Temporal-consistent Segmentation of Echocardiography with Co-learning from Appearance and Shape

Conference Paper · June 2020

CITATIONS
0

READS
101

7 authors, including:

Yongjin Zhou
Shenzhen University
69 PUBLICATIONS **550** CITATIONS
SEE PROFILE

Wufeng Xue
Shenzhen University
30 PUBLICATIONS **1,492** CITATIONS
SEE PROFILE

Dong Ni
Shenzhen University
151 PUBLICATIONS **2,093** CITATIONS
SEE PROFILE

Shuo Li
The University of Western Ontario
300 PUBLICATIONS **3,542** CITATIONS
SEE PROFILE

Temporal-consistent Segmentation of Echocardiography with Co-learning from Appearance and Shape

Hongrong Wei[1,2], Heng Cao[1,2], Yiqin Cao[1,2], Yongjin Zhou[3], Wufeng Xue[1,2⋆], Dong Ni[1,2], and Shuo Li[4]

[1] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[2] Medical Ultrasound Image Computing (MUSIC) Lab
xuewf@szu.edu.cn
[3] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[4] Department of Medical Imaging, Western Univeristy, London, ON, Canada

Abstract. Accurate and temporal-consistent segmentation of echocardiography is important for diagnosing cardiovascular disease. Existing methods often ignore consistency among the segmentation sequences, leading to poor ejection fraction (EF) estimation. In this paper, we propose to enhance temporal consistency of the segmentation sequences with two co-learning strategies of segmentation and tracking from ultrasonic cardiac sequences where only end diastole and end systole frames are labeled. First, we design an appearance-level co-learning (CLA) strategy to make the segmentation and tracking benefit each other and provide an eligible estimation of cardiac shapes and motion fields. Second, we design another shape-level co-learning (CLS) strategy to further improve segmentation with pseudo labels propagated from the labeled frames and to enforce the temporal consistency by shape tracking across the whole sequence. Experimental results on the largest publicly-available echocardiographic dataset (CAMUS) show the proposed method, denoted as CLAS, outperforms existing methods for segmentation and EF estimation. In particular, CLAS can give segmentations of the whole sequences with high temporal consistency, thus achieves excellent estimation of EF, with Pearson correlation coefficient 0.926 and bias of 0.1%, which is even better than the intra-observer agreement.

Keywords: Echocardiography · Segmentation · Tracking

1 Introduction

In clinical practice, echocardiography is frequently-used for assessing cardiac function and diagnosing cardiovascular disease (CVD) due to its advantages of

⋆ corresponding author

being real-time, radiation-free and low cost. However, it suffers from defects of large noise and low contrast, resulting in great difficulties for visual inspection. Manual annotation of the key structures is usually time-consuming and has large inter-/intra-observer variability. Concerning the ejection fraction (EF), the correlations of inter-/intra-observer agreement are only 0.801 and 0.896, respectively [9]. This indicates a demand of automatic methods for highly accurate segmentation, and precise estimation of EF from cardiac ultrasonic sequences.

Previous state-of-the-art methods [11,14] analyze each frames independently, and have achieved comparable performance with the inter-observer agreements for ventricle borders segmentation. However, eligible left ventricle segmentation does not necessary lead to accurate EF estimation, as we found in Table 3. A prerequisite for precise EF estimation is to preserve the temporal consistency of the segmentations for all frames under condition that only ED and ES frames are annotated.

Benefits of leveraging temporal information for cardiac image segmentation have been proved by previous works [1,4,5,8,10,12,13,15,16]. Recurrent neural network was employed to extract spatial-temporal features [5,10,15] for cardiac segmentation. Specially, [10] utilize hierarchically convolutional LSTMs for spatiotemporal feature embedding and a double-branch module for joint learning of LV segmentation and chamber view classification. Optical flow was utilized to extract the relative motion vector between consecutive frames, and then the learned motion vector was used as an complementary input to the original images to learn the segmentation by neural networks [4,8,16], or to propagate the labels from ED to ES [12]. Instead, [13] utilized the same encoder for joint learning of motion estimation and segmentation on cardiac magnetic resonance images, demonstrating the effectiveness of motion learning as a regularizer for the segmentation task. However, the motion estimation between frames that have large time interval in [13] may lead to large motion estimated error for echocardiographic images that have blurry (or even disappeared) boundaries. Although the temporal information in these methods could help relieve temporally inconsistent segmentation results of the whole sequence, they still suffer from two weaknesses that will be solved in our work: 1) requirements of densely labeled cardiac sequences [4,5,15,16], and 2) lack of explicit and effective enforcement of temporal consistency of the consecutive frames [4,5,8,10,12,13,15,16].

In this paper, we overcome these weaknesses and propose a novel temporally consistent segmentation method called Co-Learning of segmentation and tracking on Appearance and Shape level (CLAS) for sparsely labeled echocardiographic sequence. CLAS leverages temporal information by two co-learning strategies to improve the segmentation accuracy and enforce temporal consistency. Our contributions lies in three folds:

1) We design an appearance-level co-learning (CLA) strategy for segmentation and tracking, to collaboratively learn eligible shapes and motion fields from both labeled and unlabeled frames of the whole sequence. These results act as a warm start for further enhancement of temporal consistency.

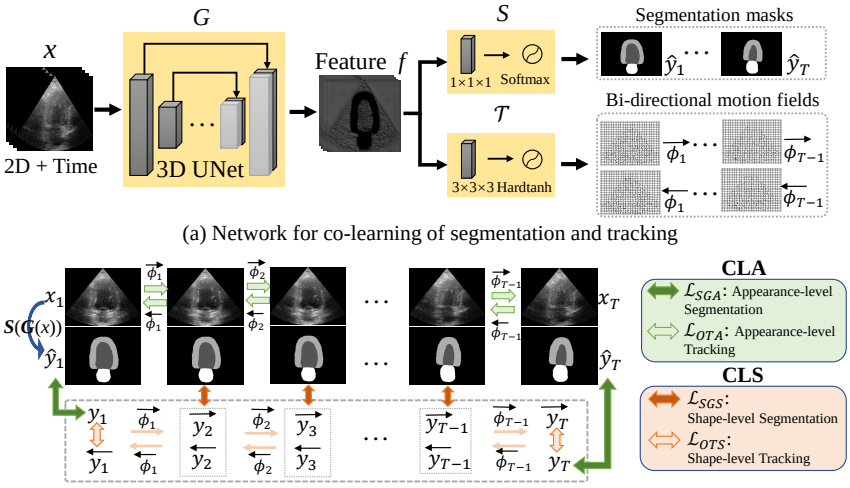

Fig. 1. Overview of the proposed CLAS, which explores the mutual benefits of segmentation and tracking with the network architecture (a) on appearance and shape levels (b) to explicitly enhance the temporal consistency of the segmentation.

2) We design a shape-level co-learning (CLS) strategy to enhance the temporal consistency of the segmentation results. Specifically, bi-direction tracking of cardiac shapes improve the previous motion fields learned from appearance. The segmentation is then further improved by pseudo labels deformed using these motion fields. The two tasks mutually benefit each other, leading to enhanced temporal consistency.

3) We validate the performance with a sparsely labeled echocardiography dataset, and achieve state-of-the-art performance that is comparable to the intra-observer variability for segmentation of endocardium (Endo) and epicardium (Epi). Especially, due to the temporal consistency enforcement, we achieve a correlation of 0.926 for EF estimation, which is obviously better than the inter-/intra-observer variability 0.801/0.896 and best reported method [9] 0.845.

2 Method

The overview of the proposed CLAS is illustrated in Fig.1. The network architecture (2.1) contains a 3D UNet for feature learning, on top of which two heads are used for co-learning of the segmentation and tracking. The co-learning operates on two levels. For CLA (2.2), supervised segmentation for the ED/ES frames and unsupervised tracking for the intermediate frames are collaboratively learned to benefit the feature learning with large amount of unlabeled data, and to provide eligible estimation of the segmentation and motion field. For CLS (2.3), pseudo labels that are deformed from the labeled ED/ES frames are used to further im-

prove the segmentation part, while the bi-directional tracking between ED and ES are used to improve the tracking, i.e., enhance the temporal consistency of the segmented cardiac sequences. The overall objective function is described in 2.4.

2.1 Co-learning Network architecture

We design a co-learning network architecture, which enables us to exploit the mutual benefits of the two tasks during the feature learning from cardiac appearance (CLA) and among shapes of ground truth labels, segmentation predictions, and those obtained in tracking (CLS). Our co-learning network differs from previous works that used a predetermined optical flow [4,8,16], or learned it directly from the compact representation [13].

Given a 2D ultrasonic cardiac sequence $\mathbf{x} = \{x_t\}_{t=1,2,...,T}$ with y_1, y_T as labels of ED ($t = 1$) and ES ($t = T$), we aim at obtaining accurate and temporally consistent segmentation masks for the whole cardiac sequence. The overall network contains three parts: 1) a basic 3D UNet G as backbone for feature extraction, 2) a segmentation head S that outputs the predicted masks $\hat{y}$ of the background, Endo, myocardium (Myo) and left atrium (LA), and 3) a tracking head $\mathcal{T}$ that outputs the bi-directional motion fields $[\overrightarrow{\phi}_t, \overleftarrow{\phi}_t](t = 1, ...T-1)$, which can be utilized to deform images or labels to adjacent frames using spatial transformation [7].

2.2 Appearance-level co-learning

We design a CLA strategy to optimize the whole network with a supervised appearance-level segmentation task and an unsupervised appearance-level object tracking task. The segmentation task learns robust shape information from the appearance of the cardiac images by G and output the shape masks of cardiac structures by optimizing the segmentation objective $\mathcal{L}_{SGA}$. The tracking task learns the bi-directional motion fields of adjacent frames from the shape information in G by optimizing the tracking objective $\mathcal{L}_{OTA}$.

The two tasks benefit each other as follows: the supervised segmentation learns from ED and ES frames and makes G encode the shape information, and acts as a noise removing filter for the tracking of cardiac structures in the noisy ultrasonic images; the tracking task learns from the whole sequence, including the intermediate unlabeled frames, help feature learning of G by feeding more training images and identifying the moving objects, thus benefits the cardiac structure segmentation.

The supervised appearance-level segmentation loss $\mathcal{L}_{SGA}$ is defined by a combination of cross-entropy (CE) loss and multi-class Dice loss:

$$\mathcal{L}_{SGA} = \frac{1}{2\,|\Omega|} \sum_{t\in\{1,T\}} \sum_{c=1}^{C} \Big(- y^{c,t} \cdot \log \hat{y}^{c,t} + \frac{1}{|C|}\big(1 - \frac{|y^{c,t} \cdot \hat{y}^{c,t}|}{|y^{c,t}| + |\hat{y}^{c,t}|}\big)\Big) \qquad (1)$$

where $c \in \{background, Endo, Myo, LA\}$, in echocardiography, y and $\hat{y}$ are ground truth and predicted probability respectively.

The unsupervised appearance-level object tracking loss $\mathcal{L}_{OTA}$ is defined by a combination of local cross-correlation (CC, [2]) and smooth loss $\mathcal{L}_{sm}$ [3]:

$$\mathcal{L}_{OTA} = \frac{-1}{2(T-1)} \sum_{t=1}^{T-1} \big(CC(x_{t+1}, x_t \circ \overrightarrow{\phi}_t) + CC(x_t, x_{t+1} \circ \overleftarrow{\phi}_t)\big) + \frac{\gamma}{2(T-1)} \sum_{t=1}^{T-1} \big(\mathcal{L}_{sm}(\overrightarrow{\phi}_t) + \mathcal{L}_{sm}(\overleftarrow{\phi}_t)\big) \tag{2}$$

where $\circ$ denotes spatial transformation, γ is a regularization parameter for the smooth term. Minimizing $\mathcal{L}_{OTA}$ encourages primary object tracking through the whole cardiac sequence.

In spite of the above mentioned benefits, there is still no explicit enforcement of temporal consistency for the predicted masks across the whole sequences. Therefore, we propose a shape-level co-learning scheme in next section.

2.3 Shape-level co-learning

We further design a CLS strategy to enforce temporal consistency of the previously estimated shape masks and motion fields. Specifically, we introduce the CLS strategy that includes: 1) unsupervised segmentation of the intermediate frames with pseudo labels that are deformed from the true masks of ED/ES, and 2) semi-supervised bi-directional tracking of the shape between ED and ES.

The shape-level tracking operates on cardiac shapes that are free of noise and background, therefore improves the motion fields learned in CLA. The enhanced motion fields can improve the segmentation performance by providing high quality pseudo labels for the intermediate frames. The improved segmentation can bring further improvement for the shape-level tracking by shape embedding in G. In the end, this mutual improvement of segmentaion and tracking can effectively enforce the temporal consistency of the predicted segmentation masks.

We first generate pseudo labels $\overrightarrow{y}^{c,t}, (t = 2, ..., T)$ by forward deformation of the true ED label $y^{c,1}$ using a sequential motion fields as: $\overrightarrow{y}^{c,t} = [[y^{c,1} \circ \overrightarrow{\phi}_1] \circ \overrightarrow{\phi}_2] \circ ... \circ \overrightarrow{\phi}_{t-1}$. Similarly, $\overleftarrow{y}^{c,t}, (t = 1, ..., T-1)$ can be obtained from the true ES label $y^{c,T}$ by backward deformation. With these pseudo labels, the shape-level segmentation helps improve the segmentation network using the large amount of intermediate frames $(t = 2, ...T-1)$:

$$\mathcal{L}_{SGS} = \frac{1}{2(T-2)|C||\Omega|} \sum_{t=2}^{T-1} \sum_{c=1}^{C} (2 - \frac{|\overrightarrow{y}^{c,t} \cdot \hat{y}^{c,t}|}{|\overrightarrow{y}^{c,t}| + |\hat{y}^{c,t}|} - \frac{|\overleftarrow{y}^{c,t} \cdot \hat{y}^{c,t}|}{|\overleftarrow{y}^{c,t}| + |\hat{y}^{c,t}|}) \tag{3}$$

The shape-level tracking helps improve the learned motion fields by matching the ED/ES labels with the pseudo labels $\overleftarrow{y}_1$, and $\overrightarrow{y}_T$:

$$\mathcal{L}_{OTS} = \frac{1}{2|C||\Omega|}\sum_{c=1}^{C}(2 - \frac{|\overrightarrow{y}^{c,T} \cdot y^{c,T}|}{|\overrightarrow{y}^{c,T}| + |y^{c,T}|} - \frac{|\overleftarrow{y}^{c,1} \cdot y^{c,1}|}{|\overleftarrow{y}^{c,1}| + |y^{c,1}|}) \tag{4}$$

2.4 Overall objective

We optimize the network in two stages. For stage one, CLA is used:

$$\mathcal{L}_{stage1} = \mathcal{L}_{SGA} + \mathcal{L}_{OTA} \tag{5}$$

Minimizing $\mathcal{L}_{stage1}$ results in eligible segmentation masks and motion fields of the whole sequence that acts as a warm start for CLS. Then, for stage two, both CLA and CLS are employed:

$$\mathcal{L}_{stage2} = \mathcal{L}_{SGA} + \mathcal{L}_{OTA} + \alpha\mathcal{L}_{SGS} + \beta\mathcal{L}_{OTS} \tag{6}$$

where α, β are trade-off parameters.

3 Experiments and analysis

3.1 Experimental Configuration

Dataset. CAMUS [9] consists of 500 patients' 2D echocardiography from ED to ES phase with two-chamber (2CH) and four-chamber (4CH) views. Half of the patients have an EF lower than 45% (being considered at pathological risk), 19% of the images have poor quality. The ground truths of Endo, Myo, LA on ED and ES frames for both 2CH and 4CH were provided. For CAMUS challenge, 450 patients are used for model training, other 50 patients for testing, for which ground truths are not accessible. During training, ten frames starting from ED and ending with ES, were sampled with equal interval from each sequence. All the images were resized to 256×256 and the intensity was normalized to [-1,1].

Experiment setup. We train the network for 30 epochs with batch size 4, and use Adam optimizer with initial learning rates 10^{-4} for both G and S, $0.5 * 10^{-4}$ for $\mathcal{T}$, of which parameters initialization is Gaussian distribution $N(0, 10^{-5})$. Learning rate is reduce to 10^{-5} after 25 epochs. The hyperparameters α, β, and γ are set to be 0.2, 0.4 and 10, respectively. We optimize $\mathcal{L}_{stage1}$ in the first 10 epochs and $\mathcal{L}_{stage2}$ in the last 20 epochs. For testing phase, we train the model using all training data (450 patients) with same setting.

3.2 Results and Analysis

Tracking performance. We first evaluate the effects of segmentation and shape-level tracking tasks to object tracking, which is measured by the Dice coefficients between pairs of $y_{c,1}$, $\overleftarrow{y}_{c,1}$ and $y_{c,T}$, $\overrightarrow{y}_{c,T}$ for $c \in \{endo, epi, LA\}$.

Table 1. Mean Dice of 2CH and 4CH views for object tracking on training set using 10-fold cross-validation. Three methods have significant difference with $p \ll 0.001$.

Method	ED			ES		
	Endo	Epi	LA	Endo	Epi	LA
OTA	0.816	0.900	0.802	0.798	0.889	0.807
$OTA + OTS$	0.909	0.939	0.861	0.887	0.927	0.889
CLAS	**0.923**	**0.945**	**0.876**	**0.903**	**0.935**	**0.900**

Table 1 demonstrates the results of tracking performance by using appearance tracking only (OTA), two levels of tracking (OTA+OTS), and CLAS. We observed that: 1) only appearance tracking cannot lead to accurate performance given the low contrast and noisy echocardiography; 2) shape-level tracking helps eliminate the background and noisy information, thereby improving tracking of Endo, Epi and LA; 3) CLAS, with further shape-level segmentation, enforces the temporal consistency, and leads to even better tracking.

Segmentation performance. The accuracy and consistency of segmentation across the whole cardiac sequence are examined on the test set. Fig.2 shows the visualization of segmentation results by UNet and CLAS. In the left, we observe that CLAS achieves better Dice coefficients and low EF error. The larger estimation error of UNet falsely indicates the case as a pathological one (EF below 45%). In the right of Fig.2, from the curves of the areas of Endo, Myo, and LA, we can see that results from CLAS are smoother across the whole sequence than those of UNet, validating its better temporal consistency.

Table 2 shows segmentation performance on the test set in terms of Dice, Hausdorff distance (HD) and mean absolute distance (MAD). We concluded that the proposed CLAS achieves excellent segmentation performance that is comparable with the intra-observer variability, and outperforms the existing best methods (UNet) reported on the CAMUS challenge website, for all cardiac structures including Endo, Epi and LA.

Volume and EF estimation. LV volumes (EDV and ESV) and EF are very important indicator of heart disease. We calculate them from segmentation masks of ED and ES frames of 2CH and 4CH views using Simpson's biplane method of discs [6], and display the results in Table 3. We concluded that CLAS gives accurate estimation of LV volumes that are better than inter-observer variability, and EF estimation that are even better than intra-observer variability. It's also clear that CLAS outperforms existing methods. Especially for EF, CLAS delivers the highest correlation (0.926) with the ground truth and very low bias (0.1), compared to the best existing results (0.845 for UNet).

[1] https://www.creatis.insa-lyon.fr/Challenge/camus/

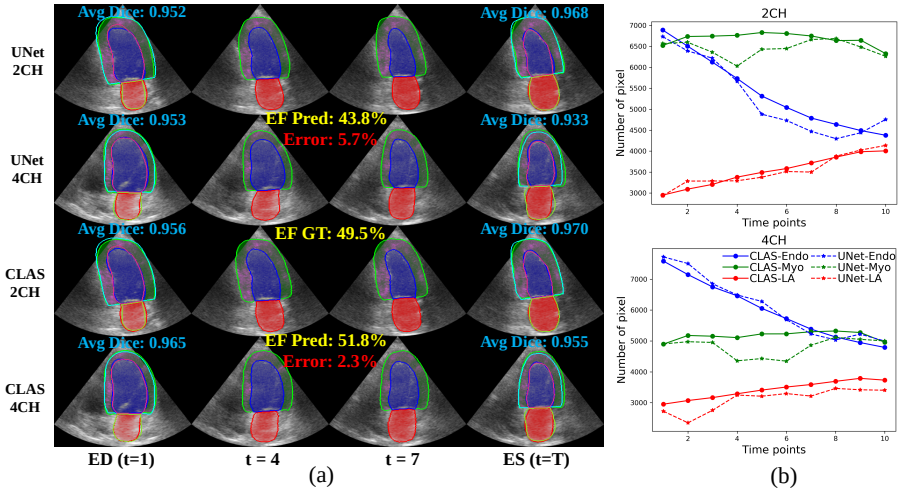

Fig. 2. Visualization of segmentation and its temporal consistency for UNet and CLAS. (a) Segmentation results of two-chamber (2CH) and four-chamber (4CH) view sequences for one patient. Dice coefficients and the predicted EF are shown. (b) Curves of predicted areas for Endo, Myo and LA in the whole cardiac sequence.

Table 2. Segmentation results on test set (50 patients). Mean performance of 2CH and 4CH views are shown. The results of UNet represent the best reported results of CAMUS challenge to date. Note that 10 poor quality patients were excluded for the inter-observer and intra-observer results [9].

ED	Endo			Epi			LA		
	Dice	HD	MAD	Dice	HD	MAD	Dice	HD	MAD
inter-observer	0.919	6.0	2.2	0.913	8.0	3.5	-	-	-
intra-observer	0.945	4.6	1.4	0.957	5.0	1.7	-	-	-
UNet[1]	0.936	5.3	1.7	0.956	5.2	1.7	0.889	5.7	2.2
ACNNs [11]	0.936	5.6	1.7	0.953	5.9	1.9	0.881	6.0	2.3
CLAS	**0.947**	**4.6**	**1.4**	**0.961**	**4.8**	**1.5**	**0.902**	**5.2**	**1.9**
ES	Endo			Epi			LA		
	Dice	HD	MAD	Dice	HD	MAD	Dice	HD	MAD
inter-observer	0.873	6.6	2.7	0.890	8.6	3.9	-	-	-
intra-observer	0.930	4.5	1.3	0.951	5.0	1.7	-	-	-
UNet[1]	0.912	5.5	1.7	0.946	5.7	1.9	0.918	5.3	2.0
ACNNs [11]	0.913	5.6	1.7	0.945	5.9	2.0	0.911	5.8	2.2
CLAS	**0.929**	**4.6**	**1.4**	**0.955**	**4.9**	**1.6**	**0.927**	**4.8**	**1.8**

Table 3. LV volume and EF estimation on test set (50 patients) of CAMUS dataset.

Methods	EDV			ESV			EF		
	corr	bias(ml)	std	corr	bias(ml)	std	corr	bias(%)	std
inter-observer	0.940	18.7	12.9	0.956	18.9	9.3	0.801	-9.1	8.1
intra-observer	0.978	-2.8	7.1	0.981	-0.1	5.8	0.896	-2.3	5.7
UNet[1]	0.926	7.2	15.6	0.960	4.4	10.2	0.845	0.1	7.3
ACNNs [11]	0.928	2.8	15.5	0.954	2.0	10.1	0.807	0.3	8.3
RAL [10]	0.952	-7.5	11.0	0.960	-3.8	9.2	0.839	-0.9	6.8
CLAS	**0.958**	-0.7	15.1	**0.979**	-0.0	8.4	**0.926**	-0.1	6.7

4 Conclusion

We proposed a co-learning model CLAS for temporal-consistent segmentation of echocardiographic sequences with sparsely labeled data. We first used appearance-level co-learning to learn an eligible prediction for the segmentation and tracking tasks from both labeled and unlabeled frames. Then we used shape-level co-learning to enhance the temporal consistency of the predicted masks across the whole sequence. This two-level co-learning strategy iteratively improved the segmentation and tracking, and made the proposed CLAS outperforms existing methods obviously in aspects of multiple cardiac structures segmentation, as well as estimation of the clinically significant measurements (EDV, ESV, and EF).

References

1. Al-Kadi, O.S.: Spatio-temporal segmentation in 3d echocardiographic sequences using fractional brownian motion. IEEE Transactions on Biomedical Engineering (2019)
2. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging **38**(8), 1788–1800 (2019)
4. Chen, S., Ma, K., Zheng, Y.: Tan: Temporal affine network for real-time left ventricle anatomical structure analysis based on 2d ultrasound videos. arXiv preprint arXiv:1904.00631 (2019)
5. Du, X., Yin, S., Tang, R., Zhang, Y., Li, S.: Cardiac-deepied: Automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network. IEEE Journal of Translational Engineering in Health and Medicine **7**, 1–10 (2019)
6. Folland, E., Parisi, A., Moynihan, P., Jones, D.R., Feldman, C.L., Tow, D.: Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. a comparison of cineangiographic and radionuclide techniques. Circulation **60**(4), 760–766 (1979)

7. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems. pp. 2017–2025 (2015)
8. Jafari, M.H., Girgis, H., Liao, Z., Behnami, D., Abdi, A., Vaseli, H., Luong, C., Rohling, R., Gin, K., Tsang, T., et al.: A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography data. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 29–37. Springer (2018)
9. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Transactions on Medical Imaging **38**(9), 2198–2210 (2019)
10. Li, M., Zhang, W., Yang, G., Wang, C., Zhang, H., Liu, H., Zheng, W., Li, S.: Recurrent aggregation learning for multi-view echocardiographic sequences segmentation. In: MICCAI. pp. 678–686. Springer (2019)
11. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., De Marvao, A., Dawes, T., O'Regan, D.P., et al.: Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. IEEE Transactions on Medical Imaging **37**(2), 384–395 (2017)
12. Pedrosa, J., Queirós, S., Bernard, O., Engvall, J., Edvardsen, T., Nagel, E., D'hooge, J.: Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based b-spline explicit active surfaces. IEEE Transactions on Medical Imaging **36**(11), 2287–2296 (2017)
13. Qin, C., Bai, W., Schlemper, J., Petersen, S.E., Piechnik, S.K., Neubauer, S., Rueckert, D.: Joint learning of motion estimation and segmentation for cardiac mr image sequences. In: MICCAI. pp. 472–480. Springer (2018)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
15. Savioli, N., Vieira, M.S., Lamata, P., Montana, G.: Automated segmentation on the entire cardiac cycle using a deep learning work-flow. In: 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS). pp. 153–158. IEEE (2018)
16. Yan, W., Wang, Y., Li, Z., Van Der Geest, R.J., Tao, Q.: Left ventricle segmentation via optical-flow-net from short-axis cine mri: preserving the temporal coherence of cardiac motion. In: MICCAI. pp. 613–621. Springer (2018)