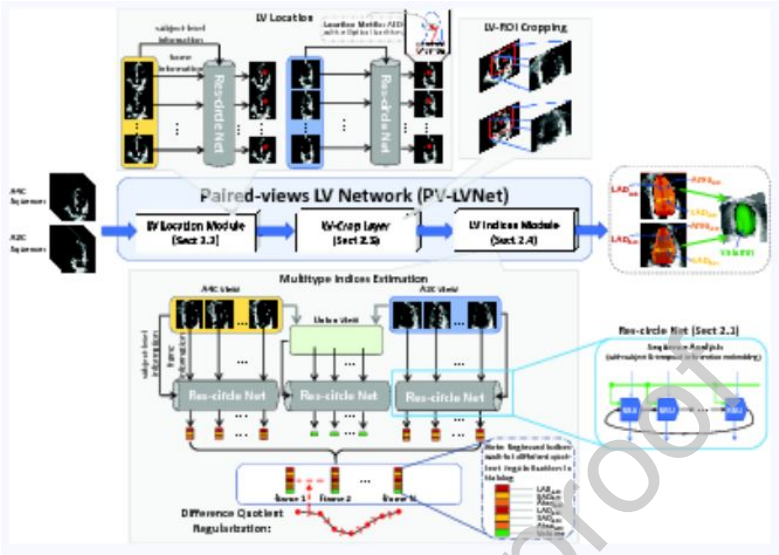
PV-LVNet Direct left ventricle multitype indices estimation from 2D echocardiograms of paired apical views with deep neural networks

Rongjun Ge, Guanyu Yang, Yang Chen, Limin Luo, Cheng Feng, Heye Zhang, Shuo Li

Highlights

- An effective method for quantifying LV from multiple dimensions and views.
- A brand-new recurrent net for embedding subject and temporal information.
- An efficient location loss function for robust location and cropping.
- A creative regularization item for enhancing sequential data evolution fitting.



PV-LVNet: Direct Left Ventricle Multitype Indices Estimation from 2D Echocardiograms of Paired Apical Views with Deep Neural Networks

Rongjun Ge[a,c,d], Guanyu Yang[a,c,d], Yang Chen[a,b,c,d,*], Limin Luo[a,c,d], Cheng Feng[e], Heye Zhang[f], Shuo Li[g,h,*]

[a] *Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China*
[b] *School of Cyber Science and Engineering, Southeast University, Nanjing, China*
[c] *Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China*
[d] *Centre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), Rennes,France*
[e] *Department of Ultrasound, The Third People's Hospital of Shenzhen, Shenzhen, China*
[f] *School of Biomedical Engineering, Sun Yat-Sen University, Guangzhou, China*
[g] *Department of Medical Imaging, Western University, London, Canada*
[h] *Digital Imaging Group of London, London, Canada*

Abstract

Accurate direct estimation of the left ventricle (LV) multitype indices from two-dimensional (2D) echocardiograms of paired apical views, i.e., paired apical four-chamber (A4C) and two-chamber (A2C), is of great significance to clinically evaluate cardiac function. It enables a comprehensive assessment from multiple dimensions and views. Yet it is extremely challenging and has never been attempted, due to significantly varied LV shape and appearance across subjects and along cardiac cycle, the complexity brought by the paired different views, unexploited inter-frame indices relatedness hampering working effect, and low image quality preventing segmentation. We propose a paired-views LV network (PV-LVNet) to automatically and directly estimate LV multitype indices from paired echo apical views. Based on a newly designed Res-circle Net, the PV-LVNet robustly locates LV and

*Corresponding author.
Email addresses: chenyang.list@seu.edu.cn (Yang Chen), sli287@uwo.ca (Shuo Li)

automatically crops LV region of interest from A4C and A2C sequence with location module and image resampling, then accurately and consistently estimates 7 different indices of multiple dimensions (1D, 2D & 3D) and views (A2C, A4C, and union of A2C+A4C) with indices module.

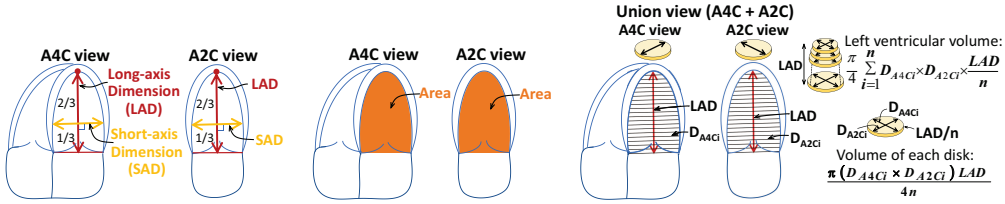
The experiments show that our method achieves high performance with accuracy up to 2.85mm mean absolute error and internal consistency up to 0.974 Cronbach's α for the cardiac indices estimation. All of these indicate that our method enables an efficient, accurate and reliable cardiac function diagnosis in clinical.

Keywords: multitype cardiac indices, direct estimation, 2D echo, paired apical views, Res-circle Net

1. Introduction

Accurate estimation for left ventricle (LV) indices (i.e., dimension, area & volume) in two-dimensional (2D) echocardiograms (echo) of paired apical views (i.e., paired apical four-chamber and two-chamber views) is of great clinical significance to cardiac function evaluation (Schiller et al., 1989; Lang et al., 2006, 2015). 2D echo is the most frequently used noninvasive modality for the diagnosis of cardiac disease because of its unique ability to provide real-time images of the beating heart, combined with its availability and portability (Lang et al., 2015; Abdi et al., 2017; Gao et al., 2017, 2018). The multitype indices of LV from 2D echo paired apical views, covering long-axis dimension (LAD), short-axis dimension (SAD), area and volume, which are measured from cavity as Fig.1, are most widely used to assess LV chamber size and contractile function (Schiller et al., 1989; Pascual et al., 2003; Lang et al., 2015). It promotes comprehensive metrics from 1D (i.e., LAD, SAD), 2D (i.e., area) and 3D (i.e., volume). Such paired orthogonal apical four-chamber (A4C) and two-chamber (A2C) views enable a better stereoscopic reproducibility of cardiac LV motion compared to the separate plane observation from single view, for further comprehensive quantitative functional analysis (Schiller et al., 1989; Ciampi and Villari, 2007).

The existing (semi-)automated cardiac indices estimation methods never refers to multitype indices in 2D echo sequences of paired apical views. These methods are mainly classified into two groups: segmentation and direct regression. However, the segmentation methods just enable limited simple index types (i.e. area) without extra interaction, and the existing direct



(a) 1D metric (b) 2D metric (c) 3D metric

Figure 1: The multitype indices from the paired apical views (A2C & A4C) are critically important for clinical diagnosis, yet extremely laborious measurement. They cover the 1D and 2D metrics of each single view, and the 3D metric of union view, for a comprehensive assessment. (a) LAD: from the apex to the middle mitral valve plane. SAD: perpendicular to the long axis, at one-third of the LAD from the mitral valve plane. (b) Area: the whole LV cavity. (c) LV volume: jointly from A4C and A2C by using the biplane method of discs (modified Simpson's rule).

methods almost all focus on a single view of cardiac magnetic resonance (CMR) causing limited observation and evaluation. Strong clinical evidence shows that the indices from echo that cover multiple dimensions and views enable a comprehensive cardiac diagnosis, yet their automated estimation is still thwarted by inherently existing challenges such as 1) LV shape and appearance in apical view significantly vary among subjects, and along the cardiac cycle. 2) Although the paired views provide complementary information, the different image structures are introduced with increased complexity. 3) Ambiguous relatedness inter frames hampers learning procedure of sequential indices from better convergence and generalization. 4) Low image quality of echo, like fuzzy border, edge dropout, acoustic shadows, etc., raises great challenges for automated methods, especially segmentation method.

1.1. Related Works

Segmentation methods aim to achieve automated LV segmentation for improving the diagnosis efficiency, however it is still an open and challenging task, due to the inherent characteristics of the 2D echo, such as low signal-to-noise ratio, edge dropout, shadows, indirect relation between pixel intensity and the physical property of the tissue, and anisotropy of ultrasonic image formation (Carneiro et al., 2012). Active contours (Debreuve et al., 2001; Malladi et al., 1995; Paragios, 2003) and deformable templates (Jacob et al., 2002; Nascimento et al., 2008) achieve good segmentation results relying on

the LV shape and appearance of the prior knowledge (Georgescu et al., 2005). By considering use of inaccurate prior knowledge and low-level handcrafted features may bound working robustness, the supervised deep learning method (Mo et al., 2018; Chen et al., 2016; Carneiro et al., 2012; Oktay et al., 2018) tries to learn information from data. The deep Poincar Map (Mo et al., 2018) coupled deep learning with the dynamic-based labeling scheme to reduce the requirement on the huge data; iMD-FCN (Chen et al., 2016) used the transfer learning from cross domains to enhance the feature representation; Carneiro et al. (2012) combined the deep belief networks, the decoupling rigid and nonrigid classifiers and the derivative-based search to increase the robustness for imaging conditions and LV shape variations; ACNNs (Oktay et al., 2018) encouraged the models to follow the global anatomical properties of the underlying anatomy via the non-linear representations of the shape learnt from the stacked convolutional autoencoder. All of these show great potential with the development of deep learning. Nevertheless, most of the working LV segmentation methods in the practical clinical diagnosis are still semi-automatic, which need time-consuming user interaction to handle a great number of medical images (Luo et al., 2018).

Direct regression methods without intermediate segmentation has undergone a great development and recognition (Ravı et al., 2017; Peng et al., 2016; Wu et al., 2017; Lathuilière et al., 2017; Pereira et al., 2018; Zhen et al., 2014a, 2015b, 2017) for better and more efficient cardiac indices estimation, but never performed on paired 2D echo apical views. By directly analyzing LV biological structure, these methods provide effective tools to automate the analysis of one single view from CMR, especially the short-axis view, and enable accurate and efficient diagnosis in clinical practice (Zhen et al., 2016). With two-phase operation, LV volume (as integration of cavity areas in short-axis view slices) is estimated on the handcrafted cardiac image representation, including Bhattacharyya coefficient between image distributions (Afshin et al., 2012, 2014), appearance features (Wang et al., 2014), multiple low level image features (Zhen et al., 2014b), as well as unsupervised features from multiscale convolutional deep belief network (Zhen et al., 2016) and supervised descriptor learning (Zhen et al., 2015a). Instead of separate representation and regression, joint learning (Xue et al., 2017a,c) captures task-relevant cardiac information for the indices estimation. For a comprehensive assessment of cardiac function, Xue et al. (2017b, 2018) achieve multitype indices estimation on short-axis view cardiac CMR. However, all of these direct methods still have the limitation on

2D echo paired apical views, due to: 1) multitype indices estimation from different views is ignored and lacked, 2) some cardiac indices in 2D echo, like volume, are often obtained jointly from paired views, and 3) LV shape in apical view is irregular and make it difficult to establish a standard preprocessing method for getting LV cropping (short-axis view CMR just need to manually find several relatively fixed landmarks).

1.2. Contributions

In this paper, we propose a paired-views LV network (PV-LVNet) to automatically achieve a high-quality estimation of LV multitype indices from 2D echo sequences of paired apical views. As shown in Fig.2, the network is built based on our newly designed Res-circle Net, and implemented with three interdependent functional parts: LV location module, image resampling and LV indices module. The Res-circle Net for sequential analysis embedded with subject's holistic characteristics and frame's temporal changes is used in both LV location and indices modules. And functionally, the LV location module with the anisotropic Euclidean distance loss shape-accordingly detects the LV center in echo apical views. The image resampling further crops the LV region of interest (LV-ROI) capable of efficiently reducing the interference of various structure from the different views. Accepting the LV-ROI, the LV indices module with the inter-frame gradient regularization and the views union effectively makes the comprehensive, accurate and internally consistent indices estimation.

The main contributions of our work include:

- For the first time, the proposed PV-LVNet enables an automatically and reliably comprehensive cardiac function clinical assessment from various dimensions and views by directly and accurately estimating LV multi-type indices on 2D echos of paired apical views.
- The newly designed Res-circle Net enables accurately and consistently estimating continuous changing centric positions and indices of LVs in echo sequence of each subject, by comprehensively combining both the subject-level base of cardiac cycle and the interrelated dynamic residual of each frame. Moreover, its residual transferring effectively reduces the gradient vanishing problem in recurrent net.
- The novel location loss in the form of anisotropic Euclidean distance (AED) guarantees robust and efficient location and cropping by matching the approximate bullet shape of LV in apical view echo.

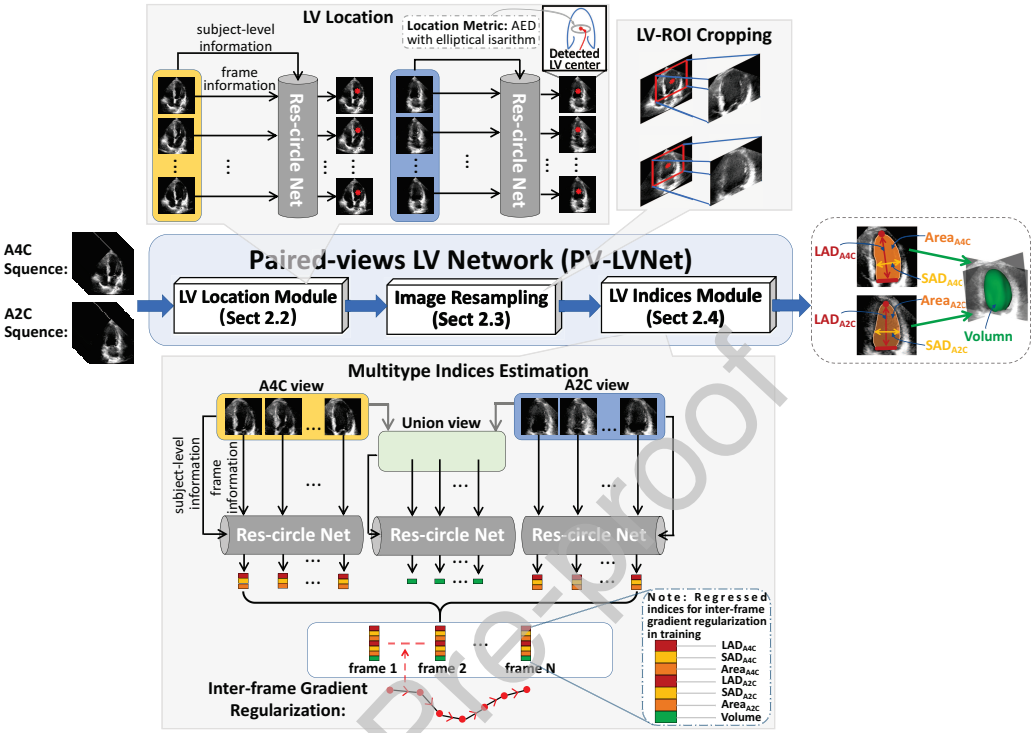


Figure 2: The PV-LVNet simultaneously estimates multitype indices of various single (A4C, A2C) and union views (A4C+A2C) from paired apical 2D echo sequences, to provide a comprehensive cardiac function assessment. Based on the Res-circle Net (Sect. 2.1), it has three interdependent parts: LV location module (Sect. 2.2) for LV location, image resamping (Sect. 2.3) for LV-ROI cropping and LV indices module (Sect. 2.4) for multitype indices estimation.

- The gradient of LV indices between adjacent frames in a cardiac cycle creatively and effectively enhances sequential indices fitting, by fully exploring inter-frame relatedness to introduce frame-by-frame evolution characteristic to regularize indices estimation.

2. Methodology

As shown in Fig.2, based on the **Res-circle Net (Sect. 2.1)** to analyze echo sequences, the PV-LVNet entirely works via three interdependent parts: **LV location module (Sect. 2.2)**, **Image Resampling** (Sect. 2.3) and **LV indices module (Sect. 2.4)** for location, cropping and indices estimation.

To enable the comprehensive and efficient echo sequence analysis, the novel Res-circle Net combines subject-level base for avoiding coarse sequential estimation from zero level and temporal dynamic residual for developing the refinement on each frame. To provide the robust LV location among views for accurate indices estimation, LV location module creatively adopts the loss in form of AED considering the LV shape in apical view echo. To automatically crop LV-ROI with the interference of various structure in paired views reduced, and build unblocked joint learning of location and indices regression, image resampling, as a differentiable transformation, is embedded. To achieve the various dimensional indices regression from single and union views, LV indices module performs not only indices-aware feature abstraction but also views union for 3D index. Moreover, to fulfill the inter-frame relatedness potential of indices for enhancing sequential data fitting, the inter-frames gradient in the time polyline of the cardiac index is used to deeply explore sequence evolution characteristics.

2.1. Res-circle Net for Analyzing Echo Sequence

The Res-circle Net combines both subject-level base and frame-level residuals for a comprehensive analysis on echo sequence. **Subject-level base** reflects the holistic characteristics among the different frames of the same subject. It gives a whole and inherent expression on the echo sequence and distinguishes different subjects. It is further extracted from the representations of all frames. **Frame-level residual** reflects interrelated temporal dynamic changes in the cardiac cycle. It enables a further refinement on each frame. It is extracted by using the inter-frame relationship among the whole cardiac cycle. The Res-circle Net captures interrelated temporal residual of each frame, then adds the residual with subject-level base together. It embeds subject and temporal information to guarantee a stable and dynamic estimation for location and indices in continuous moving and deforming LV. The net is implemented in the circle recurrent of a novel residual learning and transferring convolutional unit named as residual recurrent unit (RRU).

As shown in Fig.3, the Res-circle Net accepts current frame representation and links it to the integrated former residuals of the frames in the cycle, then adaptively updates the current frame-level residual and combine the residual with the subject-level base for a refined outputting. The Res-circle Net is achieved in the circle recurrent structure (Graves, 2012; Xue et al., 2017c) of RRU, which gives the memory characteristics of the cycle temporal

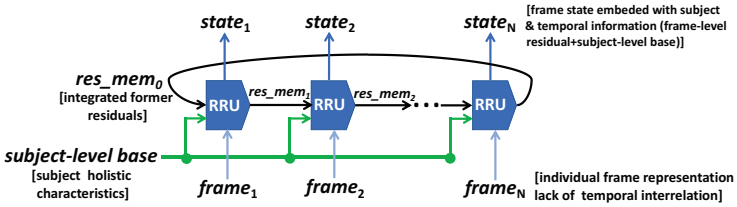


Figure 3: The Res-circle Net embeds both subject and interrelated temporal information together for comprehensive and reliable analysis on the echo sequence. It adaptively updates current dynamic change as residual by linking the current frame representation with the former memory in cycle, then adds such residual with the subject-level base together as the comprehensive state of the frame.

changes. Similar works to analyze data sequence can be seen in using LSTM of recurrent neural networks (RNN) as: Xu et al. (2018) adopt fully connected LSTM (FC-LSTM) for the dependence crossing over a long time interval; Xue et al. (2017c) further deployed circle FC-LSTM for shortening the distance between the first and last frames; and convolutional LSTM (Xingjian et al., 2015) was developed for the spatial structure in the sequence. Specially, our Res-circle Net of circle recurrent convolutional residual net is designed for temporally-spatially modeling the residuals among frames and the entirety of sequence, to the echo of dynamic and consecutive data.

The RRU has both functions of current frame state prediction and residuals memory integration, as shown in Fig.4. In the output path, the RRU provides the current state ($state_i$) for the followed regression, by adding current frame-level residual (res_i) on the subject-level base ($base$). In the hidden path, it transfers residual information (res_i) together with the formers (res_mem_{i-1}) to integrate residuals memory (res_mem_i) for the next frame. Instead of the frame-wise coarse estimation from the zero level, the net provides such a more refined way as the subject-level base reflects the stable base level of sequence and residual focuses on interrelated dynamic change of

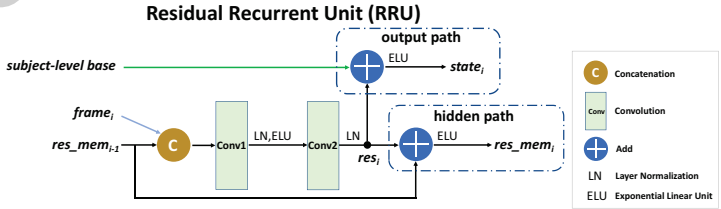


Figure 4: Residual recurrent unit (RRU) has both functions of current frame state prediction and residual transfer. In output path, the current frame-level residual is added to the subject-level base for followed regression. In hidden path, the residual information is transferred together with the formers for the next frame.

each frame. Benefited from the residual connection with subject-level base and former residuals, the net has powerful sequence analysis and temporal modeling, and meanwhile effectively reduces the gradient vanishing problem with the shortcut connection (Szegedy et al., 2017; He et al., 2016a,b).

The RRU takes both spatial structure and temporal information into account. It uses convolution process, instead of full connection in traditional RNN, to extract feature for keeping spatial correlation in the cardiac image. In recurrent way, it maps the current frame to the integrated residual memory to get its current frame-level residual. The inherent potential spatiotemporal characteristic in echo sequence is effectively mined and transmitted. Given the inputting individual frame representation $frame_i$ at each time step i, the memory res_mem_{i-1} from the previous frames, and the subject-level base $base$, RRU gets the current frame-level residual res_i for the updated memory res_mem_i and outputting state representation $state_i$, as:

$$\begin{aligned} res_i &= LN(ELU(LN((frame_i \oplus res\_mem_{i-1}) * W_1 + b_1)) * W_2 + b_2) \\ res\_mem_i &= ELU(res_i + res\_mem_{i-1}) \\ state_i &= ELU(base + res_i) \end{aligned} \tag{1}$$

where W_1 and W_2 are convolutional kernels in Conv1 and Conv2, b_1 and b_2 represents biases. $\oplus$ means concatenation, $*$ is convolution operation, and LN, ELU denote the element-wise transformations of layer normalization (Ba et al., 2016) and exponential linear unit (Clevert et al., 2015).

2.2. LV Location Module for Detecting Left Ventricle Center

LV location module aims to detect continuously moving LV center in both A4C and A2C sequences, as in Fig.5. It has four steps: 1) **CNN-loc** firstly extracts cardiac subject-level base and individual frame representations of the cardiac sequence and feeds them to the res-circle net; 2) **Res-circle Net** then models sequential LV moving in cardiac cycle for the final location, with subject's holistic position and frame's temporal changes embedded; 3) **Fully connected (FC) layer** further performs LV center coordinate regression with the output of Res-circle Net fed; And 4) **AED metric** is used to measure the regressed center with anisotropic scaling by considering approximate bullet shape of LV in echo apical views for robust location.

Advantageously, LV Location Module is benefited from the special design of **CNN-loc** and **AED location metric**, besides Res-circle Net that has been proposed in Sect. 2.1.

CNN-loc. To get expressive and task-aware representation of individual frame and entire subject on the paired echo sequences, CNN-loc consists

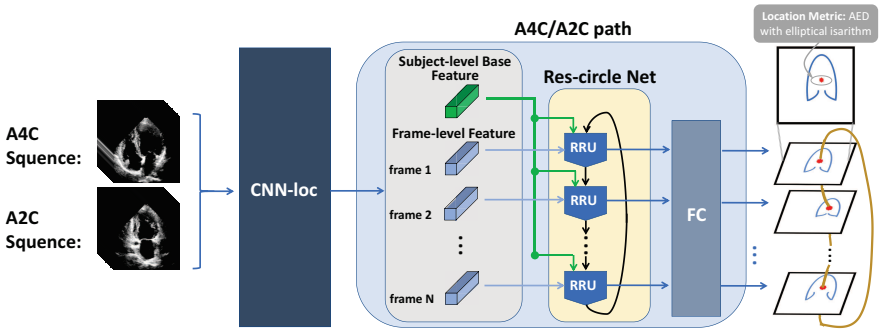


Figure 5: To achieve locating continuously changing center of LV in both A4C and A2C sequences, LV location module works via: 1) CNN-loc extracts subject-level base and frame representations for both paired views. 2) Res-circle Net captures residual information of each frame by leveraging inter-frame relationship for modeling dynamic changes, and further combine subject-level base to provide the frame state for location. 3) FC layer linearly regresses LV center coordinate. 4) The metric of anisotropic Euclidean distance (AED) ensures the robust location.

of several shared layers for general expression and two shallow paths that further refine on A4C and A2C respectively considering big view difference and enhancing robustness, as shown in Fig.6 (a). The individual cardiac distribution in each frame is extracted by the hierarchical convolutions, and the global sequence base of the subject is then captured by concatenating all these individual representations together with a further convolution operation followed so that holistically characterizes all frames. The backbone structure of CNN-lock is the stack of the successive convolutional blocks (He et al., 2016a) in Fig.6 (b), which chooses identity map for the layer input and output of the same size, or convolution of kernel size 1×1 to match dimensions. Such block promotes information propagation both forward and

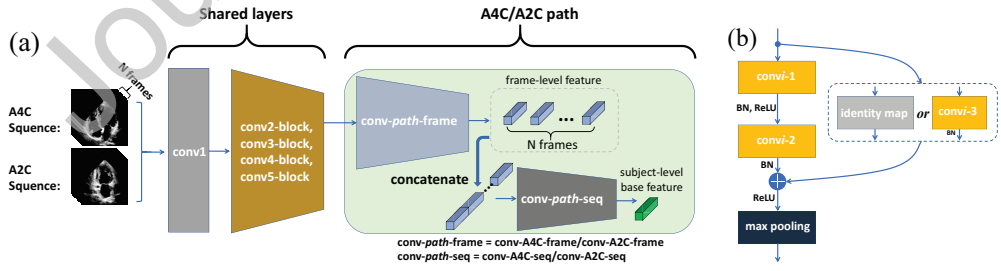
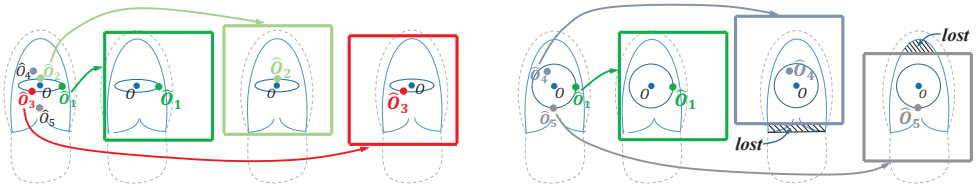


Figure 6: CNN-loc gets subject-level base and frame representation of paired echo sequences. (a) CNN-loc is composed of several shared layers and two shallow paths refined on A4C and A2C. (b) The stacked block in CNN-loc. The use of short-cut connection accelerates the net convergence and improve learning performance.



(a) AED has elliptical isarithm, enabling robust metric on LV location for ROI

(b) IED has circle isarithm, causing pool metric on LV location for ROI

Figure 7: The anisotropic Euclidean distance (AED) provides an elliptical isarithm to match the approximate bullet shape of LV in apical view echo and enable a more reasonable and robust LV location metric than the isotropic Euclidean distance (IED). (a) Considering LV shape, AED gives different scaling on the horizontal and vertical direction to construct elliptical isarithm for efficient LV location in apical view echo. (b) IED causes pool metric on LV location due to its circle isarithm.

backward and hence accelerate the net convergence and improve learning performance (Szegedy et al., 2017; Yu et al., 2017).The configurations of the stacked convolutions in CNN-loc are: $7 \times 7 \times 64$ with stride 2 for conv1, $channel = 64, 128, 128, 256, 256$ for convolutional blocks (conv{2,3,4,5}-block and conv-*path*-fram), and $3 \times 3 \times 256$ with stride 1 for conv-*path*-seq.

AED location metric. To achieve a structure matching location measurement, anisotropic Euclidean distance (AED) is deployed on the regressed center with the different metric scaling on horizontal and vertical directions, as shown in Fig.7 (a). Differently and traditionally, the location metric generally uses strict isotropic Euclidean distance (IED) in Eq. (2), where the regressed result $\hat{O} = (\hat{o}_x, \hat{o}_y)$ and the ground truth $O = (o_x, o_y)$.

$$distance_{IED} = \left\| O - \hat{O} \right\| \tag{2}$$

However, the shape of the LV is approximate to the bullet, so that the regressed points with same IED values still cause different influences to the ROI, and smaller IED does not mean a more accurate location. For example, $\hat{O}_1$ and $\hat{O}_5$ in Fig.7 (b) fall on the same circle isarithm of the IED to the LV center O, and $\hat{O}_4$ even has smaller IED than $\hat{O}_1$. But only the $\hat{O}_1$ centered square contains the entire LV cavity, while $\hat{O}_4$ and $\hat{O}_5$ lead to the weak ROIs.

In order to overcome the shortcoming in IED, AED using anisotropic scaling is a more reasonable metric that conforms to the LV shape. Comparing Figs.7 (a) with (b), $\hat{O}_4$ and $\hat{O}_5$ that have the same IED value as $\hat{O}_1$ or smaller than $\hat{O}_1$ are outside the elliptical isarithm of AED, which means getting higher AED metric. It aligns with their poor ROI quality in Fig.7 (b). Besides, the ROIs centered by the points $\hat{O}_2$ and $\hat{O}_3$ that fall on the ellipse in Fig.7 (a) have the same ROI situation as $\hat{O}_1$, that the entire

LV cavity is contained and close to the square border, and gains the same metric. Therefore, the AED introduces a more robust and effective location metric for LV. The AED calculation is given in Eq. (3).

$$distance_{AED} = \sqrt{\beta \cdot (\hat{o}_x - o_x)^2 + (1-\beta) \cdot (\hat{o}_y - o_y)^2} \tag{3}$$

2.3. Image Resampling for Cropping LV-ROI

Image resampling is implemented via spatial transform and bilinear interpolation to automatically crop LV-ROI according to the location from Sect. 2.2. Image resampling puts attention on determining the region most related to the LV. It aims to reduce the disturbance from the other pathology caused by various structure and extra chambers in different views, with the LV-ROI being cropped. Also, the LV-ROI sequence maintains the relative shapes of LVs among different frames to not destroy the inherent subject characteristics and frame-by-frame LV dynamic changes along the cardiac cycle for developing the sequential LV indices estimation of each subject. In a similar work, Dai et al. (2016) used ROI warping layer to crop feature map regions for refining further semantic segmentation. Additionally, Jaderberg et al. (2015) and Vigneault et al. (2018) used STN to spatially transform intermediate feature maps or inputting image for improving performance in classification and medical segmentation, respectively.

In our work, the image resampling transforms the images into the pattern that are centred on the predicted LV centre, and crops them to the predefined dimensions images. Given the predicted LV centre $\hat{O} = (\hat{o}_x, \hat{o}_y)$ and the source echo image I, the target LV-ROI image $I^{ROI}(\hat{O})$ is obtained by the image resampling as formulated as the differentiable linear transformation:

$$I^{ROI}(\hat{O}) = B(T(\hat{O})) \cdot I. \tag{4}$$

In Eq. (4), $T(\cdot)$ is the spatial transform that firstly translates the echo image I horizontally and vertically to be centred on $\hat{O}$ and then scales the translated image to crop a $153.6\,pixel \times 153.6\,pixel$ image (physical dimensions $79.49 \sim 115.80mm \times 79.49 \sim 115.80mm$ with pixel space $0.5175mm/pixel \sim 0.7539mm/pixel$) centred on the predicted LV centre. $B(\cdot)$ means bilinear interpolation further calculates the pixel value and produces the LV-ROI in a sufficiently fine resolution which is set as same as the original echo image, for the following indices estimation.

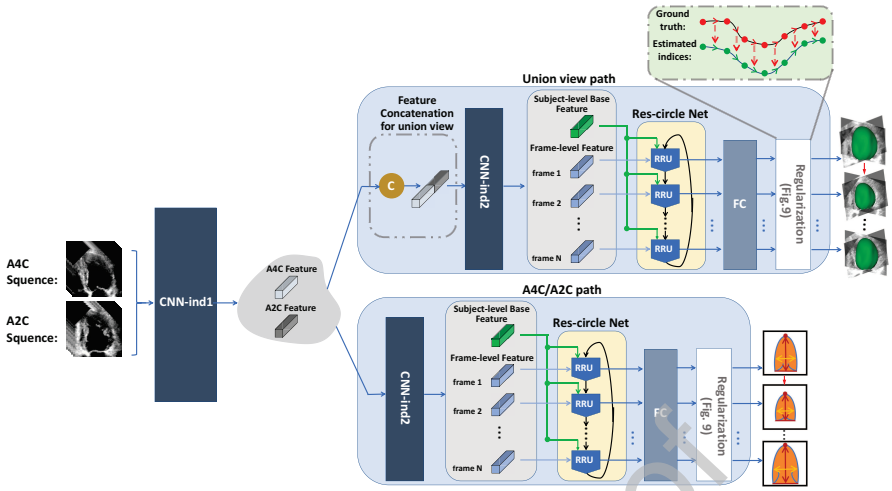


Figure 8: To estimate multitype indices from single/union views, LV indices module works via: 1) CNN-ind1+Feature Concatenation+CNN-ind2 gets feature representation on both entire subject and individual frame for all single and union views. 2) Res-circle Net models frame-by-frame dynamic residuals in the cardiac cycle by inter-frame relationship, then add them with the subject-level base of the holistic shape, for embedding subject and temporal information. 3) FC layer regresses indices with the outputs of the Res-circle Net. 4) Inter-frames gradient regularizes indices changes among frames to enhance sequential indices estimation.

2.4. LV Indices Module for Estimating Multitype Indices

LV indices module is designed to estimate multitype sequential cardiac indices in union and single views from continuously deformed LVs, as shown in Fig. 8. It consists of four components: 1) **CNN-ind1 + Feature Concatenation + CNN-ind2** makes frame and subject feature extraction, as well as union view representation. 2) **Res-circle Net** combines subject holistic shape and temporal deformation. 3) **FC layer** further regresses on the feature representation from Res-circle Net to estimate all indices. And 4) **Inter-frames Gradient** is meanwhile introduced to regularize the indices evolution among frames.

The superiority of LV indices module benefits from the special in **CNN-ind1 + Feature Concatenation + CNN-ind2** and **Inter-frames Gradient Regularization**, besides Res-circle Net demonstrate in Sect. 2.1.

CNN-ind1 + Feature Concatenation + CNN-ind2. In order to get both the frame and the subject features for all union and single views, it is further split and developed from CNN-loc that CNN-ind1 conducts the preliminary view-specialized representation on paired fed A4C and A2C ROI sequence, Feature Concatenation integrates the union view information via

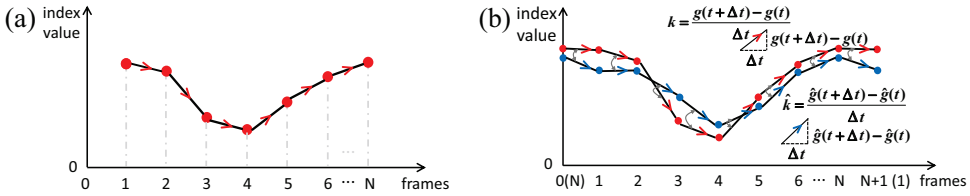


Figure 9: Inter-frames gradient regularization promotes sequential indices regression. (a) Frame-by-frame evolution of index is reflected by the polyline of index value vs. frame. (b) Inter-frames gradient regularizes frame-by-frame evolution of estimated results to strengthen sequential indices fitting. It reveals index changes among frames, and thus characterizes index evolution. Evolution is an important metric in measuring the similarity between two sequential data.

unifying the A4C and the A2C representations along the feature channel, and CNN-ind2 of the individual and the holistical features extraction is further performed on all the A4C, A2C and union view. In the procedure, the union view aims to construct the 3D spatial information from the two orthogonal views for the volume of 3D indices estimation and meanwhile further strengthening the contact among all views.

Inter-frames Gradient Regularization. For the accurate sequential indices estimation, the gradient inter frames is used with considering the evolution characteristics in the cardiac cycle. The frame-by-frame evolution of index in the cardiac cycle is shown in Fig. 9(a) with the polyline of index value vs. frame, it reflects the trend over time. And the gradient can be explored to depict these evolution characteristics of time polyline in 9(a), so that enhance the sequential indices fitting elegantly with the interrelated fluctuation regularized on the sequence of the preliminary regressed index. As shown in 9(b), it measures the slope of the secant passing through adjacent discrete points. Given the regressed result $\hat{y}^f$, and normalized the frame interval Δt as ± 1 for both adjacent frames, the inter-frames gradient $\hat{k}^f$ at each frame step is defined as:

$$\hat{k}^f = \left(\hat{k}^{f-}, \hat{k}^{f+}\right), \ and \begin{cases} \hat{k}^{f-} = \hat{y}^f - \hat{y}^{f-1} \\ \hat{k}^{f+} = -(\hat{y}^f - \hat{y}^{f+1}) \end{cases} \tag{5}$$

where $\hat{k}^{f-}$ and $\hat{k}^{f+}$ mean left and right gradient of frame f, respectively. $\hat{k}^f$ thus effectively characterizes index evolution of frame f between adjacent frames $f-1$ and $f+1$.

Therefore, the inter-frames gradient of each index among the cardiac cycle is introduced to fit the trend of polylines of regressed results and ground truth, as shown in Fig. 9(b), to enhance sequential LV indices estimation. Euclidean distance is used to calculate the gap of the change rate of each

frame between regressed results and ground truth. The fitness of sequential indices evolution is measured as Eq. (6). In addition to the fitting of each index value, such evolution of the index further gives full play to the constraint between adjacent frames, and can be used as a regularization item to strengthen sequential objects estimation.

$$Reg_{grad} = \sum_{f=1}^{N} \sum_{t} \left\| k_t^f - \hat{k}_t^f \right\|_2 \tag{6}$$

where $f \in \{1, 2, .., N\}$ for all frames in cardiac cycle, $t \in \{LAD_{A4C}, SAD_{A4C}, Area_{A4C}, LAD_{A2C}, SAD_{A2C}, Area_{A2C}, Volume\}$ for all index types.

3. Joint Loss Function for Different Tasks

The loss function in our work is designed for optimizing the two trainable modules (LV location module and indices module, while image resampling, as a powerful linear transformation, needs no training) of different tasks in the integrated PV-LVNet, so that the task-inter relevance and dependence enable the modules to mutually promote refinement of each other. The joint loss $\mathcal{L}_{joint}$ is constructed as:

$$\mathcal{L}_{joint} = \lambda_1 \mathcal{L}_{loc} + \lambda_2 \mathcal{L}_{ind} + \lambda_3 R(\theta) \tag{7}$$

where $\mathcal{L}_{loc}$ and $\mathcal{L}_{ind}$ are the loss functions of location and indices estimation, $R(\theta) = \|\theta\|_2^2$, known as Tikhonov regularization for improving the training generality, is used as the regularization item of the network parameter vector θ with l_2-norm. λ_1, λ_2 and λ_3 are set as 1000.0, 1.0 and 0.1.

The location loss function $\mathcal{L}_{loc}$ aims to guarantee a robust location of LV for LV-ROI cropping. It is constructed with AED for taking account of the approximate bullet shape of LV. The definition of $\mathcal{L}_{loc}$ is given by:

$$\mathcal{L}_{loc} = \frac{1}{N} \sum_{f=1}^{N} distance_{AED}^f \tag{8}$$

where $distance_{AED}^f$ denotes $distance_{AED}$ (defined in Eq. (3)) for the predicted centre in each frame f.

The indices loss function $\mathcal{L}_{ind}$ aims to boost high-quality indices regression. It utilizes not only the MAE of indices value estimation error in each frame but also the trend between indices of adjacent frames for both accuracy and inter consistency of the sequential indices estimation, as:

$$\mathcal{L}_{ind} = \frac{1}{N} \sum_{t} \sum_{f=1}^{N} \left| \hat{y}_t^f - y_t^f \right| + Reg_{grad} \tag{9}$$

where the first item is the MAE loss of indices, Reg_{grad} (defined as Eq.(6)) is the inter-frames gradient regularization item for indices evolution.

4. Experiment Configurations

Dataset. A dataset of 2D echos with the ground truth is used to evaluate our method, which includes 2000 echo images from 50 subjects collected from 2 hospitals. Each subject provides both paired A4C and A2C views echos, with the temporal resolution of 20 frames per cardiac cycle and the resize of 256×256. All ground truth of location and indices are manually annotated by two experienced cardiac radiologists with double-checking. In training, location labels are normalized to $[-1, 1] \times [-1, 1]$ through subtracting half of the image dimension (128) and then being divided by the image dimension (256). The labels of 1D (i.e., $LAD_{A4C}, SAD_{A4C}, LAD_{A2C}$ and SAD_{A2C}), 2D (i.e., $Area_{A4C}$ and $Area_{A2C}$) and 3D (i.e., $Volume$) metrics are normalized by LV-ROI dimension ($\frac{256}{p}$, where $\frac{1}{p} = 0.6$ is set according to prior investigation on our dataset), area ($(\frac{256}{p})^2$) and volume ($(\frac{256}{p})^3$), respectively.

Data Augmentation. To avoid the over-fitting and improve the generalization, we augment the dataset to 8000 images by three strategies as: 1) randomly rotating between $-15°$ and $15°$; 2) randomly zooming between 0.9 and 1.1 times; and 3) the combination of random rotation + zoom.

Configurations. The net is implemented by Tensorflow, and performed on NVIDIA P100 GPU. Ten-fold cross validation is employed for performance evaluation and comparison.

Evaluation Metrics. We evaluate the performance of the PV-LVNet in terms of estimation accuracy and internal consistency for multitype indices of all frames in the cardiac cycle. The evaluation is performed with two metrics including: the mean absolute error (MAE) for measuring accuracy and Cronbach's α (Cronbach, 1951) for measuring internal consistency between the estimated results and the corresponding ground truth. Denote the estimated cardiac index and ground truth of the ith subject and the fth frame as $\hat{y}_{t,i}^f$ and $y_{t,i}^f$, where $t \in \{LAD_{A4C}, SAD_{A4C}, Area_{A4C}, LAD_{A2C}, SAD_{A2C}, Area_{A2C}, Volume\}$ for index types. The **MAE** of each cardiac index is given by $MAE_t = \frac{1}{S \times N} \sum_{i=1}^{S} \sum_{f=1}^{N} \left| \hat{y}_{t,i}^f - y_{t,i}^f \right|$, where S and F are the number of subjects and frames, respectively. **Cronbach's** α of each cardiac index is calculated as $\alpha_t = 2 \cdot (1 - \frac{\sigma_{\hat{\mathcal{Y}}_t}^2 + \sigma_{\mathcal{Y}_t}^2}{\sigma_{\mathcal{X}_t}^2})$, where $\mathcal{X}_t$ is the sum of estimated indices $\hat{\mathcal{Y}}_t = \{\hat{y}_{t,1}^1, \hat{y}_{t,1}^2, \hat{y}_{t,1}^3, ..., \hat{y}_{t,S}^N\}$ and corresponding ground

truth $\mathcal{Y}_t = \{y^1_{t,1}, y^2_{t,1}, y^3_{t,1}, ..., y^N_{t,S}\}$, i.e., $\mathcal{X}_t = \hat{\mathcal{Y}}_t + \mathcal{Y}_t$. Moreover, $\sigma^2_{\hat{\mathcal{Y}}_t}$, $\sigma^2_{\mathcal{Y}_t}$ and $\sigma^2_{\mathcal{X}_t}$ are the corresponding variances for $\hat{\mathcal{Y}}_t$, $\mathcal{Y}_t$ and $\mathcal{X}_t$.

5. Results and Analysis

We conduct a set of experiments to evaluate the performance of our proposed PV-LVNet, including: 1) overall performance; 2) effectiveness of res-circle net; 3) effectiveness of anisotropic Euclidean distance location loss; 4) effectiveness of inter-frames gradient regularization; 5) performance comparison with relevant methods; 6) performance of activation function and Hyper parameter selection.

5.1. Overall Performance

As shown in the last column of Table 1, the proposed PV-LVNet achieves excellent estimation accuracy and internal consistency on all the 7 different indices, which are attributable to comprehensively analyzing sequential echos, robustly locating and cropping LV, deeply exploiting inter-frame indices relatedness. It gains extremely low MAE of $2.85mm$, $3.16mm$, $3.06mm$, $2.98mm$, $287mm^2$, $264mm^2$ and $10.7ml$ for LAD_{A2C}, SAD_{A2C}, LAD_{A4C}, SAD_{A4C}, $Area_{A2C}$, $Area_{A4C}$ and LV *volume*, as well as high Cronbach's α all exceeding 0.9, with the manually obtained ground truth.

Moreover, our proposed PV-LVNet also achieves high coincide indices estimation along the cardiac cycle, indicating powerfully modeling the LV activity. As shown in Fig. 10, it reaches extremely low normalized root-mean-square error of 1.26% (NRMSE, $NRMSE = \frac{1}{\bar{y}}\sum_{f=1}^{N}\frac{(\hat{y}^f - y^f)^2}{N}$) with the sequential ground truth, on average. Such rare few deviations strongly validate that the network effectively captures the activity pattern of sequential LVs.

Our method is also very efficient in running time. The training takes 16.36 hours with one P100 GPU. The testing takes only 0.70 seconds per subject. Clearly, our method enables a real-time solution for clinical application.

5.2. Effectiveness of Res-circle Net

As shown in Table 2, the Res-circle Net decreases the MAE by 15.7% (e.g., $15.7\% = \frac{1}{7}[\frac{3.46-2.85}{3.46} + \frac{3.64-3.16}{3.64} + \frac{3.37-3.06}{3.37} + \frac{3.24-2.98}{3.24} + \frac{336-287}{336} + \frac{321-264}{321} + \frac{15.1-10.7}{15.1}]$) and gains exceeding 0.9 Cronbach's α on all indices, compared to the the situation of being replaced by CNN in the PV-LVNet for revealing its effectiveness. By combining subject-level holistic characteristics and

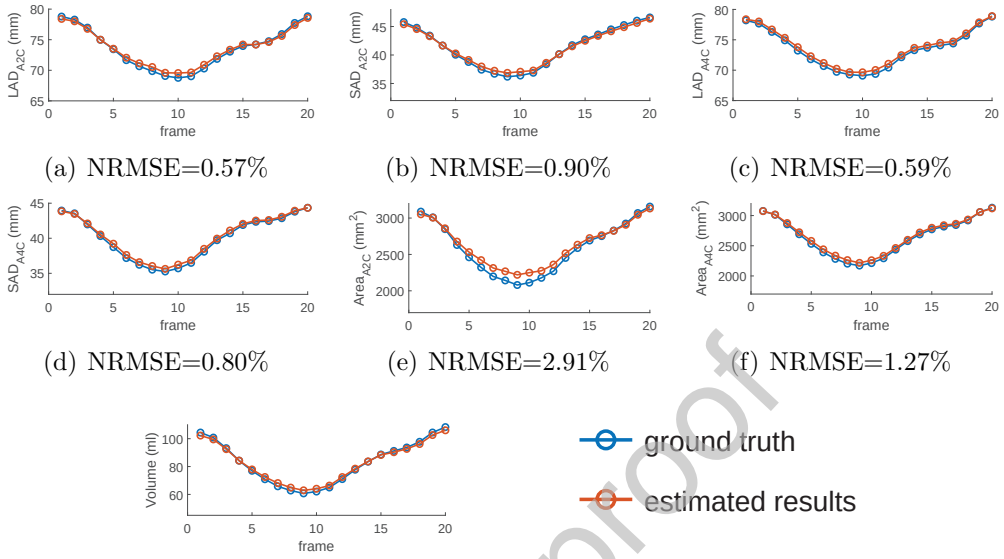
Table 1: The proposed method gains most advanced performance in the various dimensional metrics for LV of all views compared to the existing methods. It achieves higher accuracy and more excellent internal consistency, with lower MAE (18.9% ↓) and higher Cronbach's α (> 0.9) for each LV index. MAE and α are shown in each cell.

	Multi-features+RF	SDL+AKRF	MCDBN+RF	Indices-Net	U-Net	**PV-LVNet**
One-dimensional Metric (mm)						
LAD_{A2C}	3.52±3.10 0.895	3.29±2.48 0.913	3.44±3.18 0.898	3.19±2.43 0.923	/	**2.85±2.46** **0.941**
SAD_{A2C}	3.76±3.02 0.890	4.51±3.34 0.866	3.81±3.13 0.895	3.60±2.82 0.910	/	**3.16±2.68** **0.930**
LAD_{A4C}	3.86±3.48 0.864	3.73±3.05 0.904	3.93±3.38 0.863	3.29±2.42 0.896	/	**3.06±2.73** **0.932**
SAD_{A4C}	3.23+2.91 0.901	3.21±2.82 0.907	3.18±3.00 0.903	4.27±3.37 0.887	/	**2.98±2.85** **0.917**
Two-dimensional Metric (mm^2)						
$Area_{A2C}$	331±259 0.870	321±274 0.884	320±264 0.885	361±431 0.876	393±338 0.887	**287±284** **0.907**
$Area_{A4C}$	323±266 0.902	280±236 0.934	312±255 0.915	354±338 0.885	392±305 0.901	**264±228** **0.940**
Three-dimensional Metric (ml)						
$Volume$	16.1±14.2 0.918	16.4±14.6 0.922	16.1±14.0 0.925	15.3±8.7 0.938	/	**10.7±7.6** **0.974**

interrelated temporal changes existing in echo sequence, the Res-circle Net outperforms CNN which just performs independent processing for each frame, on accuracy and internal consistency. Adding subject-level base and interrelated dynamic residual of each frame together, the res-circle net enables and enhances refined sequential indices estimation by leveraging inter-frame temporal relationship and avoiding coarse estimation on each separate frame from zero level to improve accuracy. Moreover, introducing subject-level and temporal characteristics, the Res-circle Net guarantees excellent internal consistent estimation across subjects and among frames with the ground truth.

5.3. Effectiveness of AED location loss

As shown in Table 3, the AED location loss ensures developing accurate indices estimation. Compared with using IED in location, the AED location loss significantly decreases the MAEs by 21.3%, 11.0%, 13.8% and 30.5% on LAD, SAD, area and volume on average. These improvements are resulted from the fact that IED location loss effectively provides a robust and efficient location and cropping for indices estimation. It suits LV in apical view echo by adopting different scaled metrics on different directions to match



(a) NRMSE=0.57% (b) NRMSE=0.90% (c) NRMSE=0.59%

(d) NRMSE=0.80% (e) NRMSE=2.91% (f) NRMSE=1.27%

(g) NRMSE=1.85%

Figure 10: The proposed PV-LVNet effectively achieves high coincide indices estimation along the cardiac cycle to model the LV activity. The polygonal lines reflect the frame-wise value of each index for average subject. The normalized root mean square error (NRMSE) is used to measure the deviation between the polygonal lines of the estimated value and ground truth. As the results show, the network gains low NRMSE of 1.26% on average, with rarely few deviations.

the approximate bullet shape that is more strict on locations in the vertical direction than the horizontal direction, while the general IED loss can only provide a low-quality metric of no direction difference. Thus, LAD, area and volume which are extremely sensitive to vertical direction location get the highest improvements. Additionally, the SAD which is the most difficult to be estimated due to its non-independent measurement and a certain degree dependence on LAD still gets an obvious improvement of 11.0% with more accurate LAD.

5.4. Effectiveness of Inter-frames Gradient Regularization

As shown in Table 4, the inter-frames gradient regularization is capable of increasing the internal consistency of the estimated results with the ground truth. It gains higher Cronbach's α exceeding 0.9 on all indices and increased from 0.914 to 0.934 on average. By measuring the index change rate between adjacent frames, the inter-frames gradient is used to fit indices frame-by-frame evolution in sequence. So that the estimated sequential indices are

Table 2: The Res-circle Net contributes to high estimation accuracy and excellent internal consistency. It obtains lower MAE (15.7% ↓) and higher Cronbach's α (> 0.9) than being replaced by CNN.

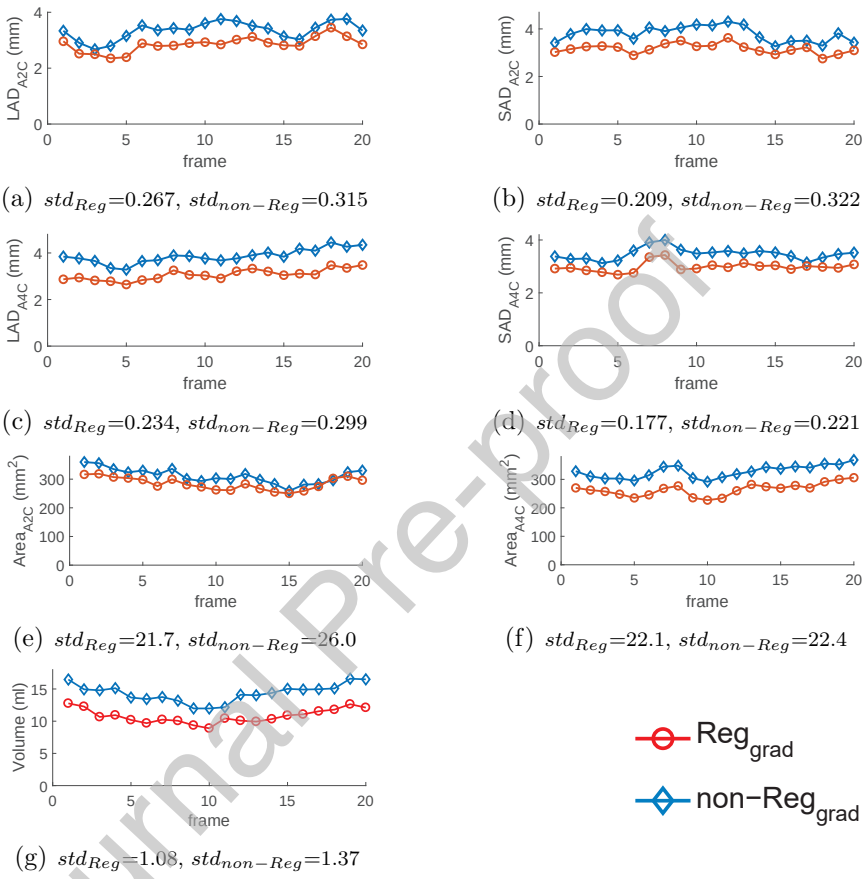
	CNN	**Res-circle Net**
One-dimensional Metric (mm)		
LAD_{A2C}	3.46±2.87	**2.85±2.46**
	0.915	**0.941**
SAD_{A2C}	3.64±2.86	**3.16±2.18**
	0.913	**0.930**
LAD_{A4C}	3.37±2.66	**3.06±2.73**
	0.893	**0.932**
SAD_{A4C}	3.24±2.65	**2.98±2.85**
	0.888	**0.917**
Two-dimensional Metric (mm^2)		
$Area_{A2C}$	336±279	**287±284**
	0.885	**0.907**
$Area_{A4C}$	321±289	**264±228**
	0.908	**0.940**
Three-dimensional Metric (ml)		
$Volume$	15.1±11.8	**10.7±7.6**
	0.935	**0.974**

regularized to get consistent variation with the ground truth.

Besides, the inter-frames gradient regularization also enhances sequential data fitting to ensure stable and accurate estimation across the whole cardiac cycle, as shown in Fig. 11. It not only gains consistently lower estimation error, but also increases the stability by 18.7% on average. The inter-frames gradient regularization mines indices inter-frame relatedness to learn the fluctuation across the cardiac cycle. Such fluctuation explicitly explores the constraints among indices of different frames to promote stable and accurate estimation and reduce pulse estimation error for sequential indices.

5.5. Performance Comparison with Relevant Methods

Our PV-LVNet achieves the most advanced performance in the various dimensional metrics for the LV of all views compared to the existing methods: 1) the two-phase direct estimation including Multi-features+RF (Zhen et al., 2014b), SDL+AKRF (Zhen et al., 2015a), MCDBN+RF (Zhen et al., 2016); 2) the end-to-end direct estimation, i,e, Indices-Net (Xue et al., 2017a); 3) the indirect estimation with segmentation U-net (Ronneberger et al., 2015). As shown in Table 1, our method significantly decreases the MAE by 18.9% on average on all indices, compared to these methods. Besides, it simultaneously maintains excellent internal consistency with the manually obtained ground truth by high Cronbach's α all exceeding 0.9.



(a) std_{Reg}=0.267, $std_{non-Reg}$=0.315

(b) std_{Reg}=0.209, $std_{non-Reg}$=0.322

(c) std_{Reg}=0.234, $std_{non-Reg}$=0.299

(d) std_{Reg}=0.177, $std_{non-Reg}$=0.221

(e) std_{Reg}=21.7, $std_{non-Reg}$=26.0

(f) std_{Reg}=22.1, $std_{non-Reg}$=22.4

(g) std_{Reg}=1.08, $std_{non-Reg}$=1.37

Figure 11: The inter-frames gradient well regularizes the network to enhance sequential data fitting. The polygonal lines record the frame-wise average MAE of each index. The standard deviation (*std*) is used to reflect the dispersion of MAE polygonal lines across a whole cardiac cycle. As the results show, using the inter-frames gradient regularization for the sequential indices decreased *std* by 18.7% compared to be removed, on average. It means stable and robust estimation on each frame. Also, the polygonal lines show consistently lower estimation error with inter-frames gradient regularization.

Table 3: The AED location loss ensures developing accurate estimation for LV indices. It brings higher estimation accuracy than the IED location loss, with lower MAE (17.3% ↓) on each type of cardiac indices.

	IED location loss	**AED location loss**
Long-axis Dimension (mm)		
LAD_{A2C}	3.89±2.89	**2.85±2.46**
LAD_{A4C}	3.62±2.38	**3.06±2.73**
$Averge$	3.76±2.64	**2.96±2.60**
Short-axis Dimension (mm)		
SAD_{A2C}	3.48±2.84	**3.16±2.68**
SAD_{A4C}	3.41±2.84	**2.98±2.85**
$Average$	3.45±2.84	**3.07±2.77**
Area (mm^2)		
$Area_{A2C}$	322±255	**287±284**
$Area_{A4C}$	314±224	**264±228**
$Average$	318±240	**274±259**
Volume (ml)		
$Volume$	15.4±15.6	**10.7±7.6**

In detail, our method is superior to the relevant methods as:

1) The proposed PV-LVNet outperforms the two-phase direct method, with the average MAE decreased by 16.2%, 12.3% and 34.0% on 1D, 2D and 3D metrics, respectively. Different from these compared methods, the proposed method jointly learns the deep task-aware information and regresses target in an end-to-end way, instead of the split handcrafted feature extraction and regression. It is obviously validated on the volume estimation. The proposed method conducts a deeper learning on the concatenated feature jointly with volume estimation, and gets 34.0% improvement.

2) The proposed PV-LVNet outperforms the existing end-to-end direct method, with the average MAE decreased by 19.4% and all Cronbach's α increased to above 0.9. All of these are own to the fact that the proposed method effectively introduces the subject holistic characteristics and temporal changes for developing an accurate, stable and consistent estimation in a coarse-to-refine way, and deeply explores inter-frame indices relatedness for enhancing sequential indices estimation. However, the compared method just conducts separate estimation on each image.

3) The proposed PV-LVNet outperforms the segmentation method, with

Table 4: The inter-frames gradient regularization increases internal consistency with the ground truth. It gains higher Cronbach's α (> 0.9) than being removed.

	non-Reg_{grad}	$\boldsymbol{Reg_{grad}}$
One-dimensional Metric		
LAD_{A2C}	0.926	**0.941**
SAD_{A2C}	0.904	**0.930**
LAD_{A4C}	0.904	**0.932**
SAD_{A4C}	0.902	**0.917**
Two-dimensional Metric		
$Area_{A2C}$	0.897	**0.907**
$Area_{A4C}$	0.918	**0.940**
Volume		
$Volume$	0.945	**0.974**

estimating 5 more indices. It efficiently explores holistic characteristics and interrelated changes among the different frames in the same subject to directly analyze LV sequence and LV biological structure for adaptively learning all cardiac indices. And U-net (Ronneberger et al., 2015) just automatically provides LV area from its segmentation while the other indices need extra interaction from the expert for apex and mitral valve plane.

In the implementation of comparison, our proposed method needs no extra interaction. Our method is fed with entire echo image and does not require post-processing, benefited from its robust processing ability. But the other direct methods need to be performed on the the pre-handcrafted region to work (Zhen et al., 2014b, 2015a, 2016; Xue et al., 2017a). The segmentation method U-net is post processed as general with maximum connected region extraction to improve its segmentation results for indices estimation.

5.6. Performance of Activation Function and Hyper Parameters Selection

Activation Function ELU vs. ReLU. As shown in Figure. 12, ELU better fits the RRU than ReLU, with the lower estimation MAE (sum of normalized multitype indices MAE). Since the activation of ELU is able to transmit not only positive value message but also negative value message among frames, which is important for stimulating the inter-frames communication. But ReLU misses the information during the negative regime because of all being forcefully pushed to zero.

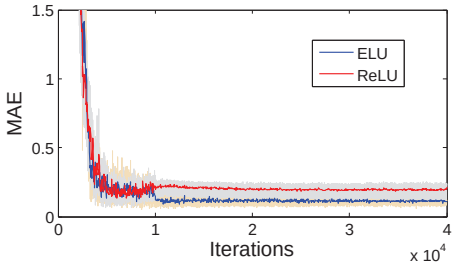


Figure 12: ELU activation outperforms ReLU in the RRU with lower testing MAE and better fitting.

Hyper Parameters Setting. As shown in Figure. 13, our Hyper parameters of $\lambda_1 = 1000$, $\lambda_2 = 1$ and $\lambda_3 = 0.1$ gain the best estimation accuracy compared to the other settings, with the lowest estimation MAE. Defaulting λ_2 for indices estimation as 1, λ_1 gets the large magnitude of 1000 for the trade-off between the trainable tasks location and indices estimation to mutually promote them; λ_3 with the small magnitude of 0.1 balances tasks training and network parameters regularization. Figure 13(a) indicates that the large λ_1 is more effective than small setting as larger ones have lower rate of accuracy decay. Specifically, λ_1 setting smaller than 1000 extremely increases the estimation error. Since the unsuitable small λ_1 decreases the location supervision of LV-ROI, which leads the indices estimation in a mess. The messed indices estimation then arbitrarily misleads the location through the joint learning and further degrades the indices accuracy in return via the chain reaction. Big is better, but not infinite. The too huge magnitude of λ_1 exceeding 1000 also has the risk of decreasing the performance. Because the too huge λ_1 weakens the effect of indices estimation in the mutual promotion, so that make the indices accuracy lower. In Figure 13(b), our choice also gets the best result. The big λ_3, as 1 and 10, have the serious problem of making the learning target unclear, so that influence the learning ability. Small λ_3 keeps the learning target clear, but the too tiny λ_3 of 0.01 and 0.001 weakens the regularization on network parameters so that reduces the generalization of the network and worsens the practical estimation.

6. Conclusions

In this paper, we proposed the PV-LVNet for the first time achieve the direct and accurate estimation of LV multitype indices (LAD_{A2C}, SAD_{A2C}, $Area_{A2C}$, LAD_{A4C}, SAD_{A4C}, $Area_{A4C}$, $Volume$) from 2D echos of paired apical views. The PV-LVNet conducts the sufficient metrics from various dimensions (1D, 2D & 3D) and views (A2C, A4C, and union of A2C+A4C)

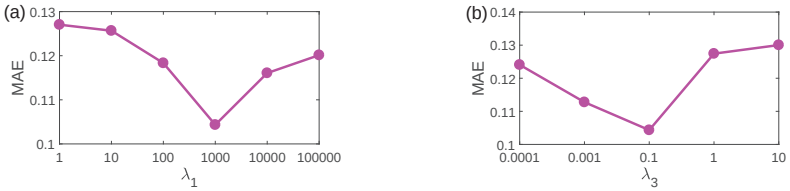


Figure 13: Our hyper-parameters setting gets the best estimation accuracy. (a) Influence of λ_1 selection. (b) Influence of λ_3 selection.

to provide a reliable comprehensive cardiac function assessment. It is built based on the Res-circle Net for sequential analysis. The Res-circle Net embeds both subject holistic characteristics and temporal changes by combining common subject-level base among frames and interrelated residuals of each frame, so that accurate and consistent location and indices estimation of LVs in echo sequence are enabled. The PV-LVNet is integrated of three interdependent parts for location, cropping and indices regression, as: 1) the LV location module utilizes AED that gives different scaled metrics on different directions as the loss to suit approximate bullet shape of LV in apical echos, so that robust and efficient location for indices estimation is ensured; 2) the Image Resampling automatically crops LV-ROI from the entire echo image, so that the interference of various structures in paired views is reduced; 3) by using inter-frames gradient regularization for exploring indices inter-frame relatedness, the LV location module fits not only each index value but also the indices evolution, so that sequential indices estimation is further enhanced. The PV-LVNet reaches high accuracy on all indices estimation and maintains excellent internal consistency with the ground truth, indicating its great potential in clinical cardiac function evaluation.

Acknowledgment

This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX17_0104); the China Scholarship Council (No. 201706090248); the States Key Project of Research and Development Plan (No. 2017YFA0104302, 2017YFC0109202 and 2017YFC0107900); the National Natural Science Foundation (No. 81530060 and 61871117); and the Science and Technology Program of Guangdong (No. 2018B030333001).

References

Abdi, A.H., Luong, C., Tsang, T., Allan, G., Nouranian, S., Jue, J., Hawley, D., et al., 2017. Automatic quality assessment of echocardiograms using

convolutional neural networks: Feasibility on the apical four-chamber view. IEEE Transactions on Medical Imaging 36, 1221–1230.

Afshin, M., Ayed, I.B., Islam, A., et al., 2012. Global assessment of cardiac function using image statistics in mri, in: MICCAI, Springer. pp. 535–543.

Afshin, M., Ayed, I.B., Punithakumar, K., Law, M.W., et al., 2014. Regional assessment of cardiac left ventricular myocardial function via mri statistical features. IEEE Transactions on Medical Imaging 33, 481–494.

Ba, J., Kiros, J., Hinton, G., 2016. Layer normalization. arXiv preprint arXiv:1607.06450 .

Carneiro, G., Nascimento, J.C., Freitas, A., 2012. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Transactions on Image Processing 21, 968–982.

Chen, H., Zheng, Y., Park, J.H., Heng, P.A., et al., 2016. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images, in: MICCAI, Springer. pp. 487–495.

Ciampi, Q., Villari, B., 2007. Role of echocardiography in diagnosis and risk stratification in heart failure with left ventricular systolic dysfunction. Cardiovascular Ultrasound 5, 1–12.

Clevert, D.A., et al., 2015. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289 .

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. psychometrika 16, 297–334.

Dai, J., He, K., Sun, J., 2016. Instance-aware semantic segmentation via multi-task network cascades, in: CVPR, pp. 3150–3158.

Debreuve, E., Barlaud, M., Aubert, G., Laurette, I., et al., 2001. Space-time segmentation using level set active contours applied to myocardial gated spect. IEEE Transactions on Medical Imaging 20, 643–659.

Gao, Z., Li, Y., Sun, Y., Yang, J., Xiong, H., et al., 2018. Motion tracking of the carotid artery wall from ultrasound image sequences: a nonlinear state-space approach. IEEE Transactions on Medical Imaging 37, 273–283.

Gao, Z., Xiong, H., Liu, X., Zhang, H., Ghista, D., Wu, W., Li, S., 2017. Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. Medical Image Analysis 37, 1–21.

Georgescu, B., Zhou, X.S., et al., 2005. Database-guided segmentation of anatomical structures with complex appearance, in: CVPR, pp. 429–436.

Graves, A., 2012. Supervised sequence labelling, in: Supervised sequence labelling with recurrent neural networks. Springer, pp. 5–13.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition, in: CVPR, pp. 770–778.

He, K., Zhang, X., et al., 2016b. Identity mappings in deep residual networks, in: European conference on computer vision, Springer. pp. 630–645.

Jacob, G., Noble, J.A., Behrenbruch, C., Kelion, A.D., Banning, A.P., 2002. A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. IEEE Transactions on Medical Imaging 21, 226–238.

Jaderberg, M., Simonyan, K., et al., 2015. Spatial transformer networks, in: Advances in neural information processing systems, pp. 2017–2025.

Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., et al., 2015. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. Journal of the American Society of Echocardiography 28, 1–39.

Lang, R.M., Bierig, M., Devereux, R.B., Flachskampf, F.A., Foster, E., Pellikka, P.A., Picard, M.H., et al., 2006. Recommendations for chamber quantification. European Journal of Echocardiography 7, 79–108.

Lathuilière, S., Juge, R., et al., 2017. Deep mixture of linear inverse regressions applied to head-pose estimation, in: CVPR, pp. 4817–4828.

Luo, G., Dong, S., Wang, K., Zuo, W., Cao, S., Zhang, H., 2018. Multi-views fusion cnn for left ventricular volumes estimation on cardiac mr images. IEEE Transactions on Biomedical Engineering 65, 1924 – 1934.

Malladi, R., Sethian, J.A., Vemuri, B.C., 1995. Shape modeling with front propagation: A level set approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 17, 158–175.

Mo, Y., Liu, F., McIlwraith, D., Yang, G., Zhang, J., He, T., Guo, Y., 2018. The deep poincaré map: A novel approach for left ventricle segmentation, in: MICCAI, Springer. pp. 561–568.

Nascimento, J.C., et al., 2008. Robust shape tracking with multiple models in ultrasound images. IEEE Transactions on Image Processing 17, 392–406.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., de Marvao, A., et al., 2018. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. IEEE Transactions on Medical Imaging 37, 384–395.

Paragios, N., 2003. A level set approach for shape-driven segmentation and tracking of the left ventricle. IEEE Transactions on Medical Imaging 22, 773–776.

Pascual, M., Pascual, D., Soria, F., Vicente, T., Hernandez, A., Tebar, F., Valdes, M., 2003. Effects of isolated obesity on systolic and diastolic left ventricular function. Heart 89, 1152–1156.

Peng, P., Lekadir, K., Gooya, A., et al., 2016. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics 29, 155–195.

Pereira, S., et al., 2018. Enhancing interpretability of automatically extracted machine learning features: application to a rbm-random forest system on brain lesion segmentation. Medical image analysis 44, 228–244.

Ravı, D., Wong, C., Deligianni, F., et al., 2017. Deep learning for health informatics. IEEE Journal of Biomedical and Health Informatics 21, 4–21.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: MICCAI, Springer. pp. 234–241.

Schiller, N.B., Shah, P.M., Crawford, M., DeMaria, A., Devereux, R., Feigenbaum, H., Gutgesell, H., Reichek, N., et al., 1989. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. Journal of the American Society of Echocardiography 2, 358–367.

Szegedy, C., Ioffe, S., et al., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, in: AAAI, pp. 4278–4284.

Vigneault, D.M., Xie, W., Ho, C.Y., et al., 2018. ω-net (omega-net): Fully automatic, multi-view cardiac mr detection, orientation, and segmentation with deep neural networks. Medical Image Analysis 48, 95–106.

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

Wu, L., Cheng, J.Z., Li, S., Lei, B., Wang, T., Ni, D., 2017. Fuiqa: Fetal ultrasound image quality assessment with deep convolutional networks. IEEE Transactions on Cybernetics 47, 1336–1349.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., , et al., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: Advances in neural information processing systems, pp. 802–810.

Xu, C., Xu, L., Gao, Z., Zhao, S., Zhang, H., Zhang, Y., et al., 2018. Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. Medical image analysis 50, 82–94.

Xue, W., Brahm, G., et al., 2018. Full left ventricle quantification via deep multitask relationships learning. Medical Image Analysis 43, 54–65.

Xue, W., Islam, A., Bhaduri, M., Li, S., 2017a. Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Transactions on Medical Imaging 36, 2057–2067.

Xue, W., Lum, A., Mercado, A., Landis, M., et al., 2017b. Full quantification of left ventricle via deep multitask learning network respecting intra-and inter-task relatedness, in: MICCAI, Springer. pp. 276–284.

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

Yu, L., Yang, X., Chen, H., Qin, J., Heng, P.A., 2017. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images, in: AAAI, pp. 66–72.

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

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

Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014a. A comparative study of methods for cardiac ventricular volume estimation, in: Annual Meeting-Radiological Society of North America (RSNA), pp. 228–244.

Zhen, X., Wang, Z., Yu, M., Li, S., 2015b. Supervised descriptor learning for multi-output regression, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1211–1218.

Zhen, X., Wang, Z., et al., 2014b. Direct estimation of cardiac bi-ventricular volumes with regression forests, in: MICCAI, Springer. pp. 586–593.

Zhen, X., Yu, M., He, X., Li, S., 2017. Multi-target regression via robust low-rank learning. IEEE transactions on pattern analysis and machine intelligence 40, 497–504.

Conflicts of interest: none