Medical Image Analysis 36 (2017) 184–196

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Direct and simultaneous estimation of cardiac four chamber volumes by multioutput sparse regression

Xiantong Zhen[a,b], Heye Zhang[c,*], Ali Islam[d], Mousumi Bhaduri[e], Ian Chan[e], Shuo Li[a,b,*]

[a] The University of Western Ontario, London, ON, Canada
[b] Digital Image Group (DIG), London, ON, Canada
[c] Institute of Biomedical and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[d] St. Joseph's Health Care, London, ON, Canada
[e] London Healthcare Sciences Centre, London, ON, Canada

ARTICLE INFO

Article history:
Received 27 April 2016
Revised 22 September 2016
Accepted 22 November 2016
Available online 30 November 2016

Keywords:
Cardiac four-chamber volume estimation
Direct estimation
Multi-output regression
Supervised descriptor learning

ABSTRACT

Cardiac four-chamber volume estimation serves as a fundamental and crucial role in clinical quantitative analysis of whole heart functions. It is a challenging task due to the huge complexity of the four chambers including great appearance variations, huge shape deformation and interference between chambers. Direct estimation has recently emerged as an effective and convenient tool for cardiac ventricular volume estimation. However, existing direct estimation methods were specifically developed for one single ventricle, i.e., left ventricle (LV), or bi-ventricles; they can not be directly used for four chamber volume estimation due to the great combinatorial variability and highly complex anatomical interdependency of the four chambers.

In this paper, we propose a new, general framework for direct and simultaneous four chamber volume estimation. We have addressed two key issues, i.e., cardiac image representation and simultaneous four chamber volume estimation, which enables accurate and efficient four-chamber volume estimation. We generate compact and discriminative image representations by supervised descriptor learning (SDL) which can remove irrelevant information and extract discriminative features. We propose direct and simultaneous four-chamber volume estimation by the multioutput sparse latent regression (MSLR), which enables jointly modeling nonlinear input-output relationships and capturing four-chamber interdependence. The proposed method is highly generalized, independent of imaging modalities, which provides a general regression framework that can be extensively used for clinical data prediction to achieve automated diagnosis. Experiments on both MR and CT images show that our method achieves high performance with a correlation coefficient of up to 0.921 with ground truth obtained manually by human experts, which is clinically significant and enables more accurate, convenient and comprehensive assessment of cardiac functions.

1. Introduction

Cardiac ventricular volumes have been widely used as a measurement of cardiac abnormalities and functions, *e.g.*, ejection fraction (EF) and stroke volume (Wang et al., 2009; Punithakumar et al., 2013; Marchesseau et al., 2013). One single ventricle, i.e, left ventricle (LV) (Suinesiaputra et al., 2014) and right ventricles (RV) (Petitjean et al., 2015), have been extensively studied due to their critical role in heart disease diagnosis. However, cardiac four chambers, which provide more comprehensive information for quantitative functional analysis of the whole heart (Zheng et al., 2008), have less researched jointly due to the great challenges; the left/right atrium (LA/RA) are anatomically connected to LV and RV, respectively, and LA/RA volumes are strongly associated with heart functions and indicate severity of diastolic dysfunctions and cardiovascular disease burden (Baur, 2008; Glatard et al., 2013). However, LA and RA have long been less studied due to the particular difficulty in measuring their volumes, which started to attract increasingly attentions recently (Fonseca et al., 2011; Tobon-Gomez et al., 2015).

Simultaneous cardiac four chamber volume estimation is extremely challenging due to the substantial topology changes and high anatomical variability of the four chambers (Pace et al., 2015). Compared to one single ventricle and bi-ventricles, the four chambers demonstrate greater combinatorial shape variabilities and appearance variations as shown in Fig. 1. Furthermore, the four

* Corresponding authors.
E-mail addresses: hy.zhang@siat.ac.cn (H. Zhang), slishuo@gmail.com (S. Li).

http://dx.doi.org/10.1016/j.media.2016.11.008
1361-8415/

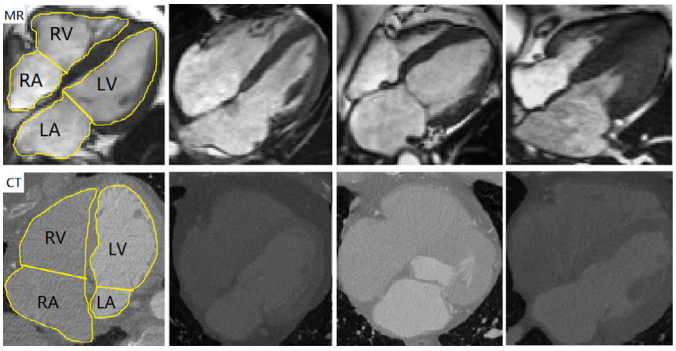

Fig. 1. The four chambers exhibit high contour complexity and low tissue contrast in both MR (top) and CT (bottom) images.

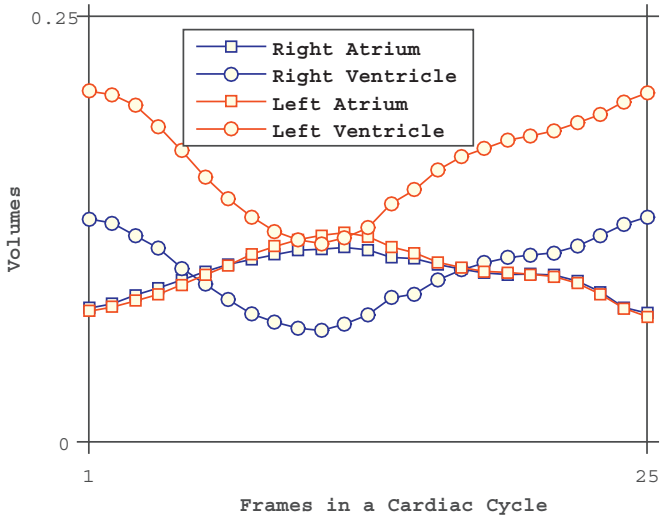

Fig. 2. The volume change of the four chambers in a cardiac cycle.

chambers are highly interdependent and temporally interfere with each other across a cardiac cycle, which poses even more challenges for simultaneous four chamber volume estimation.

Segmentation has made great progress and becomes more reliable, accurate and less time-consuming, and cardiac image segmentation has been mainly focus on the LV with few on the RV which remains unsolved (Petitjean and Dacher, 2011; Nambakhsh et al., 2013), LA and RA volume estimation has long been overlooked (Fonseca et al., 2011), while simultaneous four-chamber volume estimation has not been addressed yet. The whole heart segmentation algorithm (Zheng et al., 2008; Ecabert et al., 2008) potentially offers a solution to simultaneous four-chamber volume estimation; however, it turns out to be extremely challenging to segment the four chambers separately to obtain individual volumes due to the fact that ventricle and atrium are of intensity homogeneity with vague boundary between them, and two atriums are mostly connected with a very thin wall as shown in Fig. 1.

Moreover, even more challenges arise from dealing with multiple modalities due to huge variations between and within modalities. However, both MR and CT are routinely used in clinical practice (Donnell et al., 2006; Jolly, 2006), which yield specific advantages (Nikolaou et al., 2011) and serve complementary roles in diagnosis of cardiac diseases (Wintersperger, 2009). Cardiac CT allows for a reliable exclusion of significant coronary artery disease in proper patient populations; cardiac MR is the modality of choice in cardiac functional aspect, e.g., cardiac and valvular function and perfusion. In both medical image analysis and clinical practice, it is highly desired to have a general method that can naturally handle cardiac chamber volume estimation, irrespective of modalities (Jolly, 2006).

Direct estimation without segmentation has recently emerged as a convenient and effective tool for cardiac ventricular volume estimation. The effectiveness of direct estimation has been shown on one single ventricle, i.e., LV, and bi-ventricles, i.e., LV and RV, (Afshin et al., 2012; Zhen et al., 2014b; Wang et al., 2013, 2014b; Zhen et al., 2014a, 2016c; Mukhopadhyay et al., 2015; Kabani and El-Sakka, 2016; Hussain et al., 2016; Kong et al., 2016), which sheds light on solving even more challenging cardiac four-chamber volume estimation. However, in contrary to one single ventricle or bi-ventricles, the four chambers exhibit much larger shape deformations, greater combinatorial variability and more complex anatomical interdependency, which coupled with challenges from multi-modal imaging data makes existing direct estimation methods not directly applicable to four-chamber volume estimation. Moreover, LV/RV and LA/RA share similar change patterns respectively in a cardiac cycle as illustrated in Fig. 2, which has not been explored in existing methods. It is therefore more attractive and highly desired to simultaneously estimate four chamber volumes. By capturing the interdependency of the four chambers to leverage the shared information and anatomical dependency among the four chambers, the estimation performance can be largely improved, which more importantly ensures clinically meaningful volume estimation.

In this paper, we propose a new direct method for simultaneous four-chamber volume estimation by multioutput regression. In a general framework, we have addressed two key issues, i.e., cardiac four chamber image representation and simultaneous four chamber volume estimation.

- To tackle the great combinatorial variabilities of the four chambers, we propose to use supervised descriptor learning (SDL) (Zhen et al., 2015b) to generate compact and discriminative cardiac four-chamber image representations; the SDL can jointly remove irrelevant and redundant information and extract features directly associated with targets, i.e., four chamber volumes.
- To cope with the high complexity of the four chambers, we propose direct and simultaneous four-chamber volume estimation by the multioutput sparse latent regression (MSLR) (Zhen et al., 2016a) to handle the highly nonlinear relationship between image appearance and four chamber volumes while jointly modeling the interdependency of the four chambers; Leveraging interdependency and reciprocal validation of the four chambers, simultaneous four-chamber volume estimation is clinically very significant for more accurate and comprehensive assessment of cardiac functions.

1.1. Contributions

In this work, we have made multiple contributes in the following four aspects:

1) We achieve simultaneous cardiac four-chamber volume estimation, which avoids unreliable segmentation and enables more accurate and convenient whole heart functional analysis. The method can be conveniently extended to other clinical data prediction for automated diagnosis;
2) We formulate four-chamber volume estimation as a multioutput regression problem, which enables volume estimation to be conducted in a unified framework rather than as specific problems. Traditional challenging tasks, e.g., model personalization, can be reformulated as multioutput regression problems and solved efficiently in the same way;
3) We propose direct and simultaneous four chamber volume estimation by multioutput sparse latent regression (MSLR). The

MSLR can jointly handle the highly nonlinear relationship between image appearance and four chamber volumes while capturing the interdependency of the four chambers, which enables more accurate and clinically meaningful volume estimation;

4) We achieve a highly generalized direct estimation framework independent of modalities, and we have conducted extensive experimental evaluation on both MR and CT data, which validates the generality of the proposed methods for direct and simultaneous cardiac four-chamber volume estimation from imaging data of multiple modalities. The success of our method on both MR and CT in this work will significantly change the way to handle multi-modality image analysis and inspire new clinical applications based on direct estimation.

The preliminary test of direct and simultaneous four chamber volume estimation on a small MR dataset has been conducted in (Zhen et al., 2015a). In this journal version, we have made new contributions in terms of both methodology and experimental validation.

- We propose direct and simultaneous four chamber volume estimation by the multioutput sparse latent regression (MSLR) (Zhen et al., 2016a). In one single framework, the MSLR can effectively handle highly nonlinear relationships between image appearance and four chamber volumes by kernel regression, while explicitly modeling the interdependency of the four chambers by sparse learning. The MSLR provides a well-suited multioutput regression model for direct and simultaneous four chamber volume estimation, which enables more accurate and clinical meaningful volume estimation of the four chambers compared to existing multioutput regression models.
- We generalize the method from one single modality to multiple modalities, i.e., MR and CT. Our method achieves quantitative analysis of cardiac images in one single framework regardless of modalities. We demonstrate that our method can create discriminative representations of cardiac four-chamber images of multiple modalities and estimate volumes from images, e.g., CT, without need of clear boundaries. Our method provides a unified framework of multioutput regression for direct and simultaneous four-chamber volume estimation, which has seldom been addressed and remains unsolved yet.

1.2. Related work

Direct methods via regression has recently generated increasing interest in both computer vision and medical image analysis (Fletcher, 2013; Hara and Chellappa, 2014; Toshev and Szegedy, 2014; Torki and Elgammal, 2011; Wang et al., 2010; Joshi et al., 2010; Davis et al., 2010; Prakosa et al., 2014; Metz et al., 2012; Zhou, 2010; Criminisi et al., 2009; 2010; Criminisi and Shotton, 2013; Criminisi et al., 2013; Zhen et al., 2014a; Zhou, 2014; Kainz et al., 2015; Kabani and El-Sakka, 2016). One of the most attractive advantages of direct methods via regression is its nature to avoid conventional unreliable, computationally expensive, even intractable segmentation (Zhou et al., 2007; Metz et al., 2012; Criminisi and Shotton, 2013; Prakosa et al., 2014) and inverse problems (Jiang et al., 2011; Zettinig et al., 2014; Guzman-Rivera et al., 2014). Formulating as a regression problem not only offers a more compact and exquisite mathematical formula to circumvent the difficulty in conventional approaches but also substantially improves the performance. More importantly, regression serves as a bridge that allows to deploy advanced machine learning techniques to facilitate medical image analysis. It also provides an effective tool to automate analysis of medical imaging data and therefore enables accurate and efficient diagnosis in clinical practice (Wang and Summers, 2012).

We review representative work based on regression for medical image analysis including shape inference (Zhou, 2010), organ localization (Criminisi et al., 2013), model personalization (Zettinig et al., 2014), cell detection (Kainz et al., 2015) and cardiac ventricular volume estimation (Afshin et al., 2012; Wang et al., 2014a; Zhen et al., 2014b).

Zhou (2010) proposed a machine learning approach called shape regression machine (SRM) for efficient shape reference of an anatomical structure by boosting based regression without explicitly conducting segmentation. The SRM consists of two components, i.e., regression based object detection and regression based deformation. Due to the statistical learning of regression, the SRM can extract knowledge from the annotated database for accurate estimation. Criminisi et al. (2013) proposed a new parametrization of the anatomy localization task as a multivariate, continuous parameter estimation problem which is implemented effectively via tree based, non-linear regression, i.e., regression forests. The algorithm can localize both macroscopic anatomical regions, e.g. abdomen, thorax, trunk and smaller scale structures, e.g. heart, adrenal gland, femoral neck using a single, efficient model in CT scans. Recently, Kainz et al. (2015) proposed automated cell detection in histopathology images by regression. Cells can be detected reliably in images by predicting, for each pixel location, a monotonous function of the distance to the center of the closest cell. Cell centers can then be identified by extracting local extremums of the predicted values. Zettinig et al. (2014) proposed a data-driven estimation of cardiac electrical diffusivity from 12-lead ECG for diagnosis and treatment of dilated cardiomyopathy (DCM). Instead of solving an inverse problem to find patient-specific parameters of electrophysiology (EP) model, they learn the inverse function by formulating as a polynomial regression problem to directly estimate model parameters for specific patients. The ECG features are taken as the input of the regressor with model parameters being the output. The method achieves substantial improvement over conventional approaches.

Regression was first introduced by Zhen et al. (2014b) for direct cardiac bi-ventricular volume estimation, which consists of cardiac image representation and regression forests estimation. Multiple complementary handcrafted features including pyramid Gabor features (Zhen and Shao, 2013), histogram of gradients (HOG) (Dalal and Triggs, 2005) and image pixel intensity are carefully designed to describe shape and appearance of bi-ventricles for cardiac image representations; regression forests (Breiman, 2001) are adopted as regressors for direct bi-ventricular volume estimation.

However, existing direct methods for cardiac ventricular volume estimation (Afshin et al., 2012; Wang et al., 2014a; Zhen et al., 2014b) are carefully and specifically designed for the single LV or bi-ventricles, and use only handcrafted features, which were not able to learn optimal representations by extracting from data the discriminative information related to targets. In addition, existing direct volume estimation of both one single ventricle and bi-ventricles is developed and evaluated on imaging data of a single modality, i.e., MR, which provides higher tissue intensity contrast compared with CT. However, in the CT scans, the boundary between atrium and ventricle is almost invisible due to the low intensity contrast, which poses more challenges for segmentation methods. Our method is highly generalized and for the first time achieves direct and simultaneous four-chamber volume estimation in a unified framework, irrespective of imaging modalities.

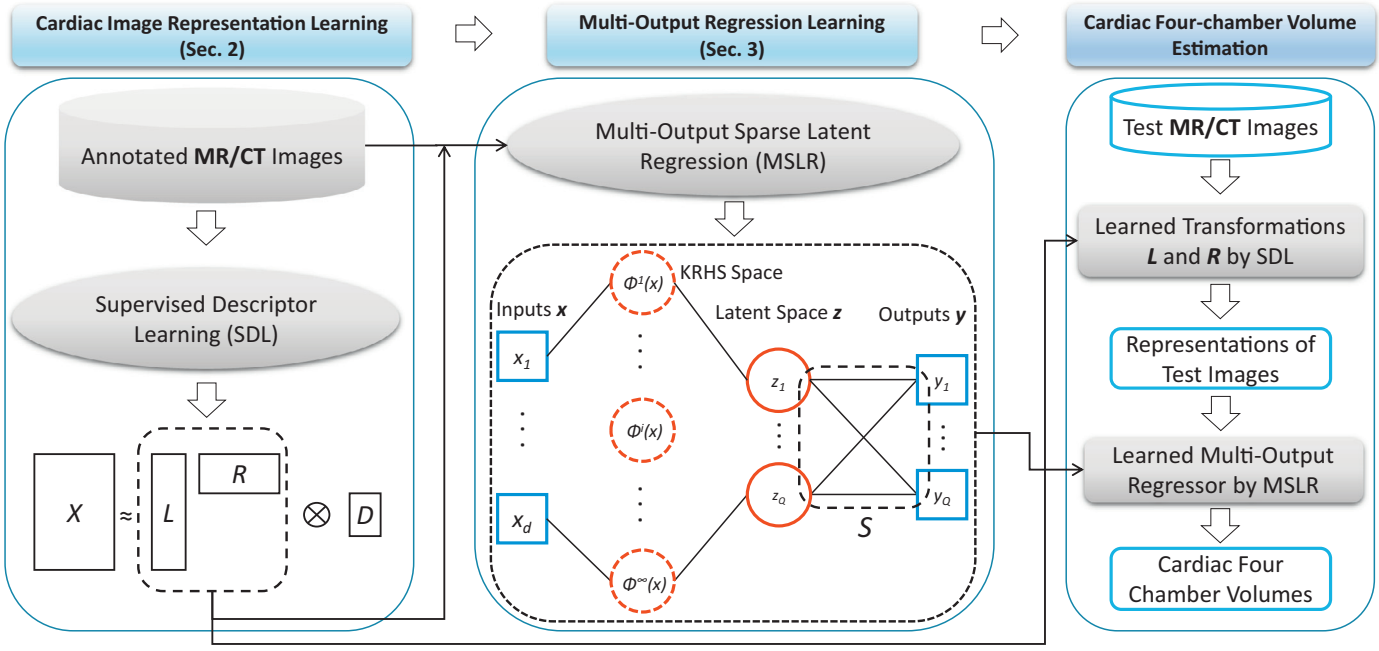

Fig. 3. The work flow of direct and simultaneous four-chamber volume estimation. In the left block, annotated training MR/CT images are used to learn two transformation matrices W and V which are used to obtain descriptors of images. In the middle block, training images associated with targets are fed to regression forests to learn multi-output regressors. In the right block, the incoming test images go through transformation matrices to obtain representations which are fed into learned regression forests to obtain four-chamber volumes.

1.3. Methodology overview

The proposed method consists of two components: cardiac image representation by supervised descriptor learning and four chamber volume estimation by multioutput sparse latent regression (MSLR) as shown in Fig. 3.

- In the training stage, two steps including cardiac image representation learning and multioutput regressor learning are conducted on annotated training data. In the first step (the leftmost block), two feature transformations L and R are learned by the proposed SDL algorithm (Section 2), which can effectively transform complex cardiac four-chamber images into compact but highly discriminative representations. In the second step (the middle block), based on image representations obtained by the SDL, a newly proposed multioutput regression model called multioutput sparse latent regression (MSLR) (Section 3) are then trained on annotated data, which can jointly handle highly nonlinear relationship between image representations and four chamber volumes while modeling interdependency of the four chambers.
- In the prediction stage (the rightmost block), the learned transformations L and R are then applied to test samples to obtain cardiac image representations which are fed into the trained MSLR for direct and simultaneous cardiac four-chamber volume estimation.

We will present the cardiac image representation by the supervised descriptor learning (SDL) in Section 2 and direct and simultaneous four chamber volume estimation by multioutput sparse latent regression (MSLR) in Section 3.

2. Cardiac image representations

Image representation plays crucial role in both computer vision (Yu et al., 2016) and medical image analysis (Zhen et al., 2015a). We propose to use the supervised descriptor learning (SDL) algorithm (Zhen et al., 2015b, 2016b) to obtain effective image representations for accurate and efficient volume estimation. The SDL can establish a compact and discriminative image representation by extracting features directly related to four-chamber volumes. The SDL is formulated as generalized low-rank approximations of matrices with a supervised manifold regularization (SMR).

The generalized low-rank approximations of matrices is able to effectively find low-dimensional and therefore compact representations for efficient estimation. The SMR incorporates the guidance of four-chamber volumes to generate target-oriented feature representations, which enables more accurate and efficient estimation. The SDL provides a best-suited feature learning algorithm for cardiac image representation by taking 2D image matrices as inputs, which can effectively explore the spatial layout information of objects in images to achieve informative representations.

To be self-contained, we briefly sketch the main derivation of the SDL algorithm for completeness. We are given a set of annotated data $\{X_1, \ldots, X_{\mathcal{N}}\}$ and the corresponding multivariate targets $\{\mathbf{y}_1, \ldots, \mathbf{y}_{\mathcal{N}}\}$, where $\mathcal{N}$ is the number of training samples and $y_i \in \mathbb{R}^d$ denote four-chamber volumes. We start with matrix representations of four chamber cardiac images, *i.e.*, $X_i \in \mathbb{R}^{M \times N}$, which could be any matrix representations, *e.g.*, raw pixel intensities. We use the gradient orientation matrix (GOM) which is constructed from pyramid histogram of gradients (PHOG) of images by stacking spatial cells in rows and orientation bins in columns. The GOM takes advantages of prior knowledge to capture characteristic spatial layout and local shape which are the key characteristics of the four chambers. The GOM is fed into the proposed SDL to learn a compact and discriminative representation of the four chambers.

2.1. Generalized low-Rank approximation of matrices

We propose to use the generalized low-rank approximation of matrices due to its appealing properties (Ye, 2005). It operates on matrix representations of images, which 1) reduces time and space costs for more accurate and efficient computation of low-rank matrices than on vectorized representations (Ye, 2005); 2) finds low-dimensional compact feature representations that can substantially

reduce complexity of multioutput regression; 3) allows to explore distinctive physical meanings, *e.g.*, spatial layout and local shape of the four chambers, to achieve optimal cardiac image representations.

We aim to find two transformations: $L \in \mathbb{R}^{M\times m}$ and $R \in \mathbb{R}^{N\times n}$ with $m \ll M$ and $n \ll N$, and $\mathcal{N}$ matrices $D_i \in \mathbb{R}^{m\times n}$ such that LD_iR^{T} is an appropriate approximation of each X_i, $i = 1, \ldots, L$. We solve the following optimization problem of minimizing the reconstruction errors:

$$\arg\min_{\substack{L,R,D_1,\ldots,D_{\mathcal{N}} \\ L^{\mathrm{T}}L=I_m, R^{\mathrm{T}}R=I_n}} \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \|X_i - LD_iR^{\mathrm{T}}\|_F^2 \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, I_m is an identity matrix of size $m \times m$ and the constraints $L^{\mathrm{T}}L = I_m$ and $R^{\mathrm{T}}R = I_n$ ensure that L and R have orthogonal columns to avoid redundancy in the approximations.

From (1), we know that D_i is the low-rank approximation of X_i in terms of the transformations of L and R, and it is worth mentioning that the matrices $D_1, \ldots, D_L$ are not required to be diagonal. It is also proven in (Ye, 2005) that given the L and R, for any i, D_i is uniquely determined by $D_i = L^{\mathrm{T}}X_iR$ which is the compact representation of X_i that will reduce regression complexity for efficient multivariate estimation. (1) only minimize the reconstruction error in the low-rank space leading to indiscriminate representations $\{D\}_{i=1}^{L}$. The regression targets, i.e., the multiple outputs can be explored to guide the feature learning to achieve more discriminative image representations.

2.2. Supervised manifold regularization

To extract features associated with four-chamber volumes for discriminative representations, we propose a supervised manifold regularization (SMR) for supervised learning. The SMR encodes the intrinsic local geometrical structure of the target space and automatically incorporates the supervision to extract target-oriented features. In contrast to the conventional manifold regularization (Belkin et al., 2006; Zhang and Zhao, 2013), the SMR makes full use of targets for supervised descriptor learning which is firstly studied for multioutput regression.

We impose discrimination on the low-rank representation $\{D_i\}_{i=1}^{L}$ by integrating the proposed SMR into (1). To this end, we first construct a complete weighted graph $G = (V, E)$ using the ϵ-neighborhood method (He and Niyogi, 2004), where V and E respectively represent $\mathcal{N}$ vertices and edges between vertices. The graph is constructed on the multivariate targets $(\mathbf{y}_1, \ldots, \mathbf{y}_{\mathcal{N}})$, *i.e.*, the four-chamber volume values, rather than on inputs in conventional manifold regularization (Belkin et al., 2006), which naturally induces the supervision. We denote $A \in \mathbb{R}^{L\times L}$ as the symmetric similarity matrix with non-negative elements corresponding to the edge weight of the graph G, where each element S_{ij} is computed by a heat kernel with parameter σ: $A_{ij} = \exp(\frac{-\|\mathbf{y}_i-\mathbf{y}_j\|^2}{2\sigma^2})$, $i, j = 1, \ldots, \mathcal{N}$. We set the diagonal elements of S to be zeros, *i.e.*, $A_{ii} = 0$. In the low-rank space, we would like to minimize the following term

$$\sum_{i,j} \|D_i - D_j\|_F^2 A_{ij}. \tag{2}$$

Since the similarity matrix A characterizes the manifold structure of the multivariate target space, low-rank approximations $\{D_i\}_{i=1}^{\mathcal{N}}$ preserve the intrinsic local geometrical structure of the target space and are therefore automatically aligned to their regression targets. The discrimination is then naturally injected into the low-rank representations $\{D_i\}_{i=1}^{L}$. An intuitive consequence of minimizing the regularization term is that in the low-dimensional space, data points with similar targets, *i.e.*, four-chamber volumes, are forced to be close while these with different targets tend to be far apart, which therefore increases the discriminative ability in new representations.

2.3. Supervised descriptor learning

By integrating the SMR term in (2) into (1), we obtain the compact objective function of generalized low-rank approximation of matrices with the supervised manifold regularization (SMR) as follows:

$$\arg\min_{\substack{L,R,D_1,\ldots,D_L \\ L^{\mathrm{T}}L=I_m, R^{\mathrm{T}}R=I_n}} \underbrace{\frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \|X_i - LD_iR^{\mathrm{T}}\|_F^2}_{\text{reconstruction errors}} + \underbrace{\beta \sum_{i,j} \|D_i - D_j\|_F^2 A_{ij}}_{\text{supervised manifold regularization}} \tag{3}$$

where $\beta \in (0, \infty)$ is the regularization parameter to balance the tradeoff between reconstruction errors and discrimination of the low-rank approximations, which also serves to keep the flexibility of the model.

In the objective function of (3), the first term guarantees the reconstruction fidelity in the low-rank approximation while the second SMR term introduces the discrimination to learn new representations. The achieved compact and discriminative feature representation of four-chamber cardiac images can not only reduce complexity in multioutput regression of four-chamber volumes while avoiding overfitting but also guarantee more accurate estimation. The objective in (3) is solved by an iterative algorithm via alternate optimization: fixing L, solve R and fixing R, solve L. Once the transformation matrices L and R are learned, for a new input matrix X_t, its descriptor is obtained by

$$D_t = L^{\top}X_tR \tag{4}$$

where D_t is the obtained lower-dimensional descriptor of the input X_t.

3. Simultaneous four-chamber volume estimation

We propose directly and simultaneously estimate four chamber volumes by multioutput sparse latent regression (MSLR), which jointly captures interdependency between the four chambers by sparse learning (Section 3.2) while handling nonlinear relationship between image appearance and four chamber volumes by kernel regression (Section 3.3).

3.1. Preliminary of multioutput regression

Direct and simultaneous four-chamber volume estimation is formulated as multioutput regression, where $\mathbf{y} = [y_1, \ldots, y_Q]^{\top} \in \mathbb{R}^Q$, where Q is the number of regression targets, i.e., the four chamber volumes. The associated image is represented by the feature descriptor $\mathbf{x} \in \mathbb{R}^d$ obtained by the SDL, where d is the dimensionality. We develop the MSLR based on the widely-used fundamental multi-task regression model

$$\mathbf{y} = W\mathbf{x} + \mathbf{b}, \tag{5}$$

where $W = [\mathbf{w}_1, \ldots, \mathbf{w}_i, \ldots, \mathbf{w}_Q]^{\top} \in \mathbb{R}^{Q\times d}$ is the regression coefficient, $\mathbf{w}_i \in \mathbb{R}^d$ is the task parameter for y_i, and $\mathbf{b} \in \mathbb{R}^Q$ is the bias.

3.2. Sparse learning of interdependency

We propose modeling interdependency of the four chambers by sparse learning. By incorporating a latent space, a structure matrix S is deployed to explicitly model interdependency via the $\ell_{2,1}$-norm based sparse learning. Due to the $\ell_{2,1}$-norm constraint, the

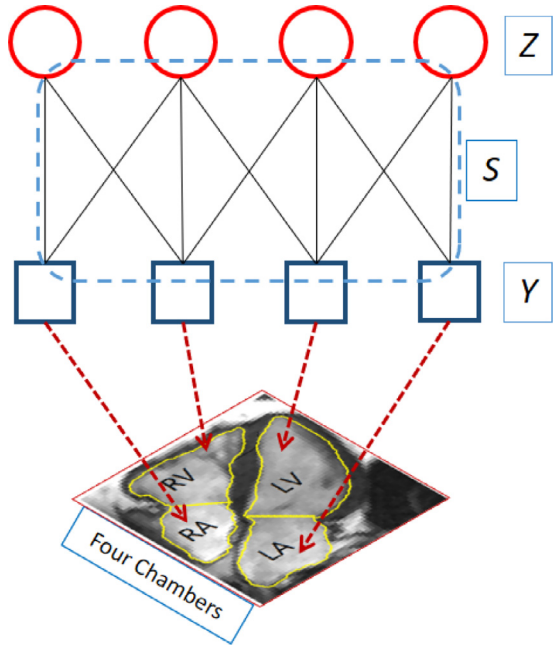

Fig. 4. The illustration of the sparse learning of the interdependency of the four chambers. Z represents the latent variables, S is structure matrix and Y denotes the outputs.

interdependency of the four chambers is encoded by sharing common variables, namely, a subset of higher-level features in the latent space, which is illustrated in Fig. 4.

Based on the least square loss function and ℓ_2 regularization, we have the following objective function w.r.t. W and S

$$\min_{W,S} \frac{1}{\mathcal{N}} ||Y - SZ||_F^2 + \lambda ||W||_F^2 + \gamma ||S^\top||_{2,1}, \quad s.t. \;\; Z = WX \tag{6}$$

where $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathcal{N}]$, $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_\mathcal{N}]$, $Z = [\mathbf{z}_1, \ldots, \mathbf{z}_i, \ldots, \mathbf{z}_\mathcal{N}] \in \mathbb{R}^{Q \times \mathcal{N}}$, $\mathbf{z}_i \in \mathbb{R}^Q$ is a variable in the latent space and $S \in \mathbb{R}^{Q \times Q}$ is named as the structure matrix; the $\ell_{2,1}$-norm constraint on the structure matrix S in (6) encourages to learn an S that is prone to column sparsity (Nie et al., 2010); the parameter γ controls the column sparsity of S and a larger γ induces higher sparsity; the bias is omitted since it is proven that the bias can be absorbed into the regression coefficient W (Nie et al., 2010). Note that unlike the work in (Nie et al., 2010) which is specifically developed for feature selection, our MSLR is a new multioutput regression model and is fundamentally different although both use an $\ell_{2,1}$ norm.

Thanks to the $\ell_{2,1}$-norm based sparse learning, regressors of four chamber volumes are encouraged to share similar parameter sparsity patterns to capture a common set of higher-level features, i.e., the latent variables in the latent space as shown in Fig. 4, which enables to encode the interdependency between the four chambers. Moreover, the performance of all regressors can be improved by leveraging knowledge across the four chambers that share subsets of common features. By deploying a structure matrix S to explicitly model interdependency of the four chambers, the MSLR can automatically learn the interdependency from data, which further improves the generality for different modalities.

We highlight that due to the incorporation of the latent space associated with the structure matrix S, the proposed MSLR brings multiple attractive merits:

- The induced latent space decouples low-level image representation and semantic four chamber volumes with W and S, which enables effectively handling high image appearance variations and huge combinatorial variability of four chamber volumes to disentangle their complex relationships.
- The structure matrix explicitly encodes the interdependency of the four chamber volumes by $\ell_{2,1}$-norm based sparse learning and is automatically inferred from data, which enables more accurate and clinically meaningful estimation.

Due to the huge appearance variations of images and the high variabilities of the four chambers, the relationship between image appearance and four chamber volumes is complicated and highly nonlinear, which cannot be handled by linear regression and demands more powerful nonlinear regressors.

3.3. Nonlinear kernel regression

The objective function in (6) admits a linear representer theorem (Kimeldorf and Wahba, 1970), which allows us to extend the model by kernelization to nonlinear multioutput regression.

Assume we have the objective function in (6) defined over a Hilbert space $\mathcal{H}$. Since S is given and fixed, the third term in (4) is constant and can be dropped. Therefore, we achieve an objective function w.r.t W

$$J_S(W) = \frac{1}{N} ||Y - SWX||_F^2 + \lambda ||W||_F^2 \tag{7}$$

The remaining two terms are quadratic, and therefore (7) is convex w.r.t. W and has a minimizer. The regularization term is strictly monotonically increasing real-valued function and the first term is a bounded loss function which is a special case of weighted least square loss and therefore is point wise (Dinuzzo and Schölkopf, 2012). According to the nonparametric Representer theorem (Schölkopf et al., 2001), (7) admits a linear representer theorem of the form

$$W = \boldsymbol{\alpha} X^\top, \tag{8}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{Q \times \mathcal{N}}$ is the coefficient matrix.

Although the objective function (6) is not jointly convex with W and S, kernelization can be derived with respect to W with a fixed S thanks to the incorporation of the latent space. This enables kernel regression to handle nonlinear relationship between image appearance and four chamber volumes, while being able to encode the interdependency of the four chambers.

The linear representer theorem (Kimeldorf and Wahba, 1970) demonstrates great advantages when $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS), which simplifies the empirical risk minimization problem from an infinite dimensional to a finite dimensional optimization problem (Kimeldorf and Wahba, 1970). Assume that we map $\mathbf{x}_i$ to $\phi(\mathbf{x}_i)$ in some RKHS of infinite dimensionality where $\phi(\cdot)$ denotes the feature map of $\mathbf{x}_i$; the mapping serves as a nonlinear feature extraction that enables to disentangle complicated relationships between image appearance and four chamber volumes. The corresponding kernel function $k(\cdot, \cdot)$ is $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$.

The objective function in (6) can be rewritten in term of traces as follows:

$$\min_{W,S} \frac{1}{\mathcal{N}} tr((Y - SWX)^\top (Y - SWX)) + \lambda tr(W^\top W) + \gamma ||S^\top||_{2,1}. \tag{9}$$

By the linear representer theorem, W admits the following form

$$W = \boldsymbol{\alpha} \Phi(X)^\top \tag{10}$$

where $\Phi(X) = [\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_i), \ldots, \phi(\mathbf{x}_\mathcal{N})]$ and $\boldsymbol{\alpha} \in \mathbb{R}^{Q \times \mathcal{N}}$.

Substituting (10) into (9) gives rise to the objective function w.r.t $\boldsymbol{\alpha}$ and S:

$$\min_{\boldsymbol{\alpha},S} \frac{1}{\mathcal{N}} tr((Y - S\boldsymbol{\alpha}K)^\top (Y - S\boldsymbol{\alpha}K)) + \lambda tr(\boldsymbol{\alpha}K\boldsymbol{\alpha}^\top) + \gamma ||S^\top||_{2,1} \tag{11}$$

where $K = \Phi(X)^\top \Phi(X)$ is the kernel matrix in the RKHS.

The latent space spanned by $Z = \boldsymbol{\alpha} K$ is obtained by the linear transformation $\boldsymbol{\alpha}$ via the Representer Theorem from the KRHS induced by a nonlinear kernel K. As a result, higher-level semantic concepts are extracted to fill the semantic gap between image representations of low-level feature descriptors and four chamber volumes, which enables efficient linear $\ell_{2,1}$-based sparse learning of S to explicitly model the interdependency of the four chambers for more accurate and clinical meaningful volume estimation. The objective in (11) can be solved for optimal solutions of $\boldsymbol{\alpha}$ and S via alternating optimization by solving one with the other fixed.

3.4. Four-chamber volume estimation

In the training stage, the MSLR is trained on labeled data with ground truth of four chamber volumes; in the prediction stage, given input images with the feature representations obtained by the SDL in Section 2, the four chamber volumes are estimated by

$$\mathbf{y}_t = S\boldsymbol{\alpha} K_t \tag{12}$$

where $K_t = \Phi(X)^\top \phi(\mathbf{x}_t)$.

4. Experiments and results

The proposed method is highly generalized and produces high performance with a correlation coefficient up to 0.921 with the ground truth obtained manually by human experts, which demonstrates its effectiveness and efficiency for direct and simultaneous four chamber volume estimation. Our method achieves a convenient clinical tool for comprehensive whole heart functional analysis.

4.1. Experimental settings

We have conducted extensive experiments on cardiac imaging data from multiple modalities including both four-chamber MR and CT images. We have also done extensive comparison to show the advantages of the proposed cardiac image representation by supervised descriptor learning (SDL) and simultaneous four chamber volume estimation by the multioutput sparse latent regression (MSLR).

4.1.1. Datasets

The cardiac MR/CT datasets contain 125 and 120 subjects, respectively, including both health and diseased cases. The subjects are collected from 3 hospitals affiliated with two health care centers (London Healthcare Science Centre and St. Joseph's HealthCare) and 2 vendors (GE and Siemens). The pathologies are diverse including regional wall motion abnormalities, myocardial hypertrophy, mildly dilated RV, atrial septal defect, LV dysfunction, mildly enlarged LV, and decreased ejection fraction (LVEF < 0.5 and RVEF < 0.4).

4.1.2. Implementation details

The performance of the proposed method is quantitatively evaluated by comparing with ground truth by manually obtained by human experts. The Pearson's r correlation coefficient between the ground truth and the estimation is used as measurement to evaluate estimation performance as in (Zhen et al., 2014b; Wang et al., 2013), and higher correlation coefficient indicates better performance. All the volumes in a cardiac cycle are used to calculate the correlation coefficient. The leave-one-subject-out cross validation is used for the evaluation.

We follow the clinical routine which practically uses 2D image slices for diagnosis, and estimate cavity areas of the four chambers in 2D MR/CT slices, and volumes are computed by integrating cavity areas in the sagittal direction. Note that we use the normalized areas as the targets, *i.e.*, the number of pixels in a chamber divided by the total number of pixels of the images. A region of interest (ROI) is placed to enclose the four chambers in an MR image according to the method in (Wang et al., 2014a). We use a three-level pyramid HOG (PHOG) obtaining a matrix of size 84×31 from an image of 64×64 pixels. We set the tradeoff parameter $\beta = 1$ to keep both the reconstruction fidelity and discriminative ability.

To show the advantage of the SDL, we have also compared with popular descriptors, *e.g.*, GIST and histogram of LBP both of which are implemented with a similar spatial pyramid to the PHOG descriptor, and dimensionality reduction methods, *e.g.*, generalized principal component analysis (GPCA) (Ye et al., 2004) and principal component analysis (PCA). We have also compared to deep learning based feature representation. Specifically, we extract features based on pre-trained model in (Simonyan and Zisserman, 2015) using the MatConvNet toolbox (Vedaldi and Lenc, 2015) and we use the first fully connected layer of 4096 dimensions as the feature representation.

To show the advantage of the MSLR, we have compared with state-of-the-art multioutput regression models including the baseline multioutput kernel ridge regression (mKRR) (Hastie et al., 2001), multi-dimensional support vector regressor (mSVR) (Sanchez-Fernandez et al., 2004), adaptive k-cluster random forests (AKRF) and kernel partial least square (KPLS) (Rosipal and Trejo, 2001). We follow the experimental settings of AKRF in the original work (Hara and Chellappa, 2014) and keep the same for all methods to establish fair comparison.

To show the advantages of our direct and simultaneous four chamber volume estimation, we have compared with two representative segmentation methods, i.e., the shape regression machine (SRM) (Zhou, 2010) and regression segmentation (RS) (Wang et al., 2015). The two methods were originally developed for one single object segmentation, and we extend them for joint four chamber segmentation.

4.2. Results

Our method for the first time achieves simultaneous four-chamber volume estimation and produces high estimation accuracy for all the four chambers despite of the great challenge of the four chambers, and substantially outperforms other state-of-the-art algorithms including both feature learning and multioutput regression.

4.2.1. Estimation effectiveness

The proposed method achieves consistently high performance for the four chambers on both the MR and CT datasets, especially for the LV with a high correlation coefficient of 0.921 for MR and 0.913 for CT, respectively, as illustrated in Figs. 5 and 6. Compared to the results in (Zhen et al., 2015a), outlier points have been largely reduced as can be observed in Fig. 5 due to the performance improvement by the new multioutput sparse latent regression model. Although the boundary between ventricle and atrium is mostly invisible and not supported by edge and region homogeneity, especially on CT images, our method can successfully predict four-chamber volumes due to the use of multioutput latent regression. Moreover, LA and RA volumes which were not measured previously due to their complex anatomical geometry are predicted by our method with high accuracy. The consistently high performance on all the four chambers of both MR and CT demonstrate the power of the adapted regression forests in direct and simultaneous four-chamber volume estimation. Moreover, the estimation can be conducted very efficient with about 20 s for a new subject.

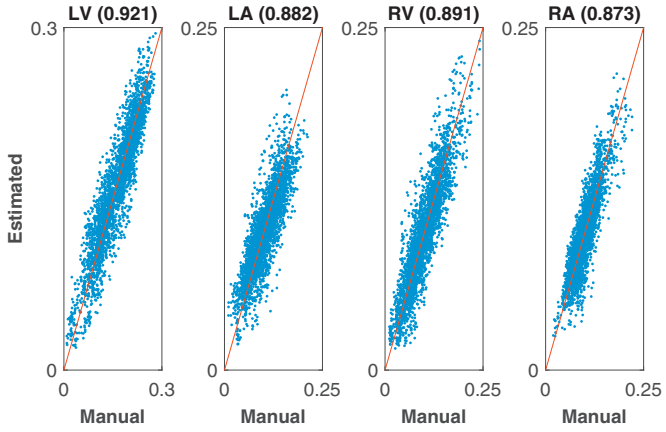

Fig. 5. The correlation coefficients between estimated and manually obtained volumes for the four chambers on the MR dataset.

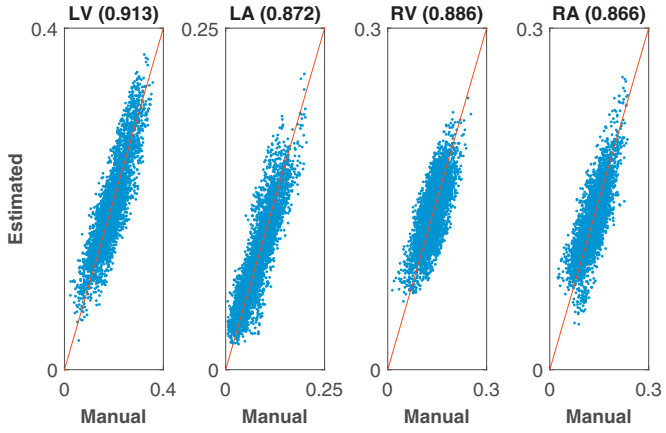

Fig. 6. The correlation coefficients between estimated and manually obtained volumes for the four chambers on the CT dataset.

The results are significant showing potential use in clinical practice (Zhen et al., 2014a; 2014b). The volume estimation errors are illustrated in Figs. 7 and 8 for MR and CT, respectively.

- Fig. 7 shows comparison between estimated volumes and ground-truth counterparts manually obtained by human experts across a cardiac cycle with respect to all the subjects on the MR dataset. Our method can produce very close estimations with low errors: 0.011 ± 0.010 (LV), 0.016 ± 0.014 (LA), 0.015 ± 0.011 (RV) and 0.018 ± 0.014 (RA) for all the four chambers averaged over all subjects. Since all the four-chamber volumes across a cardiac cycle are accurately estimated simultaneously, cardiac functions can be comprehensively and efficiently assessed in clinical diagnosis. More importantly, the change patterns of volumes in a cardiac cycle have been successfully captured: the LV and RV share the same volume change pattern from the diastole to the systole; the LA and RA share the same pattern from the systole to the diastole. The assessment of cardiac functions, e.g., ejection fraction (EF), can conduced in a more comprehensive and efficient way.
- Fig. 8 shows comparison between estimated volumes and ground-truth counterparts for every single subject on the CT dataset. Our method can produce accurate volumes close to those obtained by human experts, even though the RV and RA have very poor tissue intensity contrast without visible boundary between them. The average estimation errors are 0.012 ± 0.011 (LV), 0.016 ± 0.014 (LA), 0.015 (RV) ± 0.014 and 0.02 ± 0,017 (RA). Moreover, our method captures the huge inter-subject variations by producing very close estimation to ground truth for most of the subjects. This shows the ability of our method to accurately estimate highly varied volumes from healthy to diseased subjects, which indicates its practical use in disease cases for clinical diagnosis.

In general, the high estimation accuracy with low errors on both MR and CT demonstrates the great generality of the proposed method for cardiac four-chamber volume estimation from multi-modal imaging data, indicating its practical use in clinical routine.

4.2.2. Comparison

The advantage of the proposed method for cardiac four-chamber image representation is demonstrated by better performance than other methods on both MR and CT datasets.

- The SDL substantially outperforms both state-of-the-art descriptors, *e.g.*, PHOG, LBP and GIST as shown in Table 1, and dimensionality reduction techniques, *e.g.*, GPCA and PCA as shown in Table 2, by up to 7.2% on MR and 8.6% on CT showing the effectiveness of the SDL for feature learning in multi-output regression. The CNN based feature representations produce impressive performance, which however is inferior to the proposed SDL. This is largely due to its nature of unsupervised learning without considering the regression targets. Fine-tuning by taking into account regression targets would potentially further improve the performance. The larger advantages of our method on CT than MR over other descriptors demonstrate the effectiveness of the SDL in for effective feature learning, even from CT images with low tissue contrast. Due to the compactness (low dimensionality with $m = 40$ and $n = 20$) of the feature representation learned by the SDL, the estimation is enabled to be more efficient.
- The proposed MSLR consistently outperforms other multioutput regression models including the baseline mKRR, mSVR, KPLS and AKRF (Zhen et al., 2015a) by up to 6% as shown in Table 3. To achieve fair comparison, all those regression models use the feature descriptors learned by the SDL. The large improvement over representative multioutput regression models, especially on the more challenging CT dataset shows the effectiveness of the proposed MSLR in explicitly modeling interdependency of the four chambers for more accurate direct and simultaneous four chamber volume estimation.
- Our method outperforms two representative segmentation methods, i.e., the SRM and the RS, for both separate and joint segmentation of the four chambers as shown in Table 4. Although separate segmentation can produce better results than joint segmentation, separate segmentation is still worse the our direct and simultaneous estimation. It is also worth mentioning that the segmentation is conduced based on localization obtained from manual segmentation. This result also demonstrates the advantages of direct estimation over segmentation based methods in leveraging the relationship between the four chambers.

4.2.3. Convergence analysis

We provide experimental analysis on the convergence of the alternating optimization algorithm on both the MT and CT datasets in Fig. 9. As we can see from the figure, the alternating optimization algorithm consistently converge on the two datasets. Both the objective function value and average mean square error (aMSE), i.e., average estimation error of the four chamber volumes, decrease monotonically with the iterations. On both datasets, our algorithm converges within 20 iterations, which guarantees its efficiency.

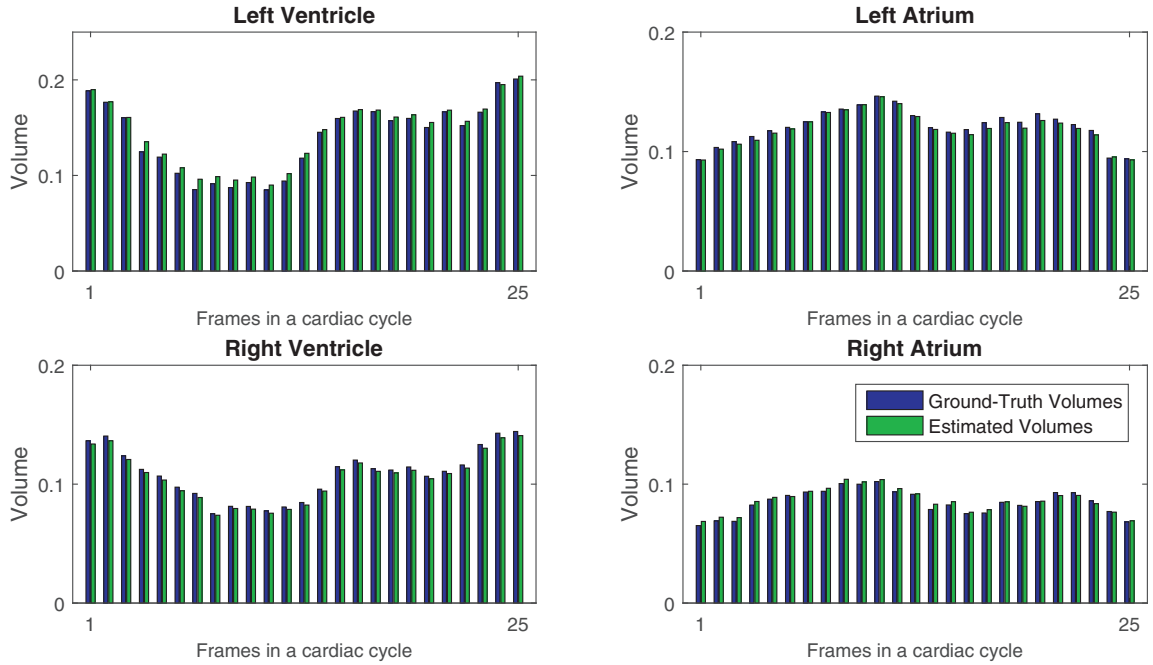

Fig. 7. The illustration of ground truth against estimation by the proposed method 25 frames in cardiac cycle averaged over all the subjects on the **MR** dataset.

Table 1
The comparison with other descriptors (correlation coefficient). *#d* indicates the dimensionality of descriptors. Bold means the best results.

	MR				CT			
Methods	LV	LA	RV	RA	LV	LA	RV	RA
MSLR (800d)	**0.921**	**0.882**	**0.891**	**0.873**	**0.913**	**0.872**	**0.886**	**0.866**
PHOG (2604d)	0.875	0.827	0.844	0.824	0.855	0.825	0.829	0.805
LBP (4872d)	0.877	0.807	0.838	0.806	0.831	0.799	0.816	0.797
GIST (4096d)	0.876	0.831	0.827	0.854	0.828	0.795	0.823	0.804
CNN (4096d)	0.884	0.839	0.842	0.831	0.866	0.814	0.833	0.821

Table 2
The comparison with other dimensionality reduction methods (correlation coefficient). *#d* indicates the dimensionality of descriptors. Bold means the best results.

	MR				CT			
Methods	LV	LA	RV	RA	LV	LA	RV	RA
MSLR (800d)	**0.921**	**0.882**	**0.891**	**0.873**	**0.913**	**0.872**	**0.886**	**0.866**
GPCA (800d)	0.899	0.848	0.855	0.836	0.878	0.841	0.845	0.825
PCA (800d)	0.884	0.822	0.836	0.821	0.865	0.820	0.823	0.805

Table 3
The comparison with other multioutput regression models (correlation coefficient). Bold means the best results.

	MR				CT			
Methods	LV	LA	RV	RA	LV	LA	RV	RA
MSLR	**0.921**	**0.882**	**0.891**	**0.873**	**0.913**	**0.872**	**0.886**	**0.866**
mKRR (Hastie et al., 2001)	0.892	0.861	0.863	0.852	0.887	0.850	0.856	0.833
mSVR (Sanchez-Fernandez et al., 2004)	0.901	0.862	0.865	0.855	0.889	0.851	0.857	0.836
KPLS (Rosipal and Trejo, 2001)	0.884	0.858	0.853	0.836	0.878	0.838	0.844	0.819
AKRF (Zhen et al., 2015a)	0.915	0.871	0.882	0.862	0.904	0.869	0.875	0.851

5. Discussion

We have successfully addressed the two key problems, i.e., cardiac image representation and simultaneous four chamber volume estimation. Our method has shown great effectiveness on data of multiple modalities including both MR and CT. The success of the proposed method lies in the advantages of supervised descriptor learning (SDL) for cardiac image representation and the newly proposed multioutput sparse latent regression (MSLR) model for simultaneous volume estimation.

5.1. The advantage of SDL for cardiac four chamber image representation

The supervised descriptor learning (SDL) algorithm shows great effectiveness to generate compact and discriminative cardiac im-

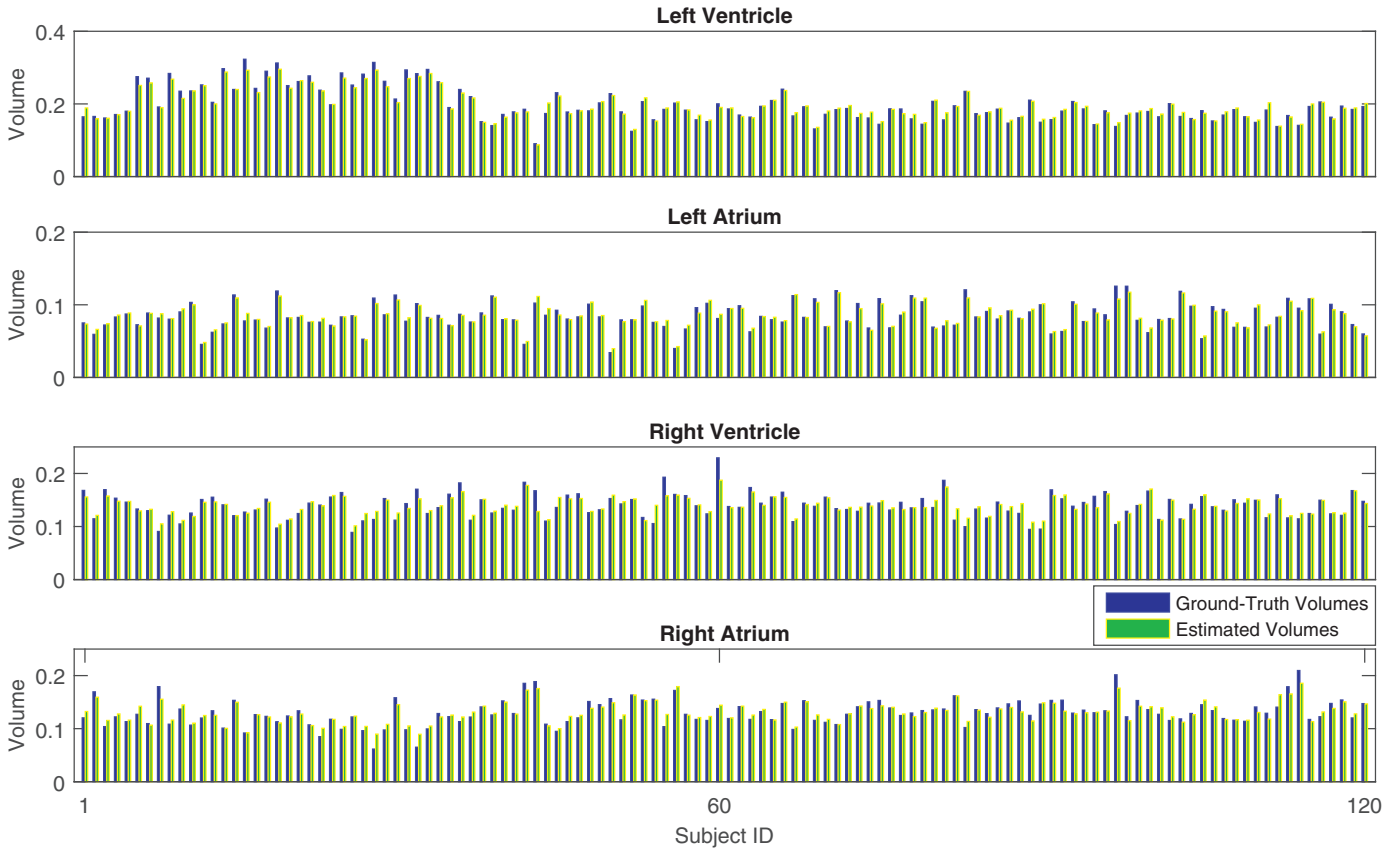

Fig. 8. The illustration of ground truth against estimation by the proposed method over 120 subjects on the **CT** dataset. Our method can accurately predict four-chamber volumes with low errors, even with huge inter-subject variations.

Table 4
The comparison with segmentation methods (correlation coefficient). Bold means the best results.

	MR				CT			
Methods	LV	LA	RV	RA	LV	LA	RV	RA
MSLR	**0.921**	**0.882**	**0.891**	**0.873**	**0.913**	**0.872**	**0.886**	**0.866**
SRM (S) (Zhou, 2010)	0.877	0.855	0.857	0.856	0.844	0.837	0.836	0.839
RS (S) (Wang et al., 2015)	0.881	0.863	0.868	0.862	0.857	0.851	0.852	0.855
SRM (J) (Zhou, 2010)	0.844	0.808	0.819	0.756	0.797	0.752	0.760	0.731
RS (J) (Wang et al., 2015)	0.851	0.816	0.829	0.772	0.809	0.764	0.768	0.749

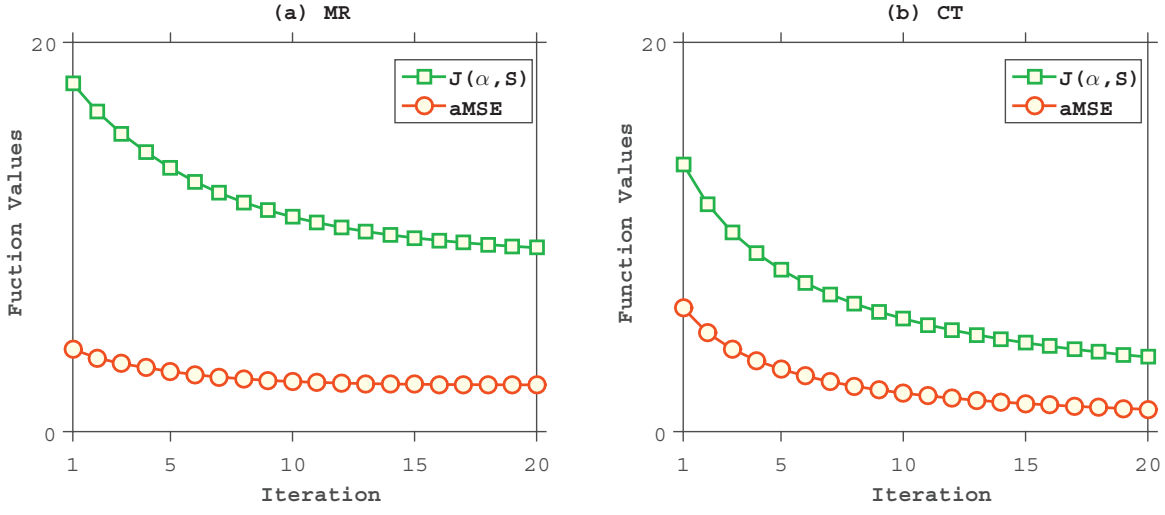

Fig. 9. The convergence of the alternating optimization algorithm on the MR (a) and CT (b) datasets.

age representation, which has been validated by comparing with other feature learning algorithms including representative dimensionality reduction algorithms: PCA and GPCA/2DPCA. By leveraging the strength of generalized low-rank approximation of matrices, the SDL takes 2D matrices as inputs, which allows to explore the distinctive physical meanings of rows and columns in the matrix; it can achieve more robust parameter estimation by largely reducing the number of parameters to learn due the use of 2D inputs. More importantly, the SDL incorporates the target information, i.e., the four chamber volumes, into the feature learning by a supervised manifold regularization, which extracts discriminative features closely related to four chamber volumes while removing redundant and irrelevant information, leading to more accurate and efficient estimation.

5.2. The advantage of MSLR for direct and simultaneous four chamber volume estimation

The effectiveness of the proposed multioutput sparse latent regression (MSLR) has been validated by the better performance than representative multioutput regression models including multi-dimensional support vector regressor (mSVR), adaptive k-cluster random forests (AKRF) and kernel partial least square (KPLS), which do not provide explicit and effective way to model interdependency of the four chambers. The advantage of the MSLR stems from that it successfully handles the high complexity of the four chambers by jointly disentangling the highly complex relationship between image representations and four-chamber volumes while capturing the interdependency between the four chambers. Our MSLR provides a general framework for direct and simultaneous four chamber volume estimation, which enables more accurate and clinically meaningful volume estimation of the four chambers by leveraging their interdependency.

5.3. Segmentation vs. direct estimation for cardiac four chamber volume estimation

Simultaneous four chamber volume estimation is meaningful and useful in terms of both methodology and clinical practice, while it remains a challenging task for existing segmentation based methods (Toth et al., 2013). Segmentation has also made great progress and becomes more reliable, accurate and less time-consuming. However, existing segmentation methods has been mainly focused on one single chamber/ventricle, mostly on left ventricle (LV), recently on right ventricle (RV) and left atrium (LA), and very few on right atrium (RA). Cardiac four chambers are anatomically connected and interdependent, and their volumes are statistically correlated. Segmenting each of the four chambers separately would not be able to capture their relationships, and tends to be extremely time-consuming by quadrupling the task. Moreover, due to their spatial interdependency and temporal deformation, it tends to be very challenging to automatically localize each of the four chambers individually, which however is a prerequisite to achieve accurate an efficient segmentation.

The significant advantages of direct estimation by multioutput regression come from its innate capability of simultaneous estimation of four chamber volumes while naturally capturing their correlations to achieve accurate estimation. Four chamber volumes could also reciprocally validate each other to achieve reliable and clinically more meaningful estimation. The performance of direct estimation by regression on bi-ventricle and the four chambers is encouraging and indicates its great potential to directly estimate anatomical structures, e.g., landmarks, contours and wall thickness, etc. and clinical measurements, e.g., ejection fraction (EF) and ventricular mass, etc. These quantities provide complementary measurement to volumes for more comprehensive assessment, which makes direct estimation more reliable and transparent to radiologists. Moreover, simultaneous four chamber volume estimation does not require the localization of each of the four chambers, and only needs a rough region of interest (ROI) enclosing all the four chambers.

6. Conclusion

In this paper, we have presented a new method for direct and simultaneous cardiac four-chamber volume estimation. The proposed method for the first time achieves simultaneous cardiac four-chamber volume estimation. By simultaneously estimating four chamber volumes using multioutput regression, our method not only enables a more efficient, accurate and comprehensive cardiac functional assessment, but also allow to cross-validate each individual of the four chambers for accurate clinical diagnosis. The proposed direct estimation framework is highly generalized independent of imaging modalities, and can not only be used for volume estimation of organs, but also be widely applied to other clinical data prediction, e.g., pattern prediction from fMRI (Chu et al., 2011) and cardiac motion prediction from single-phase CTA (Metz et al., 2012). We have conducted extensive experiments on cardiac four chamber datasets with both MR and CT modalities, and the proposed direct method achieves high estimation accuracy and performs better than representative segmentation methods.

Acknowledgment

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. Thanks to the partial support from NSFC (Grant No. 61571147).

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: Ayache Delingette, N.H, Golland, P., Mori, K. (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, Heidelberg, pp. 535–543.

Baur, L., 2008. Right atrial function: still underestimated in clinical cardiology. Int. J. Cardiovasc. Imaging (formerly Cardiac Imaging) 24 (7), 711–712.

Belkin, M., Niyogi, P., Sindhwani, V., 2006. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J. Mach. Learn. Res. 7, 2399–2434.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Chu, C., Ni, Y., Tan, G., Saunders, C.J., Ashburner, J., 2011. Kernel regression for fmri pattern prediction. NeuroImage 56 (2), 662–673.

Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K., 2013. Regression forests for efficient anatomy detection and localization in computed tomography scans. Med. Image Anal. 17 (8), 1293–1303.

Criminisi, A., Shotton, J., 2013. Decision Forests for Computer Vision and Medical Image Analysis. Springer Publishing Company, Incorporated.

Criminisi, A., Shotton, J., Bucciarelli, S., 2009. Decision Forests with Long-range Spatial Context for Organ Localization in Ct Volumes. Medical Image Computing and Computer-Assisted Intervention (MICCAI) Workshop on Probabilistic Models for Medical Image Analysis.

Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E., 2010. Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: International MICCAI Workshop on Medical Computer Vision, pp. 106–117.

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. In: IEEE conference on computer vision and pattern recognition. Vol. 1, pp. 886–893.

Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S., 2010. Population shape regression from random design data. Int. J. Comput. Vis. 90 (2), 255–266.

Dinuzzo, F., Schölkopf, B., 2012. The Representer Theorem for Hilbert Spaces: A Necessary and Sufficient Condition. In: Advances in neural information processing systems, pp. 189–196.

Donnell, T., Funka-Lea, G., Tek, H., Jolly, M.-P., Rasch, M., Setser, R., 2006. Comprehensive cardiovascular image analysis using mr and ct at siemens corporate research. Int. J. Comput. Vis. 70 (2), 165–178.

Ecabert, O., Peters, J., Schramm, H., Lorenz, C., Von Berg, J., Walker, M.J., Vembar, M., Olszewski, M.E., Subramanyan, K., Lavi, G., et al., 2008. Automatic model-based segmentation of the heart in ct images. IEEE Trans. Med. Imaging 27 (9), 1189–1201.

Fletcher, P.T., 2013. Geodesic regression and the theory of least squares on riemannian manifolds. Int. J. Comput. Vis. 105 (2), 171–185.

Fonseca, C.G., Backhaus, M., Bluemke, D.A., Britten, R.D., Do Chung, J., Cowan, B.R., Dinov, I.D., Finn, J.P., Hunter, P.J., Kadish, A.H., et al., 2011. The cardiac atlas projectÃ¬an imaging database for computational modeling and statistical atlases of the heart. Bioinformatics 27 (16), 2288–2295.

Glatard, T., Lartizien, C., Gibaud, B., Ferreira da Silva, R., Forestier, G., Cervenansky, F., Alessandrini, M., Benoit-Cattin, H., Bernard, O., Camarasu-Pop, S., et al., 2013. A virtual imaging platform for multi-modality medical image simulation. IEEE Trans. Med. Imaging 32 (1), 110–118.

Guzman-Rivera, A., Kohli, P., Glocker, B., Shotton, J., Sharp, T., Fitzgibbon, A., Izadi, S., 2014. Multi-Output Learning for Camera Relocalization. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1114–1121.

Hara, K., Chellappa, R., 2014. Growing Regression Forests by Classification: Applications to Object Pose Estimation. In: European Conference on Computer Vision, pp. 552–567.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. 2001. NY Springer.

He, X., Niyogi, P., 2004. Locality Preserving Projections. In: Advances in neural information processing systems. Vol. 16, p. 153.

Hussain, M.A., Hamarneh, G., O'Connell, T.W., Mohammed, M.F., Abugharbieh, R., 2016. Segmentation-free estimation of kidney volumes in ct with dual regression forests. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 156–163.

Jiang, M., Lv, J., Wang, C., Huang, W., Xia, L., Shou, G., 2011. A hybrid model of maximum margin clustering method and support vector regression for solving the inverse ECG problem. In: Computing in Cardiology, 2011. IEEE, pp. 457–460.

Jolly, M.-P., 2006. Automatic segmentation of the left ventricle in cardiac mr and ct images. Int. J. Comput. Vis. 70 (2), 151–163.

Joshi, S., Karthikeyan, S., Manjunath, B., Grafton, S., Kiehl, K., et al., 2010. Anatomical parts-based regression using non-negative matrix factorization. In: Computer Vision and Pattern Recognition, 2010 IEEE Conference on. IEEE, pp. 2863–2870.

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. Springer, pp. 678–686.

Kainz, P., Urschler, M., Schulter, S., Wohlhart, P., Lepetit, V., 2015. You Should Use Regression to Detect Cells. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, pp. 276–283.

Kimeldorf, G.S., Wahba, G., 1970. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Stat. 41 (2), 495–502.

Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S., 2016. Recognizing end-diastole and end-systole frames via deep temporal regression network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 264–272.

Marchesseau, S., Delingette, H., Sermesant, M., Cabrera-Lozoya, R., Tobon-Gomez, C., Moireau, P., Figueras i Ventura, R., Lekadir, K., Hernandez, A., Garreau, M., et al., 2013. Personalization of a cardiac electromechanical model using reduced order unscented kalman filtering from regional volumes. Med. Image Anal. 17 (7), 816–829.

Metz, C.T., Baka, N., Kirisli, H., Schaap, M., Klein, S., Neefjes, L.A., Mollet, N.R., Lelieveldt, B., de Bruijne, M., Niessen, W.J., et al., 2012. Regression-based cardiac motion prediction from single-phase cta. IEEE Trans. Med. Imaging 31 (6), 1311–1325.

Mukhopadhyay, A., Oksuz, I., Bevilacqua, M., Dharmakumar, R., Tsaftaris, S.A., 2015. Data-driven feature learning for myocardial segmentation of cp-bold mri. In: International Conference on Functional Imaging and Modeling of the Heart. Springer, pp. 189–197.

Nambakhsh, C.M., Peters, T.M., Islam, A., Ayed, I.B., 2013. Right ventricle segmentation with probability product kernel constraints. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N.E. (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Heidelberg, pp. 509–517.

Nie, F., Huang, H., Cai, X., Ding, C.H., 2010. Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In: Advances in neural information processing systems, pp. 1813–1821.

Nikolaou, K., Alkadhi, H., Bamberg, F., Leschka, S., Wintersperger, B.J., 2011. Mri and ct in the diagnosis of coronary artery disease: indications and applications. Insights Imaging 2 (1), 9–24.

Pace, D.F., Dalca, A.V., Geva, T., Powell, A.J., Moghari, M.H., Golland, P., 2015. Interactive Whole-heart Segmentation in Congenital Heart Disease. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, pp. 80–88.

Petitjean, C., Dacher, J.-N., 2011. A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. 15 (2), 169–184.

Petitjean, C., Zuluaga, M.A., Bai, W., Dacher, J.-N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I.B., Cardoso, M.J., Chen, H.-C., et al., 2015. Right ventricle segmentation from cardiac mri: a collation study. Med. Image Anal. 19 (1), 187–202.

Prakosa, A., Sermesant, M., Allain, P., Villain, N., Rinaldi, C.A., Rhode, K., Razavi, R., Delingette, H., Ayache, N., 2014. Cardiac electrophysiological activation pattern estimation from images using a patient-specific database of synthetic image sequences. IEEE Trans. Biomed. Eng. 61 (2), 235–245.

Punithakumar, K., Ben Ayed, I., Islam, A., Goela, A., Ross, I.G., Chong, J., Li, S., 2013. Regional heart motion abnormality detection: an information theoretic approach. Med. Image Anal. 17 (3), 311–324.

Rosipal, R., Trejo, L.J., 2001. Kernel partial least squares regression in reproducing kernel hilbert space. J. Mach. Learn. Res. 2 (Dec), 97–123.

Sanchez-Fernandez, M., de Prado-Cumplido, M., Arenas-García, J., Pérez-Cruz, F., 2004. Svm multiregression for nonlinear channel estimation in multiple-input multiple-output systems. TSP 52 (8), 2298–2307.

Schölkopf, B., Herbrich, R., Smola, A.J., 2001. A Generalized Representer Theorem. In: Computational learning theory, pp. 416–426.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-scale Image Recognition. In: International Conference on Learning Representations.

Suinesiaputra, A., Cowan, B.R., Al-Agamy, A.O., Elattar, M.A., Ayache, N., Fahmy, A.S., Khalifa, A.M., Medrano-Gracia, P., Jolly, M.-P., Kadish, A.H., et al., 2014. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images. Med. Image Anal. 18 (1), 50–62.

Tobon-Gomez, C., Geers, A.J., Peters, J., Weese, J., Pinto, K., Karim, R., Ammar, M., Daoudi, A., Margeta, J., Sandoval, Z., et al., 2015. Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets. IEEE Trans. Med. Imaging 34 (7), 1460–1473.

Torki, M., Elgammal, A., 2011. Regression from Local Features for Viewpoint and Pose Estimation. In: IEEE International Conference on Computer Vision, pp. 2603–2610.

Toshev, A., Szegedy, C., 2014. Deeppose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1653–1660.

Toth, R., Ribault, J., Gentile, J., Sperling, D., Madabhushi, A., 2013. Simultaneous segmentation of prostatic zones using active appearance models with multiple coupled levelsets. Comput. Vision Image Understand. 117 (9), 1051–1060.

Vedaldi, A., Lenc, K., 2015. Matconvnet – Convolutional Neural Networks for Matlab. In: Proceeding of the ACM Int. Conf. on Multimedia.

Wang, S., Summers, R.M., 2012. Machine learning and radiology. Med. Image Anal. 16 (5), 933–951.

Wang, V.Y., Lam, H., Ennis, D.B., Cowan, B.R., Young, A.A., Nash, M.P., 2009. Modelling passive diastolic mechanics with quantitative mri of cardiac structure and function. Med. Image Anal. 13 (5), 773–784.

Wang, Y., Fan, Y., Bhatt, P., Davatzikos, C., 2010. High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. Neuroimage 50 (4), 1519–1535.

Wang, Z., Ben Salah, M., Gu, B., Islam, A., Goela, A., Li, S., 2014a. Direct estimation of cardiac bi-ventricular volumes with an adapted bayesian formulation. IEEE TBME 1251–1260.

Wang, Z., Ben Salah, M., Gu, B., Islam, A., Goela, A., Li, S., 2014b. Direct estimation of cardiac biventricular volumes with an adapted bayesian formulation. Biomed. Eng., IEEE Trans. 61 (4), 1251–1260.

Wang, Z., Salah, M., Ayed, I., Islam, A., Goela, A., Li, S., 2013. Bi-ventricular Volume Estimation for Cardiac Functional Assessment. RSNA.

Wang, Z., Zhen, X., Tay, K., Osman, S., Romano, W., Li, S., 2015. Regression segmentation for M^3 spinal images. IEEE Trans. Med. Imaging 34 (8), 1640–1648.

Wintersperger, B.J., 2009. Complementary Role of Cardiac CT and MRI. In: Multislice CT. Springer, pp. 269–284.

Ye, J., 2005. Generalized low rank approximations of matrices. Mach. Learn. 61 (1–3), 167–191.

Ye, J., Janardan, R., Li, Q., 2004. Gpca: an efficient dimension reduction scheme for image compression and retrieval. In: ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 354–363.

Yu, M., Shao, L., Zhen, X., He, X., 2016. Local feature discriminant projection.. IEEE Trans. Pattern Anal. Mach. Intell. 38 (9), 1908–1914.

Zettinig, O., Mansi, T., Neumann, D., Georgescu, B., Rapaka, S., Seegerer, P., Kayvanpour, E., Sedaghat-Hamedani, F., Amr, A., Haas, J., et al., 2014. Data-driven estimation of cardiac electrical diffusivity from 12-lead ecg signals. Med. Image Anal. 18 (8), 1361–1376.

Zhang, Z., Zhao, K., 2013. Low-rank matrix approximation with manifold regularization. IEEE Transaction on Pattern Analysis and Machine Intelligence 35 (7), 1717–1729.

Zhen, X., Islam, A., Chan, I., Li, S., 2015a. Direct and Simultaneous Four Chamber Volume Estimation by Multi-output Regression. Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Zhen, X., Shao, L., 2013. A local descriptor based on laplacian pyramid coding for action recognition. Pattern Recognit. Lett. 34 (15), 1899–1905.

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

Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014a. A Comparative Study of Methods for Cardiac Ventricular Volume Estimation. Annual Meeting - Radiological Society of North America (RSNA).

Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014b. Direct Estimation of Cardiac Bi-ventricular Volumes with Regression Forests. Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Zhen, X., Wang, Z., Yu, M., Li, S., 2015b. Supervised Descriptor Learning for Multi--output Regression. In: IEEE Conference on Computer Vision and Pattern Recognition.

Zhen, X., Yin, Y., Bhaduri, M., Nachum, I.B., Laidley, D., Li, S., 2016a. Multi-task shape regression for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 210–218.

Zhen, X., Yu, M., Islam, A., Bhaduri, M., Chan, I., Li, S., 2016b. Descriptor Learning via Supervised Manifold Regularization for Multioutput Regression doi:10.1109/TNNLS.2016.2573260.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2008. Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging 27 (11), 1668–1681.

Zhou, S.K., 2010. Shape regression machine and efficient segmentation of left ventricle endocardium from 2d b-mode echocardiogram. Med. Image Anal. 14 (4), 563–581.

Zhou, S.K., 2014. Discriminative anatomy detection: classification vs regression. Pattern Recognit. Lett. 43, 25–38.

Zhou, S.K., Zhou, J., Comaniciu, D., 2007. A boosting regression approach to medical anatomy detection. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8.