Temporally Consistent Segmentations from Sparsely Labeled Echocardiograms Using Image Registration for Pseudo-labels Generation

Matteo Tafuro(✉), Gino Jansen, and Ivana Išgum

University of Amsterdam, Amsterdam, The Netherlands
tafuromatteo00@gmail.com, {g.e.jansen,i.isgum}@amsterdamumc.nl

Abstract. The segmentation of the left ventricle in echocardiograms is crucial for diagnosing cardiovascular diseases. However, current deep learning methods typically focus on 2D segmentations and overlook the temporal information in ultrasound sequences. This choice might be caused by the scarcity of manual annotations, which are typically limited to end-diastole and end-systole frames. Therefore, we propose a method that trains temporally consistent segmentation models from sparsely labeled echocardiograms. We leverage image registration to generate pseudo-labels for unlabeled frames enabling the training of 3D models. Using a state-of-the-art convolutional neural network, 3D nnU-Net, we delineate the left ventricle (LV) cavity, LV myocardium, and left atrium. Evaluation on the CAMUS dataset demonstrates the quality and robustness of the generated pseudo-labels, serving as effective training data for subsequent segmentation. Additionally, we evaluate the segmentation model both intrinsically, measuring accuracy and temporal consistency, and extrinsically, estimating cardiac function markers like ejection fraction and left ventricular volumes. The results show accurate delineation of the cardiac structures that evolves smoothly over time, effectively demonstrating the model's accuracy and temporal consistency.

Keywords: Left ventricle segmentation · Echocardiography · Image registration · Pseudo-labels

1 Introduction

The analysis of 2D transthoracic echocardiograms is crucial in clinical cardiology for disease diagnosis and treatment selection [2]. The analysis comprises the extraction of a number of quantitative markers of cardiac function, such as the ejection fraction (EF) and the chamber volumes [10]. Extraction of these quantitative markers requires accurate and precise delineation of the cardiac anatomy. However, manual expert annotation is a time-consuming task associated with high inter- and intra-rater variability [1]. Existing commercial solutions allow semi- or fully-automatic delineation of the cardiac structures, but they are

B. Kainz et al. (Eds.): ASMUS 2023, LNCS 14337, pp. 195–204, 2023.
https://doi.org/10.1007/978-3-031-44521-7_19

typically limited to the segmentation of the end-diastolic (ED) and end-systolic (ES) frames [14].

The focus on ED and ES frames is also reflected in most published research utilizing machine learning approaches [12]. As these methods require large and diverse datasets for training, collecting annotations of full sequences has not been the prime focus. The most commonly used public datasets for echocardiography segmentation, CAMUS [8] and EchoNet-Dynamic [11], provide manual labels[1] for the ED and ES frames only. Therefore, most current state-of-the-art (SoTA) segmentation methods rely solely on expert annotations for these two frames [12]. Despite achieving performance within the margins of intra-observer variability [15,18], these methods do not address the smooth evolution of the cardiac structures over time, leading to temporally inconsistent predictions [12].

Since preserving the temporal consistency of the segmentations is beneficial for precise EF estimation [18], several studies have addressed this issue. Some approaches combine temporal and multi-view information using 3D CNN and convolutional LSTM [9]. Others enforce temporal smoothness through post-processing [12] or leverage optical flow for segmentation accuracy improvement [3,21]. Wei et al. introduced CLAS, an end-to-end approach that combines co-learning of appearance and shape features with the generation of left ventricle (LV) pseudo-labels for the intermediate time points [18]. These LV pseudo-labels are obtained by warping the ground truth maps to other frames using optical flow. Chen et al. further added data augmentation (A-CLAS) [4], while Wei et al. introduced two auxiliary tasks, view classification and EF regression, and proposed the multi-task version of CLAS (MCLAS) [19].

Although these methods achieve temporally consistent segmentation, their reliance on co-learning and pseudo-labels makes them computationally complex. Moreover, their constrained end-to-end nature restricts their modularity. In contrast, we present a method that addresses pseudo-label generation and temporally smooth segmentation as separate components. It leverages an unsupervised image registration model to sequentially estimate the deformations between frames and generate pseudo-labels through the warping of the available segmentation maps. The generated pseudo-labels allow supervised training of arbitrary 3D (2D+time) segmentation networks. To this end, we train a 3D nnU-Net [7] to delineate the LV cavity, LV myocardium and left atrium. We evaluate the proposed approach on the public CAMUS dataset [8], demonstrating that it generates reliable pseudo-labels that bring significant benefits to the downstream segmentation task. The segmentation model exhibits remarkable accuracy in delineating cardiac structures while preserving spatiotemporal smoothness, ultimately yielding accurate EF estimations.

2 Method

To obtain accurate and temporally consistent 3D (2D+time) segmentations from a sparsely labeled dataset, the method first generates the pseudo-labels for those

[1] To aid readability, it may be worth specifying that "segmentations" and "labels" are used interchangeably throughout the paper.

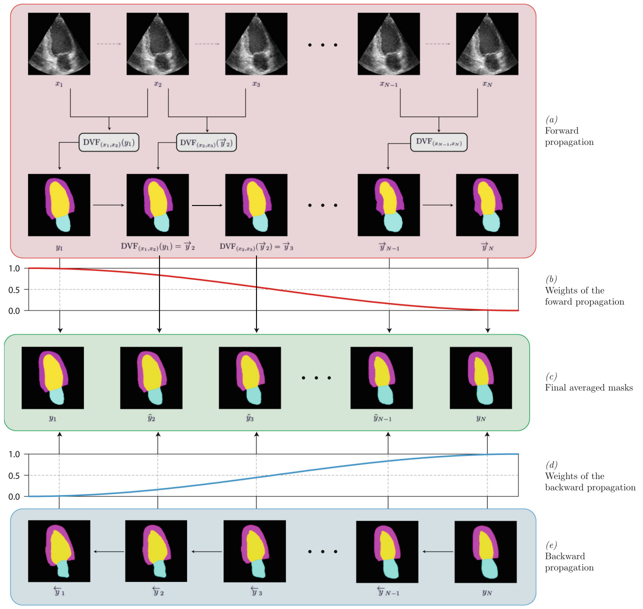
x_1
x_2
x_3
x_{N-1}
x_N
$\mathrm{DVF}_{(x_1,x_2)}(y_1)$
$\mathrm{DVF}_{(x_2,x_3)}(\overrightarrow{y}_2)$
$\mathrm{DVF}_{(x_{N-1},x_N)}$
y_1
$\mathrm{DVF}_{(x_1,x_2)}(y_1) = \overrightarrow{y}_2$
$\mathrm{DVF}_{(x_2,x_3)}(\overrightarrow{y}_2) = \overrightarrow{y}_3$
$\overrightarrow{y}_{N-1}$
$\overrightarrow{y}_N$
(a) Forward propagation
1.0
0.5
0.0
(b) Weights of the foward propagation
y_1
$\tilde{y}_2$
$\tilde{y}_3$
$\tilde{y}_{N-1}$
y_N
(c) Final averaged masks
1.0
0.5
0.0
(d) Weights of the backward propagation
$\overleftarrow{y}_1$
$\overleftarrow{y}_2$
$\overleftarrow{y}_3$
$\overleftarrow{y}_{N-1}$
y_N
(e) Backward propagation

Fig. 1. The proposed image registration-based pseudo-labels generation method. The provided segmentations are propagated from ED to ES *(a)* and from ES to ED *(e)*. The masks from the two directions are aggregated as described in Sect. 2.1 and weighted according to a sinusoidal function *(b and d)*.

frames that lack reference segmentations. This is done through the sequential application of image registration. Thereafter, the method uses these pseudo-labels to augment sparse reference annotations and train a segmentation model.

2.1 Pseudo-labels Generation

Echocardiography acquisition consists of a sequence of image frames x_t, $\forall t \in \{1, 2, .., N\}$ showing the evolution of the heart over the cardiac cycle. Given the reference segmentation for the ED and ES frames, unsupervised deformable image registration (DIR) is exploited to segment the frames lacking segmentation masks. The registration's dense displacement vector field (DVF) is employed to warp the segmentation of frame x_t (y_t) to frame x_{t+1}, resulting in a pseudo-segmentation $\overrightarrow{y}_{t+1}$ of frame x_{t+1}. Specifically, the available ED segmentation is iteratively forward-propagated through the sequence to produce $\overrightarrow{y}_t$, $\forall t \in \{1, 2, .., N\}$. Akin, backward-propagating the ES segmentation mask returns a set of $\overleftarrow{y}_t$, $\forall t \in \{1, 2, .., N\}$.

To mitigate error accumulation caused by sequential registrations, the two sets of pseudo-labels $\overrightarrow{\mathbf{y}}$ and $\overleftarrow{\mathbf{y}}$ are combined using a weighted average of their class-wise signed distance maps. Specifically, for each class and time point, a binary mask is extracted and the signed distance to its edges is computed. The resulting distance maps, $d(\overrightarrow{y}_{t,C})$ and $d(\overleftarrow{y}_{t,C})$, are then weighted-averaged to return an image with negative values outside the object, positive values inside and zero crossings at the object boundaries. Thresholding this image at zero produces the final mask. The final *bidirectional* method is illustrated in Fig. 1 and defined mathematically in Eq. 1:

$$\tilde{y}_{t,C} = \left(d(\overrightarrow{y}_{t,C}) \cdot \cos^2 \frac{\pi}{2N} t + d(\overleftarrow{y}_{t,C}) \cdot \sin^2 \frac{\pi}{2N} t \right) > 0 \tag{1}$$

where $\overleftrightarrow{y}_{t,C}$ is the binary mask corresponding to class C at time point t, $d(\cdot)$ is the distance transform operation and N is the ED-to-ES sequence length. The weights are determined according to the temporal proximity of $d(\overrightarrow{y}_{t,C})$ and $d(\overleftarrow{y}_{t,C})$ to the ED and ES reference segmentations, respectively. More specifically, they are designed to decrease from 1 to 0 in the direction of the propagation, thereby exerting more influence on the forward direction at the beginning of the sequence and on the backward direction at the end. This further mitigates error accumulation and improves the accuracy of the object representation.

In this work, an unsupervised deep learning registration framework is utilized to perform image alignment through CNNs [6]. The method exploits image similarity between fixed and moving image pairs, B-splines as the transformation model, and supports coarse-to-fine alignment. Additionally, the loss function combines the negative normalized cross correlation $\mathcal{L}_{NCC}$ with the bending energy penalty P: $\mathcal{L} = \mathcal{L}_{NCC} + \alpha P$ [13]. The regularization term P minimizes the second order derivative of local transformations, thereby enforcing global smoothness and preventing anatomically implausible image folding.

2.2 Segmentation

The reference segmentations of the echocardiograms are augmented with the pseudo-labels to provide densely labeled reference sequences. This enables the training of 3D (2D+time) segmentation models, which are designed to be trained on densely annotated data. By encoding the time dimension as the third dimension in convolutional space, a 3D model can learn spatiotemporal features that encourage temporally smooth predictions. To this end, a 3D nnU-Net is trained on the augmented dataset (*3D Dense* nnU-Net) [7].

2.3 Evaluation

Both the generated pseudo-labels and the predicted segmentations are intrinsically evaluated by overlap and boundary metrics: the DICE coefficient (DC), the mean absolute surface distance (MAD) and the 2D Hausdorff Distance (HD). The metrics are calculated per frame and subsequently averaged over an entire video. Additionally, the segmentation models are evaluated extrinsically through quantification of EF and LV volumes at end-diastole and end-systole, EDV and ESV. To aggregate dataset-level statistics for these indices, the correlation coefficient, bias and mean absolute error (MAE) are calculated between the reference and automatically obtained values. Finally, the temporal consistency of the automatic segmentation is assessed by tracking the area of a given class over time. The smoothness of a sequence is computed as the integral of the second derivative of the resulting curve (*area curve*). To account for changes in the slope of the area curve and to prevent the loss of information due to opposite bending, the second derivative is squared prior to integration. The final smoothness metric is defined in Eq. 2, with N being the ED-to-ES sequence length and $a_C(t)$ the area of class C at time point t.

$$\text{Smoothness} = \int_1^N \left(a''_C(t)\right)^2 dt, \tag{2}$$

3 Experiments

Two main experiments were conducted[2]. First, the pseudo-labels were generated and evaluated against reference segmentations. Second, the pseudo-labels were utilized to complement the original dataset and train the segmentation network.

All the models were implemented in PyTorch 1.12.1 and trained using 2 Intel Xeon Gold 6128 CPUs (6 cores, 3.40GHz) and a GeForce RTX 2080 Ti.

3.1 Data and Preprocessing

This study uses two public datasets: CAMUS [8] and TED [12]. CAMUS contains 2D echocardiograms with 2-chambers (2CH) and 4-chambers (4CH) views of half-cycle sequences (from ED to ES) of 500 patients (450 training, 50 test). Manual annotations of the LV cavity, LV myocardium and LA are provided for the ED and ES frames only. TED is a subset of CAMUS that comprises 98 full cycle 4CH sequences, with manual segmentations of the LV cavity and the LV myocardium for the *whole* cardiac cycle. 94 sequences are part of the CAMUS training set and 4 of the test set.

Prior to analysis, all images are resized to 512 × 512 px, and the pixel spacing is scaled proportionally to preserve the anisotropic nature of the data.

3.2 Pseudo-label Generation

The DIR model was trained on the CAMUS training set after leaving out the overlapping 94 TED echocardiograms, resulting in a set of 806 echo sequences. Successively, the frame-wise alignment quality was evaluated against these 94 left-out TED sequences. The DIR network was trained on every intra-patient combination of two frames from the registration training set. The training was performed in 10,000 iterations and used a batch size of 32, the AMSGrad variant of the ADAM optimizer and a learning rate of 10^{-3}. Hyperparameters such as

[2] The code is publicly available at https://github.com/matteo-tafuro/temporally-consistent-echosegmentation.

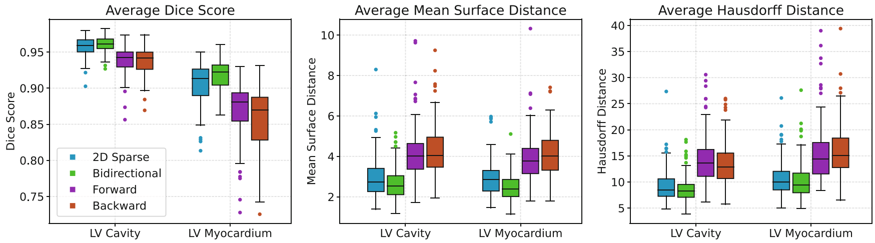

Fig. 2. Comparison of the pseudo-labels quality in terms of geometric metrics evaluated on the densely annotated TED dataset.

the size, the number of kernels and the B-spline grid spacing were determined in preliminary experiments by testing values between 2 and 128. Optimal results were obtained with 32 kernels of size 32 × 32, a grid spacing of 32 and a regularization hyperparameter of 1.0 to prevent folding. Coarse-to-fine registration did not improve performance, hence simple one-stage alignment was employed.

Figure 2 demonstrates the performance of pseudo-label generation using different approaches. Pseudo-labels were compared with predictions from a SoTA 2D nnU-Net trained on the original sparsely labeled CAMUS dataset (*2D Sparse nnU-Net*). Figure 3 highlights the effectiveness of our label propagation method in generating temporally consistent pseudo-labeled segmentation maps, promoting coherent feature learning during the segmentation step.

3.3 Segmentation

The *3D Dense* nnU-Net was trained and tested on the sparsely labeled CAMUS datasets augmented with pseudo-labels, allowing direct comparison with related works. In addition, the *3D Dense* model was evaluated against two baselines: a 2D nnU-Net trained on the sparsely labeled CAMUS dataset (*2D sparse* nnU-Net) and a 2D nnU-Net trained on the *augmented* CAMUS dataset (*2D Dense* nnU-Net). Each nnU-Net was trained for 1,000 epochs, using 5-fold cross-validation with an interleaved test setup. After training, the framework automatically selected the best U-Net configuration. Finally, three SoTA CLAS-based methods [4,18,19] were included for comparison. The models were compared in terms of (i) accuracy of the LV cavity, LV myocardium and LA segmentation at ED and ES; (ii) estimation of EF, EDV, and ESV; (iii) temporal smoothness.

The average segmentation performance on the ED and ES frames of the test set is listed in Table 1; the results of the EDV, ESV and EF estimation are displayed in Table 2; the observed temporal consistency of frame-by-frame predictions is shown in Fig. 4; finally, the area curve of a test patient is depicted in Fig. 5 along with the corresponding ED and ES predictions.

Table 1. Average segmentation results at ED and ES on the (sparsely annotated) CAMUS test set. The intra-observer variability results (in blue) are taken from the official CAMUS website and are not provided for the left atrium. The best value per column is indicated in bold.

	ED									ES								
	LV Cavity			LV Myocardium			LA			LV Cavity			LV Myocardium			LA		
	DC	HD	MAD	DC	HD	MAD	DC	HD	MAD	DC	HD	MAD	DC	HD	MAD	DC	HD	MAD
Intra-observer	0.945	4.6	1.4	0.957	5.0	1.7	–	–	–	0.930	4.5	1.3	0.951	5.0	1.7	–	–	–
CLAS [18]	0.947	4.6	1.4	0.961	4.8	1.5	0.902	5.2	**1.9**	0.929	4.6	1.4	0.955	4.9	1.6	0.927	4.8	1.8
A-CLAS [4]	0.942	–	–	0.955	–	–	0.887	–	–	0.923	–	–	0.950	–	–	0.916	–	–
2D Sparse	**0.955**	**4.1**	**1.2**	**0.965**	4.4	**1.4**	**0.906**	**4.9**	**1.9**	0.938	**4.0**	**1.2**	**0.959**	**4.3**	**1.5**	**0.937**	**4.3**	**1.5**
2D Dense	0.950	4.2	1.3	0.963	**4.3**	**1.4**	0.902	5.0	2.0	0.934	4.2	1.3	0.957	4.5	**1.5**	0.933	4.5	1.7
3D Dense	0.952	4.2	1.3	0.961	4.6	1.5	0.899	5.2	2.0	**0.939**	**4.0**	**1.2**	0.958	4.8	**1.5**	0.932	4.7	1.6

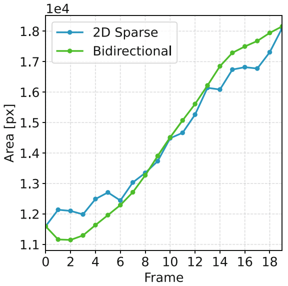

Fig. 3. Left atrium area over time from the pseudolabels of `patient0010` (4CH).

Table 2. LV volume and EF estimation on the CAMUS test set. The intra-observer variability is indicated in blue, and the best column-wise value is displayed in bold.

Methods	EDV			ESV			EF		
	Corr	Bias	MAE	Corr	Bias	MAE	Corr	Bias	MAE
Intra-observer	0.978	−2.8	6.5	0.981	−0.1	4.5	0.896	−2.3	4.7
CLAS [18]	0.958	−0.7	7.7	0.979	**0.0**	4.4	0.926	**−0.1**	**4.0**
A-CLAS [4]	0.969	–	–	0.983	–	–	0.883	–	–
MCLAS [19]	0.975	−1.0	–	0.983	−1.2	–	**0.946**	1.0	–
2D Sparse	0.972	**0.0**	6.0	0.980	−0.6	4.8	0.827	1.3	5.0
2D Dense	0.972	0.4	5.7	**0.986**	−0.3	4.2	0.841	1.3	4.6
3D Dense	**0.978**	−1.4	**4.8**	**0.986**	−0.1	**4.0**	0.859	**−0.1**	4.6

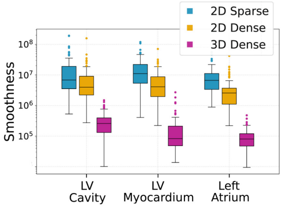

Fig. 4. Temporal smoothness of the CAMUS test set predictions in terms of the metric from Eq. 2 (lower values, higher smoothness). Note the logarithmic y-axis.

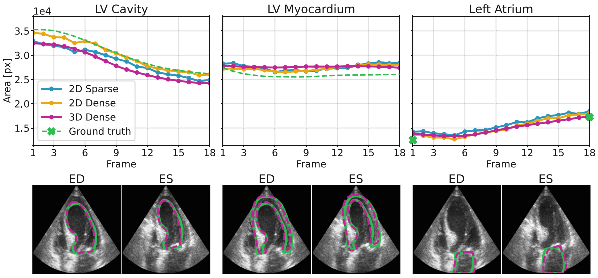

Fig. 5. Evaluation of the temporal consistency on `patient0002` from the test set. *Top row:* area curves. *Bottom row:* predictions at ED and ES. The green contours refers to the ground truth and the magenta outline is the prediction of the *3D Dense* model.

4 Discussion and Conclusion

This paper presented a method for temporally consistent segmentation of echocardiography using sparsely labeled data. The method exploits pseudo-labels generated by the use of DIR to complement the original set of sparsely annotated frames and allow the training of a 3D nnU-Net.

The analysis of the generated pseudo-label revealed the benefits of bidirectional over unidirectional label propagation. Results on the subsequent ED and ES segmentation task demonstrate that exploiting the pseudo-labels retains or improves the performance of the model trained on the sparsely labeled dataset, thereby endorsing their quality for downstream applications. The geometric metrics show that all three evaluated models perform *at least* as well as the SoTA methods, achieving a level of accuracy on par with intra-observer variability. However, evaluation of the temporal smoothness showed that the *2D Dense* model outperforms the *2D Sparse* model and that the *3D Dense*, in turn, outperforms both. For quantification of LV volumes, the *3D Dense* model outperforms all SoTA methods with EDV and ESV values closely matching intra-observer variability. EF estimation, however, is less remarkable. Yet, we argue that our method's very low bias and MAE akin to intra-rater variability advocate sufficiently good estimations of the measure.

A more notable limitation of our approach is its exclusive focus on the systolic function. Longer sequences can be analyzed by identifying and extracting the systolic phase from the entire heart cycle [4], but this would still preclude the characterization of the diastolic function, which is relevant to various heart diseases [16]. To this end, related studies have investigated the extraction of more meaningful temporal features [21] and the application of cyclical self-supervision [5]. As a direct extension of this work, future research could explore the efficacy of registering unlabeled frames to the same image (specifically, the ED or ES ground truth) as an alternative to the sequential approach. This could limit error accumulation and potentially extend our method to encompass full- or multi-cycle sequences. However, this may be detrimental to the temporal consistency of the pseudo-labels and thus to the downstream segmentation and quantification.

Figure 5 shows that *3D Dense* model results in slightly offset quantitative indices from ground truth and 2D models, especially at ED and ES. Examination of other patients indicates that the model does not favor over- or under-segmentation. Rather, Fig. 5 suggests the presence of uncertain boundaries in the data. Disagreements between manual and automatic segmentations arise when the endocardium is occluded, or when the LV myocardium and/or the LA extend beyond the field of view. In these cases, the ambiguous position of the structures likely influences the creation of manual annotations. Accordingly, the ambiguity is reflected in the predictions of the models, resulting in the observed discrepancy. Future work could model this randomness in order to convey the reliability of a given estimation. Extensions of this study may also attempt to limit the aforementioned uncertainty, for instance by selectively choosing high-quality pseudo-labels for training, or by leveraging distinct loss functions (or weighting schemes) for ground truth and pseudo-labeled frames [17,20].

In conclusion, our approach achieves accurate segmentation comparable to SoTA methods while offering remarkable temporal consistency. Unlike end-to-end frameworks such as CLAS [4,18,19], our approach separates pseudo-label generation and segmentation, offering flexibility and modularity.

References

1. Armstrong, A.C., et al.: Quality control and reproducibility in M-mode, two-dimensional, and speckle tracking echocardiography acquisition and analysis: the CARDIA study, year 25 examination experience. Echocardiography **32**(8), 1233–1240 (2014). https://doi.org/10.1111/echo.12832
2. Chen, C., et al.: Deep learning for cardiac image segmentation: a review. Front. Cardiovasc. Med. **7** (2020). https://doi.org/10.3389/fcvm.2020.00025
3. Chen, S., Ma, K., Zheng, Y.: Tan: Temporal affine network for real-time left ventricle anatomical structure analysis based on 2D ultrasound videos. ArXiv (2019). https://doi.org/10.48550/ARXIV.1904.00631
4. Chen, Y., Zhang, X., Haggerty, C.M., Stough, J.V.: Assessing the generalizability of temporally coherent echocardiography video segmentation. In: Išgum, I., Landman, B.A. (eds.) Medical Imaging 2021: Image Processing. vol. 11596, p. 115961O. International Society for Optics and Photonics, SPIE (2021). https://doi.org/10.1117/12.2580874
5. Dai, W., Li, X., Ding, X., Cheng, K.T.: Cyclical self-supervision for semi-supervised ejection fraction prediction from echocardiogram videos. IEEE Trans. Med. Imaging **42**(5), 1446–1461 (2023). https://doi.org/10.1109/TMI.2022.3229136
6. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Med. Image Anal. **52**, 128–143 (2019). https://doi.org/10.1016/j.media.2018.11.010
7. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2020). https://doi.org/10.1038/s41592-020-01008-z
8. Leclerc, S., et al.: Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans. Med. Imaging **38**(9), 2198–2210 (2019). https://doi.org/10.1109/tmi.2019.2900516
9. Li, M., Wang, C., Zhang, H., Yang, G.: MV-RAN: multiview recurrent aggregation network for echocardiographic sequences segmentation and full cardiac cycle analysis. Comput. Biol. Med. **120**, 103728 (2020). https://doi.org/10.1016/j.compbiomed.2020.103728
10. Moal, O., et al.: Explicit and automatic ejection fraction assessment on 2D cardiac ultrasound with a deep learning-based approach. Comput. Biol. Med. **146**, 105637 (2022). https://doi.org/10.1016/j.compbiomed.2022.105637
11. Ouyang, D., et al.: Video-based AI for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020). https://doi.org/10.1038/s41586-020-2145-8
12. Painchaud, N., Duchateau, N., Bernard, O., Jodoin, P.M.: Echocardiography segmentation with enforced temporal consistency. IEEE Trans. Med. Imaging **41**(10), 2867–2878 (2022). https://doi.org/10.1109/TMI.2022.3173669
13. Rueckert, D.: Nonrigid registration using free-form deformations: application to breast MRI images. IEEE Trans. Med. Imaging **18**(8), 712–721 (1999). https://doi.org/10.1109/42.796284

14. Schuuring, M.J., Išgum, I., Cosyns, B., Chamuleau, S.A.J., Bouma, B.J.: Routine echocardiography and artificial intelligence solutions. Front. Cardiovasc. Med. **8**, 648877 (2021)
15. Sfakianakis, C., Simantiris, G., Tziritas, G.: GUDU: geometrically-constrained ultrasound data augmentation in U-net for echocardiography semantic segmentation. Biomed. Signal Process. Control **82**, 104557 (2023). https://doi.org/10.1016/j.bspc.2022.104557
16. Thomas, L., Marwick, T.H., Popescu, B.A., Donal, E., Badano, L.P.: Left atrial structure and function, and left ventricular diastolic dysfunction: JACC state-of-the-art review. J. Am. Coll. Cardiol. **73**(15), 1961–1977 (2019). https://doi.org/10.1016/j.jacc.2019.01.059
17. Wang, C., et al.: Pseudo-labeled auto-curriculum learning for semi-supervised keypoint localization (2022)
18. Wei, H., et al.: 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
19. 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). https://doi.org/10.1016/j.media.2022.102686
20. Xia, Y., et al.: 3D semi-supervised learning with uncertainty-aware multi-view co-training (2020)
21. 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 Inform. **26**(12), 6105–6115 (2022). https://doi.org/10.1109/JBHI.2022.3221429