Wall Thickness Estimation from Short Axis Ultrasound Images via Temporal Compatible Deformation Learning

Ang Zhang[1,2,3], Guijuan Peng[4], Jialan Zheng[1,2,3], Jun Cheng[1,2,3], Xiaohua Liu[4], Qian Liu[4], Yuanyuan Sheng[4], Yingqi Zheng[4], Yumei Yang[4], Jie Deng[4], Yingying Liu[4(✉)], Wufeng Xue[1,2,3(✉)], and Dong Ni[1,2,3]

[1] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen, China
xuewf@szu.edu.cn

[2] Medical Ultrasound Image Computing (MUSIC) Laboratory, Shenzhen University, Shenzhen, China

[3] Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, China

[4] Department of Ultrasound, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, China
yingyingliu@ext.jnu.edu.cn

Abstract. Structural parameters of the heart, such as left ventricular wall thickness (LVWT), have important clinical significance for cardiac disease. In clinical practice, it requires tedious labor work to be obtained manually from ultrasound images and results in large variations between experts. Great challenges exist to automatize this procedure: the myocardium boundary is sensitive to heavy noise and can lead to irregular boundaries; the temporal dynamics in the ultrasound video are not well retained. In this paper, we propose a Temporally Compatible Deformation learning network, named *TC-Deformer*, to detect the myocardium boundaries and estimate LVWT automatically. Specifically, we first propose a two-stage deformation learning network to estimate the myocardium boundaries by deforming a prior myocardium template. A global affine transformation is first learned to shift and scale the template. Then a dense deformation field is learned to adjust locally the template to match the myocardium boundaries. Second, to make the deformation learning of different frames become compatible in the temporal dynamics, we adopt the mean parameters of affine transformation for all frames and propose a bi-direction deformation learning to guarantee that the deformation fields across the whole sequences can be applied to both the myocardium boundaries and the ultrasound images. Experimental results on an ultrasound dataset of 201 participants show that the proposed method can achieve good boundary detection of basal, middle, and apical myocardium, and lead to accurate estimation of the LVWT, with a mean absolute error of less than 1.00 mm. When compared with

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 213–222, 2023.
https://doi.org/10.1007/978-3-031-43987-2_21

human methods, our TC-Deformer performs better than the junior cardiologists and is on par with the middle-level cardiologists.

Keywords: Wall thickness · Segmentation · Deformation learning

1 Introduction

Cardiovascular disease is a leading cause of death in the world. Accurate quantification of left ventricular wall thicknesses (LVWT) from ultrasound images is among the most clinically important and significant indices for evaluating cardiac function and diagnosis of cardiac diseases [2,6]. Figure 1 illustrates short axis (SAX) ultrasound images of basal, middle, and apical myocardium, with the corresponding LVWTs according to the 16-segment myocardium model. In clinical practice, obtaining reliable clinical information mainly depends on radiologists to manually draw the contours of the endocardium and epicardium of the left ventricle (LV). It is time-consuming and laborious. Efforts have been devoted to the automatic estimation of LVWTs, where great challenge exists. First, the myocardium boundary is sensitive to heavy noise, especially for the apical and basal SAX images, and can lead to irregular boundaries and undermine the estimation of LVWTs. Second, the temporal dynamics in the ultrasound video are difficult to be modeled, leading to prediction results that are not well compatible with the temporal dynamics of the whole video.

Existing work can be divided into two categories: segmentation-based and direct-regression methods. The direct-regression methods to learn the regress LVWTs from cardiac images directly without identifying the contours first. [5,15] proposed end-to-end cardiac index quantification frameworks based on cascaded convolutional autoencoders and regression networks, using only the values of cardiac indices for supervision. [4] proposed a two-stage network that learns the LV contours first and then estimates the LV indices with a new network. [16] proposed a residual recurrent neural network further improves the estimation by modeling the temporal and spatial of the LV myocardium to achieve accurate frame-by-frame LVWT estimation. However, these models lack explicit temporal dynamic modeling of the whole sequence.

Segmentation-based methods segment the myocardium first and then calculate cardiac parameters. To the best of our knowledge, existing segmentation work mainly focus on apical views to evaluate the ejection fraction, and rare work exists for short-axis views. [14] utilizes the underlying motion information to assist in improving segmentation results by accurately predicting optical flow fields. [12,13] proposed appearance-level and shape-level co-learning (CLAS) to enhance the temporal consistency of the predicted masks across the whole sequence and accuracy. This method effectively improves the accuracy and consistency of myocardial segmentation. [1] proposed to introduce residual structure into U-net and [3] proposed a hybrid framework combining a convolutional encoder-decoder structure and a transformer. [7,17] proposed a multi-attention mechanism to guide the network to capture features effectively while suppressing

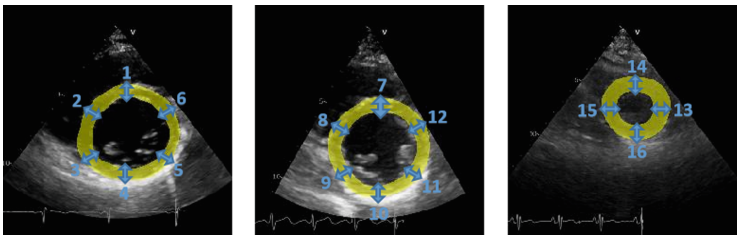

Fig. 1. Illustration of LVWT to be quantified for short-axis view cardiac image and 16 segments.Basal (left), middle (middle), and apical (right). Regional wall thicknesses (black arrows). 1–16: 16 segments.

noise, and integrated deep supervision mechanism and spatial pyramid feature fusion to enhance feature extraction. However, these models are not robust to the heavy noise in the SAX images, which may lead to irregular boundaries and undermines the estimation of LVWT.

To overcome the above mention challenges and inspired [8] where template transformer was employed for image segmentation, we propose a novel Temporal-Compatible Deformation learning network for myocardium boundary detection from ultrasound SAX images. The primary contributions of this paper are as follows. 1) To overcome the irregular boundaries caused by the heavy noise, we propose a two-stage deformation learning network for myocardium boundary detection. A global affine transformation and a local deformation are used to deform the prior myocardium template to match the myocardium boundary. 2) To make the template-deformed myocardium boundaries compatible across the whole sequence, we propose a bi-direction deformation learning to guarantee that the deformation fields across the whole sequences can be applied to both the myocardium boundaries and the ultrasound images. 3) The proposed TC-deformer achieves excellent performance for LVWT estimation, with an error of less than 1.00 mm, and is comparable with middle-level cardiologists.

2 Methods

The structures of the myocardium in ultrasound SAX images generally follow a circular ring shape, which is a vital characteristic of prior knowledge in short-axis myocardial segmentation, especially for ultrasound images with heavy noisy myocardium boundaries. In this paper, we propose a novel method, Temporal Compatibility Deformation Learning, named *TC-Deformer*, to achieve accurate and plausible myocardium contours and LVWTs estimation. The details are described as follows.

2.1 Deformation Learning

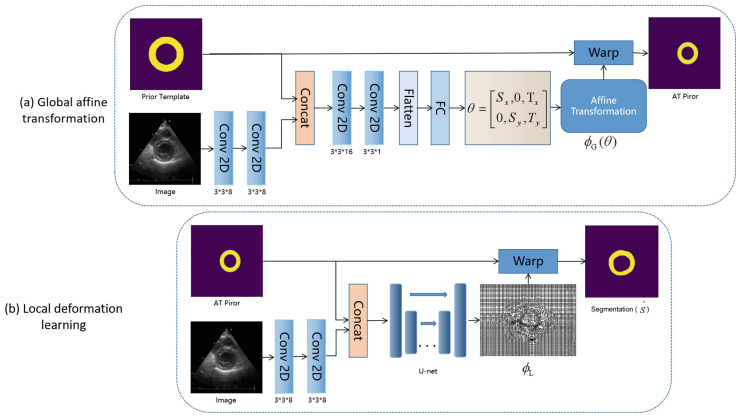

Fig. 2. (a)Overview of the proposed global affine transformation. (b)Overview of the proposed local deformation learning.

Global Affine Transformation. Our deformation learning consists of two stages: global affine transformation and local deformation learning. In this subsection, we will describe our global affine transformation. Considering the diversity of the cardiac structure and data gaps from different machines, we compute the mean value of the thickness measurements and the center position of the circle according to the annotated information to generate the prior template. As shown in Fig. 2(a), the prior template (P) is first concatenated with the convolution features extracted from the SAX image to learn the global affine parameters $\theta = \{(\Delta x, \Delta y), s\}$, which represent the shift and scale from the prior template to the myocardium boundaries in the SAX image. Then we get the affine prior (AT) $S_{AT} = \phi_G(P)$ by warping the prior template (P) with the global affine parameters θ. The loss function is as follows:

$$Loss_{AT} = -S \cdot logS_{AT} + (1 - \frac{|S \cdot S_{AT}|}{|S| + |S_{AT}|}) \tag{1}$$

where S is the ground truth of the myocardium segmentation. However, the shapes warping from the prior template with global affine parameters are far from precise due to the individual variation. So, we introduce the local deformation learning to get a precise myocardium shape with the learned dense deformation field.

Local Deformation Learning. In this part, we aim to learn a dense deformation field to adjust the AT prior locally to match the myocardium boundaries. As shown in Fig. 2(b), the AT prior is first concatenated with the image convolution feature to learn a dense deformation field $\phi_L \in R^{256\times256\times2}$, which represents

the pixel-level displacement along both horizontal and vertical directions. Our local deformation learning considers the prior shape, the prior position, and the image feature simultaneously, which can help the network learn a more precise deformation field and get a local adjustment of the template.

Finally, we take the segmentation $\hat{S} = \phi_L(S_{AT})$ warping from the AT prior with the dense deformation filed as the final myocardium segmentation result. The loss function is as follows:

$$Loss_{seg} = -S \cdot log\hat{S} + (1 - \frac{|S \cdot \hat{S}|}{|S| + |\hat{S}|}) \tag{2}$$

where S is the ground truth of the myocardium segmentation.

2.2 TC-Deformer

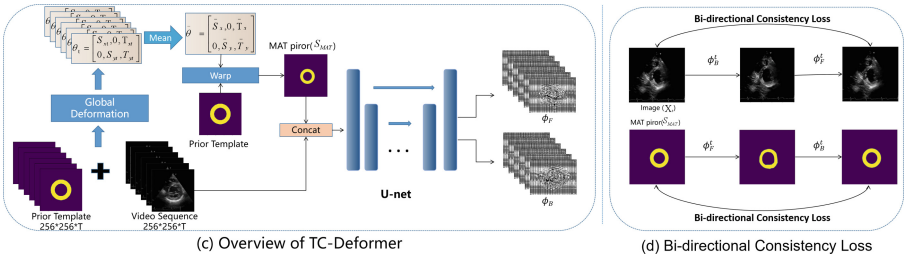

Fig. 3. (c) Overview of the proposed TC-Deformer network framework. (d) Schematic diagram of bi-directional consistency loss for images and segmentation.

In this work, only frames at the end-systolic (ED) and end-diastolic (ES) phases are annotated by cardiologists for each ultrasound SAX video. To make the template-deformed myocardium boundaries compatible across the whole sequence, we propose a bi-direction deformation learning to guarantee that the deformation fields across the whole sequence can be applied to both the myocardium boundaries and the ultrasound images in the sequence.

As shown in Fig. 3, first of all, we use the mean $\bar{\theta}$ of the affine transformation parameters $\{\theta_1, \theta_2, \theta_3, ..., \theta_{T-2}, \theta_{T-1}, \theta_T\}$ of all frames in the sequence as a video-level parameter and obtain the mean affine transformation prior template (MAT prior). Next, the MAT prior is combined with the sequential images to learn a series of dense bi-direction deformation fields. Let ϕ_F^t be the forward deformation and ϕ_B^t the backward deformation for the frame t. Our bi-direction deformation learning (as shown in Fig. 3(b)) aims to constrain that for each frame X_t, after forward deformation and backward deformation, we can still obtain the original images. We adopt the structural similarity metric (SSIM) [11] in image quality assessment to quantify the deformation error. For the image cycle, the loss function is as follows:

$$Loss_{imgcycle} = \sum_{t=1}^{T} \left(1 - SSIM(X_t, \phi_F^t\left(\phi_B^t\left(X_t\right)\right)\right) \tag{3}$$

Similarly, for the MAT prior S_{MAT}, we can have the shape consistency constraint when applied to the bi-direction deformation procedure:

$$Loss_{shapecycle} = \sum_{t=1}^{T}(1 - \frac{|S_{MAT} \cdot \ \phi_B^t\left(\phi_F^t\left(S_{MAT}\right)\right)|}{|S_{MAT}| + |\phi_B^t\left(\phi_F^t\left(S_{MAT}\right)\right)|}) \tag{4}$$

As the temporally compatible deformation started with a common template, we assume that when warped backward, all the frames will have a similar appearance. So, we introduce a centralization loss, to minimize the deviation between those backward deformed frames:

$$Loss_{cent} = \sum_{t=1}^{T} |\phi_B^t(X_t) - \bar{\phi}_B(X_t)|^2 \tag{5}$$

where T represents the time.

The total loss function of our TC-Deformer is as follows:

$$Loss_{total} = Loss_{seg} + \alpha Loss_{cent} + \beta Loss_{imgcycle} + \gamma Loss_{shapecycle} \tag{6}$$

where α, β and γ is the hyper-parameters.

After myocardial segmentation, we use neural networks to determine two key points, combined with the centroid of the segmentation, to divided it into 16-segments according to the 16-segment model of the American Heart Association and calculated the corresponding LVWTs.

3 Experiments and Results

3.1 Dataset Description and Experimental Setup

Datasets. In this experiment, we trained our method on ultrasound SAX videos of 141 participants and tested with 60 participants. All the data was collected with the GE Vivid E95 from the Shenzhen People's Hospital and this study was approved by local institutional review boards. For each participant, videos of the basal, middle, and apical SAX views were acquired with multiple cardiac cycles. The mask of the myocardium at the ED and ES phases from one cardiac cycle was annotated by experts (including two senior, three middle-level, and three junior cardiologists). We compare the LVWT of each group with the average wall thickness of the senior doctors to get the average of the errors for different doctors as result. For the training dataset, the senior cardiologists conducted quality control for junior and middle-level cardiologists. All images were annotated by experienced doctors using the Pair annotation software package [9](https://www.aipair.com.cn/en/, Version 2.7, RayShape, Shenzhen, China).

Experimental Setup. We resized the images to the same size 256×256 and used the Adam optimization strategy during model training. The training is in two stages and the initial learning rate was 0.0001, the total epoch number is 100, and the batch size is 4. The hyperparameters α, β, and γ were set to be 0.1, 0.5, and 0.5, respectively, according to a small validation set. The models are implemented with PyTorch on the NVIDIA A100 Tensor Core GPU.

3.2 Results

Table 1. Quantitative comparison results of the average Dice, Hausdorff distance(HD), FLOPs.

	Dice	HD (mm)	FLOPs (G)
Unet [10]	0.827	3.73	**40.46**
CLAS [12]	0.842	2.98	208.47
Deformer	0.852	**2.64**	41.46
TC-Deformer	**0.854**	2.92	41.46

Table 1 shows the segmentation performance on the test set in terms of Dice, Hausdorff distance (HD), and floating-point operations per second (FLOPs). We can conclude that the proposed method achieves excellent segmentation performance and outperforms U-net and the state-of-the-art CLAS, while costing much less computation. Figure 4 shows the segmentation results of myocardium compared with four methods. It indicates that our method has a more reasonable myocardium shape than others, which is important to LVWT.

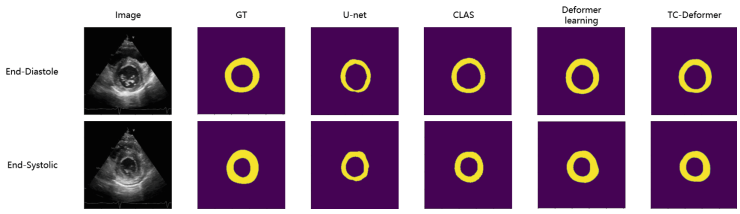

Fig. 4. Visualization results of four segmentation methods.

Figure 5 shows the segmentation results of the myocardium in one cardiac cycle. It indicates that our method can obtain smoother contours for the middle frames than CLAS, implying that the temporally compatible deformation learning has a better temporal consistency.

Table 2 shows the MAE of the measurements in LVWT for middle-level and junior cardiologists, as while as for the proposed TD-Deformer. It indicates our results of the measurements are better than the junior groups and comparable

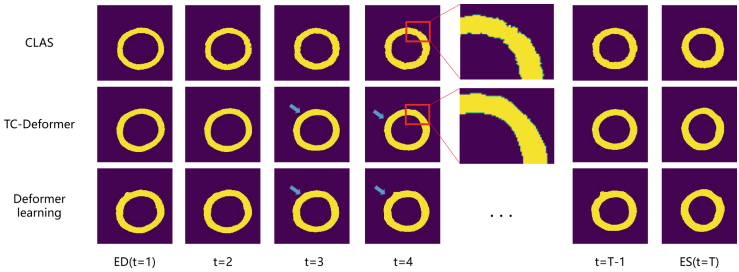

Fig. 5. Visualization of the segmentation results for one cardiac cycle.

with the middle-level group. The error of LVWT estimation is less than 1.00 mm. Figure 6 shows the absolute error of the predicted results for 16 segments. We can conclude that for all 16 segments, the prediction results of TC-Deformer are stable and accurate.

Table 2. The MAE of LVWT (mm)

	Basal					
	1	2	3	4	5	6
Mid-level	0.86	**0.67**	0.93	0.89	**0.87**	0.85
Junior	1.02	0.87	1.03	**0.87**	0.91	0.90
TC-Deformer	**0.80**	0.80	**0.86**	0.91	0.92	**0.85**
	Middle					
	7	8	9	10	11	12
Mid-level	**0.67**	0.95	**0.86**	0.87	1.04	0.83
Junior	0.73	0.95	1.02	**0.81**	1.02	0.93
TC-Deformer	1.02	**0.82**	0.90	1.04	**0.82**	**0.83**
	Apical				16-segments	
	13	14	15	16	mean	
Mid-level	0.90	0.89	0.96	**0.97**	**0.88**	
Junior	1.20	1.22	1.08	1.36	0.99	
TC-Deformer	**0.87**	**0.97**	**0.94**	1.18	0.91	

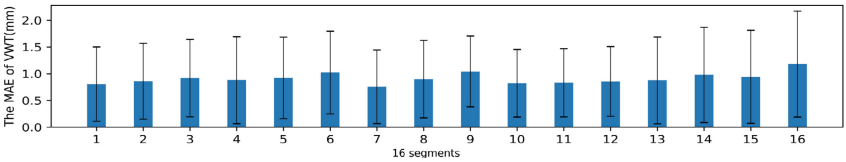

Fig. 6. The MAE of prediction LVWT indices as well as their corresponding standard deviation.

4 Conclusion

In this paper, We propose a Temporally Compatible Deformation learning network, named TC-Deformer, to detect the myocardium boundaries and estimate regional left ventricle wall thickness. Our method is designed to avoid the irregular contours that can happen in ultrasound images with heavy noise and can incorporate the temporal dynamics of the myocardium in one cardiac cycle into the deformation field learning. When validated with a dataset of 201 patients, our method achieves less than 1.00 mm estimation error for all 16 myocardium segments and outperforms existing state-of-the-art methods.

Acknowledgement. The work is partially supported by the Natural Science Foundation of China (62171290), the Shenzhen Science and Technology Program (20220810145705001, JCYJ20190808115419619, SGDX20201103095613036), Medical Scientific Research Foundation of Guangdong Province (No. A2021370).

References

1. Amer, A., Ye, X., Janan, F.: ResDUnet: a deep learning-based left ventricle segmentation method for echocardiography. IEEE Access **9**, 159755–159763 (2021)
2. Chen, L., Su, Y., Yang, X., Li, C., Yu, J.: Clinical study on LVO-based evaluation of left ventricular wall thickness and volume of AHCM patients. J. Radiat. Res. Appl. Sci. **16**(2), 100545 (2023)
3. Deng, K., et al.: TransBridge: a lightweight transformer for left ventricle segmentation in echocardiography. In: Noble, J.A., Aylward, S., Grimwood, A., Min, Z., Lee, S.-L., Hu, Y. (eds.) ASMUS 2021. LNCS, vol. 12967, pp. 63–72. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87583-1_7
4. Du, X., Tang, R., Yin, S., Zhang, Y., Li, S.: Direct segmentation-based full quantification for left ventricle via deep multi-task regression learning network. IEEE J. Biomed. Health Inf. **23**(3), 942–948 (2018)
5. Ge, R., et al.: PV-LVNet: direct left ventricle multitype indices estimation from 2d echocardiograms of paired apical views with deep neural networks. Med. Image Anal. **58**, 101554 (2019)
6. Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B., Neubauer, S.: The role of cardiovascular magnetic resonance imaging in heart failure. J. Am. Coll. Cardiol. **54**(15), 1407–1424 (2009)
7. Leclerc, S., et al.: LU-Net: a multistage attention network to improve the robustness of segmentation of left ventricular structures in 2-D echocardiography. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **67**(12), 2519–2530 (2020)

8. Lee, M.C.H., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M.: TeTrIS: template transformer networks for image segmentation with shape priors. IEEE Trans. Med. Imaging **38**(11), 2596–2606 (2019)
9. Liang, J., et al.: Sketch guided and progressive growing GAN for realistic and editable ultrasound image synthesis. Med. Image Anal. **79**, 102461 (2022)
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
11. Wang, Z.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
12. Wei, H., Cao, H., Cao, Y., Zhou, Y., Xue, W., Ni, D., Li, S.: Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 623–632. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_60
13. Wei, H., Ma, J., Zhou, Y., Xue, W., Ni, D.: Co-learning of appearance and shape for precise ejection fraction estimation from echocardiographic sequences. Med. Image Anal. **84**, 102686 (2023)
14. Xue, W., Cao, H., Ma, J., Bai, T., Wang, T., Ni, D.: Improved segmentation of echocardiography with orientation-congruency of optical flow and motion-enhanced segmentation. IEEE J. Biomed. Health Inf. **26**(12), 6105–6115 (2022)
15. Xue, W., Islam, A., Bhaduri, M., Li, S.: Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Trans. Med. Imaging **36**(10), 2057–2067 (2017)
16. Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S.: Direct estimation of regional wall thicknesses via residual recurrent neural network. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 505–516. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_40
17. Zeng, Y., et al.: MAEF-Net: multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography. Ultrasonics **127**, 106855 (2023)