Direct Estimation of Regional Wall Thicknesses via Residual Recurrent Neural Network

Wufeng Xue, Ilanit Ben Nachum, Sachin Pandey, James Warrington,
Stephanie Leung, and Shuo Li[(✉)]

Department of Medical Imaging, Western University, London, ON, Canada
xwolfs@hotmail.com, slishuo@gmail.com

Abstract. Accurate estimation of regional wall thicknesses (RWT) of left ventricular (LV) myocardium from cardiac MR sequences is of significant importance for identification and diagnosis of cardiac disease. Existing RWT estimation still relies on segmentation of LV myocardium, which requires strong prior information and user interaction. No work has been devoted into direct estimation of RWT from cardiac MR images due to the diverse shapes and structures for various subjects and cardiac diseases, as well as the complex regional deformation of LV myocardium during the systole and diastole phases of the cardiac cycle. In this paper, we present a newly proposed Residual Recurrent Neural Network (ResRNN) that fully leverages the spatial and temporal dynamics of LV myocardium to achieve accurate frame-wise RWT estimation. Our ResRNN comprises two paths: (1) a feed forward convolution neural network (CNN) for effective and robust CNN embedding learning of various cardiac images and preliminary estimation of RWT from each frame itself independently, and (2) a recurrent neural network (RNN) for further improving the estimation by modeling spatial and temporal dynamics of LV myocardium. For the RNN path, we design for cardiac sequences a Circle-RNN to eliminate the effect of null hidden input for the first time-step. Our ResRNN is capable of obtaining accurate estimation of cardiac RWT with Mean Absolute Error of 1.44 mm (less than 1-pixel error) when validated on cardiac MR sequences of 145 subjects, evidencing its great potential in clinical cardiac function assessment.

Keywords: Regional wall thickness · Residual recurrent neural network · Spatial dependency · Temporal dependency · Circle-RNN

1 Introduction

Estimation of regional wall thicknesses (RWT) of left ventricle (LV) myocardium is of significant importance for early identification and diagnosis of cardiac disease [8,12,15]. Figure 1 demonstrates the RWT to be estimated for a short-axis view cardiac image. A traditional way for RWT estimation is to segment the LV myocardium from other structures first and then measure the corresponding RWT of each region. However, existing segmentation methods for cardiac
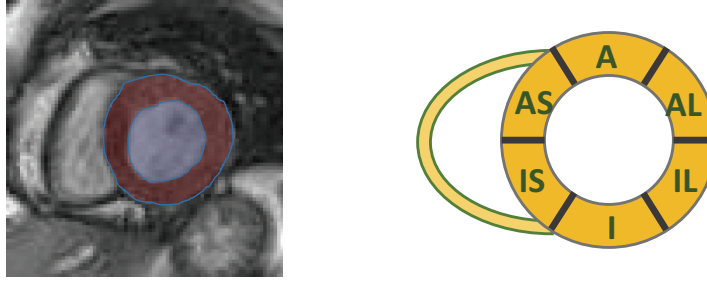
Fig. 1. Illustration of RWT for short-axis view cardiac MR image. Left: cardiac image with the contours of myocardium delineated. Right: the 6 segments model for mid-slice cardiac left ventricle. A: anterior; AS: anterospetal; IS: inferoseptal; I: inferior; IL: inferolateral; AL: anterolateral.

images [3,12,13] require strong prior information and user interaction to obtain reliable results, which may hinder them from efficient clinical application. An alternative way is to circumvent this segmentation step and estimate RWT from cardiac images directly. Direct estimation of cardiac volumes [1,2,17,18,20–22] have achieved great success in recent years, while direct estimation of RWT has never been explored due to the diversity of cardiac shape and structures for various subjects and various diseases, as well as the complication of regional myocardium deformation through the cardiac cycle.

In this work, we provide a method to estimate the frame-wise RWT from cardiac MR sequences through a newly proposed Residual Recurrent Neural Network (ResRNN). This ResRNN contains two paths: (1) a CNN path for low dimension embedding to robustly represent cardiac images of diverse structures, and preliminary estimation of RWT independently from the embedding of each frame itself, and (2) an RNN path for residual estimation from neighboring frames by leveraging the temporal and spatial dynamic deformation in cardiac sequences simultaneously. In the RNN path, a temporal RNN is applied to the features of temporally neighboring frames for modeling the complex long-range temporal dependencies, and another spatial RNN is applied to the predicted results of spatially neighboring wall thicknesses for modeling the mutual dependencies among these wall thicknesses. For the RNN module, a new Circle-RNN is designed to eliminate the effect of null initial hidden states by characterizing the periodicity of cardiac sequences and the circular spatial layout of cardiac RWT. With image represented by the CNN embeddings, the dynamic deformation of myocardium and the diversity of cardiac shape are well captured by the temporal and spatial RNN, thus leading to accurate estimation of RWT.

1.1 Related Work

Segmentation-Based and Direct Methods for Cardiac Volumes Estimation. Currently, the most related work to RWT estimation is cardiac volumes estimation, which falls into two categories: segmentation-based methods [3,12, 13] and direct methods [1,2,17,18,20–22]. Segmentation-based methods rely on

the premise of cardiac segmentation, which is still a great challenge due to the diverse structure of cardiac image and therefore requires strong prior information and user interaction [3,12,13].

To circumvent these limitations, direct methods without segmentation have grown in popularity in cardiac volumes estimation [1,2,17,18,20–22]. In these methods, hand-crafted features extracted from cardiac images are fed into regression models such as random forest (RF), adaptive K-clustering RF (AKRF), Bayesian model, and neural networks, to predict cardiac volumes. The employed features include Bhattacharyya coefficient between image distributions [1,2], appearance features [18], multiple low level image features [21], as well as features from multiscale convolutional deep belief network (MCDNB) [22] and supervised descriptor learning (SDL) [20]. Although these methods obtained effective performance, two limitations still exist: (1) they followed two separated phases, i.e., *feature extraction+volumes regression*, and no feedback exists between them to make them compatible with each other; (2) neither the temporal dependencies nor the spatial dependencies are taken into account, while the dependencies are important for dynamic modeling of cardiac sequence. In the present work, we provide an elegant solution for direct RWT estimation with an end-to-end architecture that respects both temporal and spatial dependencies.

Recurrent Neural Network. Recurrent neural network, especially when the long short-term memory units (LSTM) are deployed, is specialized in long-range temporal dynamic modeling and spatial context modeling. Promising results have been achieved by RNN in a wide spectrum of applications including language modeling [19], object recognition [11], visual recognition and description [5,10], and also medical image analysis [9,14,16]. In the work of cardiac image segmentation [14], an RNN was employed to capture the spatial changes of cardiac structure (represented as low dimensional CNN embeddings) in cardiac sequences. In [9], an RNN with LSTM was employed to model the temporal dependencies in cardiac MR sequences to identify the end-diastole and end-systole frames across a cardiac cycle. In [16], an RNN was trained to describe the contexts of detected disease in Chest X-Rays. These methods only explored one of the spatial or temporal dependencies while in cardiac sequences, both are important for robust dynamic modeling.

To fully explore the dependencies that exist in cardiac sequences during RWT estimation, two RNN modules are deployed in our work accounting for the temporal and spatial dependencies simultaneously. Besides, we propose a Circle-RNN for periodic cardiac sequences to better serve this aim, avoiding the effect of the null initial hidden input in existing RNN.

1.2   Contributions

The contributions of our work include:

– An effective end-to-end method that has great potential in clinical cardiac function assessment is proposed for direct cardiac RWT estimation, which has never been investigated previously.

– The newly proposed two-path ResRNN endows the network with abilities to robustly represent complex cardiac structure, and effectively model the capricious spatial and temporal dynamic deformation.
– The temporal RNN that accounts for the temporal deformation of the cardiac shape, and the spatial RNN that accounts for the smoothness of the LV myocardium shape, enable ResRNN to estimate collaboratively RWT of all frames and all regions by leveraging the temporal and spatial dependencies in cardiac MR sequences, rather than to estimate independently each cardiac RWT from a single image.
– A Circle-RNN designed for characterizing the periodicity of cardiac sequences and the circular spatial layout of cardiac RWT is proposed to eliminate the effect of the null hidden input for the first time step, to incorporate both the future and the past information in the dynamic modeling, and to treat every frame in the sequence with equally long-term dependencies.

2 RWT Estimation via ResRNN

2.1 Problem Formulation

For a set of cardiac MR sequences $\mathcal{X} = \{X_f^s\}$, where $s = 1 \cdots S$ indexes the subject and $f = 1 \cdots F$ indexes the frame in a cardiac cycle, we aim to estimate the frame-wise values of RWT $\mathcal{Y} = \{y_{f,l}^s\}$ for all the frames, where $l = 1 \cdots 6$ indexes the spatial location of each RWT (see Fig. 1, from IS to AS in counter-clockwise order). We consider in this work the mid-cavity of LV myocardium in short axis view, which is divided into six segments according to the AHA 17 segments model [4]. The objective function can be formulated as:

$$\min_{\theta} \frac{1}{2S \times F} \sum_f \sum_s \|y_f^s - \mathbf{Q}(X_f^s|\theta)\|_2^2 + \lambda \mathcal{R}(\theta) \tag{1}$$

where $\mathbf{Q}$ is the network, θ is the parameter vector to be learned, and $\mathcal{R}(\theta)$ regularizes the parameter vector.

2.2 Overview of the Proposed Method

To estimate the frame-wise RWT from cardiac MR sequence, we build a new architecture of network: ResRNN. As shown in Fig. 2, two paths are included in ResRNN: with the CNN path $\mathbf{Q_{CNN}}$, each frame in the sequence is independently processed by the CNN network, forming a low dimensional embedding of the cardiac images, from which the RWT is preliminarily estimated with another fully-connected layer; with the RNN path $\mathbf{Q_{RNN}}$, two RNN modules are deployed to model the temporal dependencies between neighboring frames and the spatial dependencies between neighboring RWTs, so as to further correct the residual of the preliminary estimation. The RNN path shares the same CNN embedding with the CNN path. The final RWT estimation is computed as:

$$\mathbf{Q}(X_f^s|\theta) = \mathbf{Q}_{CNN}(X_f^s|\theta) + \mathbf{Q}_{RNN}(X_f^s|\theta) \tag{2}$$
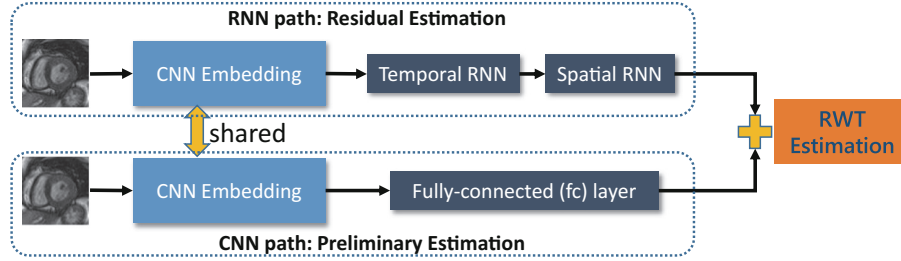
Fig. 2. Overview of ResRNN. A CNN path (the lower part) is employed to learn CNN embedding for the cardiac image and further obtain a preliminary estimation of RWT from these features. An RNN path (the upper part) is employed to model the temporal and spatial dependencies based on the CNN embedding of the cardiac sequences to estimate the residual of the preliminary estimation.
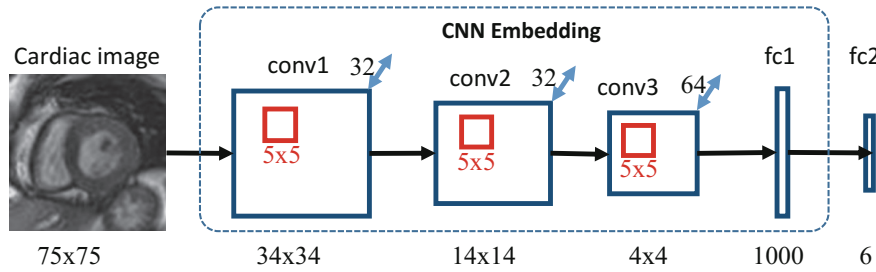
Fig. 3. Diagram of the CNN path. Each convolution layer (conv) is followed by a rectified linear unit (ReLU) layer and a max-pooling layer. The CNN embedding part represents the cardiac image as a low dimensional feature vector, which is shared by both the CNN and RNN paths.

2.3   Preliminary Estimation with the CNN Path

The diagram of our CNN path is shown in Fig. 3. Three convolution (conv1 $\sim$ 3) and one fully-connected (fc1) layers are deployed to obtain the low dimensional CNN embedding e_f^s of cardiac images. The second fully-connected layer (fc2) estimates a preliminary results $y_f^{s,CNN}$ of RWT from the CNN embeddings.

$$y_f^{s,CNN} = \theta_{fc2:w} e_f^s + \theta_{fc2:b} \tag{3}$$

where $\theta_{fc2:w}$ and $\theta_{fc2:b}$ are the weight matrix and bias of the fc2 layer.

As a feed forward neural network, the CNN path bears a notable limitation that it relies on the assumption of independence among samples, which does not hold for cardiac sequence. The dependencies among cardiac sequences have to be modeled to further reduce the residual of the CNN estimation.

## 2.4   Residual Estimation with the RNN Path

The diagram of the RNN path is shown in Fig. 4. Based on the CNN embedding obtained with the CNN path, temporal and spatial RNN are employed to
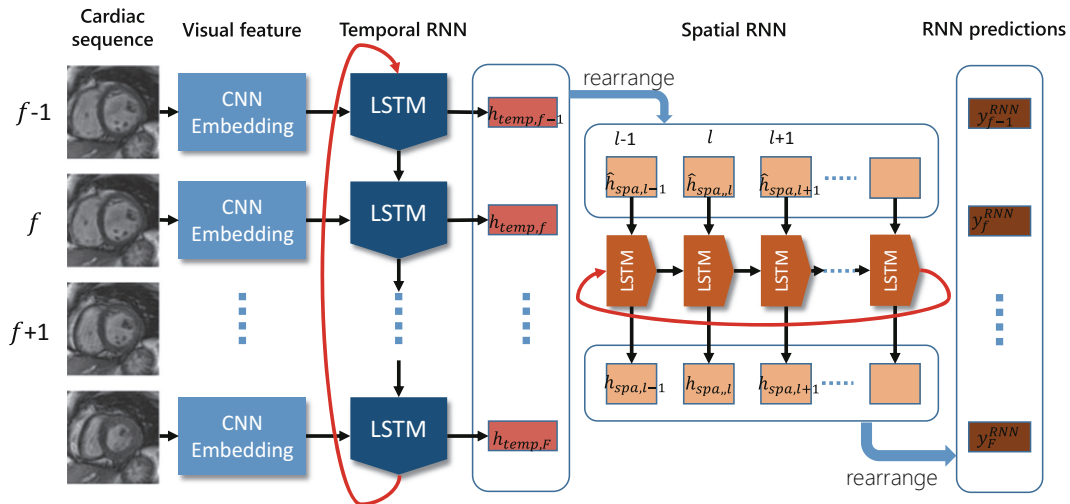
Fig. 4. Diagram of the RNN path. The temporal dependencies among neighboring frames are modeled by the temporal-RNN with CNN embedding as input. The spatial dependencies among neighboring RWT are modeled by the spatial-RNN with rearranged output of temporal-RNN as input. (Color figure online)

effectively model the dependencies existing among the RWT of all frames. In this section, we first introduce the memory unit LSTM that we use in the RNN path, and then describe the temporal and spatial RNN.

LSTM. In order to learn the long-term dynamics in sequential data and avoid the gradient vanishing/exploding problem in traditional RNN, LSTM unit [6], as shown in Fig. 5, was introduced into RNN. The input gate, output gate, forget gate and the memory cell allow the network to learn when to forget previous hidden states and when to update current hidden states given current input. This strategy enables LSTM to adaptively memorize and access information
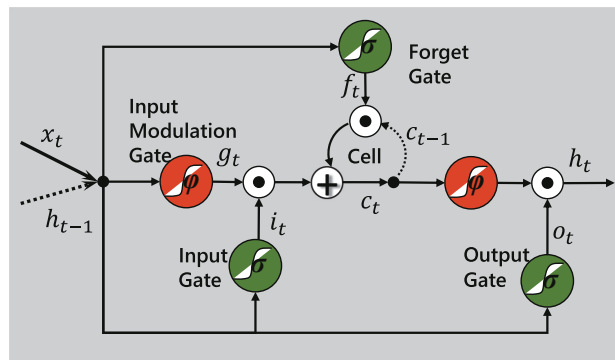
Fig. 5. A diagram of LSTM unit in RNN, which is capable of adaptively modeling dynamic deformations in cardiac sequences due to the input gate, forget gate, and output gate.

long term ago. The LSTM computations for time step t given the current input x_t, the previous hidden states h_{t-1} and memory states c_{t-1}, are as follows [6]:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
g_t &= \varphi(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot \varphi(c_t)
\end{aligned}
\tag{4}
$$

where $\sigma(\cdot)$ and $\varphi(\cdot)$ are element-wise sigmoid and hyperbolic tangent non-linearity functions, $\odot$ are element-wise product. The first three equations map the current input and previous hidden states to the input gate i_t, the forget gate f_t and the output gate o_t, to adaptively control the information flow. Ws are the weight matrices to be learned and bs are the corresponding bias terms.

Temporal RNN and Spatial RNN. With CNN embedding e_f^s, we first deploy a temporal RNN with the frame index in a cardiac sequences as time step to predict the values of RWT $h_{temp,f}^s \in \mathcal{R}^6$ for each frame f taking account of the dependencies between neighboring frames (See *Temporal RNN* in Fig. 4).

$$
h_{temp,f}^s = LSTM(e_f^s, h_{temp,f-1}^s), f = 1 \ldots F
\tag{5}
$$

Based on the prediction of temporal RNN, we again deploy a spatial RNN with spatial location as time step to predict RWT $h_{spa,l}^s \in \mathcal{R}^F$ of one specific location l for all the frames in the sequences (See *Spatial RNN* in Fig. 4). We rearrange column vectors $[h_{temp,1}^s, h_{temp,2}^s, \ldots, h_{temp,F}^s]$ to row vectors $[\hat{h}_{spa,1}^{s,T}; \hat{h}_{spa,2}^{s,T}, \ldots, \hat{h}_{spa,6}^{s,T}]$. Then we have

$$
h_{spa,l}^s = LSTM(\hat{h}_{spa,l}^s, h_{spa,l-1}^s), l = 1 \ldots 6
\tag{6}
$$

By rearranging these row vectors $[h_{spa,1}^{s,T}; h_{spa,2}^{s,T}; \ldots, h_{spa,6}^{s,T}]$ back to column vectors $[y_1^{s,RNN}, y_2^{s,RNN}, \ldots, y_F^{s,RNN}]$, we get the RNN predictions for frame-wise RWT.

Circle-RNN. As can be seen from Eqs. 5 and 6, three limitations exist for RNN: (1) for the first time step ($f = 1$ or $l = 1$), there is no value for previous hidden units, which influences the prediction of the first frame; (2) only past information can be used to determine the output of current time step, while future information is also equally important; (3) for the first few time steps, long-term dependency model can not be well built from the limited past information, leading to unfair treatment of different frames.

We provide an elegant solution to overcome these limitations for cardiac MR sequence: Circle-RNN, which connects the output of the last frame to the hidden
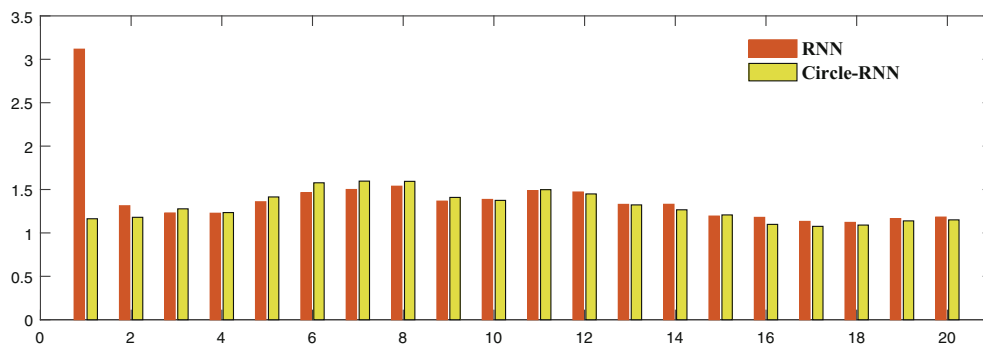
Fig. 6. Average estimation error (mm) of different frames by RNN and Circle-RNN for cardiac RWT. Only temporal RNN is employed here for comparison. The estimation error for the first frame is clearly reduced by Circle-RNN.

input of the first frame, as the red arrows show in Fig. 4. Within Circle-RNN, Eqs. 5 and 6 become:

$$h^s_{temp,f} = LSTM(e^s_f, h^s_{temp,mod(f-1-1,F)+1}), f = 1 \ldots F \tag{7}$$

$$h^s_{spa,l} = LSTM(\hat{h}^s_{spa,l}, h^s_{spa,mod(l-1-1,6)+1}), l = 1 \ldots 6 \tag{8}$$

where $mod(a, b)$ computes the modulus. Circle-RNN can be easily optimized with the BPTT algorithm [6]. To avoid infinite information loop within this Circle-RNN, we introduce a parameter *depth* to control how many rounds the information flow in our Circle-RNN. Figure 6 shows the error reduction of Circle-RNN over RNN when predicting cardiac RWT from the CNN embedding with only temporal RNN employed.

3   Experiments

3.1   Dataset and Implementations

A dataset of short-axis cine MR images paired with manually obtained ground truth values of RWT is constructed to evaluate the performance of our method, which includes 2900 images from 145 subjects. These subjects are collected from 3 hospitals affiliated with two health care centers and 2 vendors (GE and Siemens). Each subject contains 20 frames throughout a cardiac cycle. In each frame, the middle slice is selected following the standard AHA prescriptions [4] for validation of the proposed cardiac RWT estimation method. The ground truth values of RWT are manually obtained for each image.

In our experiments, two landmarks, i.e., junctions of the right ventricular wall with the left ventricular, are manually marked for each cardiac image to provide reference for ROI cropping and the LV myocardial segments division. The cropped images are resized to dimension of 80×80. All values of RWT are normalized to the range of [0,1] according to the image dimension (80). Five-fold cross validation is employed for performance evaluation and comparison. Mean

Table 1. Estimation error comparison of CNN, RNN and ResRNN for cardiac RWT estimation in terms of MAE (Mean±standard deviation) (mm). WT-IS means wall thickness for the inferoseptal segment of myocardium, similar for the rest. Best results are highlighted in bold.

Method	CNN	RNN (plain)	RNN (circle)	ResRNN (plain)	**ResRNN (circle)**
WT-IS	1.28 ± 1.10	1.30 ± 1.11	1.31 ± 1.04	1.26 ± 1.15	$\mathbf{1.22 \pm 1.02}$
WT-I	1.56 ± 1.27	1.60 ± 1.28	1.54 ± 1.26	1.55 ± 1.24	$\mathbf{1.47 \pm 1.22}$
WT-IL	1.81 ± 1.51	1.99 ± 1.66	1.86 ± 1.60	1.80 ± 1.48	$\mathbf{1.78 \pm 1.53}$
WT-AL	1.64 ± 1.30	1.70 ± 1.42	1.66 ± 1.35	$\mathbf{1.60 \pm 1.29}$	1.60 ± 1.31
WT-A	1.35 ± 1.09	1.36 ± 1.18	1.30 ± 1.09	1.40 ± 1.20	$\mathbf{1.31 \pm 1.08}$
WT-AS	1.26 ± 0.80	1.39 ± 1.06	1.29 ± 1.01	1.34 ± 1.10	$\mathbf{1.25 \pm 0.97}$
Average	1.49 ± 0.76	1.56 ± 0.83	1.49 ± 0.80	1.49 ± 0.81	$\mathbf{1.44 \pm 0.74}$

absolute error (MAE) between estimation and the ground truth is computed to evaluate the performance. The network is implemented by Caffe [7] with SGD solver. Learning rate and weight decay are set to (0.05, 0.0005). 'step' learning policy is employed with gamma and step size being (0.5, 2500) and momentum 0.9. The depth of Circle-RNN is set as the number of time steps, i.e., 20 for the temporal RNN and 6 for the spatial RNN. Data augmentation is conducted by randomly cropping images of size 75×75 from the original image.

3.2   Performance Comparison: ResRNN vs. CNN and RNN

The advantages of the proposed ResRNN are firstly demonstrated by comparing performance of three different network architectures: (1) CNN, i.e. the CNN path as shown in Fig. 3; (2) RNN, i.e. the RNN path as shown in Fig. 4, and (3) the proposed ResRNN. For RNN and ResRNN, both the original RNN (the plain RNN) and Circle-RNN are employed for comparison.

From the average estimation error shown in Table 1, we can observe the followings. (1) The two-path ResRNN outperforms CNN and RNN, with either plain or circle RNN deployed, which can be ascribed to the complementarity of the preliminary estimation from each frame itself and the residual estimation that modeling the dependencies of cardiac sequence. (2) When Circle-RNN, rather than the plain RNN, is deployed, lower estimation error can be obtained by RNN or ResRNN, due to the fact that Circle-RNN is capable of memorizing useful cardiac dynamic information for the first time step. In the following experiments, Circle-RNN is deployed in ResRNN.

3.3   Performance Comparison: ResRNN vs. State-of-the-Art

To demonstrate the advantages of our proposed method over segmentation based [3] and two-phase direct methods [20–22], we apply these methods to our database for cardiac RWT estimation. For the direct methods, the same five-fold cross-validation protocol is used. As can be observed in Table 2 and Fig. 7, the

Table 2. Estimation error comparison of the proposed ResRNN with segmentation based method and existing direct methods for RWT estimation in terms of MAE (Mean± standard deviation) (mm). Best results are highlighted in bold.

Method	Max Flow [3]	Multi-features+RF [21]	SDL+AKRF [20]	MCDBN+RF [22]	ResRNN
WT-IS	1.53 ± 1.73	1.70 ± 1.47	1.98 ± 1.58	1.78 ± 1.40	$\mathbf{1.22 \pm 1.02}$
WT-I	3.23 ± 2.83	1.71 ± 1.34	1.67 ± 1.40	1.68 ± 1.41	$\mathbf{1.47 \pm 1.22}$
WT-IL	4.15 ± 3.17	1.97 ± 1.54	1.88 ± 1.63	1.92 ± 1.45	$\mathbf{1.78 \pm 1.53}$
WT-AL	5.08 ± 3.95	1.82 ± 1.41	1.87 ± 1.55	1.66 ± 1.20	$\mathbf{1.60 \pm 1.31}$
WT-A	3.47 ± 3.25	1.55 ± 1.33	1.65 ± 1.45	$\mathbf{1.20 \pm 1.01}$	1.31 ± 1.08
WT-AS	1.76 ± 1.80	1.68 ± 1.43	2.04 ± 1.59	1.63 ± 1.23	$\mathbf{1.25 \pm 0.97}$
Average	3.21 ± 1.98	1.73 ± 0.97	1.85 ± 1.03	1.65 ± 0.77	$\mathbf{1.44 \pm 0.74}$
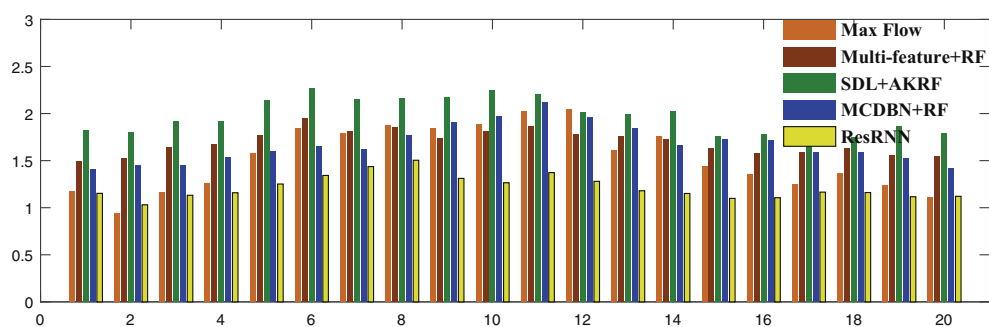
Fig. 7. Average frame-wise estimation error (mm) of WT-IS over all subjects for ResRNN and all other competitors. The proposed ResRNN delivers clearly lower estimation error than other methods.

proposed ResRNN demonstrates great advantages over existing segmentation-based and two-phase direct methods for cardiac RWT estimation.

From Table 2, we can see that the proposed ResRNN estimates cardiac RWT with high accuracy (average MAE of 1.44 mm) and outperforms all competitors. Specifically, it outperforms the Max Flow method with 55.14% MAE reduction. Note that Max Flow obtained high dice metric of 0.9182 for LV cavity segmentation when applied to our database. In fact, the dependency on manual segmentation of the first frame makes the estimation error of Max Flow increase as the estimated frame is far from the first frame within the cardiac cycle (see the frame-wise estimation error of Max Flow in Fig. 7). ResRNN outperforms the best of the direct methods (MCDBN+RF [22]) with a clear error reduction: 12.73%. Within the two-phase framework, the hand-crafted multifeatures, the features obtained by supervised learning, and MCDBN features employed in existing methods all fail to beat ResRNN, which evidences the benefits of the network architecture in ResRNN over the two-phase direct methods. Figure 7 also shows that ResRNN estimates each frame with consistently lower MAE.

Besides, we can draw that all the deep neural network based methods in Table 1 outperform existing two-phase direct methods in Table 2. This further

confirms the argument that independent feature extraction and regression cannot make the two phases maximumly compatible with each other. The end-to-end learning procedure of neural network integrates both phases together and leads to clearly improved performance.

4   Conclusions

In this paper, we propose an effective network architecture ResRNN for the task of cardiac RWT estimation, which has never been explored before. In ResRNN, a CNN path is employed to estimate from each cardiac image independently the preliminary results of RWT, and an RNN path is employed to compensate the residual of CNN estimations with temporal and spatial dependencies being accounted by Circle-RNN. Validation on a data set of cardiac MR sequences from 145 subjects demonstrates that the proposed ResRNN is capable of estimating cardiac RWT values with performance better than state-of-the-art methods, and is of great potential in clinical cardiac function assessment.

References

1. Afshin, M., Ayed, I.B., Islam, A., Goela, A., Peters, T.M., Li, S.: Global assessment of cardiac function using image statistics in MRI. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 535–543. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33418-4_66
2. Afshin, M., Ben Ayed, I., Punithakumar, K., Law, M., Islam, A., Goela, A., Peters, T.M., Li, S.: Regional assessment of cardiac left ventricular myocardial function via MRI statistical features. IEEE Trans. Med. Imaging **33**(2), 481–494 (2014)
3. Ayed, I.B., Chen, H.M., Punithakumar, K., Ross, I., Li, S.: Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure. Med. Image Anal. **16**(1), 87–100 (2012)
4. 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.: 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 (2002)
5. Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T.: Long-term recurrent convolutional networks for visual recognition and description. In: IEEE CVPR, pp. 2625–2634 (2015)
6. Graves, A.: Supervised sequence labelling. In: Graves, A. (ed.) Supervised Sequence Labelling with Recurrent Neural Networks. SCI, vol. 385, pp. 5–13. Springer, Heidelberg (2012). doi:10.1007/978-3-642-24797-2_2
7. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: ACM International Conference on Multimedia, pp. 675–678. ACM (2014)

8. Kawel, N., Turkbey, E.B., Carr, J.J., Eng, J., Gomes, A.S., Hundley, W.G., Johnson, C., Masri, S.C., Prince, M.R., van der Geest, R.J., et al.: Normal left ventricular myocardial thickness for middle-aged and older subjects with steady-state free precession cardiac magnetic resonance the multi-ethnic study of atherosclerosis. Circ.: Cardiovasc. Imaging **5**(4), 500–508 (2012)

9. Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S.: Recognizing end-diastole and end-systole frames via deep temporal regression network. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 264–272. Springer, Cham (2016). doi:10.1007/978-3-319-46726-9_31

10. Li, Y., Lan, C., Xing, J., Zeng, W., Yuan, C., Liu, J.: Online human action detection using joint classification-regression recurrent neural networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 203–220. Springer, Cham (2016). doi:10.1007/978-3-319-46478-7_13

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

12. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magn. Reson. Mater. Phys. Biol. Med. **29**(2), 155–195 (2016)

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

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

15. Puntmann, V.O., Gebker, R., Duckett, S., Mirelis, J., Schnackenburg, B., Graefe, M., Razavi, R., Fleck, E., Nagel, E.: Left ventricular chamber dimensions and wall thickness by cardiovascular magnetic resonance: comparison with transthoracic echocardiography. Eur. Heart J.-Cardiovasc. Imaging **14**(3), 240–246 (2013)

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

17. Wang, H., et al.: Prediction of clinical information from cardiac MRI using manifold learning. In: van Assen, H., Bovendeerd, P., Delhaas, T. (eds.) FIMH 2015. LNCS, vol. 9126, pp. 91–98. Springer, Cham (2015). doi:10.1007/978-3-319-20309-6_11

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

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

20. Zhen, X., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct and simultaneous four-chamber volume estimation by multi-output regression. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 669–676. Springer, Cham (2015). doi:10.1007/978-3-319-24553-9_82

21. Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct estimation of cardiac bi-ventricular volumes with regression forests. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 586–593. Springer, Cham (2014). doi:10.1007/978-3-319-10470-6_73

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