Real-Time Automatic Ejection Fraction and Foreshortening Detection Using Deep Learning

Erik Smistad, Andreas Østvik, Ivar Mjåland Salte, Daniela Melichova, Thuy Mi Nguyen, Kristina Haugaa, Harald Brunvand, Thor Edvardsen, Sarah Leclerc, Olivier Bernard, Bjørnar Grenne, and Lasse Løvstakken

***Abstract*— Volume and ejection fraction (EF) measurements of the left ventricle (LV) in 2-D echocardiography are associated with a high uncertainty not only due to interobserver variability of the manual measurement, but also due to ultrasound acquisition errors such as apical foreshortening. In this work, a real-time and fully automated EF measurement and foreshortening detection method is proposed. The method uses several deep learning components, such as view classification, cardiac cycle timing, segmentation and landmark extraction, to measure the amount of foreshortening, LV volume, and EF. A data set of 500 patients from an outpatient clinic was used to train the deep neural networks, while a separate data set of 100 patients from another clinic was used for evaluation, where LV volume and EF were measured by an expert using clinical protocols and software. A quantitative analysis using 3-D ultrasound showed that EF is considerably affected by apical foreshortening, and that the proposed method can detect and quantify the amount of apical foreshortening. The bias and standard deviation of the automatic EF measurements were −3.6 ± 8.1%, while the mean absolute difference was measured at 7.2% which are all within the interobserver variability and comparable with related studies. The proposed real-time pipeline allows for a continuous acquisition and measurement workflow without user interaction, and has the potential to significantly reduce the time spent on the analysis and measurement error due to foreshortening, while providing quantitative volume measurements in the everyday echo lab.**

***Index Terms*— Deep learning, echocardiography, ejection fraction (EF), foreshortening.**

I. Introduction

LEFT ventricle (LV) ejection fraction (EF) and volume measurements are important clinical indices in cardiology. When using 2-D echocardiography, an established protocol describes which images to use for measurement, which frames during the heart cycle should be selected for tracing, and how to properly trace the heart wall (endocardium) [1]. Despite the existence of a standard protocol, these measurements are associated with a high interobserver variability. This variability is known to be caused by differences in manual frame selection and endocardial tracing, all done after image acquisition. In this context, automatic measurements without user intervention have the potential to limit the interobserver variability as well as to reduce time spent on analysis.

An important and often overlooked challenge in echocardiography is finding the properly aligned views without any foreshortening of the LV. Apical foreshortening occurs when the operator places the ultrasound probe in a suboptimal position such that the imaging plane does not cut through the true apex of the LV, as shown in Fig. 1. Foreshortening is a common problem in routine 2-D cardiac ultrasound resulting in inaccurate volume and thus EF measurements, as emphasized in a recent study by the European Association of Cardiovascular Imaging (EASCVI) and the American Society of Echocardiography (ASE) Standardization Task Force [2]. Using a foreshortened imaging plane results in an underestimation of the volume and inaccurate EF. It also has a major impact on deformation measurements such as global longitudinal and regional strain. Since apical foreshortening is introduced when the operator is scanning, it is required to have a detection method that runs in real time while scanning

Manuscript received January 13, 2020; accepted March 11, 2020. Date of publication March 16, 2020; date of current version November 23, 2020. This work was supported in part by the Research Council of Norway under Project 237887 and in part by the Norwegian Health Association, South-Eastern Norway Regional Health Authority, and the National Programme for Clinical Therapy Research (KLINBEFORSK) under Project 2017207. *(Corresponding author: Erik Smistad.)*

Erik Smistad and Andreas Østvik are with the Centre for Innovative Ultrasound Solutions, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, 7491 Trondheim, Norway, and also with the SINTEF Medical Technology, 7052 Trondheim, Norway (e-mail: ersmistad@gmail.com).

Ivar Mjåland Salte, Daniela Melichova, and Thuy Mi Nguyen are with the Department of Medicine, Sørlandet Hospital at Kristiansand, 4615 Kristiansand, Norway, and also with the Faculty of Medicine, University of Oslo, 0372 Oslo, Norway.

Kristina Haugaa is with the Department of Cardiology, Oslo University Hospital at Rikshospitalet, 0372 Oslo, Norway.

Harald Brunvand is with the Department of Medicine, Sørlandet Hospital at Kristiansand, 4615 Kristiansand, Norway.

Thor Edvardsen is with the Department of Cardiology, Oslo University Hospital at Rikshospitalet, 0372 Oslo, Norway, and also with the Faculty of Medicine, University of Oslo, 0372 Oslo, Norway.

Sarah Leclerc and Olivier Bernard are with the CREATIS, University of Lyon, 69007 Lyon, France.

Bjørnar Grenne is with the Centre for Innovative Ultrasound Solutions, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, 7491 Trondheim, Norway, and also with the Clinic of Cardiology, St. Olavs Hospital, 7030 Trondheim, Norway.

Lasse Løvstakken is with the Department of Circulation and Medical Imaging, Centre for Innovative Ultrasound Solutions, Norwegian University of Science and Technology, 7491 Trondheim, Norway.

Digital Object Identifier 10.1109/TUFFC.2020.2981037

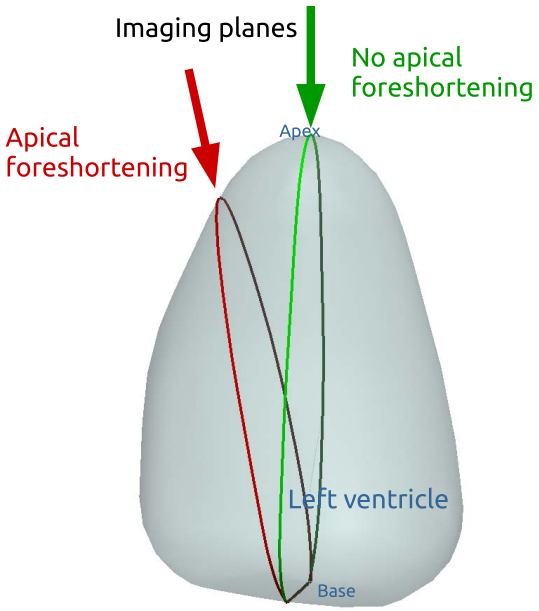

Fig. 1. Illustration of apical foreshortening in 2-D ultrasound imaging of the LV. The green imaging plane cuts through the true apex, while the red imaging plane does not. Using the red plane for volume estimation will result in an underestimation of the LV volume.

in order to help the operator acquire a more optimal view of the LV.

In this work, a fully automated real-time EF and foreshortening detection method is proposed with the potential of reducing inter- and intraobserver variability, measurement time, and acquisition errors such as foreshortening. The observer variability and the frequency of acquisition errors are expected to be much higher for inexperienced users. Thus, the proposed method may be an important tool for enabling and improving the performance of inexperienced users in echocardiography. The method uses several deep learning components, such as view classification, cardiac cycle timing, and segmentation and landmark extraction, to measure the amount of foreshortening, LV volume, and EF.

A. Related Work

Segmentation of the endocardial border is fundamental to automate LV measurements. This has been an active research area for several decades for both 2-D and 3-D ultrasound. Due to reverberation noise, poor contrast, and heterogeneous tissue intensity of cardiac ultrasound images, this has been a challenging task. Traditionally, LV segmentation methods have used some form of shape prior (e.g., active contours, deformable models, active shape models, and active appearance models) to constrain the segmentation [3]. Recently, the advances in deep learning have pushed the field of LV segmentation even further in the work of [4] for 2-D and [5] for 3-D. In both these studies, a U-net [6] type of architecture was used to achieve state-of-the-art performance.

Measurements such as volume and EF require images from two time points of the cardiac cycle; end-diastole (ED) and end-systole (ES). According to [1], ED is defined either as the frame after mitral valve closure, or as the time when the LV volume is the largest. ES is defined as either the frame after aortic valve closure, or the moment of smallest LV volume. Electrocardiogram (ECG) may be used to identify ED and ES; however, estimating ES solely on ECG is unreliable [7]. There is substantial previous work on the estimation of ED and ES directly from B-mode images, using approaches such as deformable models with speckle tracking [8], manifold learning [9], dimensionality reduction [10], and also recently deep neural networks [11]–[13].

LV volume and EF can be estimated with a single ultrasound view, but the current recommendation is to use both the apical four-chamber (A4C) and the apical two-chamber (A2C) views [1]. Thus, automatic view classification is needed to separate image frames of the A4C and A2C views, as well as other irrelevant views. Highly accurate view classification has been demonstrated by several groups using deep neural networks [14]–[16].

To the best of author's knowledge, there are no published methods on direct detection and quantification of apical foreshortening. However, there are studies on using deep learning to automatically estimate the overall image quality of an ultrasound image as done by Abdi *et al.* [17] using a regression network to quality assess A4C views. In this study, the image quality was estimated using five categories, where one of the criteria for a high-quality image was the absence of foreshortening. For this quality estimation, Abdi *et al.* [17] use a single image to determine the overall image quality, but foreshortening is hard to see in a still image as it is mainly characterized by a shorter LV long axis and a false apex which is moving in the image over time [2]. Our method instead uses segmented images from both A4C and A2C views in both ED and ES time points of the cardiac cycle to directly measure the amount of foreshortening.

Methods for fully automatic volume and EF measurements were proposed by Zhang *et al.* [16], although limited details are provided on their automatic pipeline. Our group proposed a fully automatic real-time method in 2018 [18]. Jafari *et al.* [19] presented a similar approach in 2019, optimized for mobile devices, although without view classification, requiring users to specify which view is being scanned, and therefore not fully automatic. Silva *et al.* [20] proposed to do automatic EF estimation as a direct classification problem instead, dividing EF into four categories.

This article is a substantial extension of our previous work [18], adding a novel apical foreshortening detection method, improved ED and ES detection using a separate neural network, accuracy evaluation on a separate data set, where EF and volumes were measured using clinical practice and software, and a more comprehensive analysis of the results.

II. Methods

This section starts by describing the key components needed for the foreshortening detection and EF measurements and how they were optimized for real-time performance. These components include view classification, cardiac cycle timing, LV segmentation, and landmark extraction. Then, the proposed apical foreshortening method is described, and finally we describe how a real-time application for automatic foreshortening

detection and EF measurements was created and a description of the data sets used for evaluating the methods.

A. View Classification

In this work, the cardiac view classification (CVC) network of Østvik *et al.* [15] was used, which has demonstrated an accuracy up to 98% and real-time inference. This network can recognize eight different cardiac views, A4C, A2C, apical long axis, parasternal short axis, parasternal long axis, subcostal four chamber, subcostal vena cava, and unknown views. The precision and recall for A4C view sequences were 99%, and 98% for A2C. The network is composed of inception blocks [21] and a dense connectivity pattern [22] where the input of the inception block is concatenated with the output and processed into a transition module with 1×1 convolution layer bottlenecks and 2×2 max pooling. Each convolutional layer is followed by batch normalization and parametric rectified linear unit (PReLU) activation functions, and a global average pooling layer is used after the final convolution layer. Input to the network is a 128×128 B-mode image, and the output is a normalized logit for each class through a softmax function. Only A4C or A2C views with a logit output higher than 0.75 were accepted in the auto measurement pipeline. Also, a valid view has to be maintained for at least 2 s. For robustness, the view classification output was averaged over the last ten frames.

B. Timing of ED and ES

From an accurate segmentation, ED and ES could be estimated by calculating the LV area, and then selecting ED as the frame, where the LV area is at its maximum and ES when it is at its minimum. This method has been used in several previous publications [16], [18], [19] and we will refer to it as the LV area method. This method is however very sensitive to any mistakes in the segmentation during the cardiac cycle and thus might require outlier detection methods. Due to this challenge, we have used the neural network method by Fiorito *et al.* [13] instead to directly estimate ED and ES from B-mode images. This method has shown an accuracy of -5.5 ± 28.2 ms and -0.6 ± 31.8 ms on ED and ES, respectively, and mean absolute error of 1.53 and 1.55 frames from ED/ES reference. While the deep learning approaches of [11] and [12] use 2-D convolutional layers with long short-term memory (LSTM) layers on fixed length sequences of 20 and 30 frames, respectively, the neural network of Fiorito *et al.* [13] uses 3-D convolutional layers with LSTM layers and has been trained and tested on variable length sequences. The image input sequence is processed through five 3-D convolutional layers with an increasing number of filters all with size $3 \times 3 \times 3$, except the first which has a spatial filter size of 7×7 and three in the temporal dimension. The final feature vectors are processed through two LSTM layers of size 32. Finally, a 1-D convolutional layer with a sigmoid activation is applied on the temporal axis, generating one prediction per time step and effectively applying temporal smoothing on the predictions.

Network input is a sequence of N image frames which were resized to 80×128, while the output is a sequence of N scalars with values between 0 and 1, where 0 indicates that the image is from the systolic phase and one diastolic phase. ED and ES were identified as the crossover points from 1 to 0 and 0 to 1, respectively. Based on our experience, this network needs a sequence of sufficient length to accurately predict ED and ES, thus the minimum of N was set to 30 frames. One can argue that a few frames should be enough to see if the ventricle expands or shrinks, but the rate of change in the cardiac cycle is not constant. Around ED and ES, the rate of change is especially small. This can be one reason why a large temporal window is needed. Another reason might be that various pathologies of the heart can seriously affect the movement of tissue and valves during the cardiac cycle, leading to complex movement patterns and thereby making it even more difficult to predict ED/ES using only B-mode images. As the length of the input sequence also affects the inference time, the maximum of N was set empirically to 80. Thus, up to 80 of the latest frames in the current image buffer were used to find ED and ES.

C. Left Ventricle Segmentation

The segmentation network used in this work was first described by Smistad *et al.* [23] and later used as the U-Net 1 in the CAMUS study of Leclerc *et al.* [4]. This U-net architecture consists of an encoder and a decoder stage which have several layers of 3×3 2-D convolutional filters with ReLU activation functions. In the encoder stage, the input image is processed by an increasing number of filters followed by max pooling subsampling after the convolution layers. Reaching a final spatial size of 8×8, the decoder stage increases the spatial size gradually by upsampling and convolution stages with decreasing number of filters. In addition, the network has multiple skip connections from the encoder to the decoder stage to recover the fine-grained spatial details which may be lost after max pooling. The network was designed for real-time performance by keeping the number of layers and convolutions as low as possible and using a simple 2×2 repeat upsampling instead of transposed convolution/deconvolution for the decoder. The result is a network with about two million parameters which can do segmentation in a matter of milliseconds. Network input is a single image resized to 256×256 pixels, and the output is an image of the same size as the input with four channels. Each channel is a normalized logit for each class by softmax activation. The output thus represents an image segmentation with four classes: background, LV lumen, myocardium, or left atrium.

D. Landmark Extraction

Contours of the LV were extracted using morphological erosion on the segmentation. These contours were used to extract three basal landmarks (left, right, and mid $\vec{B}$) and one apex landmark $\vec{A}$, as shown in Fig. 2. An overview of the symbols used and their meaning can be found in Table I. A contour point was determined to be at the base if any pixels immediately below was segmented as the left atrium. From these base contour points, the left, right, and mid base landmarks were extracted. The apex landmark was the contour point furthest away from the base mid landmark.

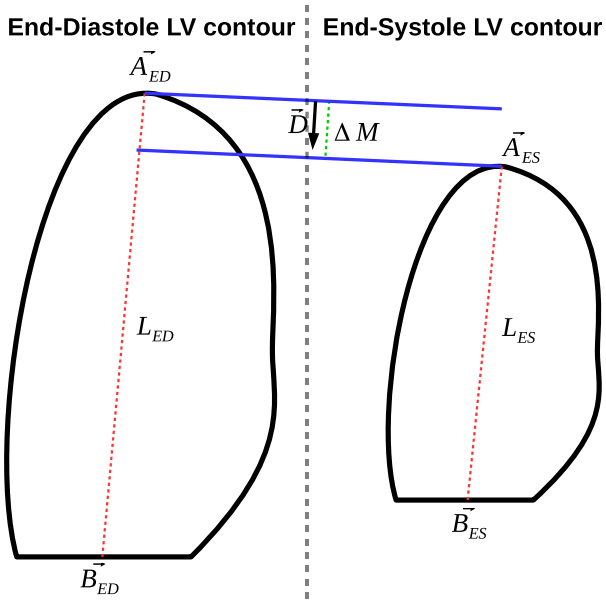

Fig. 2. Illustration of an extracted LV contour, and the apex and base mid landmarks $\vec{A}$, $\vec{B}$ in both ED and ES. These contours and landmarks are used to calculate the proposed apical foreshortening metrics: LV length difference ΔL and longitudinal apex movement ΔM.

TABLE I
OVERVIEW OF SYMBOLS USED

Symbol	Meaning
$L_{t,v}$	LV length at time t in view v
ΔL_t	LV length difference at t
ΔM_v	Longitudinal apex motion in view v
$\vec{A}_{t,v}$	Apex position at time t in view v
$\vec{B}_{t,v}$	Base mid point position at time t in view v
$\vec{D}$	Longitudinal LV length direction
α_v	Angle between apex movement vector and the longitudinal LV direction $\vec{D}$ in view v
$d_v(i)$	LV diameter at step i in view v
V_t	Volume at time t

E. Apical Foreshortening Detection

Ideally, if there is no foreshortening, the LV length should be the same in both views and the apex should be stationary throughout the heart cycle. However, in real life, some LV length difference and apex movement are expected even when there is no apical foreshortening [2]. Based on these assumptions, an apical foreshortening quantification method is proposed. The method involves two different metrics which use the landmarks derived from the LV segmentation as shown in Fig. 2.

1) The difference in LV length L between the A4C and A2C views for both ED and ES

$$\Delta L_{\mathrm{ED}} = |L_{\mathrm{ED,A4C}} - L_{\mathrm{ED,A2C}}| \quad (1)$$

$$\Delta L_{\mathrm{ES}} = |L_{\mathrm{ES,A4C}} - L_{\mathrm{ES,A2C}}| \quad (2)$$

2) Longitudinal apex movement from ES to ED in each view

$$\Delta M_{\mathrm{A4C}} = |\vec{A}_{\mathrm{ES,A4C}} - \vec{A}_{\mathrm{ED,A4C}}| \cos\alpha_{\mathrm{A4C}} \quad (3)$$

$$\Delta M_{\mathrm{A2C}} = |\vec{A}_{\mathrm{ES,A2C}} - \vec{A}_{\mathrm{ED,A2C}}| \cos\alpha_{\mathrm{A2C}} \quad (4)$$

where α is the angle between the apex movement vector from ES to ED and the longitudinal direction $\vec{D}$ of the LV, which is calculated as the vector from apex $\vec{A}$ to base $\vec{B}$ summed over ED and ES

$$\vec{D} = (\vec{A}_{\mathrm{ED}} - \vec{B}_{\mathrm{ED}}) + (\vec{A}_{\mathrm{ES}} - \vec{B}_{\mathrm{ES}}) \quad (5)$$

$$\cos\alpha = \frac{\vec{D}}{|\vec{D}|} \cdot \frac{\vec{A}_{\mathrm{ES}} - \vec{A}_{\mathrm{ED}}}{|\vec{A}_{\mathrm{ES}} - \vec{A}_{\mathrm{ED}}|} \quad (6)$$

F. EF Estimation

The volumes were calculated using Simpson's method with 20 disks, as this is the recommended clinical practice [1]. In this method, the diameter d is calculated at 20 steps perpendicular to the mid-axis of the LV. This is done in both the A4C and A2C views. The volume V was then calculated as

$$V = \frac{\max(L_{\mathrm{A4C}}, L_{\mathrm{A2C}})}{20} \sum_{i=1}^{20} \pi \frac{d_{\mathrm{A4C}}(i) d_{\mathrm{A2C}}(i)}{4} \quad (7)$$

where L is the length of the LV. Both the ED and ES volumes (V_{ED} and V_{ES}) were calculated. From these volumes, EF was calculated in percentage as

$$\mathrm{EF} = 100 \frac{V_{\mathrm{ED}} - V_{\mathrm{ES}}}{V_{\mathrm{ED}}} \quad (8)$$

G. Real-Time Pipeline

The methods were implemented using the medical high-performance computing and visualization framework FAST [24]. This framework has been developed with focus on data streaming and real-time image processing using parallel and GPU computing. For the neural network inference, FAST can use Google's TensorFlow, Intel's OpenVINO, or NVIDIA's TensorRT for high-speed inference [25]. In this work, the TensorFlow backend was used. Fig. 3 shows the pipeline used for real-time processing. The different processing steps are spread out over four main threads to enable concurrent processing in real time. Thread 1 maintains a connection with the ultrasound scanner and receives 2-D images from the scanner in real time. When processing stored images, this thread reads data from disk and outputs them in the interval they were originally captured in. Thread 2 reads the current image from thread 1 and runs the view classification neural network for each frame. A4C and A2C images are stored in separate buffers, and any other views are discarded. This thread executes every time a new frame is available from thread 1. Any frames that are older than 2 s are discarded from the buffers. Thread 3 executes at regular intervals, two times per second. While images are received, processed, and visualized at real-time frame rates, this thread only executes at 2 Hz to allow the A2C/A4C image buffers to build up to the size needed for the timing network to find ED and ES. This thread first runs the timing network on the most recent 80 frames, or less, of the current image view buffer. If the timing algorithm is able to find an ED and ES frame, the segmentation network is executed for both these two frames and the result is stored. Thread 4 uses the ED and ES ultrasound images and segmentations to

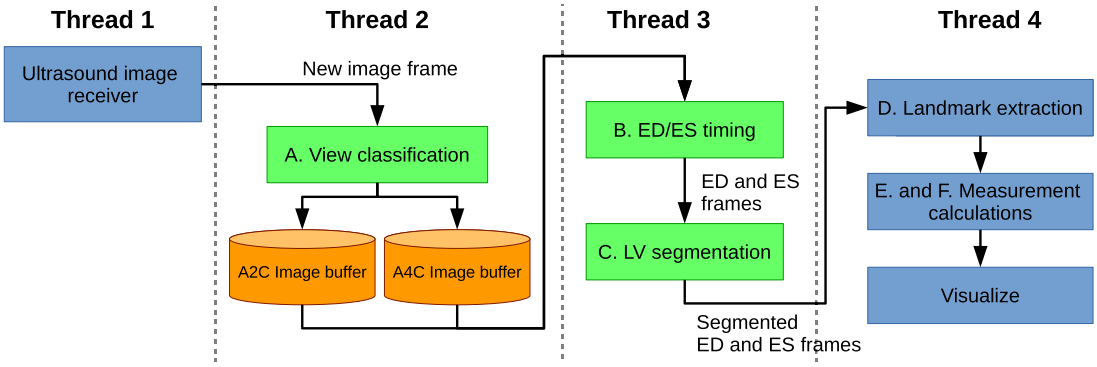

Fig. 3. This block diagram shows the real-time application pipeline, and how data flows between each processing step and thread. The green blocks are the three neural networks used in the pipeline.

first find the landmarks and then performs the measurements. Finally, the results are visualized as shown in Fig. 4 which also shows how the resulting application works in practice. The user simply starts by placing the probe and looks for the A4C or A2C view, when one of these views is found the user can optimize the view using the view quality feedback, the segmentation, and also the foreshortening information shown on screen while scanning. When satisfied with one of the views, the user can move the probe to find the next view, optimize it, and view the measurements without any manual user input. A video demonstrating the real-time application is available online in the multimedia material and on YouTube https://youtu.be/MYeM-qPD_yk

H. Data

1) Training Data: The neural networks for ED/ES timing and LV segmentation were trained using the publicly available CAMUS data set [4] with A4C and A2C 2-D ultrasound recordings of 500 patients from a French outpatient clinic. The view classification network was trained using 2-D ultrasound recordings from multiple views of 500 patients from a Norwegian population study and data from a Norwegian outpatient clinic [15]. None of these training data were used for the evaluation and testing in this article. The evaluation data are described in the next two paragraphs.

2) Apical Foreshortening Evaluation: A 3-D ultrasound data set was used to evaluate if the two proposed foreshortening metrics can quantify the amount of apical foreshortening. For this purpose, the open CETUS data set [26] was used, which contains several 3-D echocardiography recordings with ground truth segmentations of both ED and ES. This annotated 3-D data set enables us to: 1) know the true volume of the LV; 2) locate the true anatomical apex; and 3) extract a 2-D ultrasound image with an arbitrary amount of foreshortening. The CETUS data set consists of recordings from 45 patients in total, where 15 of these have ground truth segmentations available publicly. Since 3-D ultrasound images have a lower image quality than 2-D, the five recordings with the best image quality were selected for analysis with an EF ranging from 14% to 51% according to calculations done by the database organizers. The optimal four- and two-chamber planes were manually found defining the apex and base points in both ED and ES. The 3-D volumes were then sliced systematically about these planes to create 2-D images at different angles, simulating a tilt effect as shown in Fig. 1, and thus different amounts of apical foreshortening. Each slice runs through the two base points, while the distance from the apex was gradually increased at a tilt angle of 1°.

3) Volume and EF Evaluation: A data set of 100 patients was collected from a clinical database of patients diagnosed with acute myocardial infarction or de novo heart failure at a Norwegian hospital. The study was approved by the regional ethics committee (ref. 2013/573) and written consent was given by all patients. The images were acquired using a GE Vingmed Vivid (E7, E9 or E95) scanner. Patients were included consecutively regardless of image quality. All exams were performed in clinically stable patients with sinus rhythm. Images were analyzed by a single clinician using clinical best practice as defined in [1] with the clinical software EchoPAC (GE Vingmed, Horten, Norway), thereby, ensuring that the proposed automated method is compared with actual clinical practice measurements techniques. Also, this data set was not used for training any of the neural networks, ensuring that there is no data overfitting involved on the final clinical measurements.

III. Results

A. Apical Foreshortening Detection

Fig. 5 shows the absolute error in EF when the amount of apical foreshortening was gradually increased by tilting the view plane as shown in Fig. 1. The plot shows the mean and deviation of the absolute difference for both A4C and A2C views as a function of the amount of tilt in degrees in both directions. This shows clearly that the EF error increases with the amount of foreshortening. Similar plots were created for the two proposed foreshortening metrics. In Fig. 6, the apex motion and LV length difference metrics are plotted as a function of the amount of apical foreshortening in degrees. The same trends can be seen in these plots; as the amount of apical foreshortening increases, the two metrics also increase. This strongly indicates that given an accurate segmentation, the two metrics can in fact detect foreshortening. Also, note that even when there is no foreshortening (x-axis = 0°), there is some apex motion (0.5–3.5 mm) and LV length

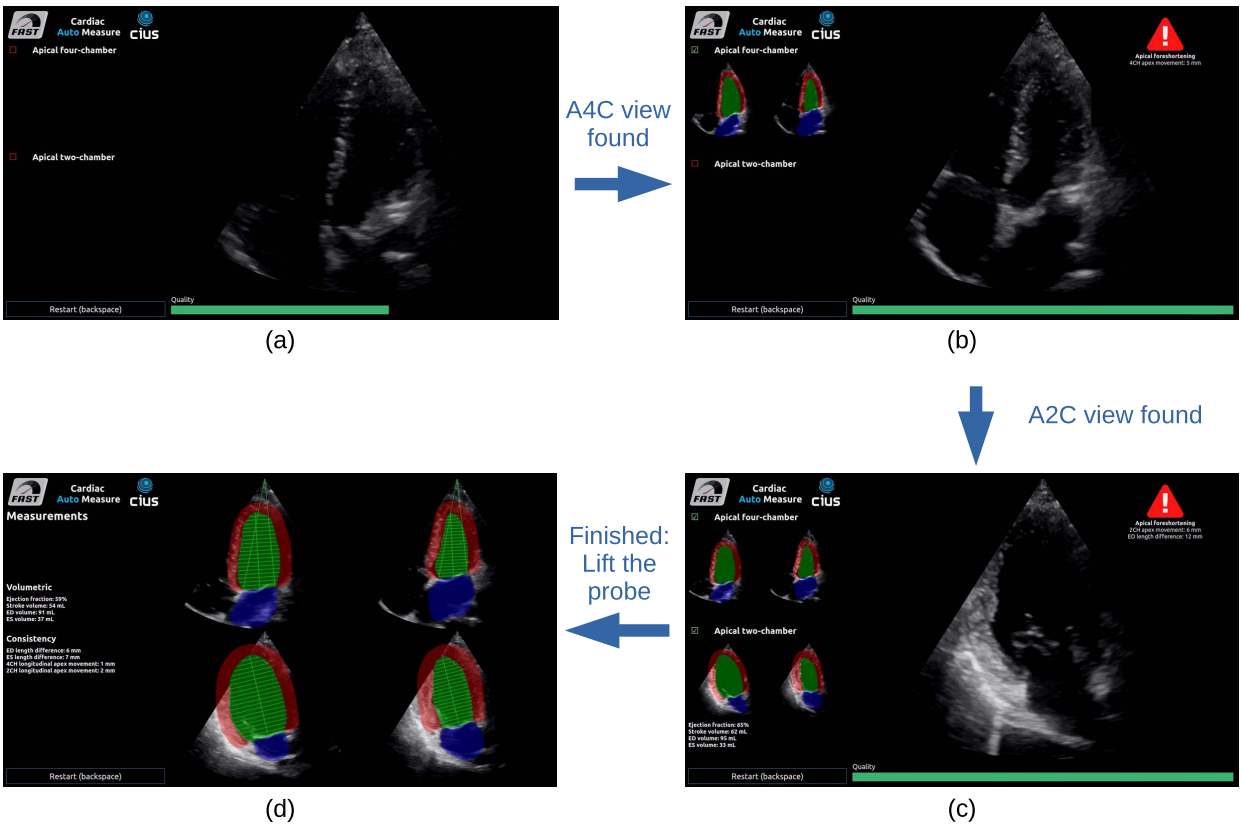

Fig. 4. Four screenshots of the real-time implementation of the full pipeline enabling users to do volume and EF measurements without any manual input. The users simply need to find and optimize the A4C and A2C views. The segmented ED and ES images of both A4C and A2C views are displayed on the left side when scanning and may be updated at any time by repositioning the ultrasound probe as seen in (b) and (c). A quality bar at the bottom gives users feedback on the view quality [see (a)–(c)], and if apical foreshortening is detected, a warning with quantitative information is displayed in the top right corner as seen in (b) and (c). A video demonstrating the real-time method is available online in the multimedia material and on YouTube https://youtu.be/MYeM-qPD_yk. (a) Start by placing the probe and look for A4C using the view quality feedback at the bottom. (b) Optimize the A4C view using view quality feedback, segmentations, and foreshortening info. (c) Optimize the A2C view as well, view the measurements, and redo the A4C view if needed. (d) View a summary of the measurements along with segmentation in both views in both ED and ES.

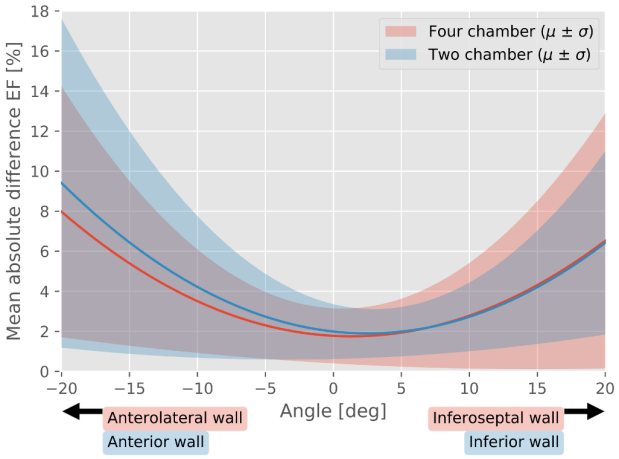

Fig. 5. Absolute error in EF when the image plane is gradually tilted with respect to the defined four- and two-chamber base point axes, respectively. The trend lines correspond to a mean and deviation of five patients. The red lines and area correspond to the four-chamber view, while the blue correspond to the two-chamber view.

difference (0–3 mm). This highlights that in real life some apex motion and LV length difference is to be expected, but still the minimum values are obtained where there is no foreshortening.

An additional experiment was performed to validate how the proposed method works in practice with the proposed pipeline and the 100 patients 2-D data set. On average, the foreshortening metrics with the automatic pipeline on this data set were 3.9 ± 4.2 and 3.4 ± 4.3 mm for the LV length difference in ED and ES, respectively, and 2.4 ± 2.1 and 2.8 ± 2.3 mm for the apex motion in A4C and A2C, respectively. In this experiment, one expert performed a qualitative assessment by categorizing recordings of the 100 patients into three categories of no/low, moderate, and significant foreshortening. A histogram showing that this classification versus the automatic foreshortening metrics is shown in Fig. 7, where each color represents the expert's qualitative assessment. In six of the patients, there was a clear anatomical error in the segmentation, and these are marked in blue in the histogram. For this histogram, a combination F of the four foreshortening measurements was made

$$F = 4\max(\Delta M_{\mathrm{A4C}}, \Delta M_{\mathrm{A2C}}) + \max(\Delta L_{\mathrm{ED}}, \Delta L_{\mathrm{ES}}) \quad (9)$$

where the factor 4 was introduced by observing in Fig. 6 that the LV length difference increased much faster than the apex motion. This experiment indicates that although there is a clear trend of moderate and significant foreshortening, as well as

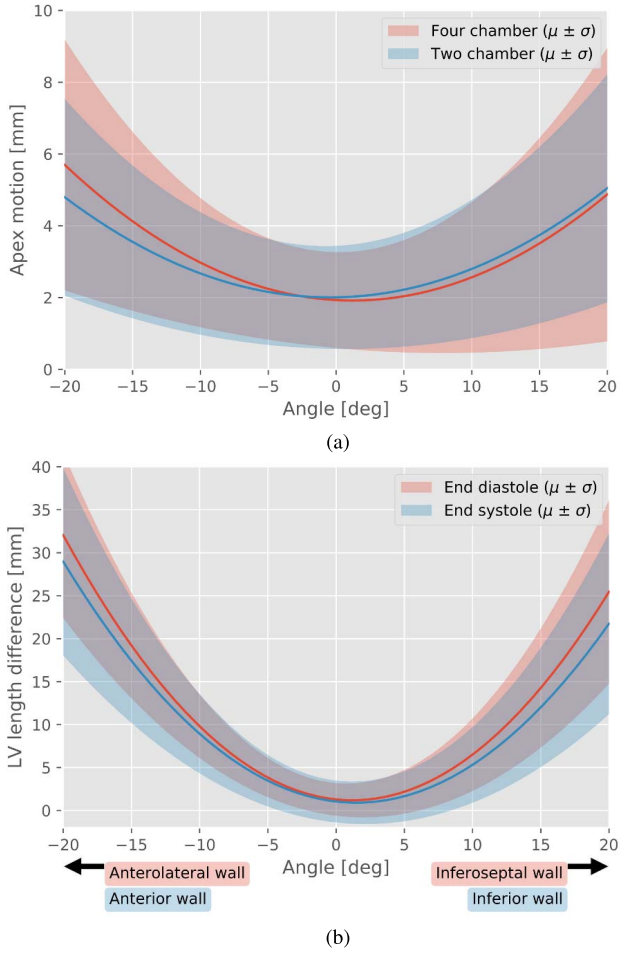

Fig. 6. Apical foreshortening detection metrics when the degree of apical foreshortening was gradually increased. The trend lines correspond to a mean and deviation of five patients, where 3-D volumes and corresponding LV meshes were used to simulate the effect of apical foreshortening by moving the image planes systematically about the true apex of each mesh as shown in Fig. 1. (a) Apex motion ΔM_{A4C} and ΔM_{A2C} versus foreshortening angle. The red lines and area correspond to the four chamber view, while the blue correspond to the two chamber view. (b) LV length difference ΔL_{ED} and ΔL_{ES} versus foreshortening angle. The red lines and area correspond to ED, while the blue correspond to ES.

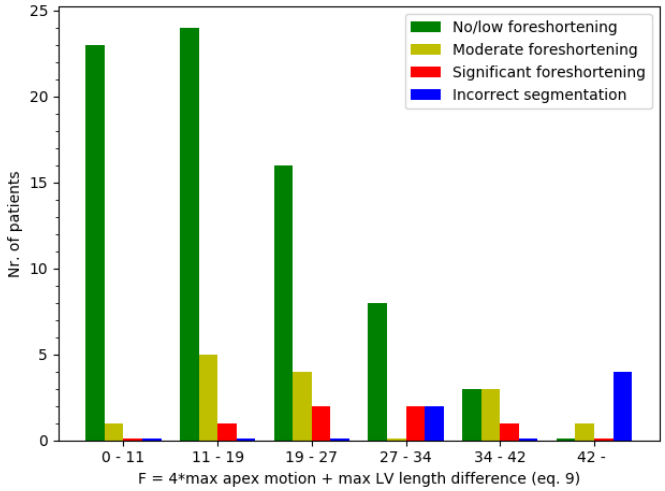

Fig. 7. Histogram with six bins showing the proposed foreshortening detection method compared with a qualitative foreshortening analysis performed by one expert on the 100 patient 2-D data set. The different colors represent the expert's assessment of each patient, and the *x*-axis is *F*, a combination of the two proposed foreshortening measurements extracted from the neural network segmentation of the ultrasound images, [see (9)]. In six of the patients, there was a clear anatomical error in the segmentation, and these are marked in blue in the histogram.

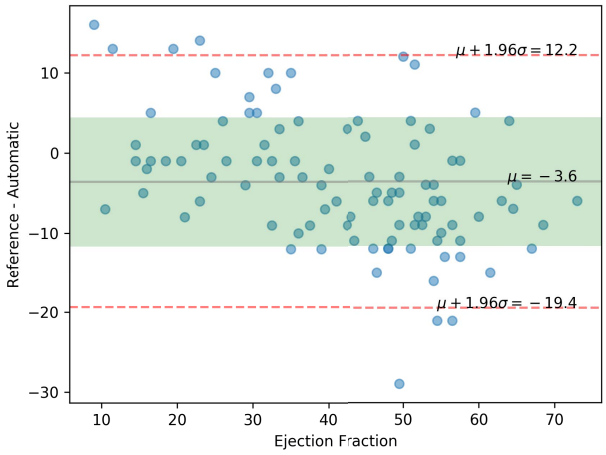

Fig. 8. Bland–Altman plot of EF with reference minus the proposed automatic method for 100 patients. The green field marks the standard deviation around the mean, while the red lines are the 95% limits of agreement.

anatomically incorrect segmentations, having larger values of apex motion and LV length difference, there is no clear cutoff.

B. Measurement Accuracy

The accuracy of automatic ED volume, ES volume, and EF measurements compared with the clinical data set of 100 patients is collected in Table II. Accuracy was measured using a Bland–Altman analysis, where the bias and standard deviation were calculated as the reference minus the proposed automatic method. Fig. 8 shows a Bland–Altman plot of EF. The mean and median absolute difference were also measured. For comparison, reported volume and EF accuracy numbers were taken from relevant articles (see [4], [16], [19]) and included in Table II. While the studies of Zhang *et al.* [16] and Jafari *et al.* [19] used the LV area method to find ED and ES, the proposed method uses a separate neural network to find ED and ES. To study the effect of using the LV area method instead on the same data, an additional experiment was performed using the proposed method with the LV area method instead. The results of this experiment are included in the second row of Table II.

In the CAMUS study of Leclerc *et al.* [4], predefined ED and ES frames were used. The methods in this study were thus not fully automatic in calculating volume and EF, and the exact same frames were used for the reference measurements as the automatic measurements. While for the other studies, including the proposed method, the frames used for reference and automatic measurements can be different. We also applied the proposed fully automatic method on the CAMUS data set producing volume and EF measurements seen in the third row of Table II. As these data are publicly available, others may also use this data for direct comparison. Note, however, that EF was not measured using clinical software in CAMUS, as was done for the 100 patients data set.

TABLE II
MEASUREMENT ACCURACY OF THE PROPOSED METHOD COMPARED WITH CLINICAL MEASUREMENTS IN ECHOPAC. BIAS WAS MEASURED AS THE REFERENCE MINUS THE AUTOMATIC MEASUREMENTS. REPORTED EF AND VOLUME ACCURACY TAKEN FROM ARTICLES OF RELEVANT WORK ARE ALSO PROVIDED AT THE BOTTOM FOR COMPARISON

Study	Level of automation	Test dataset	Measurement	Bias	Std. dev.	Mean abs. difference	Median abs. difference
Proposed method	Fully automatic.	IMPROVE 100 patients	V_{ED} (mL)	40.1	31.8	40.1	31.0
			V_{ES} (mL)	29.2	24.8	29.3	23.0
			EF (%)	-3.6	8.1	7.2	6.0
Proposed method	Fully automatic. LV area ED/ES detection.	IMPROVE 100 patients	V_{ED} (mL)	23.6	24.9	25.2	16.5
			V_{ES} (mL)	39.9	32.3	39.9	31.0
			EF (%)	-18.2	8.8	18.3	18.0
Proposed method	Fully automatic.	CAMUS 406 patients	V_{ED} (mL)	9.7	13.6	12.3	10.0
			V_{ES} (mL)	3.9	9.7	7.5	5.0
			EF (%)	1.8	8.9	6.7	5.0
Jafari et al. [19]	Semi automatic: No view classification. LV area ED/ES detection.	427 patients	EF (%)	NA	NA	7.9	6.2
Zhang et al. [16]	Fully automatic. LV area ED/ES detection with outlier rejection.	6407 echocardio-grams	V_{ED} (mL)	NA	28.6	NA	NA
			V_{ES} (mL)	NA	19.9	NA	NA
			EF (%)	NA	10.2	NA	9.7
Leclerc et al. [4] U-net 1	Semi automatic: No view classification. No ED/ES detection. Poor quality images were excluded.	CAMUS 406 patients	V_{ED} (mL)	-8.3	12.6	10.9	NA
			V_{ES} (mL)	-4.9	9.9	8.2	NA
			EF (%)	-0.5	7.7	5.6	NA

TABLE III
AVERAGE RUNTIME PERFORMANCE OF EACH STEP IN THE PIPELINE AND THE OVERALL FPS PROCESSED

Step	Runtime (ms)
View classification neural network	12.1 ± 0.3
ED/ES detection neural network	149.5 ± 17.8
LV Segmentation neural network	13.4 ± 0.6
Landmark extraction and measurement	4.0 ± 1.3
Step	**Frames per second (FPS)**
All	42.6 ± 7.8

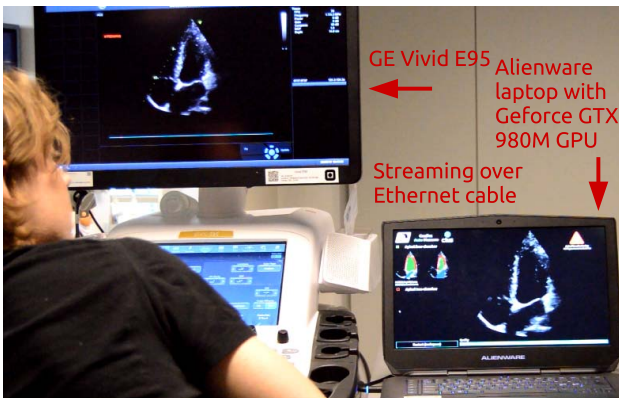

Fig. 9. Hardware setup used for real-life testing. Ultrasound images are streamed in real-time from a GE Vivid E95 scanner to a laptop with a GPU which runs the proposed method.

C. Runtime Performance

Since real-time performance is essential for enabling users to optimize their ultrasound acquisition with regard to foreshortening and EF measurement, execution runtimes were measured and collected in Table III. The runtimes were measured using an Alienware laptop with an Intel i7-6700 CPU and an NVIDIA GTX 980M GPU with 8 GB of memory, and the real-time hardware setup is shown in Fig. 9. The overall runtime was measured in frames per second (FPS). When a frame is received in thread 2, its timestamp is recorded. When thread 3 is about to process the buffer of image frames, the FPS is calculated as the number of frames in the buffer divided by the duration from the first to the last frame in the buffer. From this table, we see that real-time performance was achieved with an average FPS of 42. Additionally, the runtime of each of the three neural networks: view classification, ED/ES timing, and LV segmentation were measured. These runtimes include the time needed for normalizing the image intensities and resizing the images to fit the input layer size. Note that the runtimes were calculated while the entire application was running. Thus, the neural networks, visualizations, and other computations are all executing simultaneously which affects the runtime of each component. The runtime of the neural networks reported here may therefore be higher than if the network runtimes were measured independently.

IV. DISCUSSION

The apical foreshortening plots in Figs. 5 and 6 show first that EF is considerably affected by apical foreshortening,

which is a known issue in echocardiography [1], [2], and second that the proposed LV length difference and longitudinal apex movement metrics have the potential of detecting and quantifying apical foreshortening. The proposed foreshortening detection is therefore a promising tool for improving the quality of volume and EF measurements in echocardiography. Still, inducing apex motion and LV length difference through synthetic image slicing in 3-D is suboptimal, due to dependence of myocard motion and difference in image quality. The histogram in Fig. 7 shows the foreshortening detection method's agreement with one expert's qualitative assessment of the amount of foreshortening in the 100 patients data set. Although there is a clear trend of moderate and significant foreshortening having larger values of apex motion and LV length difference, there is no clear cutoff. Still, a limitation with this data set is that only 6 of 100 patients had significant foreshortening. The acquisitions in this data set were acquired as part of a research project by experienced cardiologists. Thus, the frequency of significant foreshortening is expected to be higher for a normal busy outpatient clinic and more inexperienced operators. Since the current segmentation method only processes one frame at a time, the segmentation contour do not necessarily follow the same physical contour over time. This is especially important for the apex motion metric, and therefore we believe a more temporal consistent segmentation will improve the foreshortening detection results. It can also be observed that the metrics can be used to detect anatomically incorrect LV segmentations. A follow-up study is necessary to see if the method can help clinicians to get more accurate and consistent measurements while scanning the patient. Furthermore, the trend in the plots in Fig. 6 indicates that these metrics might even be used to correct the volume and EF measurements. However, substantially more data are needed to properly establish the true relationship between the measurement errors and the foreshortening metrics.

As foreshortening is introduced during acquisition, it is essential to have a detection method that runs in real time while scanning in order to help the operator optimize the view of the LV. To this end, the runtime results showed that by utilizing a mid-end GPU, fully automatic foreshortening detection and EF measurements can be performed in real time with an average FPS of 42. Input image resolution of the neural networks is one factor that affects the runtime of these networks. Thus, we have kept the input image resolution as low as possible while maintaining a high accuracy for each task. This is why the input resolution differ for each network: 256×256, 128×128, and 128×80 for the segmentation, view classification, and timing, respectively. The timing of ED and ES was the most time-consuming task, and cannot be executed in real time in its present form as it has to process a long sequence of frames. This was solved in this study by executing the timing network in a separate thread at regular intervals. Runtime may be further improved by for instance using dynamic recurrent neural networks by stateful LSTMs which can process a single frame at a time while keeping an internal state.

The volume accuracy measurements of the proposed automatic pipeline show that there is a considerable bias in the volume estimation (40 mL EDV and 29 mL ESV) meaning that the method generally underestimates the volume. This is most likely due to the fact that two different experts have analyzed the data used for training the segmentation network (CAMUS data set), and the data used in this article for evaluation. The additional experiment using the proposed fully automatic method on the CAMUS data set shows an improvement on almost all metrics, thus, the accuracy differences are most likely due to interobserver variability. Fig. 10 is an example from the CAMUS study which illustrates the large tracing differences that can exist between three different experts. Clearly, the three different tracings in this example would result in large volume differences. This illustrates the need for establishing a common data set for machine learning created by consensus of multiple experts. The bias difference between EDV and ESV is thus translated to a bias in EF, resulting in a slight overestimation of EF by the proposed automatic method. Still the standard deviation in EF is low, lower than the interobserver variability measured on the CAMUS data set [4] (8.1–11.0), but a little bit higher than the automatic segmentation methods (7.1–7.7). However, in the CAMUS comparison study, segmentation and EF measurements were performed on the same predefined ED and ES frames, the measurements are thus not fully automatic. Also, both training and test data were annotated by the same expert, and poor quality images were excluded from the evaluation. These three factors are most likely the cause of the higher variability in EF in this study versus the CAMUS study. The semiautomatic method of Jafari *et al.* [19] achieved slightly worse results, and they presented no variability measures, making it hard to compare the robustness of the two methods. Our EF results compared favorably with the work of Zhang *et al.* [16] which reported a median absolute difference of 9.7 and a standard deviation of 10.2. This is most likely because of their lower segmentation accuracy (0.89 versus 0.92 in Dice score) and the fact that they also, like Jafari *et al.* [19], select ED and ES frames based on the LV segmentation area. However, Zhang *et al.* [16] used an outlier detection method to deal with this issue, while Jafari *et al.* [19] did not report such. Thus, the LV area method may not be optimal for measuring EF automatically, since it is sensitive to errors in the segmentation. This might change

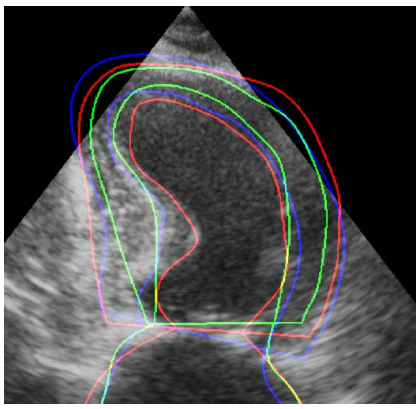

Fig. 10. Example of three expert's tracings of the LV from the CAMUS study [4] illustrating the need for establishing a common data set for machine learning created by consensus of multiple experts.

in the future if a more accurate segmentation is achieved over the entire heart cycle. Another major difference to consider between these studies in automatic EF measurements is that our training and evaluation data are two completely separate data sets from different clinics and annotated by different experts using different software, while [16] and [19] have used data from the same data set for training and evaluation.

An end-to-end neural network approach for automatic estimation of EF directly from ultrasound images is feasible, as demonstrated [20]. The downside of an end-to-end approach is that it results in a black-box method in which clinicians cannot visually inspect, verify, and correct the EF measurements. Also, with an end-to-end method, it would not be straightforward to give real-time visual quality assurance feedback as the proposed method does.

The same applies for foreshortening detection; it might be possible to train an end-to-end neural network to detect and quantify foreshortening, but it would suffer from the same black-box dilemma. The proposed method that uses multiple neural networks has the advantage of being based on real physical quantities, explainable, visualizable, and very fast.

V. Conclusion

This study presented a method for real-time automatic EF measurements and quantification of apical foreshortening in 2-D cardiac ultrasound using deep neural networks. A quantitative evaluation using 3-D ultrasound showed that EF is considerably affected by foreshortening, and that the method can measure the amount of foreshortening and thereby a promising tool for improving ultrasound image acquisitions while scanning. The automatic EF measurements were showed to be within interobserver variability and comparable with other related work on automatic EF.

References

[1] R. M. Lang *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. Amer. Soc. Echocardiography*, vol. 28, no. 1, pp. 1–39, Jan. 2015.

[2] S. Ünlü *et al.*, "Impact of apical foreshortening on deformation measurements: A report from the EACVI-ASE strain standardization task force," *Eur. Heart J.-Cardiovascular Imag.*, vol. 21, no. 3, pp. 337–343, 2020.

[3] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.

[4] S. Leclerc *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019.

[5] O. Oktay *et al.*, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 384–395, Feb. 2018.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[7] S. A. Aase, A. Stoylen, C. B. Ingul, S. Frigstad, and H. Torp, "Automatic timing of aortic valve closure in apical tissue Doppler images," *Ultrasound Med. Biol.*, vol. 32, no. 1, pp. 19–27, Jan. 2006.

[8] S. A. Aase, S. R. Snare, H. Dalen, A. Stoylen, F. Orderud, and H. Torp, "Echocardiography without electrocardiogram," *Eur. J. Echocardiography*, vol. 12, no. 1, pp. 3–10, Jan. 2011.

[9] P. Gifani, H. Behnam, A. Shalbaf, and Z. A. Sani, "Automatic detection of end-diastole and end-systole from echocardiography images using manifold learning," *Physiological Meas.*, vol. 31, no. 9, pp. 1091–1103, Sep. 2010.

[10] A. Shalbaf, Z. AlizadehSani, and H. Behnam, "Echocardiography without electrocardiogram using nonlinear dimensionality reduction methods," *J. Med. Ultrason.*, vol. 42, no. 2, pp. 137–149, Apr. 2015.

[11] B. Kong, Y. Zhan, M. Shin, T. Denny, and S. Zhang, "Recognizing end-diastole and end-systole frames via deep temporal regression network," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science), vol. 9902, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016, pp. 264–272.

[12] F. T. Dezaki *et al.*, "Deep residual recurrent neural networks for characterisation of cardiac cycle phase from echocardiograms," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, vol. 10553, 2017, pp. 100–108.

[13] A. M. Fiorito, A. Ostvik, E. Smistad, S. Leclerc, O. Bernard, and L. Lovstakken, "Detection of cardiac events in echocardiography using 3D convolutional recurrent neural networks," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.

[14] X. Gao, W. Li, M. Loomes, and L. Wang, "A fused deep learning architecture for viewpoint classification of echocardiography," *Inf. Fusion*, vol. 36, pp. 103–113, Jul. 2017.

[15] A. Østvik, E. Smistad, S. A. Aase, B. O. Haugen, and L. Lovstakken, "Real-time standard view classification in transthoracic echocardiography using convolutional neural networks," *Ultrasound Med. Biol.*, vol. 45, no. 2, pp. 374–384, Feb. 2019.

[16] J. Zhang *et al.*, "Fully automated echocardiogram interpretation in clinical practice," *Circulation*, vol. 138, no. 16, pp. 1623–1635, Oct. 2018.

[17] A. H. Abdi *et al.*, "Automatic quality assessment of echocardiograms using convolutional neural networks: Feasibility on the apical four-chamber view," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1221–1230, Jun. 2017.

[18] E. Smistad, A. Ostvik, I. M. Salte, S. Leclerc, O. Bernard, and L. Lovstakken, "Fully automatic real-time ejection fraction and MAPSE measurements in 2D echocardiography using deep neural networks," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.

[19] M. H. Jafari *et al.*, "Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 6, pp. 1027–1037, Jun. 2019.

[20] J. F. Silva, J. M. Silva, A. Guerra, S. Matos, and C. Costa, "Ejection fraction classification in transthoracic echocardiography using a deep learning approach," in *Proc. IEEE 31st Int. Symp. Comput.-Based Med. Syst. (CBMS)*, vol. 2018, Jun. 2018, pp. 123–128.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[22] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[23] E. Smistad, A. Ostvik, B. O. Haugen, and L. Lovstakken, "2D left ventricle segmentation using deep learning," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[24] E. Smistad, M. Bozorgi, and F. Lindseth, "FAST: Framework for heterogeneous medical image computing and visualization," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 11, pp. 1811–1822, Nov. 2015.

[25] E. Smistad, A. Ostvik, and A. Pedersen, "High performance neural network inference, streaming, and visualization of medical images using FAST," *IEEE Access*, vol. 7, pp. 136310–136321, 2019.

[26] O. Bernard *et al.*, "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 967–977, Apr. 2016.