images are first used to train the model, and then, the randomly selected 360 MRI images are used for fine-tuning and testing. The training set and test set are separated to ensure that the data from one patient will not appear in both training and testing simultaneously. In particular, we implemented all of the codes using Python on a Linux (Ubuntu 16.04) desktop computer with an Intel Xeon CPU E5-2650 and 64-GB DDR2 memory. The graphics card used is an Nvidia GTX1080 GPU. The deep learning libraries are implemented using Tensorflow. The hyperparameters are determined by model performance, and we adopt an SGD with 0.9 momenta. The learning rate is given by \( \eta_p = (\eta_0/(1 + \alpha p p)^\beta) \), where \( \eta_0 = 0.01 \), \( \alpha = 10 \), \( \beta = 0.75 \), and \( p \) is the training progress changing from 0 to 1 in proportion to the training step. Finally, we also balance the different losses used in the model: \( \lambda_1 = \lambda_2 = 0.01 \), and \( \lambda_3 = \lambda_4 = \lambda_5 = 1.0 \).

To measure the performance of the proposed method, we use two metrics for different cardiac indices. First, for the 2-D cardiac segmentation, the average Dice metric (Dice) is computed. Then, for the 1-D linear indices and planar indices, mean absolute error (MAE) is computed. The formal description of Dice is \( \left(2P_{\text{TE}}/P_T + P_E\right) \), where \( P_T \) denotes all pixels of manually segmented contour area and \( P_E \) denotes all pixels of automatically segmented contour area by different methods, while \( P_{\text{TE}} \) denotes the pixels of overlapping between \( P_T \) and \( P_E \). Dice always ranges between [0, 1]. The larger Dice value is, the higher consistency between manual and automated segmentation. Then, \( \text{MAE}(y, \hat{y}) = (1/N) \sum_{i=1}^{N} |y - \hat{y}| \), where \( y \in R^N \) and \( \hat{y} \in R^N \) are two vectors that denote true indices and estimated indices, respectively. To verify the effectiveness of multitask learning for the proposed method, we use two metrics to measure the multitask relationship. First, the multitask correlation matrix \( C \) for the cardiac indices, which is calculated from the cardiac indices covariate matrix [59] (the correlation matrix \( C \) only can be calculated from regression tasks and logical regression tasks according to [59]; therefore, we only calculate \( C \) from cardiac indices in our work.). In \( C \), the positive and negative values indicate positive and negative correlation, respectively. The higher absolute value indicates a strong correlation. Then, to further consider the segmentation and reverse generation tasks, we measure the performance gain of task correlation between two tasks \( T_i \) and \( T_j \), and we calculate the pairwise performance gain [60] as \( PPG_{ij} = (P_i P_j / P_i P_j)^{1/2} \), where \( P_i \), \( P_j \), \( P'_i \), and \( P'_j \) are the performances of tasks \( T_i \) and \( T_j \) when learned individually and jointly. (MAE is used to measure the PPG for all tasks. Thus, the smaller the PPG value, the greater the gain between two tasks.)

C. Experiments

To verify the effectiveness and performance of our method, we have done rich experiments as follows.

First, to evaluate the performance of our method for multitype cardiac indices estimation, we apply our method to cardiac MR images. Then, we fine-tune the trained model and apply it to a different imaging modality, i.e., CT images.

Second, to validate the ability to explore task dependencies of our adversarial reverse mapping network, we performed the following comparison and ablation studies. For comparison, we test our fully framework (Ours) on MRI, and then, we compared it with segmentation-based methods, e.g., Max Flow [18], two-phase methods, e.g., Multifeature + RF [38], SDL + AKSF [43], MCDBN + RF [41], and deep multitask learning methods, e.g., Indices-Net [2], FullLVNet [49], and DMTRL [3] due to their good results on MRI data. Then, we calculated the Dice value of 2-D segmentation results of our method and compared it with GAN [25], FCN [24], U-net [58], and V-net [26]. For ablation studies, we compare the performance from two aspects that include the different network architectures and different cardiac indices that the model predicted. For the different network architectures, both the multitask learning network and the reverse mapping network are employed. In this case, we first use \( L_{MR} \) loss to train the model \( (M + R + L_{MR}) \) without the adversarial loss. Then, we use the \( L_{MR} \) loss and the adversarial loss that only match the predicted \( I \) and real \( I \) to train the model \( (M + R + L_{MR} + I_{GAN}) \), and we train the framework without a reverse mapping network. In this case, we first use the \( L_{D_{\text{Dice}}} + L_{\text{2}} \) loss \( (M + L_{DL}) \) to train the model. Then, we use the \( L_{D_{\text{Dice}}} + L_{\text{2}} \) loss that matches only the predicted \( I \) and real \( I \) to train the model \( (M + L_{DL} + I_{GAN}) \) to train the model. For different cardiac indices, we employ the full model to predict different cardiac indices: 1) each type of cardiac indices independently (segmentation only, linear indices only, and planar indices only); 2) segmentation + linear indices; 3) segmentation + planar indices; 4) linear indices + planar indices; and 5) finally, when the model predicts only the segmentation, we also calculate the 1-D cardiac indices from the segmentation results.

Third, to show the effectiveness of our bidirectional parameter sharing scheme, we performed the following studies. First, we establish several multitask learning frameworks over different baseline deep networks. These multitask learning frameworks share the same structure of pixel-level classifier, regression network, and joint representation network but have different feature extraction layers. In particular, we employ 2-D/3-D residual nets and dense nets for feature extraction. Then, for different multitask learning frameworks that we established, we test their performance under different conditions: 1) the framework is trained directly on target data
D. Results and Discussion

1) Performance: Fig. 4 shows the segmentation results estimated by our method across the entire cardiac cycle for two representative subjects from the MRI and CT data, respectively. Red lines in each frame represent ground-truth contours. Blue lines in each frame represent the automated segmentation results using our method.

Fig. 4. Visualization of the segmentation results of 20 frames across the entire cardiac cycle for two representative subjects using MRI (top two rows) and CT (bottom two rows). Red lines in each frame represent the ground truth. Blue lines in each frame represent the automated segmentation results using our method.

without parameter sharing (No sharing); 2) the traditional one-way parameter sharing mechanism is employed (One-way-para) [61]; and 3) the bidirectional parameter sharing mechanism is employed (Bi-para). Finally, we transfer the parameters learned from the MRI to the CT. To further validate the importance of our method, we also transfer the parameters learned from the CT to the MRI.

2) Adversarial Reverse Mapping Network for Task Dependencies Learning:

a) Task dependencies: Fig. 6(a) shows the multitask correlation matrix of the cardiac indices, which is calculated from the predicted results of our method. Specifically, the myocardium area is correlated with WTs, while the cavity area is correlated with dimensions. Strong correlations exist within indices of WTs and dimensions. Furthermore, Fig. 6(b) uncovers pairwise gains (PPG) for every two tasks, where the darkness of colors indicates the strength of the gains. It is intuitive that the linear indices (WTs + Dims) with higher pairwise gains with planar indices (Areas). Meanwhile, both the linear and planar indices can achieve positive pairwise gains with the Seg and the Reverse tasks, which indicates that the Seg and reverse tasks can facilitate model learning.

b) Comparison studies: Our method achieves better performance than existing one-way methods on MRI data. For linear and planar indices estimation, our method outperforms segmentation-based method Max Flow [18], two-phase methods multiflare + RF [38], SDL + AKSF [43], MCDBN + RF [41], and deep multitask learning-based methods, e.g., Indices-Net [2], FullLVNet [49], and DMTRL [3], as shown in Table I. For the cardiac segmentation, our method outperforms than adversarial training method GAN [25], FCN [24], and V-Net [26], as shown in Table II.

Compared with Max Flow, our method achieves average MAE reductions of 56.7%, 5.28%, and 27.1% for WTs, Dims, and Areas. MaxFlow method first estimates segmentation, and then, it predicts the cardiac indices from segmentation. Although the MaxFlow method is based on the optimization of global geometry and intensity constraints, it fails to distinguish between the papillary muscles and the myocardium due to extreme deformation and almost the
same intensity of papillary muscles and the myocardium at middle frames of cardiac systolic (ES), as shown in Fig. 5. Thus, the incorrect papillary muscles segmentation leads to bad results of multitype cardiac indices. In contrast to Max Flow, our method treats segmentation as a subtask, which allows us to simultaneously estimate segmentation with other cardiac indices. Compared with the best two-phase regression method MCDBN + RF, our method achieves average MAE reductions of 6.9%, 8.28%, and 5.1% for WTs, Dims, and Areas. These methods estimate cardiac indices in two phases based on handcrafted features, while our method employs an end-to-end framework that can extract task-aware features. Our method also achieves a better performance compared to deep multitask learning methods. Indices-Net can extract task-aware features based on the jointly learning of representation and regression. FullLVNet and DMTRL have the advantages of the combination of CNN and RNN, but they heavily depend on predefined assumptions. DMTRL explores task dependencies based on a carefully designed covariance matrix. Unlike these methods, our framework explores the task dependencies based on further learning a reverse mapping from multitype cardiac indices to images. Compared with DMTRL, our method also achieves average MAE reductions of 4.32%, 5.98%, and 6.67% for WTs, Dims, and Areas (see Table I).

For the segmentation of EpiLV, EndoLV, and Myo, our method achieves the best results in all compared approaches, as shown in Table II. GAN achieves a good segmentation of EpiLV, EndoLV, and Myo based on effective adversarial training. However, it still obtains an inferior performance compared with ours.

c) Ablation studies: To prove the benefits of our model, we first test the performance of our full framework that predicts different cardiac indices. Then, we also test the performance of our framework with different network architectures. As shown in Table III, the (Independent) could not achieve satisfactory results. It increases the average MAE by 21.70%, 11.44%, and 16.07% for WTs, Dims, and Areas, while the reductions of average Dice is 1.43%, 1.50%, and 0.09% for GAN, EndoLV, and Myo. Moreover, we also cannot obtain good results when we calculate the 1-D cardiac indices from independently predicted segmentation, as shown in Fig. 7. When we perform the experiments that estimate two types of cardiac indices in each time as (Seg + Linear), (Seg + Planar), and (Planar + Linear), we also can obtain a significant boost of performance compared with (independent). This indicates that our full model (Ours) for joint estimation of multitype cardiac indices can effectively obtain task dependence and improve the overall performance.

As shown in Table III [(M + LDL) and (M + L_{DL + I-GAN})], the frameworks could not achieve satisfactory results for 1-D cardiac indices and segmentation without a reverse mapping network, regardless of which loss function is used. In this case, the frameworks are using one-way mapping structure, which could not fully explore task dependencies. In contrast, the frameworks with the reverse mapping network [(M + R + L_{M&R}), (M + R + L_{M&R + I-GAN})], and (Ours)
in Tables I and II] achieved superior results. This indicates that the reverse mapping network can explore the task dependencies via generating cardiac images from multitype cardiac indices. It is also important to choose a proper strategy to train our framework. As shown in Table III [(M + R + L_{M&R}), (M + R + L_{M&R+1-GAN}), and (Ours)], our method achieves the best performance. In contrast to (M + R + L_{M&R}) and (M + R + L_{M&R+1-GAN}) that only match the predicted I and real I, our method employs the fully adversarial training approach that matches joint distribution $p_{\theta}(X, I)$ and $q_{\phi}(X, I)$. Relying on the ability of adversarial training to assess joint distributions of complex semantic variables, our method is more powerful to explore task dependencies.

As shown in Fig. 8, the model (Ours) with our adversarial training can effectively reconstruct the cardiac image and show the correct cardiac structure. Furthermore, (Ours), MAE = 0.20, (Seg + Linear, MAE = 0.23), (Seg + Planar, MAE = 0.25), and (Linear + Planar, MAE = 0.22) achieve smaller MAE than (Independent(Seg only), MAE = 0.26) of image reconstruction. This indicates that the learning of those highly abstract features of 1-D indices would further guide the model to find the proper low-level feature ($z$) for sharper image reconstruction. Those sharp images also show that our model stores the information of $x$ into $z$ and can, thus, reconstruct the details of the image. Combined with the good
Fig. 7. Comparison of our joint estimation with segmentation-based 1-D cardiac indices calculation. (Our) denotes the results of the joint estimation. (Seg-based) denotes the results calculated via segmentation; we first employ our full model to predict the segmentation independently and then calculate the 1-D indices from the predicted segmentation.

Fig. 8. Cardiac MRI reconstructed by different frameworks. On the first row are real images sampled from a test set. The rest rows are the corresponding images reconstructed by (Ours), [Independent(Seg only)], (Seg + Linear), (Seg + Planar), (Planar + Linear), (M + Linear), and (M + Planar + Linear), respectively. The MAE values of reconstructed results are also given for different frameworks.

performance of our model in the multitype cardiac indices estimation, our adversarial reverse mapping method may be an effective way to avoid overfitting or “steganography” effect when the domains have different entropy [55].

3) Bidirectional Parameter Sharing Scheme for Knowledge Transfer: As shown in Table IV, compared with (No sharing), the four methods with our bidirectional parameter sharing mechanism (Bi-para) obtain the best performance. In particular, they achieve average MAE reductions of 15.17%, 6.02%, and 7.34% for WTs, Dims, and Areas and Dice increase of 3.33%, 2.15%, and 2.66% for EpiLV, EndoLV, and Myo. This indicates that our bidirectional parameter sharing mechanism can effectively transfer the learned task dependencies from the MRI to the CT. It is difficult to mine and represent complex task dependencies, which are high-level semantic information. Even so, compared with (One-way-para), the four methods with our bidirectional parameter sharing mechanism (Bi-para) can still achieve average MAE reductions of 7.87%, 3.46%, and 3.50% for WTs, Dims, and Areas and Dice increase of 3.45%, 1.70%, and 1.83% for EpiLV, EndoLV, and Myo. For (One-way-para) methods, even if they employ different networks, they are still trained under a one-way mapping framework from cardiac image to multitype cardiac indices regardless the learning and representation ability of the deep one-way network.

We also try an extended experiment, as shown at the bottom of Table IV, which transfers the parameters learned from the CT to the MRI. Both the (One-way-para) and (Bi-para) can enhance the performance of the MRI when there is knowledge transfer from the CT to MRI. This indicates that our method effectively meets clinical needs via transferring the proper task dependencies learned between the two different imaging modalities.

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>RESULTS OF FOUR DEEP NETWORKS (2-D/3-D RESIDUAL NET AND 2-D/3-D DENSE NET) WITH DIFFERENT CONFIGURATIONS ([NO SHARING], (ONE-WAY-PARA), AND (BI-PARA)], THEN, THE BOTTOM ROW IS THE TRANSFER LEARNING (FROM CT TO MRI) RESULTS OF THE 3-D DENSE NET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No sharing</td>
</tr>
<tr>
<td>CT to MRI</td>
<td>mAE (mm²)</td>
</tr>
<tr>
<td>2D residual net</td>
<td>WTs</td>
</tr>
<tr>
<td></td>
<td>Dims</td>
</tr>
<tr>
<td></td>
<td>Areas</td>
</tr>
<tr>
<td></td>
<td>EpiLV</td>
</tr>
<tr>
<td></td>
<td>EndoLV</td>
</tr>
<tr>
<td></td>
<td>Myo</td>
</tr>
<tr>
<td>3D residual net</td>
<td>WTs</td>
</tr>
<tr>
<td></td>
<td>Dims</td>
</tr>
<tr>
<td></td>
<td>Areas</td>
</tr>
<tr>
<td></td>
<td>EpiLV</td>
</tr>
<tr>
<td></td>
<td>EndoLV</td>
</tr>
<tr>
<td></td>
<td>Myo</td>
</tr>
<tr>
<td>2D dense net</td>
<td>WTs</td>
</tr>
<tr>
<td></td>
<td>Dims</td>
</tr>
<tr>
<td></td>
<td>Areas</td>
</tr>
<tr>
<td></td>
<td>EpiLV</td>
</tr>
<tr>
<td></td>
<td>EndoLV</td>
</tr>
<tr>
<td></td>
<td>Myo</td>
</tr>
<tr>
<td>3D dense net</td>
<td>WTs</td>
</tr>
<tr>
<td></td>
<td>Dims</td>
</tr>
<tr>
<td></td>
<td>Areas</td>
</tr>
<tr>
<td></td>
<td>EpiLV</td>
</tr>
<tr>
<td></td>
<td>EndoLV</td>
</tr>
<tr>
<td></td>
<td>Myo</td>
</tr>
<tr>
<td>CT to MRI</td>
<td>mAE (mm²)</td>
</tr>
<tr>
<td>2D dense net</td>
<td>WTs</td>
</tr>
<tr>
<td></td>
<td>Dims</td>
</tr>
<tr>
<td></td>
<td>Areas</td>
</tr>
<tr>
<td></td>
<td>EpiLV</td>
</tr>
<tr>
<td></td>
<td>EndoLV</td>
</tr>
<tr>
<td></td>
<td>Myo</td>
</tr>
</tbody>
</table>

Authorized licensed use limited to: UNIVERSITY OF WESTERN ONTARIO. Downloaded on April 18,2020 at 17:10:22 UTC from IEEE Xplore. Restrictions apply.
E. Limitations

This study has potential limitations. The study focuses on the 2-D mid-ventricular slices across one cardiac cycle. However, this might be limited to 3-D cardiac images in clinical application. Furthermore, the 3-D image-based cardiac functions, i.e., volume and ejection fraction, could not be obtained directly. In the future work, it is necessary to generalize the proposed method to 3-D and 3D + t cardiac images.

IV. CONCLUSION

In this study, we have developed a multitask learning framework based on adversarial reverse mapping for estimating multitype cardiac indices in different imaging modalities, e.g., MRI and CT. The proposed framework has an established reverse mapping network that explores task dependencies by learning the mapping from multitype cardiac indices to cardiac images via adversarial training. A bidirectional parameter sharing scheme then transfers the parameters from both the multitask learning network and the proposed reverse mapping network. Experimental results show that our method not only accurately estimates multitype cardiac indices in MRI but also performs well for knowledge transfer from MRI to CT.

REFERENCES


Shu Zhao received the Ph.D. degree in computer science from Anhui University, Hefei, China, in 2007. She is currently a Professor with the School of Computer Science and Technology, Anhui University. Her current research interests include quotient space theory, granular computing, social networks, and machine learning.

Heye Zhang received the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2007. He is currently a Full Professor with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. His research focuses on health informatics computing and machine learning.

Yanping Zhang received the Ph.D. degree from Anhui University, Hefei, China, in 2005. She is currently a Professor with Anhui University. She is a Principal Investigator of some 973 projects and the Leader of the National Natural Science Foundations of China and Anhui Province. Her main research interests include computational intelligence, quotient space theory, artificial neural networks, intelligent information processing, and machine learning.

Shuo Li received the Ph.D. degree in computer science and software engineering from Concordia University, Montreal, QC, Canada, in 2006. He was with GE (Healthcare), Chicago, IL, USA, until 2015. He is an Academic Artificial Intelligence Scientist in both medical imaging and computer science & software engineering. He has been the Founder and the Scientific Director of the Digital Imaging Group of London (DIG), London, U.K., since 2007. He has been an Associate Professor with Western University, London, ON, Canada, since 2015. As a primary supervisor, he has mentored over 60 trainees, including post-doctoral fellows, Ph.D./M.Sc. students, undergraduates, and visiting scholars. Since 2015, he has published over 120 journal articles and conference papers, six books, and two awarded patents. His research focuses on the artificial intelligence systems solving the clinical and fundamental data analytics problems with an emphasis on the innovations of our learning schemes.

Dr. Li serves as a Board Member of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) and the General Chair of the MICCAI Conference 2022. He has been serving on the editorial board of multiple premier journals in the field of medical image analysis and pattern recognition. He has been frequently invited to review grant applications for several international funding agencies. He has been actively serving multiple professional societies in various leadership capabilities.