Multi-Scale Deep Networks and Regression Forests for Direct Bi-ventricular Volume Estimation

Xiantong Zhen[a], Zhijie Wang[b], Ali Islam[c], Mousumi Bhaduri[d], Ian Chan[d], and Shuo Li[a,b*]

[a]*Department of Medical Biophysics, The University of Western Ontario, London, ON, Canada.*
[b]*GE Healthcare, London, ON, Canada.*
[c]*St. Joseph's Health Care, London, ON, Canada.*
[d]*London Healthcare Sciences Centre, London, ON, Canada.*

Abstract

Direct estimation of cardiac ventricular volumes has become increasingly popular and important in cardiac function analysis due to its effectiveness and efficiency by avoiding an intermediate segmentation step. However, existing methods rely on either intensive user inputs or problematic assumptions. To realize the full capacities of direct estimation, this paper presents a general, fully learning-based framework for direct bi-ventricular volume estimation, which removes user inputs and unreliable assumptions. We formulate bi-ventricular volume estimation as a general regression framework which consists of two main full learning stages: unsupervised cardiac image representation learning by multi-scale deep networks and direct bi-ventricular volume estimation by random forests.

By leveraging strengths of generative and discriminant learning, the proposed method produces high correlations of around 0.92 with ground truth by human experts for both the left and right ventricles using a leave-one-subject-out cross validation, and largely outperforms existing direct methods on a larger dataset of 100 subjects including both healthy and diseased cases with twice the number of subjects used in previous methods. More importantly, the proposed method can not only be practically used in clinical cardiac function analysis but also be easily extended to other organ volume estimation tasks.

Keywords: Direct volume estimation, multi-scale deep networks, random forests, regression.

1. Introduction

Accurate and automatic assessment of cardiac functions plays an increasingly important role in diagnosis and prognosis of heart diseases, one of the leading causes of death (Wang and Amini, 2012). 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). Conventional volume estimation methods usually rely on an intermediate step of segmentation. However, segmentation itself is an extremely challenging problem which in clinical practise physicians are not interested in. Direct estimation methods which remove the segmentation step become attractive in cardiac function diagnosis and ventricular estimation due to its efficiency and clinical significance (Afshin et al., 2012a,b; Wang et al., 2014; Afshin et al., 2014; Zettinig et al., 2014; Zhen et al., 2014d; Wang et al., 2013; Zhen et al., 2015a). A comprehensive study of methods for cardiac ventricular volume estimation has been conducted in (Zhen et al., 2014c) showing that direct estimation methods provide more accurate estimation than segmentation-based methods for both LV and RV volumes. More importantly, direct estimation allows us to leverage both the state-of-the-art machine learning techniques and increasingly large amount of labeled and unlabeled imaging data. In many applications, the performance of machine learning-based automatic detection and diagnosis systems has shown to be comparable to that of a well-trained and experienced radiologist (Wang and Summers, 2012). Direct methods also enable us to richly explore data statistics which therefore helps provide more meaningful and clinically significant volume estimation.

However, existing direct methods suffer from many drawbacks, *e.g.*, dependence on user inputs and unreliable assumptions, which severely restrict their applications. In this paper, towards the full capacity of direct estimation, we propose a general framework for bi-ventricular volume estimation. A preliminary conference version of this work appeared in MICCAI 2014 (Zhen et al., 2014d). In this journal version we extend in two main aspects: 1) we replace the handcrafted features in (Zhen et al., 2014d) by feature engineering (Bengio et al., 2013) with representation learned by our multi-scale deep networks. In contrast to feature engineering, the unsupervised learning by deep networks allows us to both make full use of plenty of unlabeled data which is largely available in medical imaging and faithfully detect data-driven features for informative image representations in specific tasks; 2) we provide a wider investigation on a larger dataset of 100 subjects with 6000 images than previously used. Such as a large dataset encompasses the huge inter-subject variabilities and therefore provides a more comprehensive and clinically meaningful validation.

*Corresponding author (shuo.li@ge.com).

1.1. Direct methods

Direct diagnosis has proven to be effective and efficient in assessment of cardiac function abnormality automatically without intermediate steps. Afshin et al. (Afshin et al., 2014) present a direct method for regional assessment of the left ventricular (LV) myocardial function via classification. Without delineating regional boundaries of the LV, they obtain an abnormality assessment of each standard cardiac segment in real-time with more accurate results than segmentation-based methods. A set of statistical MRI features, i.e., the Bhattacharyya coefficients (Afshin et al., 2014) based on a measure of similarity between image distributions are built for all the regional segments and all subsequent frames. The statistical features are related to the proportion of blood within each segment and can characterize segmental contraction. Unfortunately, the method requires user inputs for one frame of each subject and is restricted to the LV, which limits its further application in clinical use.

In addition to MR imaging, direct methods have also been developed in other modalities such as electrocardiograms (ECG) for diagnosis of cardiac function abnormalities. Recently, 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 propose to 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.

Direct estimation of cardiac volumes has started to generate increasing interest due to the avoidance of intermediate segmentation (Afshin et al., 2012a; Wang et al., 2014; Zhen et al., 2014d). To directly estimate the ejection fraction of the LV, global image statistics similar to (Afshin et al., 2012b, 2014) are used to calculate LV volumes in (Afshin et al., 2012a). A key limitation is that intensive user inputs including two boxes, *i.e.*, one inside the LV cavity and one enclosing the cavity, are required. Moreover, the method is restricted to the LV due to the strong assumption of correlation between the considered statistics and LV cavity areas and therefore can not be generalized to the right ventricle (RV) or bi-ventricles, *i.e.*, the LV and RV.

Joint analysis of cardiac bi-ventricles within a single framework is of great significance to cardiac function assessment and disease diagnosis (Lötjönen et al., 2004; Hu et al., 2005; Fritz et al., 2006; Lu et al., 2011; Wang et al., 2013, 2014; Zhen et al., 2014d), while posing great challenge for traditional segmentation based techniques (Cocosco et al., 2008). The first attempt to direct bi-ventricular volume estimation was provided in (Wang et al., 2014) which employs a Bayesian model. Given an input MR image with bi-ventricles, the model searches similar images in a set of templates with manually segmented LV/RV, and the similarity is measured by computing the distance on simple handcrafted features. The volume of cardiac ventricles in the input image is then simply calculated as the weighted average over the templates. The method did not fully model the statistical relationship between image features and cardiac ventricular volumes and is far from a capable tool of direct estimation.

The method suffers from several severe drawbacks. It 1) relies on a simplified assumption to model the relationship between volumes of the LV and RV in a cardiac cycle by an empirical linear function: $Vol(RV) = aVol(LV) + b$, 2) does not generalize well on more diverse datasets with large number of subjects, and 3) is computationally expensive to match a test point against all templates which are all the training samples, and typically requires uniformly-sampled training data for accurate results (Huang et al., 2011).

To deal with the above-mentioned issues, we proposed using random forests (Breiman, 2001) for direct bi-ventricular volume estimation in our conference version (Zhen et al., 2014d), in which bi-ventricular volume estimation is formulated as a regression problem. Multiple complementary features including pyramid Gabor features (PGF) (Zhen and Shao, 2013), histogram of gradients (HOG) (Dalal and Triggs, 2005) and intensity are carefully designed to represent cardiac MR images. Like (Afshin et al., 2012a; Wang et al., 2014), the method still relies on handcrafted feature representation by feature engineering (Bengio et al., 2013), which cannot be adapted to learn optimal representation from data and are unable to extract the discriminative information related to the specific domain (Bengio et al., 2013).

In general, existing direct methods (Afshin et al., 2012a; Wang et al., 2014) are still far from being a satisfactory tool for cardiac volume estimation. They might perform well on datasets of a small number of subjects due to overfitting while the performance cannot be guaranteed on datasets of a large number of subjects which present greater variability of images therefore raise more challenge (Petitjean and Dacher, 2011).

Towards full capacities of direct estimation, we propose a general fully learned-based framework for joint bi-ventricular volume estimation which removes user inputs and unnecessary assumptions. The framework consists of two main stages: cardiac image representation by multi-scale convolutional deep networks and joint bi-ventricular volume estimation by random forests. The multi-scale convolutional deep networks take advantages of deep learning as powerful tool for unsupervised representation learning and leverage the abundant unlabeled data. Random forests as discriminative learning can effectively capture the relationship between image appearance and bi-ventricular volumes, and more importantly they are able to automatically extract the most discriminate features for each ventricle due to the innate nature of feature selection.

1.2. Cardiac image representation learning

Cardiac image representation serves as a fundamental role in cardiac function analysis and is typically obtained by feature engineering in the existing methods (Montillo et al., 2004; Qian et al., 2006; Zheng et al., 2008; Chen et al., 2010; Garcia-Barnes et al., 2010; Afshin et al., 2012a; Punithakumar et al., 2013; Wang et al., 2014; Zhen et al., 2014d). Feature engineering relying on handcrafted features is labor-intensive and

Figure 1: The illustration of MR images with bi-ventricles.

suffers from the weakness of being unable to extract and organize representative information from the data (Bengio et al., 2013; Tang et al., 2014). Bi-ventricles exhibit great variability of cardiac images from gray levels to structure shapes as shown in Fig. 1. The shape of the two ventricles varies across patients, over time and along the long axis, which makes it extremely difficult for accurate analysis of bi-ventricular volumes. Since bi-ventricular volume estimation poses more challenge than previous tasks (Afshin et al., 2012a), these handcrafted image representations by feature engineering are not able to capture the variability of bi-ventricles due to ignoring the specific domain knowledge in data, especially on dataset of larger number of subjects.

A data-driven representation directly learned from unlabeled data is highly desirable and can capture underlying explanatory factors (Bengio et al., 2013; Zhen et al., 2014b). As highlighted in (Bengio et al., 2013) that being less dependant on feature engineering, data-driven representation learning is able to capture high variability of images, especially for cardiac images with bi-ventricles presenting combinatorial variations.

As a powerful state-of-the-art unsupervised feature learning technique, deep learning algorithms suit medical applications well due to the availability of abundant sample images without labels and have recently started to generate increasing attentions in medical image analysis (Cireşan et al., 2013; Carneiro and Nascimento, 2013; Prasoon et al., 2013), Convolutional deep learning networks (Krizhevsky et al., 2012; Turaga et al., 2010; LeCun et al., 1998; Lee et al., 2009; Petersen et al., 2014), one of the most representative deep learning algorithms, are effective techniques to retain topological structure, *e.g.*, 2D layout of pixels, and can be used to capture the anatomical structure of bi-ventricles in cardiac images.

To handle the great variability of bi-ventricles, we propose using a *multi-scale* convolutional deep belief network (MCDBN) to learn cardiac image representations (Shin et al., 2013; Petersen et al., 2014). By combining the strengths of both multi-scale analysis (Zhen et al., 2013; Shao et al., 2014; Zhen et al., 2014a) and deep learning, the MCDBN fits well for cardiac image representation with bi-ventricles. The MCDBN benefits representation learning of cardiac images in three folds: It **1**) can, to a large extent, retain structural layouts in cardiac images which are the most important features of bi-ventricles; **2**) can be efficiently trained due to the weight sharing which also allows us to use a large set of unlabeled data; **3**) can detect sufficient complementary features by multi-scale filtering which provides a rich and effective representation of bi-ventricles. The MCDBN leverages the strength of convolution DBN in unsupervised representation learning and more importantly allows us to use a large amount of unlabeled data which is abundantly available in medical image analysis.

1.3. Bi-ventricular volume estimation

The challenges that arise from the complex functional and geometrical interference and interdependency between the LV and RV lead to a high-dimensional representation. Regression forests (Breiman, 2001) are capable of modeling complex relationships between high-dimensional input features and continuous outputs. They have been successfully applied to various computer vision tasks (Shotton et al., 2013; Gall et al., 2011; Zhen et al., 2015b), and recently started to attract interest in medical image analysis (Criminisi et al., 2011). We choose random forests to work on top of image representations from MCDBN to fulfill bi-ventricular volume estimation due to the strong ability of feature selection and efficient implementation. Random forests are well-suited to direct and joint bi-ventricular volume estimation, which has been shown in our preliminary work (Zhen et al., 2014d).

Random forests (Breiman, 2001) are an ensemble of decision trees which combine the ideas of bagging and the random feature selection which benefit our learning-based bi-ventricular volume estimation in three folds: They **1)** effectively deal with the high-dimensional representation due to innate ability to select discriminative features (Criminisi and Shotton, 2013); **2)** avoid overfitting while providing accurate prediction by injecting randomness (Biau, 2012); **3)** are specifically fit for bi-ventricular volume estimation due to the intrinsic nature of feature selection (Breiman, 2001). With the above properties, random forests offer a prime regressor for direct and joint bi-ventricular volume estimation.

1.4. Contributions

We propose a general, fully learning-based framework to realize the full capacities of direct estimation of cardiac ventricular volumes by combining the strengths of both generative (for representation) and discriminant (for regression) learning. Specifically, **1)** we propose using multi-scale convolutional deep networks for unsupervised cardiac image representations learning from unlabeled data; and **2)** bi-ventricular volume estimation is formulated as a regression problem and random forests are employed for efficient volume estimation. Our methods provide a new framework from the perspective of regression for cardiac ventricular volume estimation which can also be used for other organ volume estimation and extensive model parameter estimation problems, e.g., model personalization (Marchesseau et al., 2013; Zettinig et al., 2014).

2. Methodology

The flowchart of the general framework is illustrated in Fig. 2. In the left block, the multi-scale convolutional deep belief networks learn a set of multi-scale data-driven feature detectors from totally unlabeled data for cardiac image representations. In the right block, random forests are trained on

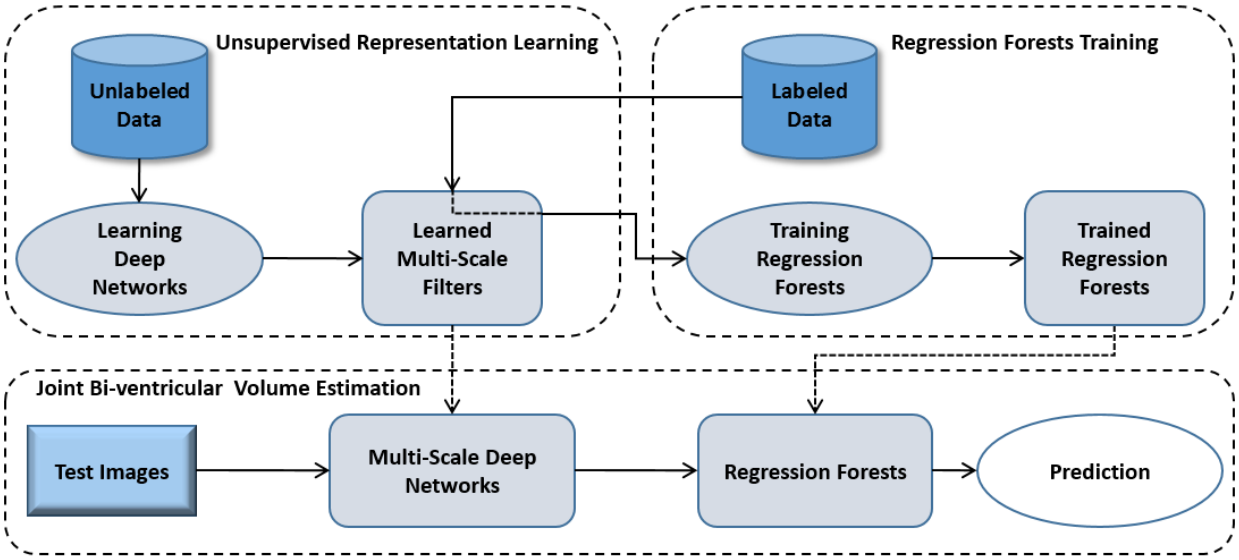

Figure 2: The flowchart of the proposed unsupervised feature learning and random forest regression. [Left block]: Unsupervised cardiac image representation learning by multi-scale deep networks from a unlabeled dataset. [Right block]: Training regression forests and on labeled data. [Bottom block]: Joint bi-ventricular volume estimation with the trained regressors.

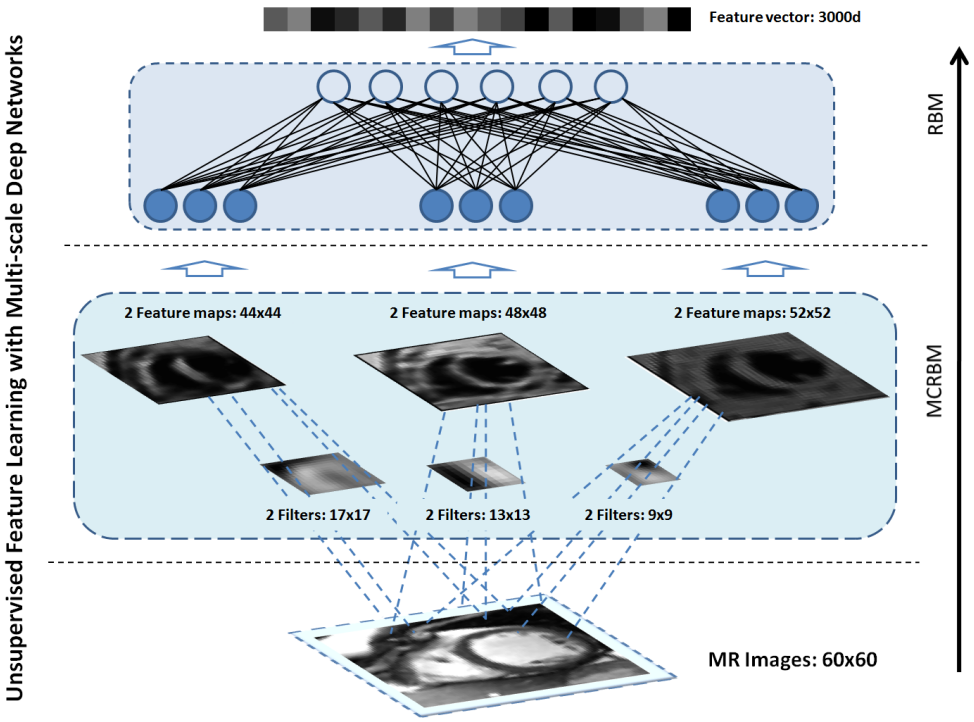

Figure 3: The schematic diagram of unsupervised feature learning with the proposed multi-scale deep networks. The three blocks from bottom to top are the input MR images, a multi-scale convolutional RBM and an RBM.

labeled data for bi-ventricular volume estimation. In the bottom block, test images go through learned feature detectors and trained regressors successively and bi-ventricular volumes are jointly estimated.

2.1. Representation learning by multi-scale deep networks

The proposed *multi-scale* convolutional deep belief network (MCDBN) is a three-layer deep network composed of a multi-scale convolutional restricted Boltzmann machine (MCRBM) and an RBM as shown in Fig. 3. Our MCDBN is inspired by convolutional deep belief nets (CDBNs) introduced by Lee et al. (Lee et al., 2009). By combining convolutional filters with deep belief nets, CDBNs can encode local structures of images and therefore achieve more descriptive representations. However, only a single scale of filters are learned in CDBNs, which limits the ability of feature detection. By learning multi-scale filters, our MCDBN further enhances the effectiveness of representation learning especially for cardiac images with bi-ventricles that exhibit complex geometrical variations.

2.1.1. Restricted Boltzmann machine (RBM)

A restricted Boltzmann machine (RBM) is a two-layer, bipartite, undirected graphical model. RBMs are fully connected with a group of binary hidden nodes $\mathbf{h}$, a group of visible nodes, which are either binary or real-valued, and symmetric connections between $\mathbf{h}$ and $\mathbf{v}$ represented by a weight matrix W. The weight W and biases $\mathbf{b}$ and $\mathbf{c}$ of an RBM can be learned by contrastive convergence (CD) (Hinton, 2002).

The RBMs can be stacked to form a deep belief network (DBN), a generative model with multiple layers, which demonstrates high ability of representation learning. In DBN, two adjacent layers are fully connected and no nodes in the same layer are connected. The DBN can be trained in a greedy layerwise way by treating each layer as an RBM (Hinton et al., 2006; Bengio et al., 2007), which works well in practice. Unlike DBNs which treat all the pixels in an image equally, the CDBNs model the topology of images by operating convolutional kernels on local neighborhoods and can naturally preserve image topological information (LeCun et al., 1998) which is extremely important especially for medical images with anatomical structures.

2.1.2. Multi-scale CDBNs

A convolutional RBM (CRBM) proposed by Lee et al. (2009) is composed of two layers, but the weights of the connections between visual and hidden layers are shared among all the locations in an image. The input visual layer consists of an array of $N_V \times N_V$ which in this work is real-valued. The hidden layer is composed of K groups each of which is a binary array of $N_H \times N_H$. Each of the K group in the hidden layer is associated with an $N_W \times N_W$ filter.

The energy function for a CRBM is defined as

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{k=1}^{K} \sum_{i,j=1}^{N_H} \sum_{r,s=1}^{N_W} h_{ij}^k W_{rs}^k v_{i+r-1, j+s-1} \qquad (1)$$

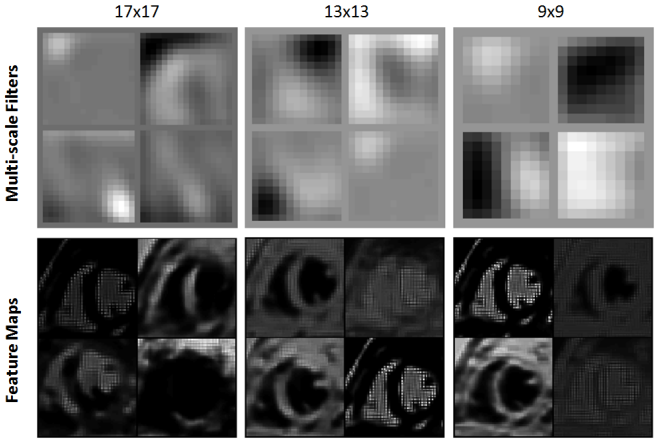

Figure 4: Illustration of the filters (feature detectors) of different sizes associated with the feature maps. In the top row are the leaned filters and in the bottom row are the corresponding feature maps.

$$- \sum_{k=1}^{K} b_k \sum_{i,j=1}^{N_H} h_{ij}^k - c \sum_{i,j=1}^{N_V} v_{ij}, \qquad (2)$$

where b_k is the bias for each group and c is the bias shared by all visible nodes. The energy function can be represented in terms of convolution as

$$E(\mathbf{v}, \mathbf{h}) = \sum_{k=1}^{K} h^k \bullet (\tilde{W} * v) - \sum_{k=1}^{K} b_k \sum_{i,j} h_{i,j}^k - c \sum_{i,j} v_{ij} \qquad (3)$$

In contrast to the original CRBM, we propose multi-scale CRBM (MCRBM) with filters of different sizes, which means we have $S \times K$ filters with S the number of scales.

By stacking an RBM on top of the proposed MCRBM, we obtain a three-layer network, *i.e.*, the multi-scale convolutional deep belief network (MCDBN). Totally unlabeled cardiac MR images are fed into the MCRBM to learn a set of multi-scale filters, *i.e.*, feature detectors. The feature maps from CRBM go further through an RBM to obtain more compact representations.

2.1.3. Learning multi-scale filters

The proposed MCDBN provides an effective representation learning of cardiac images, which creates a solid basis for subsequent bi-ventricular volume estimation. The complexities of bi-ventricles residing in different scales can be effectively captured by the multi-scale filters learned by the proposed deep networks. Larger scales of filters extract the simplex structure of the LV, while finer filters with smaller sizes are capable of detecting the sophisticated crescent-shaped of the RV which is much more complex than the LV.

Specifically, we learn 2 filters for each of 3 scales: 17×17, 13×13 and 9×9. Fig. 4 illustrates the learned multi-scale filters (in top row) and the corresponding feature maps (in the bottom row) outputted from the filters. As can be seen that the learned filters have successfully detected oriented and localized edges (Poultney et al., 2006; Lee et al., 2009) which

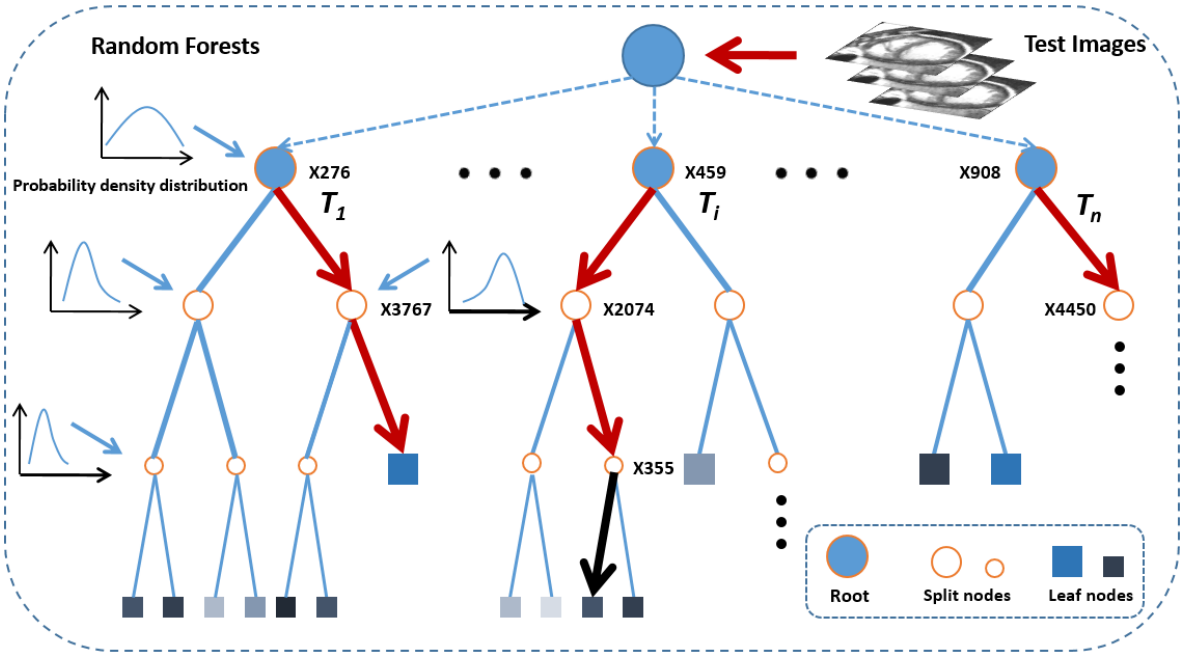

Figure 5: Illustrated are the random forests comprised of n decision trees $\{T_1, \ldots, T_i, \ldots, T_n\}$ learned from the training set. Test images can be quickly predicted by several simple comparison operations. $X\#$ indicates the #-th feature that are selected for the associated splitting notes. Test images go through each tree in the forests and the results outputted from all the trees are combined as the final prediction.

are the most representative features of cardiac ventricles. The learned feature maps intensify the main shapes and contours of bi-ventricles while removing the unrelated regions such as background, which provides an informative and discriminative representation, especially for images with bi-ventricles.

2.2. Regression forests

Mathematically, given a multivariate input $\mathbf{v}$ which in our case is the feature vector extracted from an image, our aim is to associate with a continuous multi-variate label $\mathbf{y}$, *i.e.*, the bi-ventricular cavity areas in images.

2.2.1. Training

We build decision trees using the adapted algorithm from (Breiman, 2001). Each internal node j of a tree is associated with a split function. The training process is to construct each tree with a randomly selected training subset. Note that only a subset of features are used for training each decision tree, which are fixed for prediction.

The split function at a split node j is formulated as a function with binary outputs

$$h(\mathbf{v}, \boldsymbol{\theta}_j) : R^d \times \mathcal{T} \rightarrow \{0, 1\} \tag{4}$$

where $\mathbf{v}$ is the input feature vector, $\mathcal{T}$ represents the space of all split parameters (Criminisi and Shotton, 2013), and $\boldsymbol{\theta}_j$ is the function parameter associated with the j-th node and can be trained by minimizing a least-squares error function I (Breiman, 2001) at the j-th split node:

$$\boldsymbol{\theta}_j = \arg\max_{\boldsymbol{\theta} \in \mathcal{T}} I(S_j, \boldsymbol{\theta}) \tag{5}$$

where S_j is a subset of training samples associated with the j-th node. The data point $\mathbf{v}$ arriving at the split node is sent to its left or right child node according to the result of the split function. An example of trained random forests is shown in Fig. 5 which will be used for prediction of new input images.

2.2.2. Prediction

As shown in Fig. 5, we pass a test image $\mathbf{v}$ through each tree starting from the root of each decision tree T_i, send to the left/right child by applying the split function, and stop when $\mathbf{v}$ reaches a leaf node of the tree. The simple comparison operation on each split node makes the prediction extremely fast and efficient. Given the t-th tree in a forest, the associated leaf output takes the form of a density probability function $p_t(\mathbf{y}|\mathbf{v})$. The forest output is the average over all tree outputs

$$p(\mathbf{y}|\mathbf{v}) = \frac{1}{T} \sum_{t=1}^{T} p_t(\mathbf{y}|\mathbf{v}) \tag{6}$$

where $\mathbf{y}$ is the continuous multi-variate label and T is the number of trees in the forest.

2.2.3. Feature selection

Random forests have the innate ability to select the most representative features closely related to cardiac volumes. The feature selection can be reflected by the feature importance ranked by random forests in the training stage (Breiman, 2001). Features are assigned with different values of importance according to their discriminative ability which is measured by regression

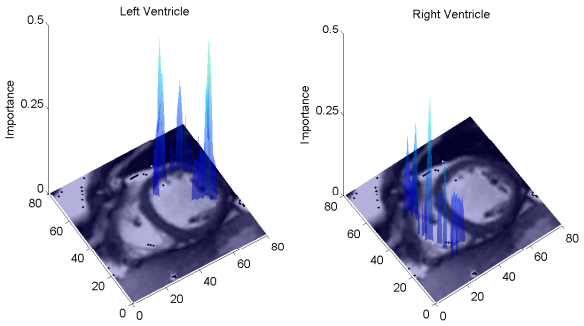

Figure 6: Illustration of feature importance learned by random forests. The magnitudes at corresponding locations indicate the importance of the features for volume estimation of the left and right ventricles, respectively. Features with large values of magnitudes are ranked as more important than features with small values.

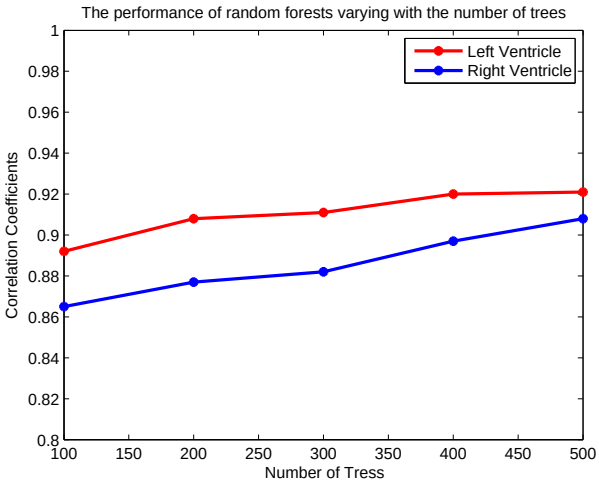

Figure 7: The performance of random forests with different numbers of trees.

errors. Features with larger values are ranked as more important than features with small values. Our experiments have indicated that random forests have a strong capacity of feature selection for the estimation of the LV and RV volumes. The results are shown in Fig. 6 from which we can easily observe that regions that are closely related to the LV and RV, respectively are successfully detected with high importance by random forests, while insignificant regions are largely removed. More importantly, the selected features mostly fall on the key regions such as edges, boundaries of the ventricles that can distinguish different cardiac ventricular volumes.

2.2.4. Tree numbers

The effects of different numbers of trees in random forests are illustrated in Fig 7. The performance keeps going up with the number of trees from 100 to 500. The computational cost will also increase with the number of trees. In our experiments, we use 500 trees as the final setting to keep the balance between performance and computational burden.

3. Experiments and results

3.1. Datasets and settings

In our experiments, two sets of subjects with 2D short-axis cine MR images are used including 2820 unlabeled images from 47 subjects for unsupervised feature learning and 6000 labeled images from 100 subjects for the validation of bi-ventricular volume estimation. The subjects are collected from 3 hospitals affiliated with two health care centers (London Healthcare Centre and St. Joseph's HealthCare) and 2 vendors (GE and Simens) including both health and diseased cases. The pathologies are extremely diverse including regional wall motion abnormalities, myocardial hypertrophy, mildly dilated RV, atrial septal defect, LV dysfunction, mildly enlarged LV, decreased ejection fraction in most cases, etc. Each subject contains 20 frames throughout a cardiac cycle. In each frame, three representative slices, i.e., apical, mid-cavity and basal, are selected following the standard AHA prescriptions (Cerqueira et al., 2002) for validation, and their manual segmentations are used as the benchmark.

To benchmark with existing direct methods (Wang et al., 2014; Zhen et al., 2014d), we estimate cavity areas of the LV and RV in MR images, and the volumes are computed by integrating LV/RV cavity areas along the sagittal direction perpendicular to the short axis. A single cropped region of interest (ROI) rather than two individual ones is placed to enclose the LV and RV in an MR image, which can be obtained automatically (Petitjean and Dacher, 2011). The cropped images are then resized into 60×60 pixels as the inputs.

The MCDBN has been conducted on the unlabeled dataset of 47 subjects to obtain filters which are used to create feature representations of all images. For volume estimation, we employ a leave-one subject-out validation approach, i.e., 100-fold cross validation on the labeled dataset of 100 subjects. The performance is evaluated by comparing with the golden standard manual segmentation using absolute estimation errors and correlation coefficients. The correlation coefficient is used to measure the linear correlation between ground truth and direct estimation.

3.2. Implementation details

For the MCDBN, we follow the original work (Lee et al., 2008, 2009; Hinton et al., 2006; Hinton, 2010) using a learning rate of 0.0001 for RBMs and sparsity of 0.01 for CRBMs. As suggested in (Hinton, 2010), we start with a momentum of 0.5 and once the large initial progress in the reduction of the reconstruction error has settled down to gentle progress, increase the momentum to 0.9. We use 2 filters of each scale to create features maps as the inputs to the RBMs. The final feature vector as the image representation is of 3000 dimensions which is fed into regression forests for bi-ventricular volume estimation.

For random forests, we set the number of features used in splitting note as 500, and since the number of trees in forests is

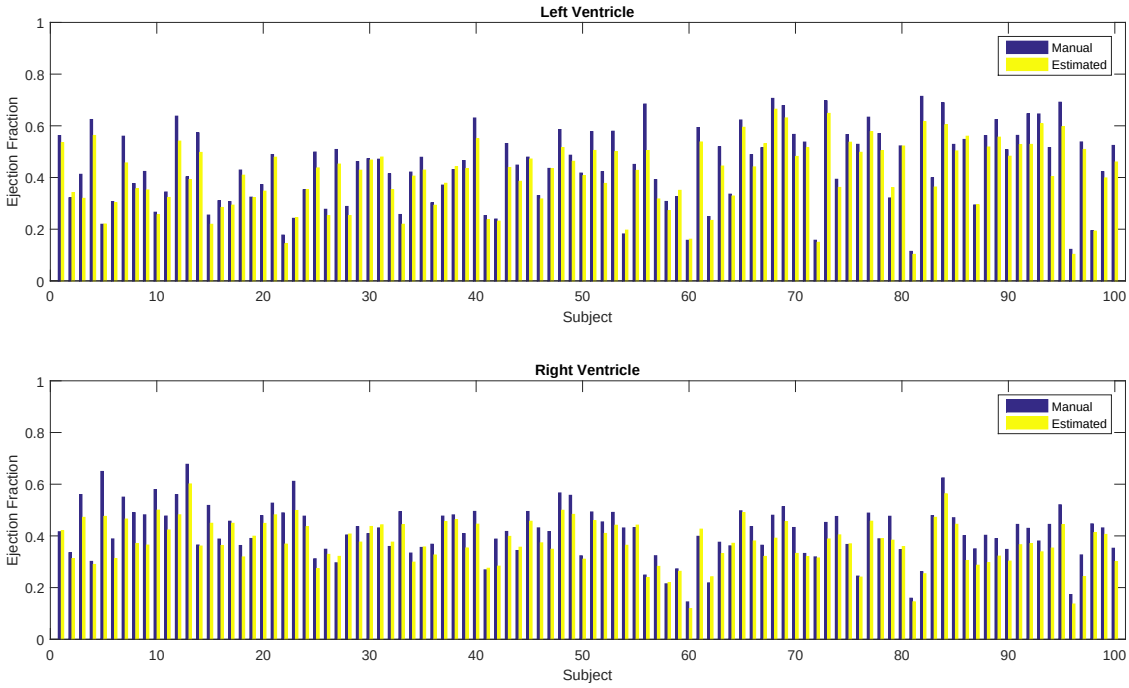

Figure 9: Comparison of EFs obtained by manual segmentation (blue) and the proposed method (red) for the LV (upper) and RV (bottom), respectively.

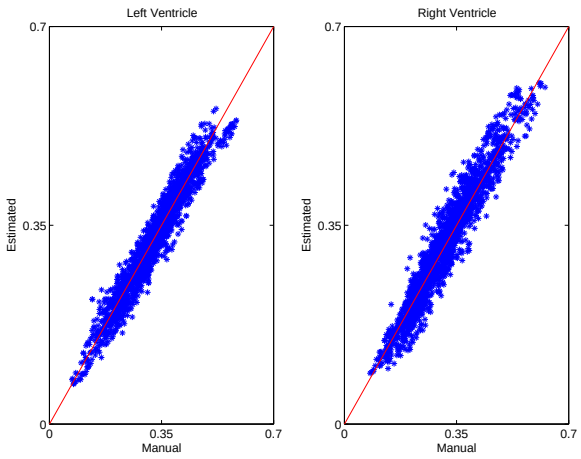

Figure 8: The correlation coefficients between the volumes obtained by the proposed method and manual segmentation for the LV and RV, respectively.

a key parameter which determines the regression performance, we experiment with different values to investigate its effects.

3.3. Estimation results

We apply the proposed methods to cardiac bi-ventricular volume estimation and calculation of ejection fractions (EF) with the estimated volumes. The EF is an important cardiac functional parameter and predictor of prognosis and widely used in clinical analysis. The EF is computed by

$$EF = \frac{V_d - V_s}{V_d}, \tag{7}$$

where V_d and V_s denote the largest (end-diastolic) and the smallest (end-systolic) volumes of a ventricle in a cardiac cycle, respectively.

3.3.1. Bi-ventricular volumes

The effectiveness of the proposed method is demonstrated by the outstanding performance on a large dataset with 6000 MR images from 100 subjects for both LV and RV. The estimated volumes by the proposed method are measured by comparing with those by manual segmentation using a leave-one-subject-out cross validation. The correlations between estimated and manually obtained volumes are depicted in Fig. 8 for the LV and RV, respectively. Despite of the challenges in joint estimation of bi-ventricular volumes, the proposed method achieves a correlation coefficient of **0.921** for the LV, and can yield **0.908** for the RV which has much greater geometrical complexity than the LV. The high estimation accuracies with low volume estimation errors, *i.e.*, 0.010 ± 0.013 (LV) and 0.014 ± 0.012 (RV), indicate that the proposed method can be practically used in clinical cardiac function analysis.

3.3.2. Ejection fraction

As shown in Fig. 9, we compare EFs obtained from estimated volumes with those from manual segmentation by human experts. The EFs from the estimated volumes provide very close

Table 1: The correlation coefficients with different sizes of learned filters.

Filter sizes	17×17	13×13	9×9	Multiple scales
LV	0.899	0.885	0.873	**0.921**
RV	0.869	0.875	0.889	**0.908**

approximations for both the LV and RV with a large proportion equal to their counterparts obtained manually. Compared to segmentation based methods, the estimation errors of EFs are relatively low: 0.0387 ± 0.0330 and 0.0455 ± 0.0347 for the LV and RV, respectively. The high estimation accuracies on both the LV and RV will be extremely important for assessment of cardiac functions, which indicates the potential use of the proposed method for cardiac disease diagnosis.

3.4. Parameter evaluation

To look further into the proposed method, we have also investigated the performance of the learned filters with different scales used in our experiments.

3.4.1. Multi-scale filters

The advantage of the proposed multi-scale convolutional deep network is validated by comparing with results from three different sizes of filters as shown in Table 1. More importantly, our results further show the complementary properties of different sizes of filters in the cardiac image representation. On the LV, larger scales of filters show to be superior over smaller ones, while on the RV, smaller scales of filters perform better. This is reasonable since structure of the RV is more complex than the LV, smaller scales of filters are more able to capture these localized complex features on the RV. The multi-scale filters outperform each of single scale filters. The results confirm that representation learning by the proposed multi-scale convolutional deep networks overcomes the great variability and geometrical complexity of bi-ventricles, and validate the effectiveness of multi-scale convolutional deep networks for cardiac image representation learning in bi-ventricular volume estimation.

3.5. Comparison

The performance of the proposed method can be further demonstrated by the comparison with existing direct estimation methods and segmentation-based methods. Both bi-ventricular volumes and ejection fraction are used as the measurements in the comparison.

3.5.1. Comparison with existing direct methods

The advantage of the proposed method is further shown by comparing with existing direct estimation methods: the Bayesian model (Wang et al., 2014), and multiple features (Zhen et al., 2014d) under the unified experimental settings. As shown in Table 2, the proposed method largely outperforms the Baysian model and multiple features both for the LV and RV by up to **0.085** in terms of correlation coefficients and the volume estimation errors are much -up to 37.5%- lower than both of them. As shown in Table 3, the proposed methods produces lower estimation errors of ejection fractions for bi-ventricles than both the Bayesian model and multiple features.

Despite of that the RV is extremely challenging due to its complex geometrical structures, the proposed method performs much better than the methods in (Wang et al., 2014; Zhen et al., 2014d) using handcrafted features, which shows the effectiveness of the proposed multi-scale deep networks for cardiac image representation learning. The better performance than the Bayesian model in (Wang et al., 2014) shows the advantages of the proposed method by formulating volume estimation as a regression problem which incorporates learning stages. The superb performance over multiple features in (Zhen et al., 2014d) demonstrates the effectiveness of unsupervised representation learning by the proposed multi-scale deep networks. Moreover, the outstanding performance indicates that the proposed method can be practically used in clinical cardiac diagnosis. Note that the Bayesian model has achieved high accuracy probably because of using a relative small size of dataset with only 56 subjects which can not capture the huge variability of cardiac ventricles. In addition, the Bayesian model relies on the unproven assumption that LV and RV volumes are linearly correlated during a cardiac cycle, which however does not always hold in diseased cases. Our dataset contains 100 subjects (nearly twice sizes of their dataset) exhibit sufficiently inter- and intra-subject variabilities, especially due to the presence of diverse pathologies.

3.5.2. Comparison with segmentation-based methods

To demonstrate the superiority of the proposed direct estimation method over conventional segmentation-based methods, following the work (Wang et al., 2013), we conduct a comparison with two representative segmentation methods including level set (Ayed et al., 2009a) and graph cut (Ayed et al., 2009b). As shown in Table 2 and 3, on our dataset the proposed method largely outperforms graph cut and level set in terms of both bi-ventricular volumes and ejections fractions. Moreover, neither level set or graph cut are applicable to the RV for volume estimation. Our method can be flexibly used for either a single ventricle or joint bi-ventricles.

4. Discussion

Direct estimation, in general, replaces tedious and unreliable segmentation, and focuses on the ultimate goal of volume estimation. The proposed direct estimation method not only solves a dual estimation problem but also for the first time formulates it as a regression framework. This new framework substantially outperforms existing segmentation based and direct methods, and more importantly offers a more compact and exquisite mathematical formulation of regression, which is flexible and easily extendable to other applications.

The effectiveness of the proposed method stems from the two key incorporated components: unsupervised representation learning and random forests regression which are suited well

Table 2: The comparison of estimation errors for bi-ventricular volumes.

Methods	Correlation coefficients		Volume estimation errors	
	LV	RV	LV	RV
Our method	**0.921**	**0.908**	**0.010 ± 0.011**	**0.014 ± 0.012**
Baysian model (Wang et al., 2014)	0.861	0.823	0.016 ± 0.019	0.018 ± 0.013
Multiple features (Zhen et al., 2014d)	0.870	0.853	0.012 ± 0.011	0.017 ± 0.016
Level set (Ayed et al., 2009a)	0.803	-	0.036 ± 0.025	-
Graph cut (Ayed et al., 2009b)	0.835	-	0.029 ± 0.027	-

Table 3: The comparison of estimation errors for ejection fraction.

Methods	Level set	Graph cut	Baysian model	Multiple features	Our method
LV	0.110	0.097	0.055	0.047	**0.0387**
RV	-	-	0.071	0.059	**0.0455**

to cardiac ventricular volume estimation and medical image analysis. Since unlabeled imaging data is always abundantly available in medical image analysis, unsupervised representation learning allows us to use a large of amount of unlabeled data to obtain more faithful and informative data-driven representations. Random forests provide an effective tool for cardiac bi-ventricular volume estimation due to its innate ability of feature selection mechanism and computational efficiency by subsampling and boosting.

The proposed method shows advantages over both conventional segmentation based methods and other recently-proposed direct estimation methods.

4.1. Proposed method vs. segmentation based methods

The proposed method possesses attractive advantages over previous cardiac ventricular volume estimation using segmentation based methods. Cardiac ventricular volume estimation has been a challenging task. Most of conventional segmentation based methods rely on the unreliable assumption that cardiac ventricles are supported by edges and region intensity homogeneity. However, edges of cardiac ventricles are not always consistently visible along the entire contour due to overlapping of anatomical structures and noise, etc., and the homogeneity is severely violated due to the complex image textures and appearances, especially with the presence of pathology. In addition, existing automatic segmentation methods are mostly restricted to the LV. RV segmentation remains an unsolved problem not to mention multiple ventricles, *e.g.*, bi-ventricles and four chambers. Our direct estimation discovers the relationship between image appearances and cardiac volumes by statistical learning from annotated data, which significantly outperforms two typical segmentation methods: level set and graph cut as indicated in experimental results. In contrast to segmentation based methods, our method as direct estimation 1) removes the intermediate and always tedious segmentation by either manual or automatic methods, achieving more accurate and efficient estimation 2) can naturally handle the cases without consistently strong edge and of region inhomogeneity, guaranteeing the applicability in clinical practise;; and 3) is able to flexibly deal with both single and multiple ventricles, showing great generality to other organ volume estimation.

4.2. Proposed method vs. existing direct methods

The proposed method also demonstrates more merits in contrast to other recently-proposed direct estimation methods. Our method is a fully automatic without relying on any assumptions, user inputs and initialization. By leveraging advanced machine learning techniques, *i.e.*, unsupervised representation learning and random forests regression, our method achieves the fully capacity of direct estimation, showing much better performance on the largest datset as indicated in experimental results. 1) Instead of using ineffective handcrafted features in previous methods, our method learns data-driven representations which can effectively capture the characteristics of objects, *e.g.*, cardiac ventricles, to achieve optimal representations for volume estimation; 2) By formulating bi-ventricular volume estimation as a regression problem, our method achieves more accurate estimation than previous methods of no learning stages, *e.g.*, template matching (Wang et al., 2014); and 3) By using powerful random forests for regression, our method can effectively handle large-scale datasets without overfitting and promises the practical use of our method in clinical applications.

4.3. Critical analysis

While the proposed method has been validated on a large dataset with highly diverse cases, the pros and cons of the proposed method come from the same fact that no explicit contour is provided. Without relying on contouring, the proposed direct estimation method is able to effectively handle challenges in conventional segmentation based methods. Meanwhile, it would pose a learning curve to those who are used to contouring for validation which, however, can be easily shortened by providing more measurement through relating to human perception and intuition: 1) Due to the temporal coherence, plotting

the volume variation cross a cardiac cycle can provide direct mutual validation of volumes in different frames; and 2) Due to the anatomical constraints between adjacent slices along the long axis, plotting the area distributions of all slices can also help check the results.

5. Conclusion

In this paper, we have presented a fully learning based method for direct estimation of cardiac bi-ventricular volumes. Our method takes advantages of both generative and discriminant learning, *i.e.*, unsupervised representation learning and supervised volume estimation. We have evaluated it on a large set of dataset with 6000 images from 100 subjects using a leave-one-subject-out cross validation. The proposed method produces high correlations with ground truth and outperforms existing direct estimation methods, showing its effectiveness for cardiac ventricular volume estimation. The comparison with segmentation based methods demonstrates the superiority of our direct method for cardiac bi-ventricular volume estimation.

More importantly, discriminant learning via random forests regression allows to deploy advanced machine learning techniques to facilitate cardiac functional analysis, which provides an effective tool to automate analysis of medical imagining data and therefore enables accurate and efficient diagnosis in clinical practise (Wang and Summers, 2012). The proposed estimation framework can be easily extended to other applications for volume estimation, such as estimating the volume of a tumor (Bolte et al., 2007; Heckel et al., 2014) which is still dependent on segmentation, and can also be used for extensive model parameter estimation problems, *e.g.*, model personalization (Zettinig et al., 2014; Marchesseau et al., 2013).

Acknowledgment

Computations were performed using the data analytics Cloud at SHARCNET (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.

References

Afshin, M., Ayed, I. B., Islam, A., Goela, A., Peters, T. M., Li, S., 2012a. Global assessment of cardiac function using image statistics in mri. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. pp. 535–543.

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

Afshin, M., Ben Ayed, I., Reza Sadeghi Neshat, H., Punithakumar, K., Goela, A., Islam, A., Peters, T., Li, S., 2012b. Automated assessment of regional left ventricular function for cardiac mri with minimal user interaction. In: Annual Meeting - Radiological Society of North America (RSNA).

Ayed, B. I., Li, S., Ross, I., 2009a. Embedding overlap priors in variational left ventricle tracking. IEEE Transactions on Medical Imaging 28 (12), 1902–1913.

Ayed, B. I., Punithakumar, K., Li, S., Islam, A., Chong, J., 2009b. Left ventricle segmentation via graph cut distribution matching. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009. Springer, pp. 901–909.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intellegnence 35, 1798–1828.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al., 2007. Greedy layer-wise training of deep networks. In: Advances in neural information processing systems. Vol. 19. p. 153.

Biau, G., 2012. Analysis of a random forests model. The Journal of Machine Learning Research 98888, 1063–1095.

Bolte, H., Jahnke, T., Schäfer, F., Wenke, R., Hoffmann, B., Freitag-Wolf, S., Dicken, V., Kuhnigk, J., Lohmann, J., Voss, S., et al., 2007. Interobserver-variability of lung nodule volumetry considering different segmentation algorithms and observer training levels. European journal of radiology 64 (2), 285–295.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Carneiro, G., Nascimento, J., 2013. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. IEEE Transactions on Pattern Analysis and Machine Intellegnence.

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 (4), 539–542.

Chen, T., Wang, X., Chung, S., Metaxas, D., Axel, L., 2010. Automated 3d motion tracking using gabor filter bank, robust point matching, and deformable models. IEEE Transactions on Medical Imaging 29 (1), 1–11.

Cireşan, D. C., Giusti, A., Gambardella, L. M., Schmidhuber, J., 2013. Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, pp. 411–418.

Cocosco, C. A., Niessen, W. J., Netsch, T., Vonken, E.-j., Lund, G., Stork, A., Viergever, M. A., 2008. Automatic image-driven segmentation of the ventricles in cardiac cine mri. Journal of Magnetic Resonance Imaging 28 (2), 366–374.

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

Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E., 2011. Regression forests for efficient anatomy detection and localization in ct studies. In: MICCAI Workshop on Medical Computer Vision. pp. 106–117.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 1. pp. 886–893.

Fritz, D., Rinck, D., Dillmann, R., Scheuering, M., 2006. Segmentation of the left and right cardiac ventricle using a combined bi-temporal statistical model. In: Medical Imaging. International Society for Optics and Photonics, pp. 614121–614121.

Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V., 2011. Hough forests for object detection, tracking, and action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (11), 2188–2202.

Garcia-Barnes, J., Gil, D., Badiella, L., Hernandez-Sabate, A., Carreras, F., Pujades, S., Martí, E., 2010. A normalized framework for the design of feature spaces assessing the left ventricular function. IEEE Transactions on Medical Imaging 29 (3), 733–745.

Heckel, F., Meine, H., Moltz, J., Kuhnigk, J., Heverhagen, J., Kiebling, A., Buerke, B., Hahn, H., 2014. Segmentation-based partial volume correction for volume estimation of solid lesions in ct. IEEE Transactions on Medical Imaging 22, 462 – 480.

Hinton, G., 2010. A practical guide to training restricted boltzmann machines. Momentum 9 (1), 926.

Hinton, G. E., 2002. Training products of experts by minimizing contrastive divergence. Neural computation 14 (8), 1771–1800.

Hinton, G. E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural computation 18 (7), 1527–1554.

Hu, Z., Metaxas, D., Axel, L., 2005. Computational modeling and simulation of heart ventricular mechanics from tagged mri. Functional Imaging and Modeling of the Heart, 881–883.

Huang, D., Storer, M., De la Torre, F., Bischof, H., 2011. Supervised local subspace learning for continuous head pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2921–2928.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86 (11), 2278–2324.

Lee, H., Ekanadham, C., Ng, A. Y., 2008. Sparse deep belief net model for visual area v2. In: Advances in neural information processing systems. pp. 873–880.

Lee, H., Grosse, R., Ranganath, R., Ng, A. Y., 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning. ACM, pp. 609–616.

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 (3), 371–386.

Lu, X., Wang, Y., Georgescu, B., Littman, A., Comaniciu, D., 2011. Automatic delineation of left and right ventricles in cardiac mri sequences using a joint ventricular model. In: Functional Imaging and Modeling of the Heart. Springer, pp. 250–258.

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. Medical image analysis 17 (7), 816–829.

Montillo, A., Metaxas, D., Axel, L., 2004. Extracting tissue deformation using gabor filter banks. In: Medical Imaging 2004. International Society for Optics and Photonics, pp. 1–9.

Petersen, K., Nielsen, M., Diao, P., Karssemeijer, N., Lillholm, M., 2014. Breast tissue segmentation and mammographic risk scoring using deep learning. In: Breast Imaging. Springer, pp. 88–94.

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

Poultney, C., Chopra, S., Cun, Y. L., et al., 2006. Efficient learning of sparse representations with an energy-based model. In: Advances in neural information processing systems. pp. 1137–1144.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M., 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, pp. 246–253.

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. Medical image analysis 17 (3), 311–324.

Qian, Z., Metaxas, D. N., Axel, L., 2006. Extraction and tracking of mri tagging sheets using a 3d gabor filter bank. In: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE. IEEE, pp. 711–714.

Shao, L., Zhen, X., Tao, D., Li, X., 2014. Spatio-temporal laplacian pyramid coding for action recognition. IEEE Transactions on Cybernetics 44 (6), 817–827.

Shin, H.-C., Orton, M. R., Collins, D. J., Doran, S. J., Leach, M. O., 2013. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8), 1930–1943.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R., 2013. Real-time human pose recognition in parts from single depth images. Communications of the ACM 56 (1), 116–124.

Tang, J., Shao, L., Zhen, X., 2014. Robust point pattern matching based on spectral context. Pattern Recognition 47 (3), 1469–1484.

Turaga, S. C., Murray, J. F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., Seung, H. S., 2010. Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Computation 22 (2), 511–538.

Wang, H., Amini, A. A., 2012. Cardiac motion and deformation recovery from mri: a review. IEEE Transactions on Medical Imaging 31 (2), 487–503.

Wang, S., Summers, R. M., 2012. Machine learning and radiology. Medical image analysis 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. Medical image analysis 13 (5), 773–784.

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

Wang, Z., Salah, M., Ayed, I., Islam, A., Goela, A., Li, S., 2013. Bi-ventricular volume estimation for cardiac functional assessment. In: Annual Meeting - Radiological Society of North America (RSNA).

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. Medical image analysis.

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

Zhen, X., Shao, L., Li, X., 2014a. Action recognition by spatio-temporal oriented energies. Information Sciences 281, 295–309.

Zhen, X., Shao, L., Tao, D., Li, X., 2013. Embedding motion and structure features for action recognition. IEEE Transactions on Circuits and Systems for Video Technology 23 (7), 1182–1190.

Zhen, X., Shao, L., Zheng, F., 2014b. Discriminative embedding via image-to-class distances. In: British Machine Vision Conference.

Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S., 2015a. Direct volume estimation without segmentation. In: SPIE Medical Imaging. International Society for Optics and Photonics, pp. 94132G–94132G.

Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014c. A comparative study of methods for cardiac ventricular volume estimation. In: Annual Meeting - Radiological Society of North America (RSNA).

Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014d. Direct estimation of cardiac bi-ventricular volumes with regression forests. In: Accepted by Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014.

Zhen, X., Wang, Z., Yu, M., Li, S., June 2015b. Supervised descriptor learning for multi-output regression. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

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 Transactions on Medical Imaging 27 (11), 1668–1681.