Full Left Ventricle Quantification via Deep Multitask Relationships Learning

Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, Shuo Li

PII: S1361-8415(17)30136-6
DOI: 10.1016/j.media.2017.09.005
Reference: MEDIMA 1298

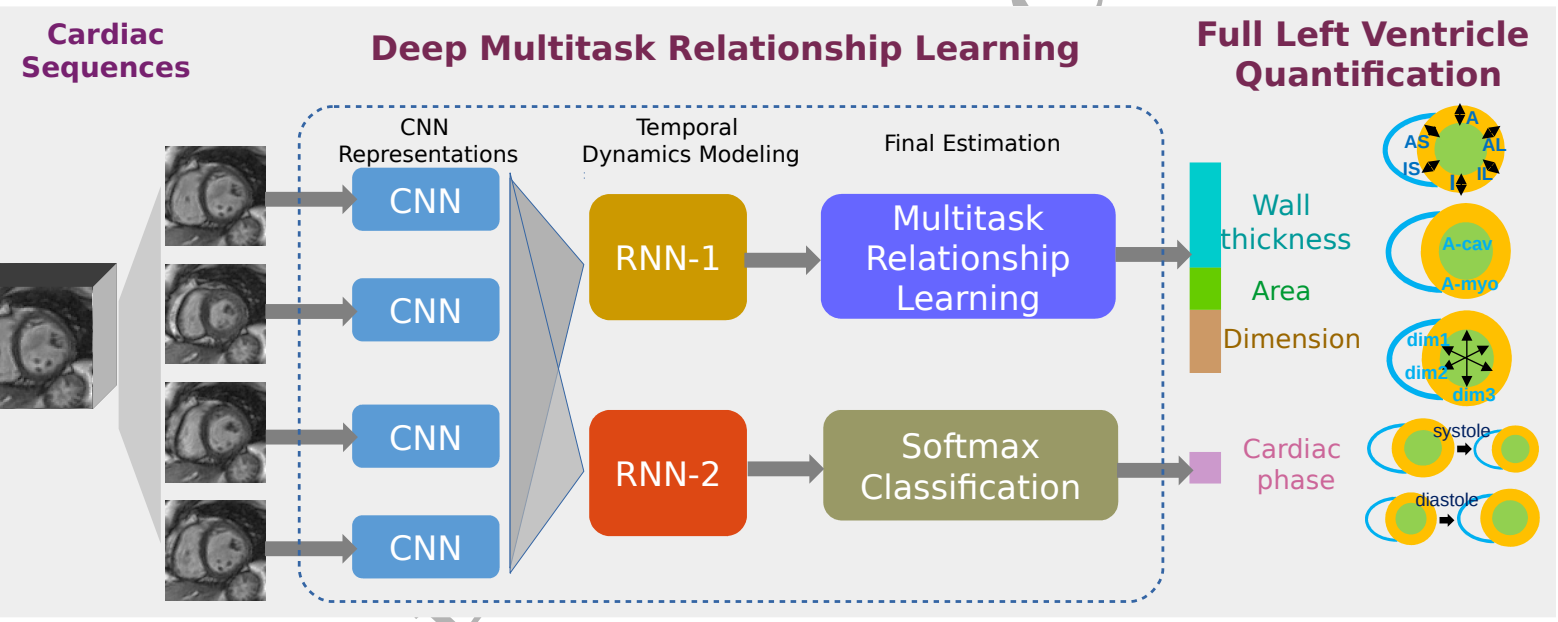
To appear in: *Medical Image Analysis*

Received date: 23 June 2017
Revised date: 25 August 2017
Accepted date: 18 September 2017

Please cite this article as: Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, Shuo Li, Full Left Ventricle Quantification via Deep Multitask Relationships Learning, *Medical Image Analysis* (2017), doi: 10.1016/j.media.2017.09.005

Highlights

- An effective end-to-end integrated framework for full quantification of cardiac LV.
- A brand-new multitask relationship learning method for deep neural networks.
- Proof of the convexity and proposing of an efficient algorithm for the framework.



Full Left Ventricle Quantification via Deep Multitask Relationships Learning

Wufeng Xue[a,b], Gary Brahm[a,b], Sachin Pandey[a,b], Stephanie Leung[a,b], Shuo Li[a,b,*]

[a] *Dept. of Medical Imaging, Western University, London, ON, Canada*
[b] *Digital Image Group (DIG), London, ON, Canada*

Abstract

Cardiac left ventricle (LV) quantification is among the most clinically important tasks for identification and diagnosis of cardiac disease. However, it is still a task of great challenge due to the high variability of cardiac structure across subjects and the complexity of temporal dynamics of cardiac sequences. Full quantification, i.e., to simultaneously quantify all LV indices including two areas (cavity and myocardium), six regional wall thicknesses (RWT), three LV dimensions, and one phase (Diastole or Systole), is even more challenging since the ambiguous correlations existing among these indices may impinge upon the convergence and generalization of the learning procedure. In this paper, we propose a deep multitask relationship learning network (DMTRL) for full LV quantification. The proposed DMTRL first obtains expressive and robust cardiac representations with a deep convolution neural network (CNN); then models the temporal dynamics of cardiac sequences effectively with two parallel recurrent neural network (RNN) modules. After that, it estimates the three types of LV indices under a Bayesian framework that is capable of learning multitask relationships automatically, and estimates the cardiac phase with a softmax classifier. The CNN representation, RNN temporal modeling, Bayesian multitask relationship learning, and softmax classifier establish an effective and integrated network which can be learned in an end-to-end manner. The obtained task covariance matrix captures the correlations existing among these indices, therefore leads to accurate estimation of LV indices and cardiac phase.

Experiments on MR sequences of 145 subjects show that DMTRL achieves high accurate prediction, with average mean absolute error of 180 mm^2, 1.39 mm, 2.51 mm for areas, RWT, dimensions and error rate of 8.2% for the phase classification. This endows our method a great potential in comprehensive clinical assessment of global, regional and dynamic cardiac function.

Keywords: multitask learning, left ventricle quantification, multitask relationship, Bayesian framework
2010 MSC: 00-01, 99-00

1. Introduction

Accurate quantification of left ventricle (LV) from cardiac imaging is among the most clinically important and most frequently demanded tasks for identification and diagnosis of cardiac diseases (Karamitsos et al., 2009). To provide a comprehensive global and regional cardiac function assessment, full quantification of cardiac LV is required, which simultaneously quantifies, for every frame in the whole cardiac cycle, multiple types of cardiac indices, such as cavity and myocardium areas, regional wall thicknesses, LV dimension and cardiac phase, as shown in Fig. 1.

In clinical practice, LV quantification is completed by manually delineation of the borders of cardiac myocardium, which is time-consuming and tedious. Great efforts have been devoted into automatizing this procedure. However, obtaining reliable and accurate full quantification of cardiac LV is still an extremely challenging ambition due to 1) the inhomogeneity of cardiac appearance within each image and the high variability of cardiac structure across subjects and in presence of various pathologies, 2) the complicated global/regional temporal deformation of myocardium during the cardiac cycle, and 3) the ambiguous relationships among different indices which impinge upon the convergence and generalization of the learning procedure. Existing efforts in cardiac quantification have been limited to one index only, i.e, the LV cavity area[1], which relates to ejection fraction and is the easiest index to estimate among the above mentioned ones. Other indices have been neglected for a long period, despite their critical role in regional and dynamic cardiac function assessment.

In order to achieve a reliable and accurate solution, we propose a Deep MultiTask Relationship Learning network (DMTRL) for full quantification of LV from short-axis view cardiac MR images. The proposed DMTRL is capable of robustly representing cardiac images, effectively capturing temporal dynamics of cardiac sequences, and ex-

*Corresponding author
Email address: slishuo@gmail.com (Shuo Li)

[1] LV volume can be calculated by integration of LV cavity areas along the long axis, thus is also included in LV cavity area.

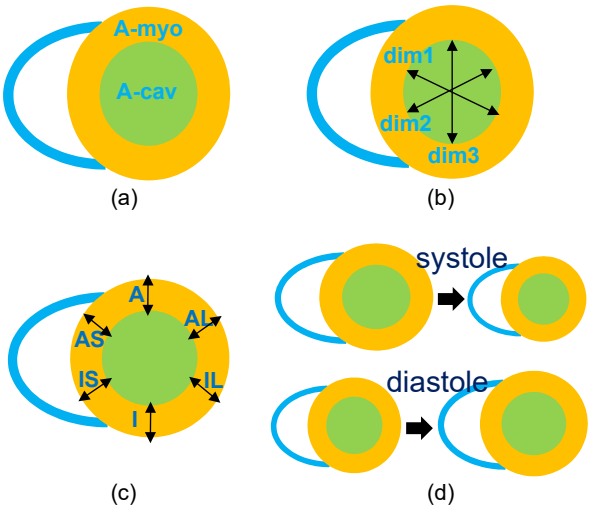


Figure 1: Illustration of LV indices to be quantified for short-axis view cardiac image. (a) Cavity (green) and myocardium (yellow) areas. (b) directional dimensions of cavity (black arrows). (c) Regional wall thicknesses (black arrows). A: anterior; AS: anterospetal; IS: inferoseptal; I: inferior; IL: inferolateral; AL: anterolateral. (d) Phase (systole or diastole).

plicitly modeling the correlations among different cardiac indices.

1.1. Existing work on cardiac LV quantification

Existing work on cardiac LV quantification fall into three categories: 1) manual quantification; 2) segmentation-based quantification; and 3) direct regression methods.

Manual quantification. In clinical practice, obtaining reliable quantification is subjected to manually contouring the borders of myocardium (Suinesiaputra et al., 2015). However, manually contouring is a laborious and time-consuming procedure, and prone to high intra- and inter-observer variability. Besides, manually contouring of myocardium is typically limited to the end diastolic (ED) and end systolic (ES) frames, which makes it insufficient for dynamic function analysis during the whole cardiac cycle. Persistent efforts have been dedicated to automatizing cardiac quantification.

Segmentation-based quantification. Segmentation methods aim to segment out cardiac myocardium from the complex surroundings, with the borders of epicardium and endocardium clearly delineated. Numerous methods (Peng et al., 2016; Petitjean and Dacher, 2011) have been proposed for cardiac LV segmentation, including categories of image-driven techniques, deformable models, and training based methods. To increase robustness and accuracy, these methods usually require the use of prior information and user interaction. Prior information can consist of assumptions such as anatomical assumptions making use of the circular geometry of the LV(Gupta et al., 1993; Wu et al., 2013), intensity histogram distribution (Pednekar et al., 2006), or statistical shape model (Lötjönen et al., 2004). User interaction can consist of pointing out the center of the LV cavity (Gupta et al., 1993; Nachtomy et al., 1998), or manually tracing the ventricle border of the first frame (Ayed et al., 2012). Inaccurate prior information and strong user interaction may prevent LV segmentation methods from efficient clinical application.

Direct regression methods. To circumvent these limitations, direct methods without segmentation have grown in popularity in cardiac volumes estimation, and obtained effective performance benefiting from machine learning algorithms. These methods can be further ranged into two classes: two-phase volume-only methods (Afshin et al., 2012, 2014; Wang et al., 2014; Zhen et al., 2014, 2016, 2015) and end-to-end deep neural network based methods (Kabani and El-Sakka, 2016; Xue et al., 2017a,b,c).

The former ones usually follow a common two-phase framework: cardiac image representation and indices prediction. Cardiac images are usually represented by hand-crafted features or those obtained by unsupervised learning, which include Bhattacharyya coefficient between image distributions (Afshin et al., 2012, 2014), appearance features (Wang et al., 2014), multiple low level image features (Zhen et al., 2014), as well as features from multiscale convolution deep belief network (MCDBN) (Zhen et al., 2016) and supervised descriptor learning (SDL) (Zhen et al., 2015). Cardiac volumes are then estimated from these features with a separated regression model. Despite their effectiveness, these methods suffer from the following limitations: 1) The vulnerable hand-crafted features are not capable of capturing task-relevant cardiac structures robustly; 2) Independently handling each frame without temporal dynamic modeling can not guarantee the consistency of the estimated volumes across the whole cardiac cycle; 3) The two phases are learned separately, instead of jointly, therefore cannot adapt to each other well; 4) These methods focus on volumes estimation only.

The later ones employ deep neural networks in an end-to-end manner for cardiac indices estimation. However, these methods either lack temporal dynamic modeling of the cardiac sequences, or incapable of handling multitask learning problems and characterizing the task relationship for better generalization. Therefore, only inferior performance were obtained.

1.2. Methodology overview

The proposed DMTRL consists of four tightly integrated modules. A deep convolution neural network (CNN) is firstly designed for cardiac image representation, then two parallel recurrent neural network (RNN) are deployed for temporal dynamic modeling of cardiac sequences. The outputs of one RNN module are fed into a Bayesian based multitask relationship learning module for LV indices estimation, and the other one fed into a softmax classifier for cardiac phase identification.

- To address the *cardiac appearance challenge*, we represent cardiac images with robust and task-aware features that learned from a newly designed deep CNN network. The obtained CNN features form a compact low-dimensional manifold for cardiac images, and are more expressive of cardiac structures than hand-crafted ones.

- To tackle the *temporal deformation challenge*, we model the dependencies between neighboring frames with long short-term units (LSTM) based RNN modules. The LSTM units can selectively memorize and forget previous information when transforming the CNN features into the RNN features, therefore deformation robust predictions can be obtained.

- To cope with the *multitask relationship challenge*, we design on top of the CNN and RNN modules a Bayesian framework for multitask relationship learning in the final estimations of LV indices. Under this framework, we model the prior of all tasks with a matrix-variate normal distribution, which is able of explicitly capturing the relationships of the three types of LV indices through a column covariance matrix.

Convolution neural network. Our newly designed CNN takes advantages of the hierarchical representation (Zeiler and Fergus, 2014) as well as the supervised mechanism, resulting in expressive task-aware representations for cardiac images. CNNs have demonstrated great success in medical image analysis (Greenspan et al., 2016). In cardiac image analysis, CNN, as well as other deep models have been employed in cardiac segmentation with optional refinement by deformable model (Avendi et al., 2016) and level set (Ngo and Carneiro, 2013; Ngo et al., 2017). Fully convolution network (FCN) was applied to cardiac segmentation (Tran, 2016) due to its success in semantic segmentation of natural image (Shelhamer et al., 2017). In the ACDC challenge of STACOM 2017, different CNN structures were proposed, including the 2D/3D U-net (Patravali et al., 2017), grid-like CNN (Zotti et al., 2017), dilated CNN (Wolterink et al., 2017) and encoder-decoder architecture (Mortazi et al., 2017). Besides, a parallel coarse and fine network was proposed in Tan et al. (2017) for LV segmentation in the polar space domain. A 3D-CNN model with deep supervised mechanism (Dou et al., 2017) was proposed for segmentation of volumetric images.

These different CNN models usually handles each frame independently and is not capable of modeling the temporal dynamics of cardiac sequences, which is very important to quantification of all frames in the whole cardiac cycle.

Recurrent neural network. In our work, two RNN modules are designed on top of the CNN to model the relevant temporal dependencies for LV indices and cardiac phase, respectively. With their outputs, accurate LV indices and cardiac phase can be obtained across all frames, despite the deformation of cardiac structure. RNN, especially when LSTM units are deployed, is specialized in long-range temporal dynamic modelling and spatial context modeling, and has delivered promising results in language modelling (Zhang et al., 2015), object recognition (Liang et al., 2015), visual recognition and description (Donahue et al., 2015; Li et al., 2016), and also medical image analysis (Shin et al., 2016; Kong et al., 2016; Poudel et al., 2016). In the work of cardiac image segmentation (Poudel et al., 2016), an RNN was employed to capture inter-slice spatial dependencies of cardiac structure. In Kong et al. (2016), an RNN with LSTM was employed to model the temporal dependencies in cardiac MR sequences. In Shin et al. (2016), an RNN was trained to describe the contexts of detected disease in chest X-Rays.

Multitask relationship learning. We incorporate multitask relationship learning into our deep neural network, leveraging both the powerful representation learning and the effective multitask relationship learning. Multitask relationship learning (Zhang and Yeung, 2010; Zhou et al., 2011; Lin and Zhou, 2016) aims at improving the generalization performance of a set of related machine learning tasks by transferring knowledge among these tasks, and suites well our aim of full LV quantification. However, on one hand, most existing work learn task relationships based on fixed representation, which may not be optimal for the tasks of interest or relationships learning. On the other hand, deep neural network provides an effective way of expressive representation learning, while the task relationship learning in deep models has not been well explored. Zhang et al. (2014) proposed a task-constrained deep model for facial landmark detection, with task-wise early stopping to facilitate learning convergence. Yin and Liu (2017) utilized side tasks as regularization to disentangle the variations from the learnt identity features of face recognition and assigned the loss weights of each side task through a dynamic-weighting scheme. Taking side tasks as regularization for the main task cannot fully exploiting the multitask relationships to improve the generalization of deep neural network. Combining deep neural network and multitask relationship learning makes perfect complementation to each other.

Two methods exist for tasks relationships modeling: assumption-based regularization and data-driven task covariance learning. *The former ones* involve assumptions about task relationships, such as shared set of features via sparsity induced norms (Liu et al., 2009; Gong et al., 2012; Xue et al., 2017b), shared low dimensional subspace via low-rank regularization (Chen et al., 2014), and clustering structures of multiple task (Jacob et al., 2009). Such assumptions are not always accurate and suitable for all tasks. *The later ones* (Zhang and Yeung, 2010; Lin and Zhou, 2016) model the multitask learning problem under a Bayesian framework and learn task relationships automatically from the training data through a multitask covariance matrix. The learned matrix is able of representing

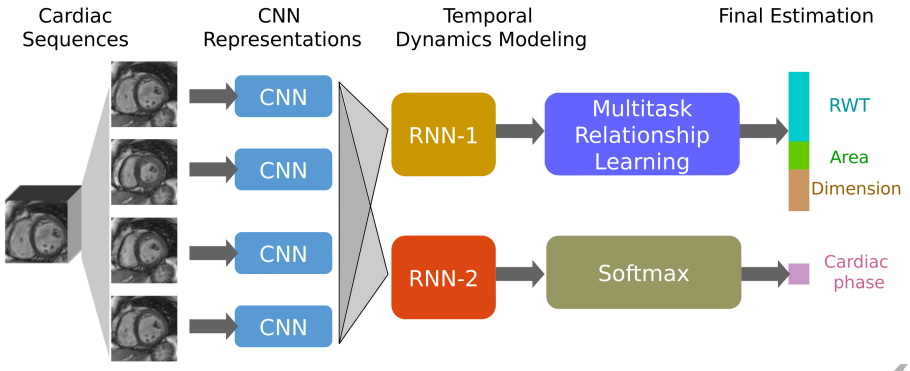


Figure 2: Overview of DMTRL, which combines a deep convolution neural network (CNN) for cardiac image representation, two parallel recurrent neural network (RNN) for temporal dynamic modeling of cardiac sequences, a Bayesian based multitask relationship learning module for LV indices estimation, and a softmax classifier for cardiac phase identification.

positive, negative and no correlation between tasks. Bearing these consideration, we combine the data-driven task covariance matrix learning with our neural network and propose an integrated deep multitask relationship learning framework for the task of full LV quantification.

1.3. Contributions

The main contributions of this work are as follows:

- We propose for full quantification of cardiac LV an effective end-to-end integrated framework, which provides for the first time an reliable solution for global and regional dynamic cardiac function assessment. The proposed framework obtains robust and accurate prediction for three types of LV indices and one cardiac phase on cardiac MR images of 145 subjects.
- We propose a brand-new multitask relationship learning method for deep neural networks, which for the first time enables generalization improvement of the deep neural networks through *automatically* leveraging the correlations among different tasks. The proposed method can be easily applied to other deep multitask models.
- We prove the convexity of the proposed objective function for deep multitask relationship learning, and propose an efficient alternating gradient descent algorithm for its optimization, with the convergence proved. This algorithm enables an end-to-end deep multitask relationship learning and thus leads to improved generalization of the network.

In this work, we advance our preliminary attempt on full LV quantification (Xue et al., 2017b) in the following aspects: 1) conduct more comprehensive review on cardiac LV quantification, providing a panorama of existing work; 2) automate the multitask relationships modeling with accurate Bayesian-based multitask covariance matrix learning, instead of the predefined task-dependent assumptions; and 3) carry out more extensive experiments on performance analysis and comparison, validating the great advantages of the proposed DMTRL over existing state-of-art methods.

The rest of this paper is organized as follows: In Section 2 we first present details of the four modules of DMTRL in 2.1, 2.2, 2.3, 2.4. Then we give the algorithm of optimizing the proposed DMTRL framework in 2.5. Experimental settings and details about dataset are introduced in Section 3. Results and detailed discussions are shown in Section 4. Section 5 concludes the paper.

2. Deep MultiTask Relationship Learning (DMTRL)

The overview of our approach is shown in Figure 2. We model full quantification of cardiac LV as a deep multitask learning problem, with four key modules specially designed for cardiac MR images. Hereafter, we use $\mathcal{X} = \{X^{s,f}\}$ as the input cardiac images, where $s = 1 \cdots n_S$ indicates the subject and $f = 1 \cdots n_F$ indicates the frame. The LV indices and cardiac phase are denoted as $\{y_t^{s,f}\}$, where $t \in \{area, dim, rwt, phase\}$, and $y_t^{s,f} \in \mathcal{R}^{k_t}$, with k_t the dimensionality of $y_t^{s,f}$.

2.1. Cardiac image representation by deep CNN

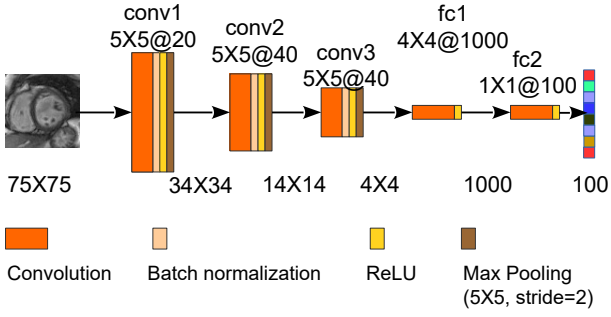


Figure 3: Architecture of the newly designed CNN for cardiac MR images. The operation type of each layer is encoded in colors. The size and number of convolution kernel are shown above the diagram, while the dimensionality of output is shown below. Our CNN is adequate for robust representation of cardiac images with a small size of network parameters.

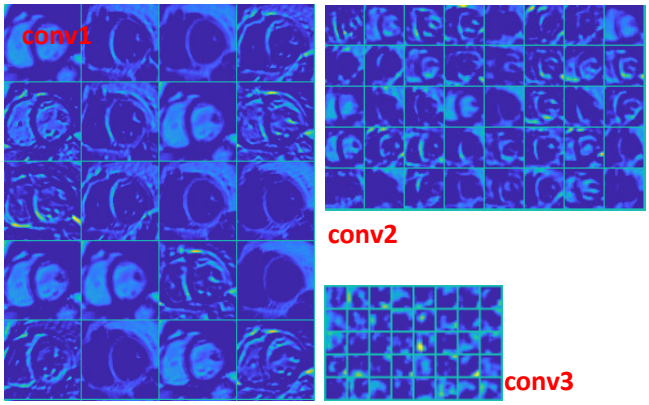


Figure 4: Visualization of feature maps obtained by our CNN for a example cardiac image.

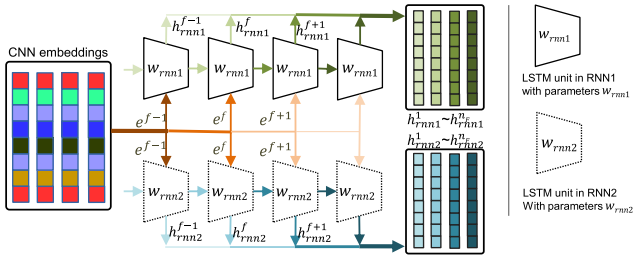


Figure 5: Architecture of the two RNN modules, which share the CNN representation as input and capture respectively the related features and temporal dynamics of cardiac sequences for LV indices (top) and cardiac phase (bottom). The two RNN modules differ in the parameters of LSTM.

To obtain expressive representations of cardiac images, we design for cardiac images a new CNN which is adequate to capture the variations of cardiac structures and appearances with a small size of network parameters (∼1M). Given the fact that all cardiac images share approximately the same spatial layout, which is of less variation than natural images, we configure a relatively low number of filters for each convolution layer to avoid model redundancy. As for the kernel size of convolution and pooling, 5×5, instead of the frequently used 3×3, is deployed to introduce more shift invariance. Dropout and batch normalization are adopted to alleviate the training procedure. As can be seen in our experiments, our CNN is very effective for cardiac images even without transfer learning. As a representation learning network, our CNN maps each cardiac image $X^{s,f}$ into a fixed-length low dimensional vector:

$$e^{s,f} = \mathbf{f}_{cnn}(X^{s,f}|w_{cnn}), \qquad (1)$$

where $e^{s,f} \in \mathcal{R}^{100}$, and w_{cnn} is the set of parameters in our CNN. The architecture of our CNN is illustrated in Figure 3.

Our CNN captures effectively the structures of cardiac images, as shown in Figure 4. The 20 filters of layer conv1 captures low level visual features of cardiac images, including low frequency appearance of the ventricles, textures of the ventricles, cardiac myocardium as well as the textures in the background. With the low level features of cardiac structure in conv1 feature maps as input, conv2 layer extracts more complex cardiac structures by combining low level features. The high frequency noise or textures, which are not related to the cardiac indices, are discarded in this procedure. The conv3 feature maps, with a resolution of 10×10, cannot keep the detailed structures. Instead, they extract the indices-related features in each local area from the structures of all conv2 feature maps during the training procedure.

2.2. Dynamic modelling of cardiac sequences by RNN

Accurate modeling of cardiac temporal dynamics improves the quantification accuracy of current frame with information from neighboring frames. The four tasks in this work can be divided into two types: 1) the three types of LV indices, which are mainly related to the spatial structure of cardiac LV in each frame, and 2) cardiac phase, which is mainly related to the structure difference between successive frames. Therefore, we design two RNN modules, as shown in Figure 5, with each of them extracting the related features and modeling the corresponding temporal dynamics. The two RNN modules have the same architecture and share the same CNN embeddings as input. They differ in the parameters that are trained with different supervised information. RNN-1 module will be trained to predict the LV indices, while RNN-2 modules will be trained to identify cardiac phase. The outputs of RNN modules are

$$\{h_m^{s,1},,,,h_m^{s,n_F}\} = \mathbf{f}_{rnn}([e^{s,1},...e^{s,n_F}]|w_m), \qquad (2)$$

where $m \in \{rnn1, rnn2\}$ and w_m is the set of parameters in the RNN modules.

In our implementation, we utilize in RNN modules the LSTM unit (Graves, 2012), which is capable of learning the long-term dynamics in sequential data and avoiding the gradient vanishing/exploding problem in traditional RNN. The input gate, output gate, forget gate and the memory cell in LSTM allow the network to learn when to forget previous hidden states and when to update current hidden states given current input. This strategy enables LSTM to adaptively memorize and access information long term ago. The computation details of LSTM can be found in (Graves, 2012).

2.3. Multitask relationship learning for LV indices

We introduces multitask relationship learning to effectively improve the generalization of our deep multitask neural network. Among the three types of LV indices to be estimated, strong or weak correlations exist between and within each type of indices, as demonstrated in the correlation matrix in Figure 6: 1) within task correlation can be observed for the different outputs of RWT and those of cavity dimensions; 2) between tasks correlation can be observed for the three tasks, no matter positive or negative correlation. Therefore, we completely leverage these

correlations with the ambition of improving quantification accuracy for all LV indices.

2.3.1. Probabilistic formulation

With the outputs of RNN-1, all the LV indices can be estimated with a linear regression model:

$$\hat{y}_t^{s,f} = \mathbf{w}_t^T h_{rnn1}^{s,f} + b_t,\ t \in \{area, dim, rwt\} \tag{3}$$

where $\mathbf{w}_t$ and b_t are the weight and bias term of the linear model for task t. For simplicity, the bias term will be ignored in the following. Under the Gaussian assumption, the likelihood for $y_t^{s,f}$ given $h_{rnn1}^{s,f}$ and $\mathbf{w}_t$ is:

$$p(y_t^{s,f}|h_{rnn1}^{s,f}, \mathbf{w}_t) = \mathcal{N}(\mathbf{w}_t^T h_{rnn1}^{s,f}, \epsilon_t^2 \mathbf{I}_{k_t}) \tag{4}$$

where $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ denotes the multivariate normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$, ϵ_t^2 is the variance of the estimation error for task t, and $\mathbf{I}_{k_t}$ is the identity matrix. Unless otherwise specified, $t \in \{area, dim, rwt\}$ in the following sections.

To model the multitask relationship, we incorporate a prior on the weight matrix of the three regression tasks $\mathbf{W} = (\mathbf{w}_{rwt}, \mathbf{w}_{area}, \mathbf{w}_{dim})$ as follows:

$$p(\mathbf{W}) \propto (\prod_t \mathcal{N}^{k_t}(\mathbf{0}_d, \xi_t^2 \mathbf{I}_d)) q(\mathbf{W}) \tag{5}$$

where d is the dimensionality of $h_{rnn1}^{s,f}$, $\mathbf{0_d}$ and $\mathbf{I}_d$ are the zero and identity matrix of size $d \times d$. The first term is the extension of ridge prior to the multitask learning scenario and controls the model complexity of each task $\mathbf{w}_t$. The second term encodes the task relationship, i.e, task covariance, using a matrix-variate normal distribution.

$$q(\mathbf{W}) = p_{\mathcal{MN}}(\mathbf{W}|\mathbf{0}_{d\times K}, \mathbf{I}_d \otimes \mathbf{\Omega}) \tag{6}$$

where $K = \sum k_t$ is the total output number for the three regression tasks, $\mathcal{MN}(\mathbf{M}, \mathbf{A}\otimes\mathbf{B})$ denotes the matrix-variate normal distribution with mean $\mathbf{M} \in \mathcal{R}^{d\times K}$, row covariance matrix $\mathbf{A} \in \mathcal{R}^{d\times d}$ and column covariance matrix $B \in \mathcal{R}^{K\times K}$. Its probability distribution function is:

$$\begin{aligned} &p_{\mathcal{MN}}(\mathbf{X}|\mathbf{M}, \mathbf{A}\otimes\mathbf{B}) \\ &= \frac{\exp(-\frac{1}{2} tr(\mathbf{A}^{-1}((\mathbf{X}-\mathbf{M})\mathbf{B}^{-1}(\mathbf{X}-\mathbf{M})^{\mathbf{T}})))}{(2\pi)^{Kd/2}|\mathbf{A}|^{K/2}|\mathbf{B}|^{d/2}} \end{aligned} \tag{7}$$

$tr(\cdot)$ denotes matrix trace and $|\cdot|$ denotes matrix determinant. The row covariance matrix $\mathbf{A}$ models the relationships between the input features, while the column covariance matrix $\mathbf{B}$ models the relationships between the weight parameters of each output in all the tasks, i.e, multitask relationships. Since the input feature in our tasks is learned from cardiac images through the deep CNN and RNN modules, independence between different dimension of the feature is expected. Therefore, in Equation 6, we use $\mathbf{I}_d$ to model the feature relationships. As for the multitask relationships $\mathbf{\Omega}$, we will learn it automatically from the data.

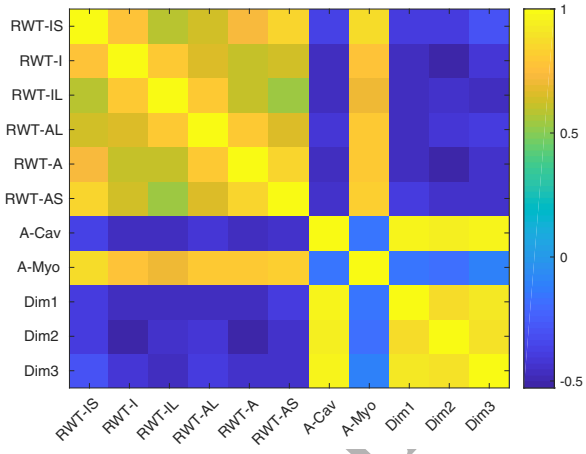


Figure 6: Multitask correlation matrix computed from the ground truth values of LV indices in our dataset. As can be seen, strong, weak and no correlations exist among these 11 LV indices.

2.3.2. Maximum a posterior

According to the Bayesian theorem, the posterior distribution of $\mathbf{W}$ given the input data $\mathcal{H}_{rnn1} = \{h_{rnn1}^{s,f}\}$ and $\mathcal{Y} = \{[y_{rwt}^{s,f}; y_{area}^{s,f}; y_{dim}^{s,f}]\}$ is proportional to the product of the prior distribution and the likelihood function:

$$p(\mathbf{W}|\mathcal{H}_{rnn1}, \mathcal{Y}) \propto p(\mathcal{Y}|\mathcal{H}_{rnn1}, \mathbf{W}) p(\mathbf{W}) \tag{8}$$

Substitute Equations 4, 5, and 6 into 8, we obtain:

$$\begin{aligned} &p(\mathbf{W}|\mathcal{H}_{rnn1}, \mathcal{Y}) \\ &\propto (\prod_{s,f}\prod_t \mathcal{N}(\mathbf{w}_t^T h_{rnn1}^{s,f}, \epsilon_t^2 \mathbf{I}_{k_t}))(\prod_t \mathcal{N}^{k_t}(\mathbf{0}_d, \xi_t^2 \mathbf{I}_d)) q(\mathbf{W}) \\ &= (\prod_{s,f}\prod_t \frac{1}{\sqrt{(2\pi)^{k_t}}\epsilon_t} \exp(-\frac{\|y_t^{s,f} - \mathbf{w}_t^T h_{rnn1}^{s,f}\|^2}{2\epsilon_t^2})) \\ &\quad (\prod_t \frac{\exp(-\|\mathbf{w}_t\|^2/(2\xi_t^2))}{\sqrt{(2\pi)^{dk_t}}\xi^{k_t}}) \frac{\exp(-\frac{1}{2} tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^{\mathbf{T}}))}{(2\pi)^{Kd/2}|\mathbf{\Omega}|^{d/2}} \end{aligned} \tag{9}$$

Take the negative logarithm and ignore the constant terms, we can obtain the following objective function to optimize:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{\Omega}} \ \mathcal{L}_{MTL} \\ &= \sum_t \sum_{s,f} \frac{1}{\epsilon_t^2} \|y_t^{s,f} - \mathbf{w}_t^T h_{rnn1}^{s,f}\|^2 + \sum_t \frac{1}{\xi_t^2} \|\mathbf{w}_t\|^2 \\ &+ tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^{\mathbf{T}}) + d \ln |\mathbf{\Omega}| \end{aligned} \tag{10}$$

The first three terms in Equation 10 are jointly convex with respect to all variables, while the last term, which penalizes the complexity of $\mathbf{\Omega}$ is concave. For feasibility, we replace this term with a constraint $tr(\mathbf{\Omega}) = 1$. Besides, taking into account the size of dataset and introducing two constants λ_1, λ_2 to mediate the weights of each term, the

optimization problem in Equation 10 can be reformulated as:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{\Omega}} \quad & \mathcal{L}_{MTL} \\ & = \frac{1}{n_S n_F}\sum_t\sum_{s,f}\|y_t^{s,f} - \mathbf{w}_t^T h_{rnn1}^{s,f}\|_2^2 \\ & + \lambda_1\sum_t\|\mathbf{w}_t\|^2 + \lambda_2 tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^{\mathbf{T}}) \\ \text{s.t.} \quad & \mathbf{\Omega}\in\mathcal{S}_+^K, tr(\mathbf{\Omega}) = 1\end{aligned} \tag{11}$$

where $\mathcal{S}_+^K$ denotes the set of K-dimensional symmetric positive semidefinite matrix. The first constrains holds due to the fact that $\mathbf{\Omega}$ is a task covariance matrix. The three terms in Equation 11 measure the prediction loss of the three regression task, the complexity of $\mathbf{W}$ and relationships between each output of these tasks.

2.4. Softmax classification for cardiac phase

With the output of RNN-2, cardiac phase can be predicted by a softmax classifier.

$$p_l^{s,f} = \frac{\exp(\mathbf{w}_{phase}^{l,T} h_{rnn2}^{s,f})}{\exp(\mathbf{w}_{phase}^{0,T} h_{rnn2}^{s,f}) + \exp(\mathbf{w}_{phase}^{1,T} h_{rnn2}^{s,f})} \tag{12}$$

where $l \in \{0,1\}$ denotes the two cardiac phase *diastole* and *systole*, $\mathbf{w}_{phase}^0$ and $\mathbf{w}_{phase}^1$ are the first and second column of $\mathbf{w}_{phase}$. Combined with penalty of model complexity, the objective function for cardiac phase prediction is:

$$\begin{aligned}\min_{\mathbf{w}_{phase}} \quad & \mathcal{L}_{phase} \\ & = -\frac{1}{n_S n_F}\sum_{s,f}\sum_l \mathbf{1}(y_{phase}^{s,f} = l)\log\frac{p_l^{s,f}}{p_0^{s,f} + p_1^{s,f}} \\ & + \lambda_1\|\mathbf{w}_{phase}\|^2\end{aligned} \tag{13}$$

2.5. Algorithm of DMTRL

The whole framework of our DMTRL can be divided into two parts: one for LV indices and the other for cardiac phase. In this section, we first prove the convexity of the two objective function when the CNN and RNN modules are fixed, and then present a two-stage iterated optimization algorithm to train our network.

Given the output of RNN-1 and RNN-2, both problems of 11 and 13 are convex with respect to their variables, as will be proved in the appendix. The objective function of multitask relationship learning $\mathcal{L}_{MTL}$ consists of three terms, while the first two terms are both convex with respect to $\mathbf{W}$. Thus we only have to prove the convexity of $tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^{\mathbf{T}})$ with respect to $\mathbf{W}$ and $\mathbf{\Omega}$. See appendix A for the proof. The objective function of the softmax classifier with weight decay in Equation 13 is also convex, as proved in Rennie (2005).

To avoid the imbalance of $\mathcal{L}_{MTL}$ and $\mathcal{L}_{phase}$ and inconsistency of their convergence speed, we adopt a two-stage training procedure: firstly, the whole network is trained with no feedback information from $\mathcal{L}_{phase}$, that is, $\mathbf{w}_{rnn2}$ and $\mathbf{w}_{phase}$ are kept frozen in this stage; secondly, the learned parameters of $\mathbf{w}_{cnn}$, $\mathbf{w}_{rnn1}$, $\mathbf{W}$ and $\mathbf{\Omega}$ are kept frozen, and $\mathbf{w}_{rnn2}$ and $\mathbf{w}_{phase}$ are learned with feedback from $\mathcal{L}_{phase}$.

Algorithm 1: Training procedure of DMTRL

Input: cardiac MR sequences: $\mathcal{X}$;
the ground truth LV indices and cardiac phase: $\mathcal{Y}_{area}$, $\mathcal{Y}_{dim}$, $\mathcal{Y}_{rwt}$, $\mathcal{Y}_{phase}$;
learning rate η, constants λ_1, λ_2.
Output: $\mathbf{w}_{cnn}$, $\mathbf{w}_{rnn1}$, $\mathbf{w}_{rnn2}$, $\mathbf{w}_{area}$, $\mathbf{w}_{dim}$, $\mathbf{w}_{rwt}$, $\mathbf{w}_{phase}$, $\mathbf{\Omega}$;

initialize ws randomly, $\mathbf{\Omega} = \mathbf{I}_K$;
`/* first stage */`
while $\mathcal{L}_{MTL}$ *not converged* **do**
  update $\mathbf{w}_{cnn}$, $\mathbf{w}_{rnn1}$ with back propagation from $\mathcal{L}_{MTL}$;
  update $\mathbf{W}$ as: $\mathbf{W} \leftarrow \mathbf{W} - \eta\bigtriangledown_{\mathbf{W}}\mathcal{L}_{MTL}$;
  update $\mathbf{\Omega}$ as: $\mathbf{\Omega} \leftarrow \frac{(\mathbf{W}^{\mathbf{T}}\mathbf{W})^{\frac{1}{2}}}{tr((\mathbf{W}^{\mathbf{T}}\mathbf{W})^{\frac{1}{2}})}$;
end
`/* second stage */`
keep $\mathbf{w}_{cnn}$, $\mathbf{w}_{rnn1}$, $\mathbf{W}$, $\mathbf{\Omega}$ frozen;
while $\mathcal{L}_{phase}$ *not converged* **do**
  update $\mathbf{w}_{rnn2}$ with back propagation from $\mathcal{L}_{phase}$;
  update $\mathbf{w}_{phase}$ as: $\mathbf{w}_{phase} \leftarrow \mathbf{w}_{phase} - \eta\bigtriangledown_{\mathbf{w}_{phase}}\mathcal{L}_{phase}$;
end
assign $\mathbf{w}_{area}$, $\mathbf{w}_{dim}$, $\mathbf{w}_{rwt}$ with the corresponding columns from $\mathbf{W}$.

In both stages, the parameters learning of neural networks, such as $\mathbf{w}_{cnn}$, $\mathbf{w}_{rnn1}$ and $\mathbf{w}_{rnn2}$, can be optimized through the usually utilized iterated stochastic gradient descent (SGD) procedure. To optimize $\mathcal{L}_{MTL}$ with respect to $\mathbf{W}$ when $\mathbf{\Omega}$ is fixed, multiple solutions can be applied, such as linear systems in the dual space (Zhang and Yeung, 2010) and FISTA (Beck and Teboulle, 2009). To be compatible with the rest of our network, we adopt SGD for optimization of $\mathbf{W}$. The derivative of $\mathcal{L}_{MTL}$ with respect to $\mathbf{W}$ given fixed $\mathbf{\Omega}$ is:

$$\begin{aligned}&\bigtriangledown_{\mathbf{W}}\mathcal{L}_{MTL} \\ &= \frac{2}{n_S n_F}\mathbf{H}_{rnn1}(\mathbf{H}_{rnn1}^T\mathbf{W} - \mathcal{Y}^T) + 2\mathbf{W}(\lambda_1\mathbf{I}_K + \lambda_2\mathbf{\Omega}^{-1})\end{aligned} \tag{14}$$

When $\mathbf{W}$ is fixed, the optimization of $\mathbf{\Omega}$ becomes:

$$\begin{aligned}\min_{\mathbf{\Omega}} \quad & tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^{\mathbf{T}}) \\ \text{s.t.} \quad & \mathbf{\Omega}\succeq 0, tr(\mathbf{\Omega}) = 1\end{aligned} \tag{15}$$

which admits to an analytical solution (Zhang and Yeung,

2010):

$$\mathbf{\Omega} = \frac{(\mathbf{W^T W})^{\frac{1}{2}}}{tr((\mathbf{W^T W})^{\frac{1}{2}})} \tag{16}$$

The optimization of $\mathcal{L}_{phase}$ is also implemented by SGD. The derivative of Equation 13 is:

$$\begin{aligned} &\nabla_{\mathbf{w}^l_{phase}} \mathcal{L}_{phase} \\ &= -\frac{1}{n_S n_F} \sum_{s,f} (\mathbf{1}(y^{s,f}_{phase} = l) - p^{s,f}_l) h^{s,f}_{rnn2} + \frac{2}{\lambda_1} \mathbf{w}^l_{phase} \end{aligned} \tag{17}$$

Put all things together, we have the whole algorithm of training our DMTRL, as shown in Algorithm 1.

3. Experiment configurations

3.1. Dataset

We collected a dataset of 2D short axis cine MR images of $n_S = 145$ subjects from 3 hospitals affiliated with two health care centers (London Healthcare Center and St. Josephs HealthCare). The subjects age from 16 yrs to 97 yrs, with average of 58.9 yrs. The pixel spacings of the MR images range from 0.6836 mm/pixel to 2.0833 mm/pixel, with mode of 1.5625 mm/pixel. Diverse pathologies are in presence including regional wall motion abnormalities, myocardial hypertrophy, mildly enlarged LV, atrial septal defect, LV dysfunction, etc. Each subject contains $n_F = 20$ frames throughout a cardiac cycle. In each frame, LV is divided into equal thirds (basal, mid-cavity, and apical) perpendicular to the long axis of the heart following the standard AHA prescription (Cerqueira et al., 2002) and a representative mid-cavity slice is selected for validation of our DMTRL.

All cardiac images undergo several preprocessing steps, including landmark labelling, rotation, ROI cropping, and resizing. The resulted images are approximately aligned with dimension of 80×80. Then, these cardiac images are manually contoured to obtained the epicardial and endocardial borders, which are double-checked by two experienced cardiac radiologists (A. Islam and M. Bhaduri). The ground truth values of LV indices and cardiac phase can be easily obtained from the two borders. The values of RWT and cavity dimensions are normalized by the image dimension, while the areas are normalized by the pixel number (6400). During evaluation, the obtained results are converted to physical thickness (in mm) and area (in mm^2) by reversing the resizing procedure and multiplying the pixel spacing for each subject.

3.2. Configurations

In our experiments, 5-fold cross validation is employed for performance evaluation and comparison. The dataset is divided into 5 groups, each containing 29 subjects. Four groups are employed to train the prediction model, and the last group is used for test. This procedure is repeated five times until the indices of all subjects are obtained. The network is implemented by Caffe (Jia et al., 2014) with SGD solver. Parameters are decided by training on the first 70% (101) subjects and validating on the following 10% (15) subjects. Learning rate are set to 0.3 for the first stage and 0.03 for the second stage. In both procedures, *step* learning policy is employed with *gamma* and *stepsize* being (0.5, 1000) and *momentum* 0.9. λ_1 and λ_2 are set to 1e-5 and 1e-4.

3.3. Evaluation Criteria

The performance of the proposed DMTRL framework is evaluated in terms of estimation accuracy for LV quantification, i.e, each LV index and cardiac phase, for all frames in the cardiac cycle. For the three types of LV indices, correlation coefficient ρ and mean absolute error (MAE) between the ground truth values and the estimated ones are computed to assess the estimation accuracy. For cardiac phase, error rate (ER) is computed. For two vectors $y \in \mathcal{R}^N$ and $\hat{y} \in \mathcal{R}^N$, they are computed as:

$$\text{MAE}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} |y^i - \hat{y}^i| \tag{18}$$

$$\rho(y, \hat{y}) = \frac{2 \sum_{i=1}^{N} (y^i - m_1)(\hat{y}^i - m_2)}{\sum_{i=1}^{N} ((y^i - m_1)^2 + (\hat{y}^i - m_1)^2)} \tag{19}$$

$$\text{ER}(y, \hat{y}) = \frac{\sum_{i=1}^{N} \mathbf{1}(y^i \neq \hat{y}^i)}{N} 100\% \tag{20}$$

where $m_1 = \frac{1}{N} \sum_{i=1}^{N} y^i$ and $m_2 = \frac{1}{N} \sum_{i=1}^{N} \hat{y}^i$.

4. Results and analysis

The effectiveness of the proposed DMTRL framework for full LV quantification is thoroughly validated in three folds. 1) Firstly, the quantification performance is examined for the three types of LV indices and cardiac phase. 2) Secondly, the role of our multitask relationship learning is probed to demonstrate its capacity in generalization improvement. 3) Thirdly, the advantages of DMTRL over existing methods are revealed by comparing it with state-of-art quantification methods, including one segmentation-based methods (Ayed et al., 2012), three two-phase volume only direct methods (Zhen et al., 2014, 2015, 2016), and one end-to-end methods that based on deep neural network (Xue et al., 2017a,b).

4.1. Performance of full LV quantification

As shown in the last row of Table. 1, DMTRL delivers accurate estimation for all the three types of LV indices and cardiac phase, which can be attributed to the integrated CNN representation, RNN temporal modeling, and multitask relationship learning. Specifically, DMTRL achieves average MAE of 180±118 mm^2, 2.51±1.58 mm, and 1.39±0.68 mm, as well as average correlation of 0.945,

Table 1: Performance of full LV quantification for existing state-of-the-art methods and the proposed DMTRL. MAE and ρ are shown in each cell. DMTRL outperforms all existing methods for the three types of LV indices and cardiac phase in terms of average MAE and ρ.

Method	Max Flow	Multi-features+RF	SDL +AKRF	MCDBN +RF	Indices-Net	FullLVNet (N/N)	FullLVNet (intra/inter)	**DMTRL (proposed)**
				Area (mm^2)				
A-cav	156±193 0.958	231±193 0.924	198±169 0.942	208±166 0.926	185±162 0.953	205±182 0.926	181±155 0.940	172±148 0.943
A-myo	339±272 0.851	291±246 0.729	286±242 0.742	269±217 0.723	223±193 0.853	204±195 0.925	199±174 0.935	189±159 0.947
Average	247±201 0.904	261±165 0.827	242±158 0.842	239±135 0.824	204±133 0.903	205±145 0.925	190±128 0.937	**180±118** **0.945**
				Dimension (mm)				
dim1	2.81±2.76 0.937	3.53±2.77 0.885	2.99±2.43 0.914	2.88±2.48 0.895	\ \	2.87±2.23 0.938	2.62±2.09 0.952	2.47±1.95 0.957
dim2	2.60±2.62 0.946	3.49±2.87 0.897	2.55±2.30 0.938	2.45±2.01 0.932	\ \	2.96±2.35 0.864	2.64±2.12 0.881	2.59±2.07 0.894
dim3	2.49±2.88 0.945	3.91±3.23 0.865	3.10±2.54 0.916	2.93±2.49 0.903	\ \	2.92±2.48 0.924	2.77±2.22 0.935	2.48±2.34 0.943
Average	2.65±2.33 **0.943**	3.64±2.61 0.882	2.88±2.03 0.923	2.75±1.90 0.910	\ \	2.92±1.89 0.901	2.68±1.64 0.917	**2.51±1.58** 0.925
				RWT (mm)				
IS	1.53±1.73 0.796	1.70±1.47 0.729	1.98±1.58 0.599	1.78±1.40 0.611	1.39±1.13 0.824	1.42±1.21 0.806	1.32±1.09 0.840	1.26±1.04 0.856
I	3.23±2.83 0.720	1.71±1.34 0.603	1.67±1.40 0.582	1.68±1.41 0.462	1.51±1.21 0.701	1.53±1.25 0.678	1.38±1.10 0.751	1.40±1.10 0.747
IL	4.15±3.17 0.743	1.97±1.54 0.483	1.88±1.63 0.515	1.92±1.45 0.435	1.65±1.36 0.671	1.74±1.43 0.618	1.57±1.35 0.691	1.59±1.29 0.693
AL	5.08±3.95 0.706	1.82±1.41 0.533	1.87±1.55 0.493	1.66±1.20 0.547	1.53±1.25 0.698	1.59±1.31 0.657	1.60±1.36 0.651	1.57±1.34 0.659
A	3.47±3.25 0.724	1.55±1.33 0.685	1.65±1.45 0.599	1.20±1.01 0.661	1.30±1.12 0.781	1.36±1.17 0.754	1.34±1.11 0.768	1.32±1.10 0.777
AS	1.76±1.80 0.785	1.68±1.43 0.777	2.04±1.59 0.626	1.63±1.23 0.726	1.28±1.00 0.871	1.43±1.24 0.821	1.26±1.10 0.864	1.25±1.01 0.877
Average	3.21±1.98 0.746	1.73±0.97 0.635	1.85±1.03 0.569	1.65±0.77 0.573	1.44±0.71 0.758	1.51±0.81 0.723	1.41±0.72 0.761	**1.39±0.68** **0.768**
				Phase (%)				
phase	\	22.2	22.6	16.17	\	13.0	10.4	**8.2**

0.925, and 0.768, for areas, dimensions and RTWs with respect to their manually obtained ground truth values. For reference, the maximums of these indices in our dataset are 4936 mm^2, 81.0 mm, 24.4 mm. The error rate for cardiac phase identification is 8.2%.

Among these indices, the area of myocardium and the RWTs involves both the endocardium and epicardium contours (usually invisible near the free lateral wall), and are more difficult to estimate than dimensions and area of cavity. However, DMTRL can still obtain an low average MAE of 1.39 mm for RWTs (which is less than the mode of pixel spacing 1.5625 mm/pixel in our dataset), and low average MAE of 189 mm^2 for myocardium area.

Figure 7 shows comparison between the estimated LV indices by DMTRL and their ground truth values across the whole cardiac cycle for an average subject in our dataset. The MR images in the sequences are of high noise, and poor lateral contrast, as shown in the top row of the figure. Even so, DMTRL provides estimations that are very close to the ground truth for all LV indices across the whole cardiac cycle. Take A-cav for example, the maximum margin between the estimated values and the true values are 360 mm^2, which is only 7.29% of the maximal cavity area in the dataset. From the curves in Figure 7, we can also draw that: 1)the temporal dynamic patterns of LV indices are successfully captured, e.g, the dimensions share the same dynamic pattern with the cavity area, while the RTWs follow the same dynamic pattern with myocardium area; 2) the regional variation pattern of myocardium is also well captured by the estimated RWTs. With these estimations, cardiac functions, both global and regional, can be efficiently assessed in clinical practice.

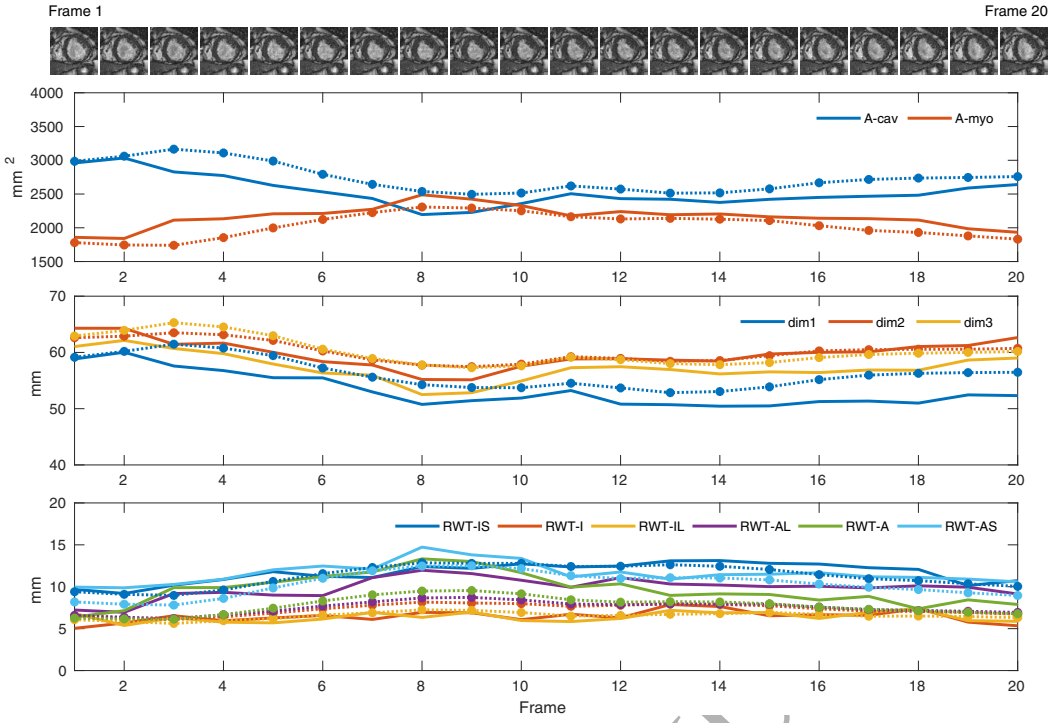


Figure 7: Example of LV indices estimation by DMTRL for an average subject across the whole cardiac cycle. The estimated results (dashed line) are very close their ground truth values (solid line in the same color) for the three types of LV indices, and captures well the temporal variation pattern of these indices. The corresponding MR images for each frame are shown in the top row for visually comparison.

4.2. Effectiveness of multitask relationship learning

The effectiveness of the multitask relationship learning can be proven by the last three columns of Table 1. For all LV indices and cardiac phase, DMTRL outperforms FullLVNet(N/N) (Xue et al., 2017b), where no task relationship is explored during the network training; and defeats FullLVNet(intra/inter) (Xue et al., 2017b), where only predefined task-dependent assumptions are incorporated during the network training. By contrast, DMTRL captures automatically the multitask relationship from the dataset through a data-driven manner, and improves the generalization and the performance of the neural network for full LV quantification.

Figure 8 shows the multitask correlation matrix $\mathbf{C}$ for the LV indices, which can be calculated from the multitask covariate matrix $\mathbf{\Omega}$ learned by DMTRL as $\mathbf{C} = \mathbf{\Omega} \oslash (d \times d^T)$, where $d = \sqrt{diag(\mathbf{\Omega})}$ and $\oslash$ denotes element-wise division. Comparing with the true correlation matrix shown in Figure 6, DMTRL is capable of characterizing well correlation among these LV indices. Specifically, myocardium area is correlated with RWTs, while cavity area is correlated with dimensions. Strong correlations exist within tasks of RWTs and dimensions, especially for neighboring indices. These uncovered correlations validate the effectiveness of our multitask relationship learning.

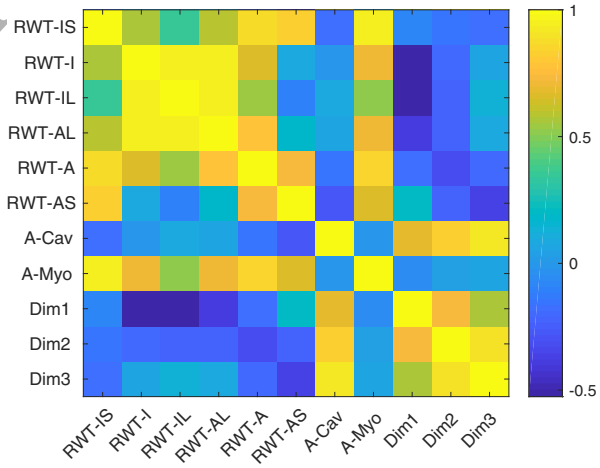


Figure 8: Multitask correlation matrix $\mathbf{C}$ learned by DMTRL from our dataset. The learned correlation matrix $\mathbf{C}$ is very close to its ground truth shown in Figure 6.

4.3. Performance Comparison

DMTRL reveals great advantages for LV indices estimation and cardiac phase identification over existing segmentation based method Max-flow (Ayed et al., 2012), two-phase direct methods (Zhen et al., 2014, 2015, 2016), and deep neural network based methods (Xue et al., 2017a,b), as shown in Table 1[2]

[2]For Max Flow, additional computations are required to get these LV indices from the segmented results; For the two-phase volume-only direct methods Zhen et al. (2014, 2015, 2016), we obtain the results for other indices and cardiac phase by applying their extracted

1) DMTRL outperforms significantly the segmentation-based method Max-flow, with average MAE reductions of 56.7%, 5.28% and 27.1% for RWTs, dimensions and areas. Max-flow achieves the best estimation for cavity area with the requirement of manually segmented myocardium of the first frame for each subject, while performs very poorly for RWTs and myocardium area which are subjected to both contours of endocardium and epicardium. Besides, its estimation errors for areas and RWTs increase clearly as the frame becomes far from the first frame within the cardiac cycle (Figure 9). This makes it incapable of capturing the variation pattern of the LV indices across the whole cardiac cycle.

2) DMTRL outperforms the best of existing two-phase non-end-to-end direct methods, with MAE reductions of 15.8%, 8.7%, 24.7% for RWTs, dimensions and areas, and 49.4% ER reduction for cardiac phase. Contrast to these methods that follow two separated phases and are based on handcrafted features, DMTRL forms an end-to-end framework, in which powerful task-aware cardiac representations can be obtained, with the cardiac dynamics well modeled and the final regression being compatible with them.

3) DMTRL also outperforms existing end-to-end deep neural network based methods. Although beging an end-to-end framework, Indices-Net employed neither cardiac temporal dynamics nor multitask relationship, while FullLVNet explored the task relationship based on predefined task-dependencies assumptions. Therefore, both lead to inferior performance to DMTRL.

4) DMTRL even helps improve the performance of phase identification. The multitask relationship learning uncovers the task relationships of LV indices and leads to better CNN representations, which then benefit the phase identification in the second training stage.

Figure 9 demonstrates the frame-wise estimation error averaged across all subjects for the three types of LV indices. It clearly shows that DMTRL can deliver more robust and accurate estimation than its competitors for all the frames in the cardiac cycle, which is important for comprehensive cardiac function analysis.

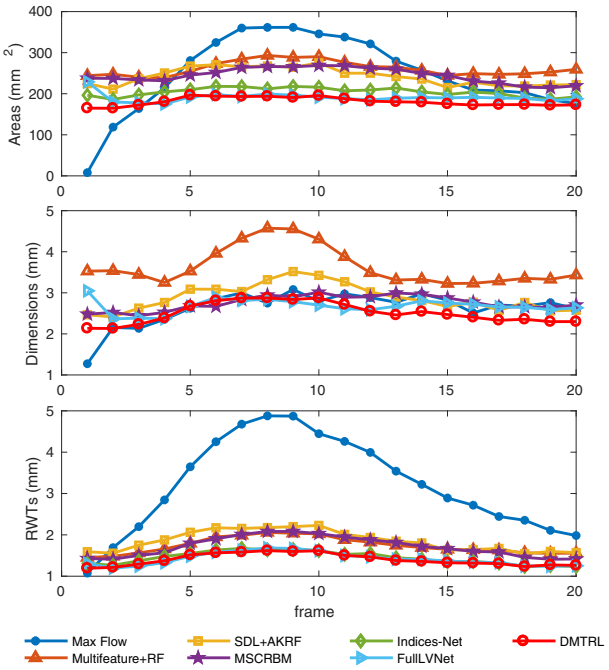


Figure 9: Average frame-wise estimation errors of the three types of LV indices for our proposed DMTRL and its competitors. DMTRL delivers consistently lower estimation error across the cardiac cycle than the segmentation-based method, two-phase direct methods and methods based on deep neural network.

5. Conclusions

In this work, we proposed a integrated deep multitask relationship learning framework DMTRL for full quantification of cardiac LV. DMTRL combines advantages of the powerful CNN-based representation for cardiac MR images, RNN-based dynamic modeling for cardiac sequences, and multitask relationship learning among different types of LV indices. It is the first time that multitask relationship leaning are investigated in deep neural networks: from the representations obtained by deep neural network, multitask correlation can be faithfully learned; with the multitask correlation incorporated during the network training, improved generalization of the network can be achieved. When validated on cardiac MR images of 145 subjects, DMTRL not only achieves accurate estimation for all the three types of LV indices and cardiac phase, but also captures successfully the variation pattern of all LV indices across the cardiac cycle, revealing its great potential of global and regional cardiac function assessment in clinical practice.

Acknowledgments

Computations were performed using the data analytics Cloud at SHARCNET (http://www.sharcnet.ca) provided through the Southern Ontario Smart Computing Innovation Platform (SOSCIP); the SOSCIP consortium is funded by the Ontario Government and the Federal Economic Development Agency for Southern Ontario. The authors also wish to thank Dr. Jinhui Qin for assistance with the computing environment.

Appendix A. Convexity of $tr(\mathbf{W\Omega^{-1}W^T})$

Denote $w_j, j = 1...K$ the jth row of $\mathbf{W}$, then

$$tr(\mathbf{W\Omega^{-1}W^T}) = \sum_{j=1}^{K} w_j \mathbf{\Omega}^{-1} w_j^T. \qquad (A.1)$$

features to an RF/AKRF model.

The Hessian of the term $w_j\mathbf{\Omega}^{-1}w_j^T$ w.r.t w_j is $\mathbf{\Omega}^{-1}$. Since $\mathbf{\Omega} \in \mathcal{S}_+^K$, for any row vectors u and $v = u\mathbf{\Omega}$, there is

$$\begin{aligned} v\mathbf{\Omega}^{-1}v^T &= u\mathbf{\Omega}\mathbf{\Omega}^{-1}\mathbf{\Omega}u^T \\ &= u\mathbf{\Omega}u^T \\ &\succeq 0 \end{aligned} \tag{A.2}$$

Therefor $\mathbf{\Omega}^{-1} \succeq 0$, and the term $w_j\mathbf{\Omega}^{-1}w_j^T$ is convex w.r.t. w_j. Convexity is preserved in summation, thus $tr(\mathbf{W}\mathbf{\Omega^{-1}}\mathbf{W^T})$ is convex w.r.t. $\mathbf{W}$.

Denote $f(t) = w_j(\mathbf{\Omega}+t\Delta)^{-1}w_j^T$, where Δ is a symmetric matrix, the convexity of $w_j\mathbf{\Omega}^{-1}w_j^T$ w.r.t. $\mathbf{\Omega}$ is equivalent to

$$\frac{d^2}{dt^2}f(t)\bigg|_{t=0} \geq 0. \tag{A.3}$$

For $f(t)$, the following holds:

$$\begin{aligned} f(t) &= w_j(\mathbf{\Omega}(\mathbf{I} + t\mathbf{\Omega}^{-1}\Delta))^{-1}w_j^T \\ &= w_j(\mathbf{I} + t\mathbf{\Omega}^{-1}\Delta)^{-1}\mathbf{\Omega}^{-1}w_j^T \\ &= w_j(\sum_{n=0}^{\infty}(-1)^n(t\mathbf{\Omega}^{-1}\Delta)^n)\mathbf{\Omega}^{-1}w_j^T \end{aligned} \tag{A.4}$$

Then we can get

$$\begin{aligned} \frac{d^2}{dt^2}f(t)\bigg|_{t=0} &= w_j(t\mathbf{\Omega}^{-1}\Delta)^2\mathbf{\Omega}^{-1}w_j^T \\ &= w_j\mathbf{\Omega}^{-1}\Delta\mathbf{\Omega}^{-1}\Delta\mathbf{\Omega}^{-1}w_j^T \\ &= \tilde{w}_j\mathbf{\Omega}^{-1}\tilde{w}_j^T \\ &\geq 0 \end{aligned} \tag{A.5}$$

where $\tilde{w}_j = w_j\mathbf{\Omega}^{-1}\Delta$. The last inequality holds because $\mathbf{\Omega}^{-1} \succeq 0$. Therefore, $w_j\mathbf{\Omega}^{-1}w_j^T$ is convex w.r.t $\mathbf{\Omega}$. Again, convexity is preserved in summation, thus $tr(\mathbf{W}\mathbf{\Omega^{-1}}\mathbf{W^T})$ is convex w.r.t. $\mathbf{\Omega}$.

Appendix B. Convergence of Algorithm 1

The two stages in Algorithm 1 both converges, as proved below.

1) $\mathcal{L}_{MTL}$: We first prove that $\mathcal{L}_{MTL}$ is bonded from below and monotonically decreases with each updating step in *line 3,4,5* of Algorithm 1.

As shown in Equation 11, the first two terms in $\mathcal{L}_{MTL}$ are mean squared errors and the norms of w_t, which are of non-negative values. The third term $tr(\mathbf{W}\mathbf{\Omega^{-1}}\mathbf{W^T})$ is also non-negative due to that $\mathbf{\Omega}^{-1}$ is a positive semidefinite matrix. Therefor $\mathcal{L}_{MTL}$ is bonded from below.

Denote the value of $\mathcal{L}_{MTL}$ after each updating step of *line 3,4,5* in the kth iteration as $\mathcal{L}_{MTL}^{k,1}$, $\mathcal{L}_{MTL}^{k,2}$, $\mathcal{L}_{MTL}^{k,3}$. From the updating rule of w_{cnn} and w_{rnn1}, we can get $\mathcal{L}_{MTL}^{k,1} \geq \mathcal{L}_{MTL}^{k-1,3}$. From the updating rule of $\mathbf{W}$ and the fact that $\mathcal{L}_{MTL}$ is convex w.r.t. $\mathbf{W}$, we have $\mathcal{L}_{MTL}^{k,2} \geq \mathcal{L}_{MTL}^{k,1}$. Since the updating rule of $\mathbf{\Omega}$ is the optimal solution of Equation 15, we have $\mathcal{L}_{MTL}^{k,3} \geq \mathcal{L}_{MTL}^{k,2}$. In this way, we can obtain the following monotonic decreasing sequences:

$$\cdots \mathcal{L}_{MTL}^{k,1}, \mathcal{L}_{MTL}^{k,2}, \mathcal{L}_{MTL}^{k,3}, \mathcal{L}_{MTL}^{k+1,1}, \mathcal{L}_{MTL}^{k+1,2}, \mathcal{L}_{MTL}^{k+1,3} \cdots . \tag{B.1}$$

Based on the monotone converge theory, $\mathcal{L}_{MTL}$, which is bonded below, converges as $k \to +\infty$.

2) $\mathcal{L}_{phase}$: The second stage of Algorithm 1 is exactly the standard training procedure of a softmax classifier on top of a RNN network and converges with the gradient-based updating rules.

References

References

Afshin, M., Ayed, I.B., Islam, A., Goela, A., Peters, T.M., Li, S., 2012. Global assessment of cardiac function using image statistics in MRI, in: MICCAI. Springer, pp. 535–543.

Afshin, M., Ben Ayed, I., Punithakumar, K., Law, M., Islam, A., Goela, A., Peters, T.M., Li, S., 2014. Regional assessment of cardiac left ventricular myocardial function via MRI statistical features. IEEE Transactions on Medical Imaging 33, 481–494.

Avendi, M., Kheradvar, A., Jafarkhani, H., 2016. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Medical image analysis 30, 108–119.

Ayed, I.B., Chen, H.M., Punithakumar, K., Ross, I., Li, S., 2012. Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure. Medical image analysis 16, 87–100.

Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences 2, 183–202.

Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., Verani, M.S., et al., 2002. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. Circulation 105, 539–542.

Chen, J., Zhou, J., Ye, J., 2014. Low-rank and sparse multi-task learning, in: Low-Rank and Sparse Modeling for Visual Analysis. Springer, pp. 151–180.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T., 2015. Long-term recurrent convolutional networks for visual recognition and description, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2625–2634.

Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., Heng, P.A., 2017. 3D deeply supervised network for automated segmentation of volumetric medical images. Medical Image Analysis 41, 40–54.

Gong, P., Ye, J., Zhang, C., 2012. Robust multi-task feature learning, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 895–903.

Graves, A., 2012. Supervised sequence labelling, in: Supervised Sequence Labelling with Recurrent Neural Networks. Springer, pp. 5–13.

Greenspan, H., van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35, 1153–1159.

Gupta, A., Von Kurowski, L., Singh, A., Geiger, D., Liang, C.C., Chiu, M.Y., Adler, L., Haacke, M., Wilson, D., 1993. Cardiac MR image segmentation using deformable models, in: Computers in Cardiology 1993, Proceedings., IEEE. pp. 747–750.

Jacob, L., Vert, J.p., Bach, F.R., 2009. Clustered multi-task learning: A convex formulation, in: Advances in neural information processing systems, pp. 745–752.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM. pp. 675–678.

Kabani, A., El-Sakka, M.R., 2016. Estimating ejection fraction and left ventricle volume using deep convolutional networks, in: International Conference Image Analysis and Recognition, pp. 678–686.

Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B., Neubauer, S., 2009. The role of cardiovascular magnetic resonance imaging in heart failure. Journal of the American College of Cardiology 54, 1407–1424.

Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S., 2016. Recognizing end-diastole and end-systole frames via deep temporal regression network, in: MICCAI, Springer. pp. 264–272.

Li, Y., Lan, C., Xing, J., Zeng, W., Yuan, C., Liu, J., 2016. Online human action detection using joint classification-regression recurrent neural networks, in: ECCV.

Liang, X., Shen, X., Xiang, D., Feng, J., Lin, L., Yan, S., 2015. Semantic object parsing with local-global long short-term memory. arXiv preprint arXiv:1511.04510 .

Lin, K., Zhou, J., 2016. Interactive multi-task relationship learning, in: International Conference on Data Mining, IEEE. pp. 241–250.

Liu, J., Ji, S., Ye, J., 2009. Multi-task feature learning via efficient l 2, 1-norm minimization, in: 25th Conference on Uncertainty in Artificial Intelligence, AUAI Press. pp. 339–348.

Lötjönen, J., Kivistö, S., Koikkalainen, J., Smutek, D., Lauerma, K., 2004. Statistical shape model of atria, ventricles and epicardium from short-and long-axis MR images. Medical image analysis 8, 371–386.

Mortazi, A., Burt, J., Bagci, U., 2017. Multi-planar deep segmentation networks for cardiac substructures from MRI and CT. STACOM .

Nachtomy, E., Cooperstein, R., Vaturi, M., Bosak, E., Vered, Z., Akselrod, S., 1998. Automatic assessment of cardiac function from short-axis mri: procedure and clinical evaluation. Magnetic resonance imaging 16, 365–376.

Ngo, T.A., Carneiro, G., 2013. Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks, in: ICIP, IEEE. pp. 695–699.

Ngo, T.A., Lu, Z., Carneiro, G., 2017. Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Medical image analysis 35, 159–171.

Patravali, J., Jain, S., Chilamkurthy, S., 2017. 2D-3D fully convolutional neural networks for cardiac MR segmentation. STACOM .

Pednekar, A., Kurkure, U., Muthupillai, R., Flamm, S., Kakadiaris, I.A., 2006. Automated left ventricular segmentation in cardiac MRI. IEEE Transactions on Biomedical Engineering 53, 1425–1428.

Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F., 2016. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine 29, 155–195.

Petitjean, C., Dacher, J.N., 2011. A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis 15, 169–184.

Poudel, R.P., Lamata, P., Montana, G., 2016. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. arXiv preprint arXiv:1608.03974 .

Rennie, J.D., 2005. Regularized logistic regression is strictly convex. Unpublished manuscript. URL people. csail. mit. edu/jrennie/writing/convexLR. pdf .

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 640–651.

Shin, H.C., Roberts, K., Lu, L., Demner-Fushman, D., Yao, J., Summers, R.M., 2016. Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. arXiv preprint arXiv:1603.08486 .

Suinesiaputra, A., Bluemke, D.A., Cowan, B.R., Friedrich, M.G., Kramer, C.M., Kwong, R., Plein, S., Schulz-Menger, J., Westenberg, J.J., Young, A.A., et al., 2015. Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. Journal of Cardiovascular Magnetic Resonance 17, 63.

Tan, L.K., Liew, Y.M., Lim, E., McLaughlin, R.A., 2017. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. Medical Image Analysis 39, 78–86.

Tran, P.V., 2016. A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv:1604.00494 .

Wang, Z., Ben Salah, M., Gu, B., Islam, A., Goela, A., Li, S., 2014. Direct estimation of cardiac biventricular volumes with an adapted bayesian formulation. IEEE TBE 61, 1251–1260.

Wolterink, J.M., Leiner, T., Viergever, M.A., Isgum, I., 2017. Automatic segmentation and disease classification using cardiac cine MR images. STACOM .

Wu, Y., Wang, Y., Jia, Y., 2013. Segmentation of the left ventricle in cardiac cine MRI using a shape-constrained snake model. Computer Vision and Image Understanding 117, 990–1003.

Xue, W., Islam, A., Bhaduri, M., Li, S., 2017a. Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Transactions on Medical Imaging .

Xue, W., Lum, A., Mercado, A., Landis, M., Warrington, J., Li, S., 2017b. Full quantification of left ventricle via deep multitask learning network respecting intra- and inter-task relatedness, in: MICCAI. Springer.

Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S., 2017c. Direct estimation of regional wall thicknesses via residual recurrent neural network, in: IPMI, Springer. pp. 505–516.

Yin, X., Liu, X., 2017. Multi-task convolutional neural network for face recognition. arXiv preprint arXiv:1702.04710 .

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks, in: ECCV, Springer. pp. 818–833.

Zhang, X., Lu, L., Lapata, M., 2015. Tree recurrent neural networks with application to language modeling. arXiv preprint arXiv:1511.00060 .

Zhang, Y., Yeung, D.Y., 2010. A convex formulation for learning task relationships in multi-task learning, in: 26th Conference on Uncertainty in Artificial Intelligence.

Zhang, Z., Luo, P., Loy, C.C., Tang, X., 2014. Facial landmark detection by deep multi-task learning, in: European Conference on Computer Vision, Springer. pp. 94–108.

Zhen, X., Islam, A., Bhaduri, M., Chan, I., Li, S., 2015. Direct and simultaneous four-chamber volume estimation by multi-output regression, in: MICCAI, Springer. pp. 669–676.

Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S., 2014. Direct estimation of cardiac bi-ventricular volumes with regression forests, in: MICCAI. Springer, pp. 586–593.

Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S., 2016. Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. Medical Image Analysis 30, 120–129.

Zhou, J., Yuan, L., Liu, J., Ye, J., 2011. A multi-task learning formulation for predicting disease progression, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 814–822.

Zotti, C., Luo, Z., Lalande, A., Humbert, O., Jodoin, P.M., 2017. Novel deep convolution neural network applied to MRI cardiac segmentation. arXiv preprint arXiv:1705.08943 .