Ultrasound in Medicine & Biology 50 (2024) 47–56

Contents lists available at ScienceDirect

Ultrasound in Medicine & Biology

journal homepage: www.elsevier.com/locate/ultrasmedbio

Original Contribution

Automated 2-D and 3-D Left Atrial Volume Measurements Using Deep Learning

Jieyu Hu [a,*], Sindre Hellum Olaisen [a], Erik Smistad [a,b], Havard Dalen [a,c,d], Lasse Lovstakken [a]

[a] Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway
[b] SINTEF Medical Technology, Trondheim, Norway
[c] Clinic of Cardiology, St. Olav's University Hospital, Trondheim, Norway
[d] Levanger Hospital, Nord-Trndelag Hospital Trust, Levanger, Norway

ARTICLE INFO

Keywords:
Echocardiography
Left atrium
Diastolic dysfunction
Deep learning

Objective: Echocardiography, a critical tool for assessing left atrial (LA) volume, often relies on manual or semi-automated measurements. This study introduces a fully automated, real-time method for measuring LA volume in both 2-D and 3-D imaging, in the aim of offering accuracy comparable to that of expert assessments while saving time and reducing operator variability.
Methods: We developed an automated pipeline comprising a network to identify the end-systole (ES) time point and robust 2-D and 3-D U-Nets for segmentation. We employed data sets of 789 2-D images and 286 3-D recordings and explored various training regimes, including recurrent networks and pseudo-labeling, to estimate volume curves.
Results: Our baseline results revealed an average volume difference of 2.9 mL for 2-D and 7.8 mL for 3-D, respectively, compared with manual methods. The application of pseudo-labeling to all frames in the cine loop generally led to more robust volume curves and notably improved ES measurement in cases with limited data.
Conclusion: Our results highlight the potential of automated LA volume estimation in clinical practice. The proposed prototype application, capable of processing real-time data from a clinical ultrasound scanner, provides valuable temporal volume curve information in the echo lab.

Introduction

Echocardiography is the primary imaging modality for diagnosing cardiac diseases. An extensive acquisition and analysis protocol that covers several 2-D views and multiple manual measurements is recommended [1]. This process is time-consuming and prone to operator subjectivity. Further, manual analysis is typically limited to one cardiac cycle, even though guidelines recommend averaging over several. By fully automating the image analysis in echocardiography, we can analyze more consistent measurements over multiple cardiac cycles and time points. Furthermore, we can perform these measurements in real time or at the bedside, much more quickly than a single manual expert measurement.

In recent years, automated image analysis has improved in robustness and accuracy. With the development and use of deep convolutional neural networks (CNNs) based on supervised learning, the vision of a fully automated analysis in echocardiography may be fulfilled. Previous works have indicated that deep learning approaches perform comparably to human experts in segmentation [2], cardiac view classification [3], ejection fraction (EF) estimation [4–6] and assessment of diastolic dysfunction [7].

In this work, we turn the focus on the assessment of the left atrium (LA). Left atrial volume is a highly prognostic indicator in various cardiac disease states [8], for example, stroke, atrial fibrillation and the development of heart failure. The current recommendation for determining LA size is at end-systole (ES) using the apical window [9]. However, this requires the recording to be perfectly aligned in terms of angle and orientation to ensure optimal images and capture of the actual volume. This process is vulnerable to error. Thus, it is recommended that images of the LA be recorded separately from those for other chambers [9]. The biplane method of disk summation (MOD) is recommended for volume estimation of 2-D LA images at ES, requiring two views and two manual contour measurements [9].

Today, 3-D echocardiography is becoming more available, and a direct 3-D volume measurement of the left atrium can be made. This is typically done using semi-automated segmentation with manual corrections. With fewer geometrical assumptions, a better association was achieved for 3-D imaging when compared with cardiac magnetic resonance (CMR) [10]. Challenges for 3-D imaging currently include a low frame rate and lower image quality, which can influence measurement accuracy.

* Corresponding author. Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Postboks 8905, 7491 Trondheim, Norway.
E-mail address: jieyu.hu@ntnu.no (J. Hu).

https://doi.org/10.1016/j.ultrasmedbio.2023.08.024
Received 20 April 2023; Revised 18 August 2023; Accepted 29 August 2023

Left atrium quantification for 2-D and 3-D imaging requires time-consuming manual work. It relies on the annotator's experience for acquisition and measurement, which leads to inter- and intra-observer variability that may limit clinical sensitivity and specificity for disease identification. A fully automated LA volume estimation method in the daily echo lab could improve efficiency, consistency and accuracy, especially for less experienced sonographers.

In this work, we introduce and evaluate a fully automated pipeline for quantifying the LA volume for 2-D and 3-D echocardiography. We aim to develop a tool to automatically quantify LA volume in the echo lab and further for large-scale analyses in echo databases, that is, for data mining and phenotyping purposes. Our main contributions are the following:

1. Solutions to fully automate the estimation of the LA volume based on deep learning from both 2-D and 3-D echocardiography
2. A training scheme that exploits the full temporal sequence of ultrasound images
3. A solution for robust LA volume curve estimation throughout the cardiac cycle
4. A comprehensive evaluation using a large echocardiographic data set with clinical expert reference measurements
5. An application prototype that measures the LA volume in real time using images streamed from a clinical ultrasound scanner

Methods

Segmentation network architectures

We evaluated different segmentation networks for our study, considering metrics such as accuracy, memory usage and inference speed. The chosen network should not only deliver high average accuracy but should also exhibit robustness to outliers. Past studies [2,11,12] have attested to the exceptional performance of the U-Net structure and its variants in medical image segmentation tasks. In particular, Leclerc et al. [2] and Smistad et al. [6,13] reported the efficacy of U-Net for cardiac ultrasound segmentation tasks for both 2-D and 3-D. Given these considerations, we selected U-Net as the baseline network architecture for this study. We incorporated three U-Net models of varying complexity and other network structures for an initial comparison.

We used a similar U-Net structure for 2-D and 3-D images, substituting 2-D convolutional layers for 3-D equivalents. The architecture includes an encoder, a decoder and standard convolutional layers with ReLU activation. We used Adam as an optimizer with an initial learning rate of 5*e*-4 and a batch size of 24. Augmentations in the form of random image rotation, flipping, non-rigid deformation and gamma intensity transformations were applied during training. The Dice metric loss function was used during training.

The 2-D images were resized to an input size of 256 × 256 pixels, while the 3-D data were resized to 64 × 64 × 64 voxels to balance inference speed and accuracy. Two variants of the baseline U-Net architecture with the same structure but different feature maps and parameters were evaluated. The lightweight U-Net is depicted in Figure 1, and the 3-D U-Net architecture is provided in Figure 2.

Alternative segmentation networks

We assessed three additional segmentation networks alongside the baseline U-Net: E-Net [14], TransU-Net [15] and nnU-Net [12]. E-Net is optimized for rapid inference, whereas nnU-Net, optimizes the U-Net by adapting the training setup to the given data [12,14]. TransU-Net combines the U-Net architecture with transformers to better capture global

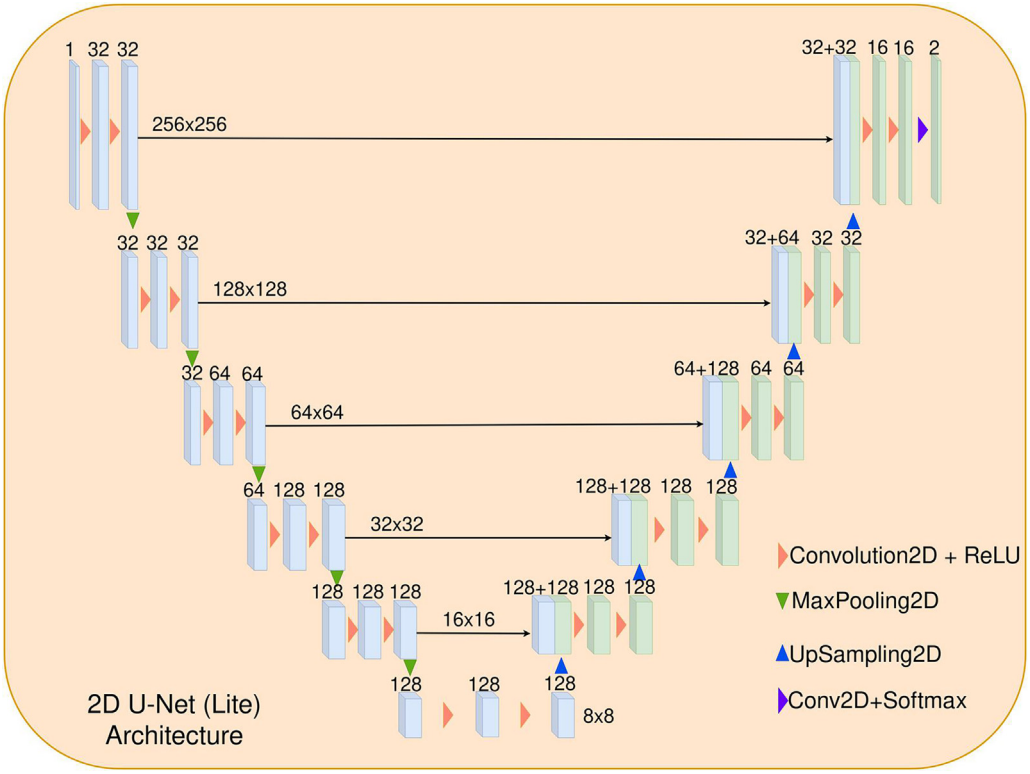

Figure 1. Two-dimensional U-Net (Lite) model. The model 2-D U-Net (Full) [2] is built with convolutional 2-D, ReLU and maximum pooling layers. The encoder follows an increasing channel sequence of [8, 16, 32, 64, 128, 256], reaching a 4 × 4 latent space with 512 channels, followed by a decreasing channel sequence in the decoder ([256, 128, 64, 32, 16, 8]), using upsampling 2-D layers. Each upsampling step concatenates with the corresponding encoder layer to retain high-resolution features. The 2-D U-Net (Lite) model, with 85% fewer parameters, reduces the encoder channels to [32, 32, 64, 128, 128] and decoder channels to [128, 128, 64, 32, 16], converging to an 8 × 8 latent space with 128 channels.

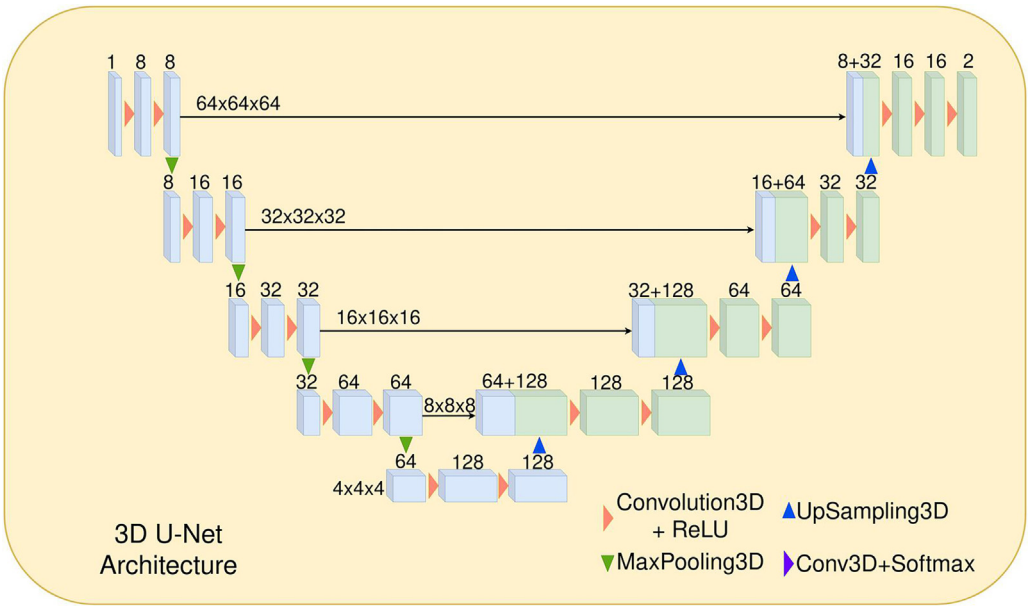

Figure 2. Architecture of the 3-D U-Net, adapted from its 2-D counterpart for lower spatial sampling (64 × 64 × 64). The encoder is constructed using Conv3D layers, ReLU activations and 3-D max-pooling layers, with a channel progression of [8, 16, 32, 64]. The decoder follows a similar pattern with [128, 64, 32, 16] channels. The model converges to a latent space of 4 × 4 dimensions with 128 channels.

information. The hybrid architecture equips TransU-Net with improved global and local information to improve predictions [15].

A fully automated pipeline

Developing a fully automated pipeline for left atrial end-systolic volume (LAESV) measurements necessitates extracting the ES frame for segmentation. Ideally, the measurement should be performed biplane, using the LA four-chamber and two-chamber cross-sections. To this end, for estimating LAESV, we compared two pipelines that can be used for both retrospective and real-time measurements: (i) a separate timing extraction network for identifying the ES frame + ES frame segmentation + LAESV volume calculation; and (ii) segmentation of all frames within the cardiac cycle + identification of ES frame as the frame with the largest volume + LAESV calculation.

The detailed steps of both pipelines are illustrated in Figure 3.

Timing network

The timing network used in this study consists of five 3-D convolutional layers with an increasing number of filters and two long short-term memory (LSTM) blocks [6]. The network was previously reported to have a mean absolute error (MAE) of 1.6 frames for ES when evaluated on left ventricular (LV) focused views.

When evaluated on LA-focused views, an MAE of 2.9 frames was found. Considering an average frame rate of 75.5 frames/s, the average error in time of 39 ms was deemed acceptable.

Volume estimation

The formula for calculating LA volume using the biplane disk summation algorithm is expressed as

$$\mathrm{LA_{vol}} = \frac{\pi}{4}\sum_{i=1}^{20} D1_i D2_i * \frac{L}{20} \tag{1}$$

where D1 and D2 denote the diameters of the sliced circular for A2C and A4C views, respectively; L is the LA long axis, divided into 20 portions; and i represents each circular slice. For correct measurement of LA volume, the guideline [9] emphasizes that the maximum length L on the segmentation mask should be taken as the true LA long axis.

In the case of 3-D data, the LA volume was calculated based on the image voxel spacing and 3-D segmentation mask. The LAESV was then determined as the maximum volume during the cardiac cycle.

Semi-supervised learning: pseudo-labeling

To tackle the challenge of manual annotation, where time restricts the practically possible number of annotated frames per examination,

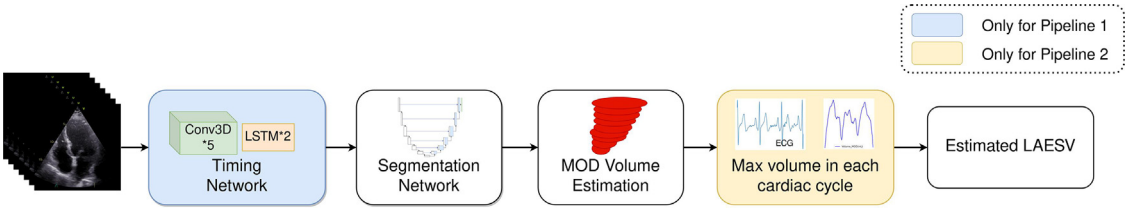

Figure 3. Pipeline 1 involves the use of two consecutive neural networks. The ultrasound videos are first sent to a timing network, predicting each cardiac cycle's end-systole moment. The output of the timing network is then sent to a segmentation network, which detects and returns a mask for the left atrium region in the image. Finally, the LAESV is calculated based on the segmentation using a disk summation approach. Pipeline 2 is based on a well-established medical fact: the LAESV is the maximum volume of the left atrium. In this pipeline, the segmentation network must segment each frame received, trace the volume and locate the frame with the maximum volume in each cardiac cycle, the LAESV. LAESV, left atrial end-systolic volume; MOD, method of disk summation.

we incorporated pseudo-labeling into our training pipeline [16]. We initially utilized a data set with annotations limited to ES frames. Using these data, we trained an initial network for LA segmentation. By using this initial network for inference, we generated a comprehensive set of annotations, effectively expanding the data set to encompass the entire cardiac cycle. These generated annotations subsequently served as the training foundation for the final segmentation network. This approach served two purposes: first, to enhance performance when data are scarce, and second, to provide annotations for all frames required for temporal networks.

To ensure the reliability of our pseudo-labeled data, we automatically identified and excluded frames with apparent segmentation errors. A smoothed LA area trace was generated, and frames that deviated significantly from the smoothed trace were identified as outliers. We performed linear interpolation between neighboring segmentation masks to replace invalid annotations. Further information on our pseudo-labeling method can be found in our previously published work [17].

Temporal networks

To improve the temporal consistency of the LA volume trace, we added Convolutional LSTM (ConvLSTM) layers into U-Net at each level in the encoder. Models with and without convLSTMs were compared qualitatively and quantitatively. Convolutional LSTM (ConvLSTM) proposed by Shi et al. [18] solved a spatiotemporal sequence forecasting problem that captured spatiotemporal correlations better than fully connected LSTM (FC-LSTM). More details on the method, additional experiments and results can be found in our previously published work [17,19].

Evaluation metrics

We assessed segmentation accuracy using the Dice score, which quantifies the overlap of the segmentation masks, and the Hausdorff distance (HD), which measures the maximum distance in millimeters between the segmentation contour and the manually traced contour. All metrics were evaluated using 10-fold cross-validation. Bland–Altman plots were used to evaluate bias and limits of agreement for the clinical volume measurements.

Real-time application

To determine the feasibility of integrating automated LA volume calculation into clinical practice, we developed an application that processes incoming 2-D or 3-D ultrasound images in real time. The application outputs a volume curve and the corresponding LA volume at ES. We used the FAST framework for real-time inference and image visualization [20].

Data material

HUNT4

The HUNT4 echo study, a part of the fourth wave of the Trndelag health population study from 2017 to 2019, consists of 2462 participants. The echo examination included standard 2-D imaging for the morphological and functional assessment of the heart, supplemented by 3-D imaging of the LV and the LA. All echocardiograms were recorded using a GE Vivid E95 system with 2-D M5S and 3-D 4V probes (GE Vingmed Ultrasound AS, Horten, Norway).

We extracted 2-D and 3-D recordings with LA-focused views for LAESV assessment. Our study comprised N = 325 individuals with 2-D biplane volume measurements, of whom N = 286 also underwent 3-D LA imaging and volume measurement. The age of participants ranged from 19 to 92, with a mean of 62.2 ± 12.1. Women constituted 45.3% of the cohort. A notable 87.7% exhibited no evident cardiac diseases. However, 12.3% presented with various abnormalities, including atrial fibrillation (N = 3), myocardial infarction (N = 6), aortic insufficiency (N = 5), left ventricular hypertrophy (N = 4), insufficient EF (N = 7) and other abnormalities. Figure 4 illustrates the distribution of LA size, presented by manual measurements of biplane LAESV.

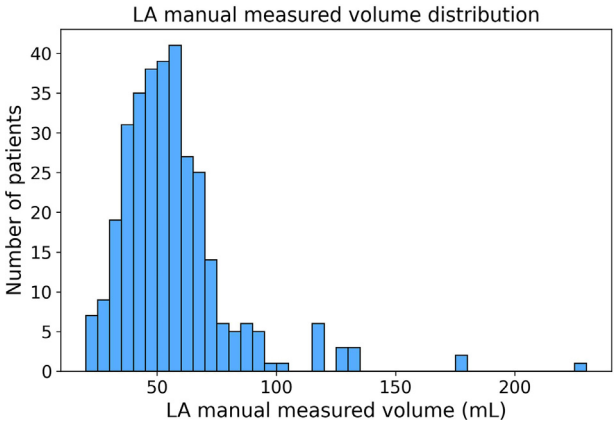

Figure 4. Distribution of left atrial size of cohort with N = 325, presented by manual measurements of biplane left atrial end-systolic volume.

We categorized all 2-D images into three classes based on their data quality: images with a clear view of the LA, images with myocardial dropouts and images affected by the reflection of the descending aorta. These categories are illustrated in Figure 5a, 5b and 5c, respectively.

Annotation extraction

The left atrial recordings were manually traced by clinical experts using the commercial GE EchoPAC analysis software (GE Vingmed Ultrasound AS). The resulting contours were overlaid on the Dicom preview image, which we automatically extracted using an image processing approach: (i) We first applied a simple color threshold to separate the colored contour from the grayscale background, and (ii) used a contour extraction method [21] to determine the LA trace coordinates, (iii) which we fitted using a smooth spline representation. We further defined the mitral annular plane as a straight line and used its midpoint as an initial point to locate the apex of the LA, which was identified as the longest distance from the midpoint to the contour. This served as the LA centerline for volume calculation. Figure 5a provides an example in which the *blue/green dotted lines* represent manually traced contours, and the red mask depicts the extracted area of the annotation.

The LA-focused 3-D echocardiographic recordings were annotated using a semi-automated GE EchoPAC tool comprising auto-tracing and manual editing. Figure 5d illustrates the GE EchoPAC tool tracing the LA in 3-D across cardiac cycles, while Figure 5e presents the sliced tracing extracted from the 3-D volume. An experienced physician validated and discarded invalid annotations, resulting in the final data set of 789 A2C and A4C LA-focused images of 325 individuals.

Results

2-D and 3-D segmentation accuracy

The data set was partitioned into training, validation and testing subsets on a per-patient basis in the ratio of 80%, 10% and 10%, respectively. Table 1 outlines the performance of the evaluated network architectures, summarized average Dice and HD metrics, number of parameters (indicating memory utilization) and inference time. The results suggest a comparable average Dice score among all networks. Although the nnU-Net slightly surpassed the baseline U-Net by 0.6 mm in HD, it also required 15 times more parameters and a more intricate training procedure. Figure 6 provides representative performance examples using the baseline U-Net for four-chamber and two-chamber views.

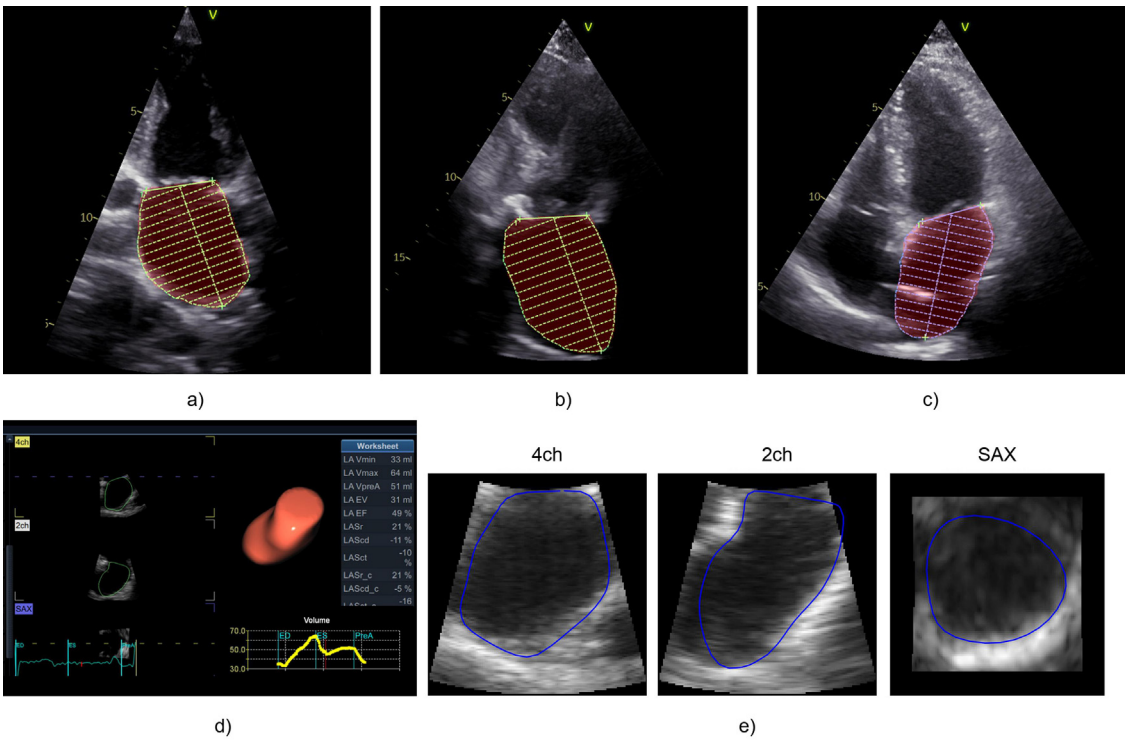

Figure 5. Representative data. (a) Normal data with a view that clearly displays the LA. (b) Dropouts with an incomplete atrial contour. (c) Descending aorta reflection with artifact. *Red masks*: Example of data extraction process; auto-extraction overlapped (*red mask*) with manually traced contour (*green or blue dotted lines*). (d) Capture from GE EchoPAC, which uses a semi-automated segmentation method to segment the LA in 3-D during the cardiac cycle. (e) Extracted 2-D contours from 3-D mesh model. LA, left atrium.

Table 1
Performance comparison of 2-D and 3-D segmentation networks

2-D segmentation network	Number of parameters[a]	Inference speed[b] (fps)	Average Dice	Average HD (mm)
E-Net	0.4M	184	0.93±0.03	5.4±2.8
U-Net 2 (Lite)	2.0M	214	0.94±0.04	4.9±2.0
U-Net 1 (Full)	13.4M	193	0.94±0.03	4.6±2.0
nnU-Net1 (2-D)	29.9M	103	0.94±0.03	4.3±1.9
TransU-Net	105.3M	98	0.94±0.03	4.6±1.8
3-D segmentation network	**Number of parameters**	**Inference speed (frames/s)**	**Average Dice Whole Sequence**	**Average Dice at ES**
3-D U-Net	2.5M	111	0.88±0.06	0.89±0.05
nnU-Net	16.5M	62	0.89±0.05	0.91±0.04

The table summarizes results evaluated via 10-fold cross-validation over the entire data set for average Dice scores and Hausdorff distances (HD). Hausdorff distance values are given in millimeters.

[a] Number of parameters learned per network (in millions).

[b] Inference speed, measured in frames per second, was obtained using an RTX A6000 GPU for a 2-D image (256 × 256 pixels) and a 3-D image (64 × 64 × 64 pixels).

For 3-D segmentation, we observed similar findings, with Dice scores slightly lower than 2-D segmentation but consistent between the lightweight 3-D U-Net and the more complex 3-D nnU-Net.

Data set size and pseudo-labeling

We conducted an experiment to investigate the influence of the quantity of annotated training data on performance, hypothesizing that pseudo-labeling could enhance performance in scenarios in which data are limited.

Figure 7 illustrates the average Dice and HD and standard deviation by including 5% to 80% of the data for training, corresponding to data sets of N = 23 to N = 368 patients. As illustrated, pseudo-labeling positively affected both metrics for small data sets and generally improved the Hausdorff metric, thus reducing the number of outliers.

Clinical measurements

In the Bland–Altman plots shown in Figure 8, we compared the LA volume measurement results using three different approaches for timing extraction: (i) using the manually annotated ES frame; (ii) using a separate neural network to extract the ES time point; and (iii) using the time of maximum volume in the cardiac cycle as ES. For all cases, the U-Net (Lite) was used for segmentation.

The average difference between the predicted LAESV and the manual measurement was 2.9 mL when employing the manual ES time point and timing network, and 3.7 mL when using the maximum volume as the ES time point. In a clinical context, the low average bias suggests the robust performance of our U-Net and training procedure. The standard deviations were 8.0, 8.5 and 10.5 mL for the manually annotated time, the timing network and the maximum volume method, respectively. No

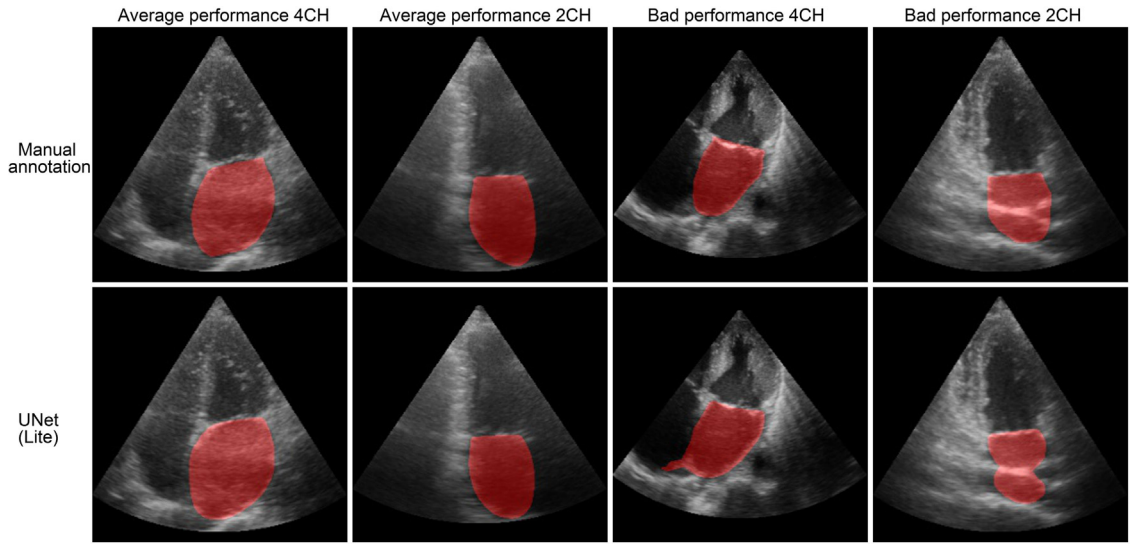

Figure 6. Examples of average and bad performance of four-chamber and two-chamber views for U-Net (Lite).

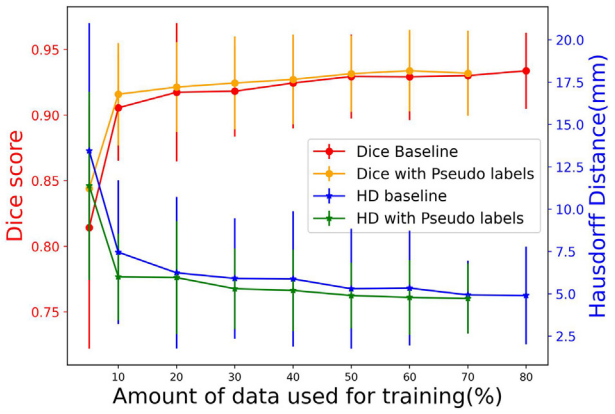

Figure 7. Test performance on 10% of data using different amounts of training data. The performance on both Dice score and Hausdorff distance (HD) increased with more training data.

proportional bias was observed. The separate timing network provided results superior to those of the maximum volume approach, particularly with respect to outliers.

In Figure 9, we compare the 3-D predicted LAESV to 3-D semi-automated manual measurements. The mean difference ± standard deviation between the 3-D predicted volume and 3-D manual measurements was 7.8 ± 9.1 mL. This mean difference is larger than the 2-D volume errors above. However, these variations align with our Dice findings for 2-D and 3-D imaging, suggesting inherent challenges with 3-D measurements.

Volume trace measurements

A representative example of volume curves using the different approaches is provided in Figure 10, including an image example where the segmentation results differ. The pseudo-labeling approach provided labels for every frame, leading to a more stable temporal volume curve.

The volume trace estimates were compared quantitatively for a separate data set of 28 participants (126 frames), where annotations were available at multiple time points (three to five time points). As seen in Figure 10, the initial model deviated from the reference at several time points. Pseudo-labeling here enhanced considerably both quantitative and qualitative performance without requiring additional manual work.

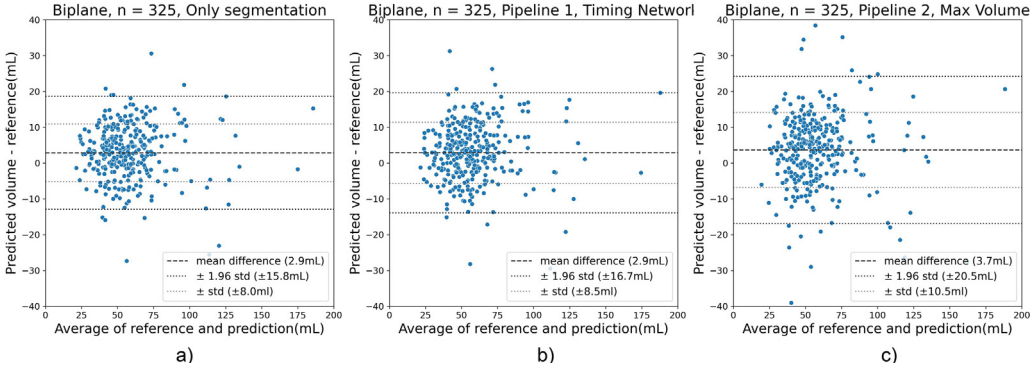

Figure 8. Findings for biplane volume estimation. (a) Using the same frame as the reference, compared between manual measurements and network results. (b) Pipeline 1, which uses a timing network to locate the frame of ES. (c) Pipeline 2, which considers the frame with the largest volume as ES. 2CH, two-chamber; 4CH, four-chamber; ES, end-systole; std, standard deviation.

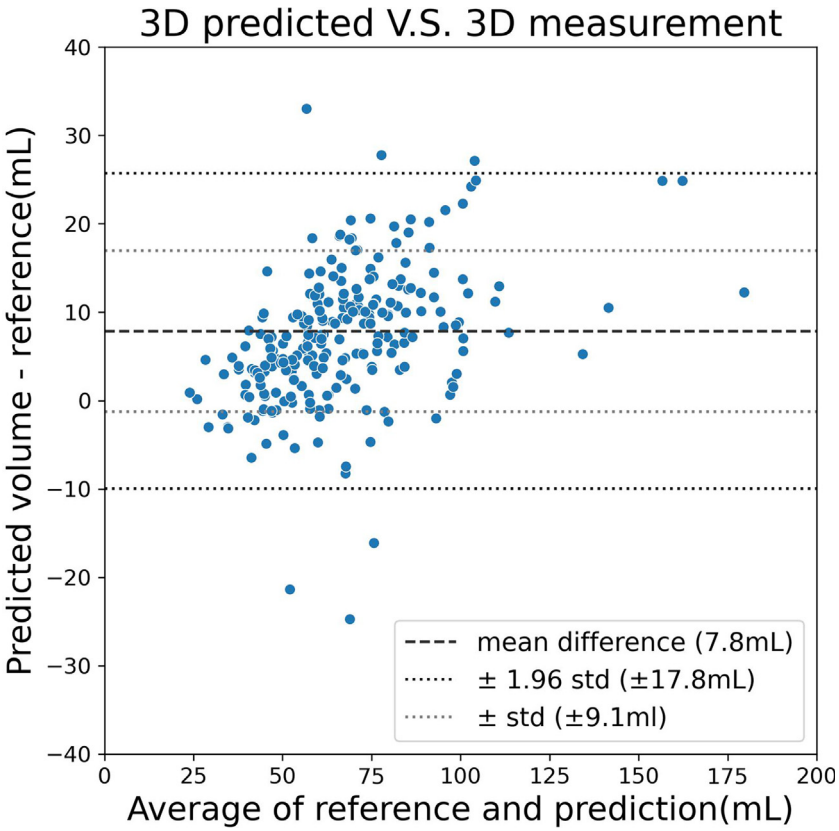

Figure 9. Comparison of 3-D volume prediction with 3-D manual measurements. std, standard deviation.

The summarized results in Table 2 indicate that pseudo-labeling, using the simpler baseline model, improved the average Dice score from 0.87 to 0.90. However, incorporating temporal blocks into the model did not enhance the metrics but further stabilized the volume trace.

Within our manually annotated data set, N = 240 patients had LAESV measured for both 2-D and 3-D recordings, enabling a comparative analysis of automated 2-D and 3-D volume estimation. In Figure 11 is a representable volume trace for a single patient, indicating that the curves exhibit similar trends and comparable values.

Real-time application

Figure 12 illustrates screen captures of both 2-D and 3-D volume measurement applications. The open source FAST framework [20] enabled rapid prototyping, while data streaming was facilitated by the open source OpenIGTLink format [22]. The software provides an opportunity to evaluate the efficacy of our approach in daily clinical practice in the echo lab.

Discussion

We developed and evaluated a deep learning approach to segment the LA in echocardiography, facilitating automated clinical volumetric measurements for both 2-D and 3-D imaging in real time. The proposed approach achieved a high degree of agreement with reference measurements, with the standard deviation within the expected range of inter-observer variability for both 2-D and 3-D segmentation [23]. Thus, our approach is suitable for further evaluation and wider clinical use. As our evaluation data set is limited in size, and the data consisted mainly of LA size within the normal range, the method should be further validated in larger external data sets and for more variation in pathology. Furthermore, because our data was acquired using a single ultrasound system vendor, additional validation would be required to generalize the findings to data sourced from other vendors.

Three-dimensional segmentation enables direct LA volumetric measurements, which could result in a more robust and easy-to-use approach. Comparison of 2-D and 3-D performance revealed a slight overestimation in volume by the 3-D approach. However, the standard deviation was at the same level as in the 2-D automated measurements.

Three-dimensional images generally have lower image quality and frame rates than 2-D images, which limits their usage in clinical diagnosis, but it represents a promising trend for the future.

Both 2-D and 3-D segmentation networks were based on the U-Net architecture. Comparison of the listed architectures revealed that our lightweight U-Net was a good compromise when considering inference speed, memory usage and accuracy. This was important for developing the real-time prototype application for the echo lab. While our light U-Net provided a high average accuracy, the more complex nnU-Net handled outliers better. An example of this was the data with a descending aorta reflection, as illustrated in Figure 5c. As we had a very limited number of such cases in our data set (N = 15), the issue may be handled by including more training data.

Pseudo-labeling was introduced to add annotations and exploit all frames in the cine loops, further enabling the training of temporal networks. As illustrated in Figure 7, pseudo-labeling improved network performance at ES, primarily for smaller data sets. Across all data sizes, the benefit of pseudo-labeling was more evident for the HD metric than the Dice score, indicating that we generally reduce outliers with less realistic pathological shapes. Pseudo-labeling contributed to more stable volume traces and should be considered for similar applications. However, the processing used to modify the most significant outliers may have the effect of interfering with the most critical frames that need manual annotation. Therefore, a targeted approach to annotation in which the most challenging frames are manually annotated should also be investigated.

Having the full LA volume traces available may be a valuable tool when studying the development of diastolic dysfunction and generally the coupling between the left ventricular performance and loading conditions [24]. Although semi-automated approaches for volume trace

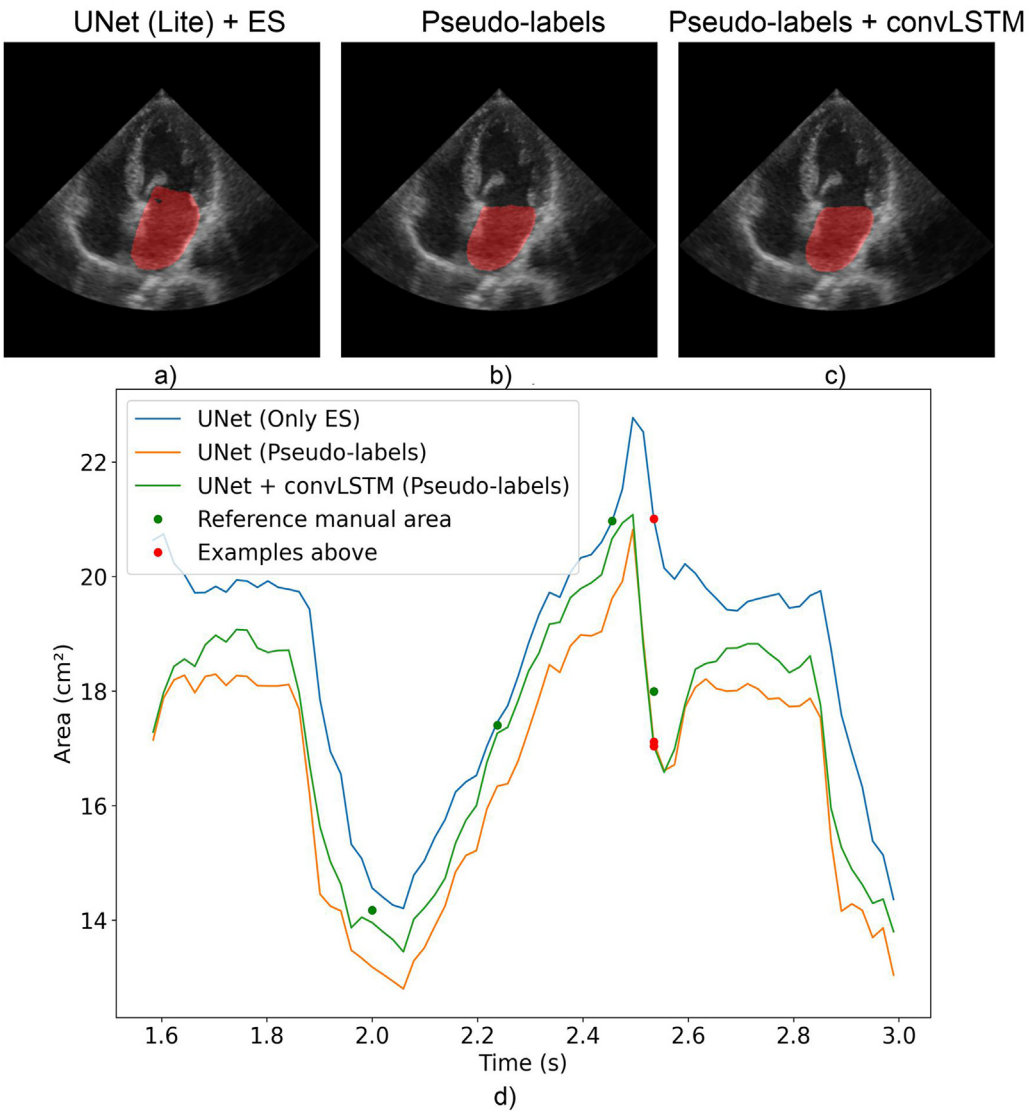

Figure 10. Segmentation results for baseline U-Net versus pseudo-labeling with and without temporal networks. (a) U-Net (Lite) trained with data annotated at ES. (b) U-Net (Lite) trained with pseudo-labels. (c) U-Net (Lite) with convLSTM layers trained with pseudo-labels. (d) Area trace of these three approaches. ES, end-systole.

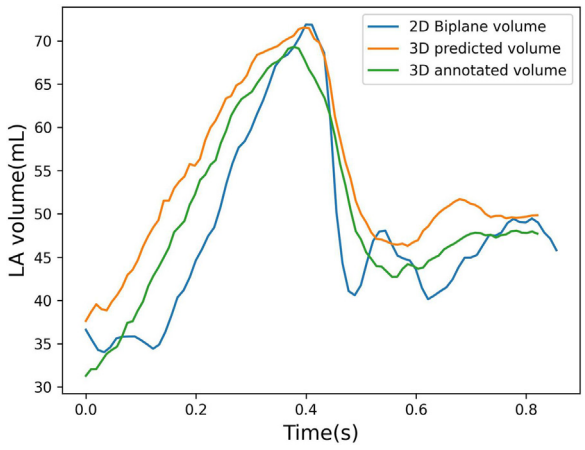

Figure 11. Left atrium volume comparison for one patient of 2-D biplane and 3-D. *Blue line*: 2-D biplane LA volume; *orange line*: 3-D-predicted LA volume; *green line*: semi-automated method—measured 3-D LA volume. LA, left atrium.

estimation are commercially available, these rely on substantial user corrections to ensure sufficient accuracy for all frames. Our fully automated workflow can deliver robust volume traces for both 2-D and 3-D LA volumes, providing valuable data for future physiological and clinical studies. Specifically, this may enable evaluation of the LA's conduit and booster functions by using parameters including the minimum and maximum volumes of the LA and the volume at the occurrence of the P wave.

In clinical practice, limited time is available, and often simple eyeballing is the basis for evaluation, or at least a prioritized selection is made for what images to acquire and measure in the extensive echo protocol. Robust automated measurements can help increase the number of quantitative measurements, and thus, the data material for clinical decision-making and research can be significantly increased. A viable clinical workflow then involves the clinicians who need to eyeball and approve the output of the automated measurement. Real-time measurements enable clinicians to optimize the view and image quality for the given measurement while scanning. Based on this work, we will evaluate our prototype application in the echo lab to document potential benefits.

Both offline and real-time approaches rely on sufficiently robust and accurate segmentation. One challenge in this respect is the subjectivity

Table 2
Performance comparison of pseudo-labeling approaches

Approach	Average Dice score	Average Hausdorff distance (mm)
Annotations only at ES[a]	0.87 ± 0.08	3.4 ± 2.1
Pseudo-labels (without post-processing)[b]	0.90 ± 0.05	3.0 ± 2.0
Pseudo-labels [c]	0.90 ± 0.04	2.7 ± 1.2
Pseudo-labels + convLSTM[d]	0.90 ± 0.05	3.1 ± 3.1

All methods were evaluated on a data set from 28 external participants, comprising 126 manually annotated frames.
ES, end-systole.
[a] Baseline model trained solely on ES frames.
[b] Use of pseudo labels without post-processing of the generated annotations.
[c] Use of pseudo labels to improve performance.
[d] Signifies the addition of convLSTM layers to the baseline model.

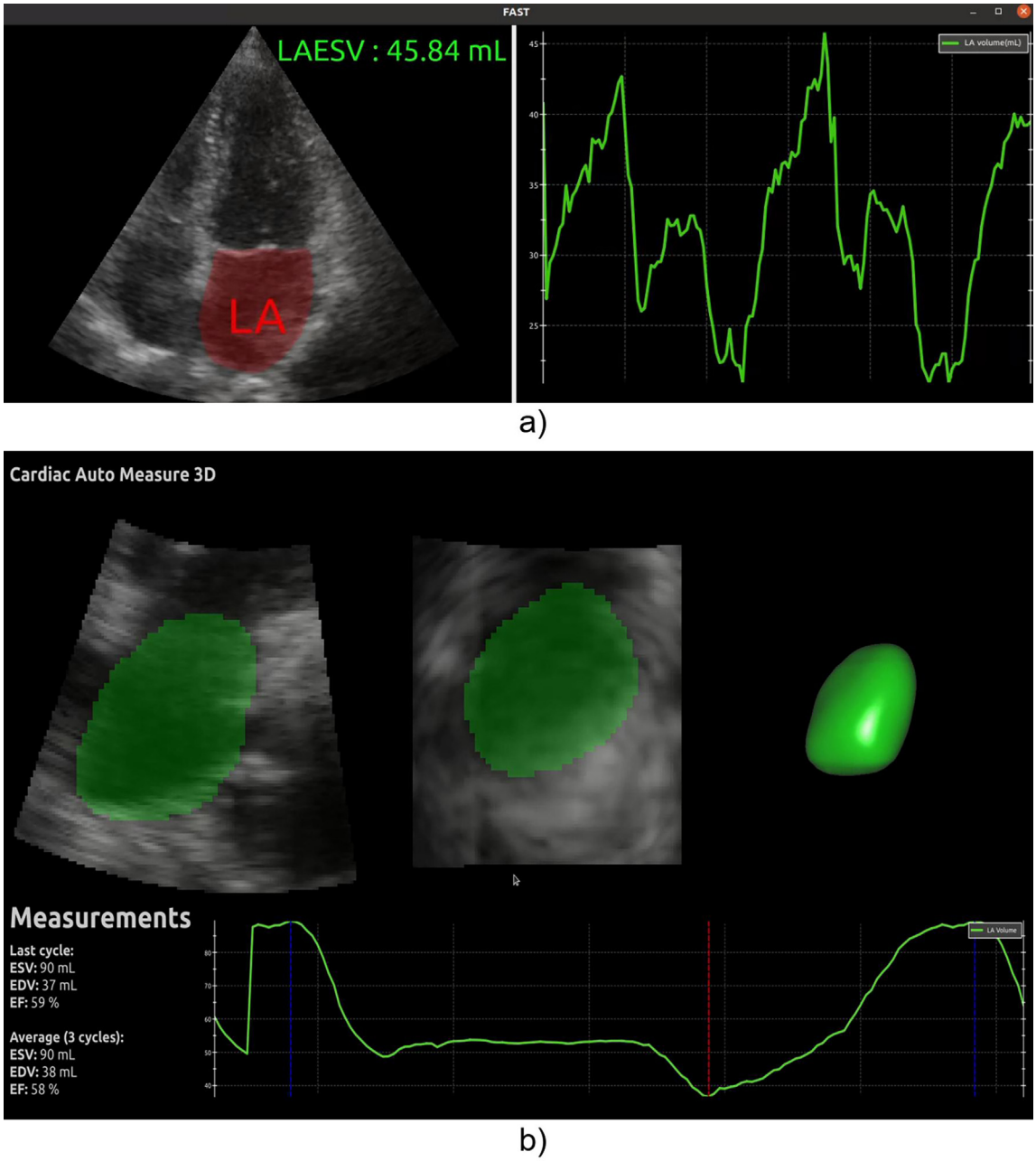

Figure 12. Real-time applications for (a) 2-D images and (b) 3-D volumes. LAESV, left atrial end-systolic volume.

in annotation, where individual user experience can influence what is perceived as correct. It is hoped that the development and possibilities of robust artificial intelligence–based automated measurements can push the agenda further to achieve consensus on how detailed echocardiographic measurements should be done.

Conclusions

Our fully automated approach robustly and accurately measures left atrial volume in echocardiography, and our lightweight U-Net implementation is suitable for real-time (>100 frames/s) and retrospective use. We validated our approach against manual measurements using common segmentation metrics and volume estimates and found its performance was comparable to that of human experts and within the expected inter-observer variability. Using pseudo-labeling to annotate all frames in the cine loops proved particularly beneficial for smaller data sets, generally decreasing outliers and yielding more stable volume curves. Further work aims to explore how automated and real-time measurement of LA volume can enhance efficiency and efficacy daily in the echo lab.

Conflict of interest

The authors declare no competing interests.

Artificial intelligence technology statement

During the preparation of this work the authors used Grammarly to check grammar, style and syntax. After using this tool/service, they reviewed and edited the content as needed and take full responsibility for the content of the publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ultrasmedbio.2023.08.024.

References

[1] Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC, et al. Guidelines for performing a comprehensive transthoracic echocardio-graphic examination in adults: Recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 2019;32:1–64.

[2] Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging 2019;38:2198–210.

[3] Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L. Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. Ultrasound Med Biol 2019;45:374–84.

[4] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature 2020;580:252–6.

[5] Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. Circ Cardiovasc Imaging 2019;12:e009303. doi: 10.1161/CIRCIMAGING.119.009303.

[6] Smistad E, Østvik A, Salte IM, Melichova D, Nguyen TM, Haugaa K, et al. Real-time automatic ejection fraction and foreshortening detection using deep learning. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2595–604.

[7] Pandey A, Kagiyama N, Yanamala N, Segar MW, Cho JS, Tokodi M, et al. Deep-learning models for the echocardiographic assessment of diastolic dysfunction. JACC Cardiovasc Imaging 2021;14:1887–900.

[8] Nagueh SF. Left ventricular diastolic function: understanding pathophysiology, diagnosis, and prognosis with echocardiography. JACC Cardiovasc Imaging 2020;13:228–44.

[9] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015;28:1–39.e14.

[10] Mor-Avi V, Yodwut C, Jenkins C, Khl H, Nesser HJ, Marwick TH, et al. Real-time 3D echocardiographic quantification of left atrial volume: multicenter study for validation with CMR. JACC Cardiovasc Imaging 2012;5:769–77.

[11] Zyuzin V, Sergey P, Mukhtarov A, Chumarnaya T, Solovyova O, Bobkova A, et al. Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network Unet. 2018 Ural Symposium on Biomedical Engineering. Yekaterinburg, Russia: Radioelectronics and Information Technology (USBEREIT); 2018. p. 76–8. doi: 10.1109/USBEREIT.2018.8384554.

[12] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nn-UNet: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203–11.

[13] Smistad E, Steinsland EN, Lvstakken L. Real-time 3D left ventricle segmentation and ejection fraction using deep learning. Proc IEEE Int Ultrason Symp 2021. doi: 10.1109/IUS52206.2021.9593301.

[14] Paszke A, Chaurasia A, Kim S, Culurciello E. ENet: a deep neural network architecture for real-time semantic segmentation, https://arxiv.org/abs/1606.02147; 2016 [accessed 22.01.21].

[15] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. TransUNet: transformers make strong encoders for medical image segmentation, https://arxiv.org/abs/2102.04306; 2021 [accessed 16.02.21].

[16] Lee DH. The simple and efficient semi-supervised learning method for deep neural networks. Workshop on challenges in representation learning. ICML; 2013. p. 896.

[17] Hu J, Smistad E, Salte IM, Dalen H, Lovstakken L. Exploiting temporal information in echocardiography for improved image segmentation. Proc IEEE Int Ultrason Symp 2022. doi: 10.1109/IUS54386.2022.9958670.

[18] Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. MIT Press; 2015. p. 802–10.

[19] Smistad E, Salte IM, Dalen H, Lovstakken L. Real-time temporal coherent left ventricle segmentation using convolutional LSTMs. Proc IEEE Int Ultrason Symp 2021:1–4. doi: 10.1109/IUS52206.2021.9593668.

[20] Smistad E, Østvik A, Pedersen A. High performance neural network inference, streaming, and visualization of medical images using FAST. IEEE Access 2019;7:136310–21.

[21] Bradski G. The OpenCV Library. Dr Dobb's. Journal Software Tools 2000;25:120–5.

[22] Tokuda J, Fischer GS, Papademetris X, Yaniv Z, Ibanez L, Cheng P, et al. OpenIGTLink: an open network protocol for image-guided therapy environment. Int J Med Robot 2009;5:423–34.

[23] Letnes JM, Eriksen-Volnes T, Nes B, Wislff U, Salvesen Dalen H. Variability of echocardiographic measures of left ventricular diastolic function: the HUNT study. Echocardiography 2021;38:901–8.

[24] Nagueh SF, Smiseth OA, Appleton CP, Byrd III BF, Dokainish H, Edvardsen T, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2016;29:277–314.