



# Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training

Mohammad H. Jafari<sup>1</sup> · Hany Girgis<sup>1,2</sup> · Nathan Van Woudenberg<sup>1</sup> · Zhibin Liao<sup>1</sup> · Robert Rohling<sup>1</sup> · Ken Gin<sup>1,2</sup> · Purang Abolmaesumi<sup>1</sup> · Terasa Tsang<sup>1,2</sup>

Received: 31 January 2019 / Accepted: 22 March 2019  
© CARS 2019

## Abstract

**Purpose** Left ventricular ejection fraction (LVEF) is one of the key metrics to assess the heart functionality, and cardiac ultrasound (echo) is a standard imaging modality for EF measurement. There is an emerging interest to exploit the point-of-care ultrasound (POCUS) usability due to low cost and ease of access. In this work, we aim to present a computationally efficient mobile application for accurate LVEF estimation.

**Methods** Our proposed mobile application for LVEF estimation runs in real time on Android mobile devices that have either a wired or wireless connection to a cardiac POCUS device. We propose a pipeline for biplane ejection fraction estimation using apical two-chamber (AP2) and apical four-chamber (AP4) echo views. A computationally efficient multi-task deep fully convolutional network is proposed for simultaneous LV segmentation and landmark detection in these views, which is integrated into the LVEF estimation pipeline. An adversarial critic model is used in the training phase to impose a shape prior on the LV segmentation output.

**Results** The system is evaluated on a dataset of 427 patients. Each patient has a pair of captured AP2 and AP4 echo studies, resulting in a total of more than 40,000 echo frames. The mobile system reaches a noticeably high average Dice score of 92% for LV segmentation, an average Euclidean distance error of 2.85 pixels for the detection of anatomical landmarks used in LVEF calculation, and a median absolute error of 6.2% for LVEF estimation compared to the expert cardiologist's annotations and measurements.

**Conclusion** The proposed system runs in real time on mobile devices. The experiments show the effectiveness of the proposed system for automatic LVEF estimation by demonstrating an adequate correlation with the cardiologist's examination.

**Keywords** Mobile application · Deep learning · Adversarial training · Cardiac ejection fraction · Image segmentation · Echocardiography

## Introduction

Cardiac ultrasound is among the most widely used imaging modalities for study of the heart. 2D Echocardiography (echo) is a basis for acquisition of a broad range of diagnostic measurements for the evaluation of cardiac structures and functions, such as cardiac output, left ventricular ejection fraction (LVEF), and diastolic function. Compared to magnetic resonance imaging (MRI) and computed tomography (CT), ultrasound imaging is less costly, non-ionizing, and available to a wider range of patients such as those with cardiac implant devices.

As technology advances, the traditional cart-based or laptop-size ultrasound machines are being replaced by

---

Hany Girgis and Nathan Van Woudenberg: Joint first author and Purang Abolmaesumi and Terasa Tsang: Joint senior authors.

---

This work is funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by the Canadian Institutes of Health Research (CIHR).

---

✉ Mohammad H. Jafari  
mohammadj@ece.ubc.ca

<sup>1</sup> The University of British Columbia, Vancouver, Canada

<sup>2</sup> Vancouver General Hospital, Vancouver, Canada

portable point-of-care ultrasound (POCUS) devices, such as Philips Lumify, Clarius and Butterfly iQ. These are often packaged as an ultrasound probe, and a wireless or wired connection to a mobile device. This trend can be ascribed to the cost-effectiveness and ease of access of POCUS devices. Specifically, in an emergency or critical care scenario, a portable device enables clinicians to perform agile preliminary exams on patients and proceed with time-critical and potentially life-saving diagnostic decisions. Recent studies [9,20,21] also show that POCUS is beneficial to anesthesia practices.

Medical image analysis has enjoyed significant progress in recent years, specifically with the emergence of deep learning techniques [15,32,34]. A comprehensive survey of deep learning-based medical image analysis can be found in [17]. Particularly, we found a rich collection of literature in ultrasound image analysis given its large demand in clinical applications, such as using deep convolutional neural networks (CNN) and recurrent neural networks (RNN) to locate fetal standard imaging planes in cine clips [5,6]; using CNN for echo image quality estimation [1]; and using neural networks to generate text description for valvular diseases from Doppler images [22]. Furthermore, on 3D ultrasound, Ghesu et al. [10] use shallow and deep sparse networks to detect and segment aortic valve. Finally, a CNN-based system for fully automated cardiac structure and function determination and disease detection can be found in [35,36].

LVEF is a key metric to assess cardiac functionality and can be derived from apical two-chamber (AP2) and apical four-chamber (AP4) 2D echo exams. On a conventional ultrasound machine, in order to perform accurate LVEF calculation, the sonographer is required to capture a high-quality cine series that includes the end-diastolic (ED) and end-systolic (ES) frames of a cardiac cycle. With the assistance of the manufacturer's built-in software, the sonographer manually provides the main axis of LV and traces its boundary in the ED and ES frames. Subsequently, the LV volume at ED and ES phases is calculated to compute the LVEF ratio.

An early attempt for solving LV segmentation in 2D echo was with the use of Deep Belief Networks (DBN) for improving the robustness of a model trained on a small dataset [4]. This method was extended in [3] by combining DBN and dynamic models for LV tracking. Nascimento et al. [23] proposed to combine manifold learning with DBN for multi-atlas LV segmentation problem. Chen et al. [7] proposed a multi-view regularized Fully Convolutional Network (FCN) [18] model for improving LV segmentation in echo images. In [26], anatomical shapes learned through a T-L network [11] were used to regularize the training of deep learning networks for LV segmentation in both 3D ultrasound and MR images. The U-net [29] architecture—a popular method for medical image segmentation is also implemented for LV segmentation in 2D echo [33]. There are a num-

ber of research groups aiming to address LV segmentation in cardiac MR and CT images, where the techniques are often adaptable to echo images. Avendi et al. [2] combined CNN, stacked auto-encoders (AE), and deformable models to handle automatic LV segmentation in short-axis cardiac MR slices, where the CNN was used to localize LV and the stacked AE was used to infer the LV shape. A two stage strategy was utilized by Zreik et al. [37] to segment LV in 3D cardiac CT volumes, where a 3D LV bounding box was determined as an aggregated predictions of three 2D CNNs, while a voxel classification CNN was used to segment LV within the bounding box. Ngo et al. [24] proposed to use one DBN for LV localization and another one to perform initial LV segmentation, where level set was introduced to refine the segmentation. Poudel et al. [27] introduced incorporation of a recurrent connection within the U-net architecture to segment LV from short-axis cardiac MR slices. Patch-based CNNs were integrated into an active contour framework for LV boundary extraction by Rupprecht et al. [30].

In comparison with a conventional ultrasound machine, a mobile POCUS device has a limited processing power, memory, and storage space, to execute a live diagnostic software. In addition, while semi-automated LV segmentation systems can improve the accuracy of LV segmentation [24], it is practically hard to manually trace or correct LV borders on the screen of a hand-held device. To address the above issues, in this work, we aim to develop an integrated mobile application that provides a computationally efficient, automated, and accurate LVEF estimation without the need for user intervention.

The proposed method aims to calculate LVEF with the use of the biplane Simpson's method [16,31]—a standard method to calculate LV volume  $v$  at ED and ES phases:

$$v = \frac{\pi}{4} \sum_{i=1}^n a_i b_i \frac{L}{n}, \quad (1)$$

where  $L$  represents the length of the ventricular cavity (*i.e.*, the longest axis of LV, to be measured by the distance between LV apex to the middle of the mitral valve annular plane), and  $a$  and  $b$  are diameters of  $n$  equal height cylinders that are apportioned by dividing LV along  $L$  into  $n$  equal sections. Note that  $L$ ,  $a$ , and  $b$  are measured from the LV segmentation map detected from the perpendicular AP2 and AP4 echo views [31]. In order to achieve an accurate LVEF estimation, the apex and middle mitral valve plane landmarks (denoted as LV landmarks) should also be accurately located in addition to an accurate LV segmentation. Therefore, we propose a novel LVEF estimation framework which involves the use of a multi-task deep learning network to simultaneously solve LV segmentation and LV landmarks detection problems.

For the purpose of fitting the application within the memory and computational constraints of a POCUS system, we implement the proposed multi-task approach as a lightweight model. In general, a lightweight network (*i.e.*, a shallower and/or a slim network) does not perform as well as their deep and wider counterparts. To alleviate this issue, we adopt an adversarial training mechanism [12,19,28] to correct higher-order inconsistencies between the expert ground truth of LV segmentation maps and the prediction maps produced by the network. The adversarial training is able to regularize the network parameters in order to reduce over-fitting and hence, improve validation accuracy [19].

To summarize, our contributions in this work are:

- the proposed framework is the first automated pipeline for LVEF estimation using POCUS mobile devices and biplane Simpson's method;
- the proposed segmentation network is implemented as a lightweight multi-task network with the performance enhanced by adversarial training.

The block diagram for the proposed system is shown in Fig. 1. A cardiac POCUS device captures echo frames live from the patient and transmits images to the mobile application. The system input could be provided by an ultra-portable hand-held ultrasound probe, or a conventional cardiac ultrasound machine through a frame grabber. The operator captures AP2 and AP4 echo views, which are the standard echo planes to study the LV. The mobile application tracks the LV region in captured frames to calculate

the LVEF. A deep learning-based segmentation network acts as the core intelligence of the mobile application. The segmentation network is trained offline on archived echo data to simultaneously segment LV area and the two LV landmark points. The detected regions are used in the pipeline to estimate LVEF based on the aforementioned Simpson's method [31].

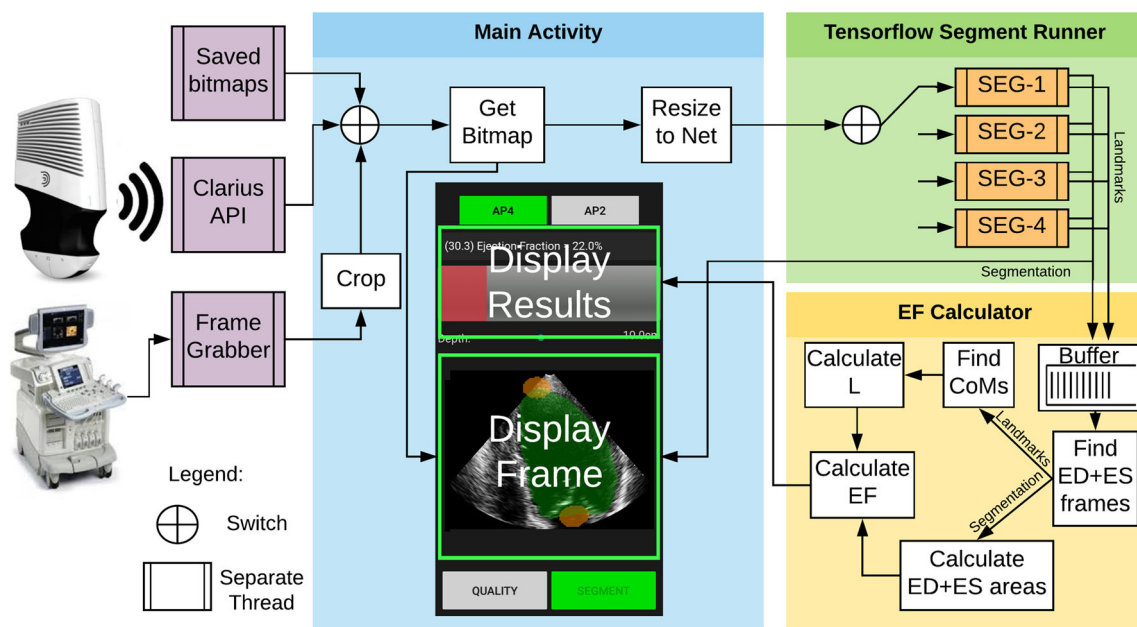
In Sect. 2, the details of the system workflow and the implemented mobile application are explained. The proposed multi-task segmentation and landmark detection methodology is discussed in Sect. 3.

## Mobile application

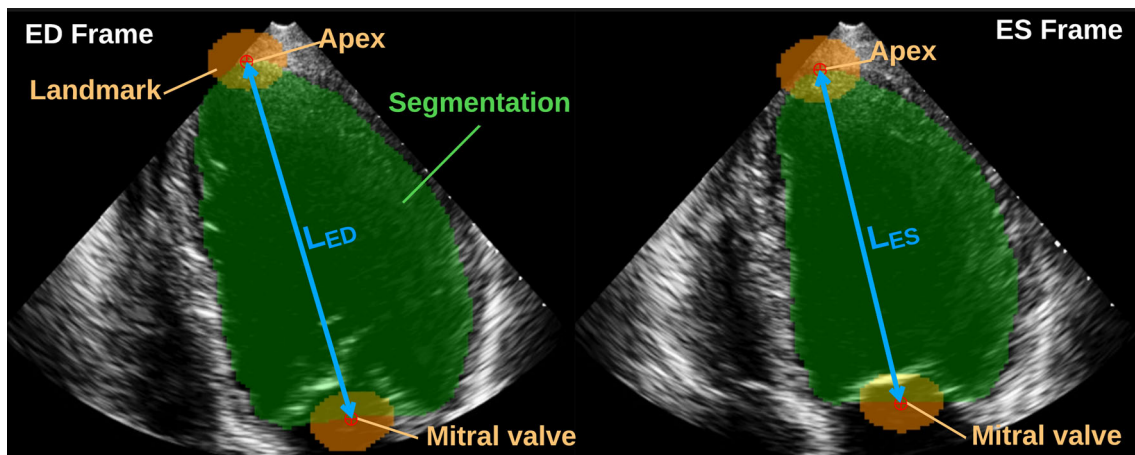
### Software pipeline

Figure 1 shows the data flow pipeline of the software. The application can be set to accept three different sources of input: bitmaps saved on the Android device, live frames from the Clarius probe streamed over a wireless connection, or live frames from a cart-based ultrasound machine streamed through a frame grabber.

The simplest way to receive input data is by storing datasets directly on the device's internal memory and then loading them frame by frame at a pre-specified frame rate. Alternatively, the wireless Clarius ultrasound probe can be used to stream live data over a wireless network. Finally, the device can also accept serial input through its USB-C port: we use an7 Epiphan AV.IO frame grabber to capture and convert



**Fig. 1** Block diagram of the proposed mobile system for real-time LV segmentation, landmark detection, and biplane ejection fraction estimation



**Fig. 2** Sample detected LV segmentation and the LV landmark points from the AP4 view. Left is the heart at the end-diastolic (ED) phase and right is the end-systolic (ES) frame. The volume ratio between ED

and ES frames calculated using the segmentation area and the landmark points is used to calculate LVEF

the output from the DVI port of any cart-based ultrasound machine, and pipe it directly into the Android device using a standard USB-C connection. When using this modality, we crop the raw frame-grabbed data so as to only include the ultrasound beam, the boundaries of which are set by the user once for each cart-based system.

Once properly connected, the application converts each full resolution ultrasound frame to a bitmap and displays it to the user in the application Graphical User Interface (GUI). After the user initiates the segmentation option, the application down-samples the raw frame data to the input dimensions of the neural network ( $128 \times 128$  pixels in our implementation). The resized frames are then sent to one of four concurrently running instances of *TensorFlow Mobile* Java inference engine. Each of these instances loads and runs the resized frames through the segmentation network, the design and training of which is described in Sect. 3. There are two outputs from the segmentation networks: the segmentation and the landmarks, shown in Fig. 2. The segmentation output (green) is a  $128 \times 128$  binary mask. The landmarks output (orange) is also a binary mask this time containing two blobs, one representing the most likely location of the LV's apex, and the other the mitral valve. The network outputs are then resized back up to the original frame dimensions, overlaid onto the original bitmap, and displayed in the application GUI. The outputs are also used to calculate LVEF, as described in Sect. 2.2.

Since the system runs in a resource limited environment (*i.e.*, on the a mobile phone), a concerted effort had to be made to achieve the desired frame rate of 30 Hz with minimal latency and no frame drop. The largest bottleneck along the data pipeline is the time it takes to run the segmentation network. We tested several networks with increasing model size in order to determine their suitability for mobile

**Table 1** Mean and standard deviation for the per-frame run times of the segmentation networks with different sizes

Base filter	# Parameters (k)	Mean (ms)	STD (ms)
4	122	94.64	14.43
8	487	194.71	30.76
16	1944	640.97	131.38

deployment. The run time statistics are shown in Table 1. The number of base filters refers to the number of the filters in the first layer of the U-net doubled after each down-sampling step.

Regardless of the network used, we need to multi-thread multiple segmentation network runners (SEGs) concurrently in order to achieve a per-frame processing time of  $1/30 \text{ Hz} = 33.3 \text{ ms}$ . In order to prevent the application from lagging, the SEGs must finish their execution before they are fed with their next frame, *i.e.*, all the per-frame processing must be completed within  $T_{\max}$ , calculated as follows:

$$T_{\max} = \frac{\# \text{ of SEGs}}{\text{FPS}} > \mu_{\text{bs}} + 2\sigma_{\text{bs}}. \quad (2)$$

In practice, we found that requiring the mean run time be two standard deviations less than  $T_{\max}$  is sufficient to prevent any noticeable lag during the program's execution. Using the data from Table 1 in Eq. (2), we determine that the minimum required number of concurrently running SEGs is four for a base filter of four, eight for a base filter of eight, and 27 for a base filter of 16. Each SEG instance requires roughly 15 MB of RAM and roughly 10% CPU usage. Additionally, the system's latency is hard capped by the run time of the network. For these reasons, we chose to



use the smallest of the tested networks, using a base filter of four.

### Ejection fraction calculation

After each run, the SEG threads send their outputs to the EF Calculator class through asynchronous callbacks, as shown in Fig. 1. The segmentation maps and landmarks are then buffered until there is enough data to find the ED and ES frames of the cardiac cycle. This is done by simply finding the maximum and minimum areas of the buffered LV segmentations, corresponding to the ED and ES frames, respectively. A 60-frame buffer is used to capture the entirety of any heart cycle above 30 bpm. The landmarks of these two frames are then used to calculate  $L_{ED}$  and  $L_{ES}$ , *i.e.*, the respective longitude of the LV measured from apex to middle of the mitral valve. This is done by finding the largest two connected components presented in the landmark output prediction map, finding the coordinates of their centers of mass (CoMs), and calculating the Euclidean distance between them, in pixels. This unit of  $L$  measurement can be converted to centimeter by dividing it by the pixel resolution, while the unit of segmentation area  $A$  can be converted to squared centimeters by dividing it with the pixel density. Note that knowledge of the ultrasound imaging depth and pixel spacing are required to make these conversions. Single-plane LV volume can then be estimated as:

$$v_s = 0.85 \frac{A^2}{L}. \quad (3)$$

Using Eq. 3 for estimating LV volumes in both ED and ES frames, we can calculate LVEF as:

$$e = \frac{V_S^{ED} - V_S^{ES}}{V_S^{ED}}. \quad (4)$$

The single-plane volume calculation shown in Eq. 3 can be performed using data from either the AP4 or AP2 view; however, we can produce a more accurate 3D volume estimation by considering both cross-sections simultaneously. Once we have captured and buffered frames from both views, we can calculate the *biplane* volume for both ED and ES frames using an adaptation of the Simpson's disk counting method [31]. First, we rotate the ED and ES frames from both AP4 and AP2 views, such that their  $L$ 's are vertically aligned. We then scale the frames such that  $L_{AP4}^{ED} = L_{AP2}^{ED}$  and  $L_{AP4}^{ES} = L_{AP2}^{ES}$ , since although the AP4 and AP2 images may appear different in scale, we know the underlying anatomy to be constant. Once properly rotated and scaled, we can apply a variant of Eq. 1, summing over the pixel length of  $L$ :

$$\begin{aligned} v_b &= \frac{\pi}{4} \sum_{i=1}^{L_{px}} a_{(i,cm)} b_{(i,cm)} \frac{L_{cm}}{L_{px}} = \frac{\pi}{4} \sum_{i=1}^{L_{px}} \frac{a_{(i,px)}}{r} \frac{b_{(i,px)}}{r} \frac{1}{r} \\ &= \frac{\pi}{4} \sum_{i=1}^{L_{px}} \frac{a_{(i,px)} b_{(i,px)}}{r^3}, \end{aligned} \quad (5)$$

where  $a_{(i,px)}$  equals to the pixel width of each horizontal pixel line in the AP4 image,  $b_{(i,px)}$  equals to the width of the AP2 lines, and  $r$  is the pixel resolution of the image. By running this calculation for both pairs of ED and ES frames, we can refine our EF estimate from Eq. 4 by using the more accurate biplane LV volume estimation.

The geometric approximation assumptions of single-plane (monoplane) and biplane area-length techniques are fairly similar. Monoplane EF estimation can be used in cases where only one of the AP2 or AP4 views is available. Gross-gasteiger et al. [13] compared the accuracy and feasibility of six commonly used 2D methods to assess LV function. Biplane Simpson method has the strongest correlation with 3D echo in LVEF, followed by the AP4 and AP2 Simpsons monoplane methods, respectively.

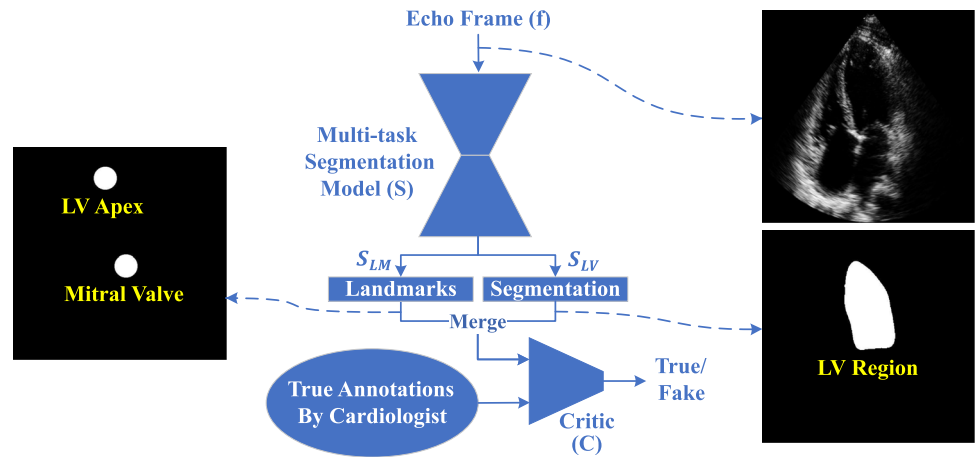
### Left ventricle segmentation and landmark detection

In this section, we discuss the details of the core intelligence of the mobile application, *i.e.*, the LV segmentation and landmark detection method. We propose a multi-task deep learning approach to simultaneously segment LV and detect the two LV landmarks. This method consists of a segmentation network ( $S$ ) and a critic network ( $C$ ), which are shown in Fig. 3. The segmentation model estimates the LV region and the two landmarks (LMs). The critic network is used in the training as an adversarial framework to improve the segmentation output.

#### Segmentation model

We implemented a network based on the U-net [29] architecture as our segmentation network. The U-net is a fully convolutional segmentation model including a down-sampling feature extraction, an up-sampling reconstruction path, and skip connections between the down-sampling and up-sampling blocks that share the same output feature size. Our U-net implementation is modified by adding two branches to its last up-sampling layer. One branch of the multi-task segmentation network predicts LV segmentation, and the other branch is used to detect the location of two LV landmarks, which are the LV apex point and middle of the mitral valve point, in both AP2 and AP4 views. We denote  $S_{LV}(f; \theta_s)$  and  $S_{LM}(f; \theta_s)$  as the functions to estimate the LV region

**Fig. 3** Architecture of the adapted deep fully convolutional network for simultaneous LV segmentation and LV landmark detection. Critic model ( $C$ ) is only used in the training phase. The trained network parameters of the multi-task segmentation network ( $S$ ) are frozen and deployed on the mobile application



and two LMs from the input frame  $f$ . The LV region and the location of the two landmark points are used in Eq. 1 for biplane EF estimation.

To train the segmentation network, we use Dice loss  $\mathcal{L}_{LV}$  to compare the LV prediction of the network with the ground truth  $p$ . A weighted binary cross-entropy  $\mathcal{L}_{LM}$  is used as the loss function for the network's landmarks detection. Detection of the centroid of the landmarks is formulated as a segmentation problem. This results in a highly unbalanced dataset, *i.e.*, there are only two points in the landmark class, compared to all other pixels of the image which belong to the background group. To rectify this unbalance distribution of classes, two solutions are applied. First, a circle with radius  $R$  is defined around each landmark point in training samples. In the test set, the centers of mass of the predicted connected components are used as the location of landmark points. Next, a class weighting approach is applied to the cross-entropy loss according to the number of samples in landmark and background classes, in order to balance against their population during the training, *i.e.*, a higher weight is given to the under-represented landmark class. In our method, a weight of  $W_c = \frac{T}{2T_c}$  is given to the class  $c$ , where  $c \in \{\text{landmark}, \text{background}\}$ ,  $T$  is the total number of pixels in a training sample, and  $T_c$  denotes the number of pixels in the class  $c$ .

### Critic model

The outputs of the multi-task segmentation network  $S$  then are fed to a critic network  $C$ . The predicted LV region and the landmark locations are element-wisely summed and re-normalized between 0 and 1, and are then passed to the critic network.

The critic network is a CNN that tries to discriminate if an annotation is done by the cardiologist (called *True*) or by the segmentation network (called *Fake*), *i.e.*,  $y = C(m; \theta_c)$ , where  $y \in \{\text{True}, \text{Fake}\}$  and  $m$  represents an annotation

map, showing LV region and landmarks. Trained by a binary cross-entropy loss ( $\mathcal{L}_C$ ), the critic network learns to discriminate the distributions of the ground truth annotations versus the outputs of the segmentation model. The critic encourages the prediction of segmentation network toward converging to the distribution of *True* masks, *i.e.*, the segmentation network produces results that are not distinguishable from the annotations done by the cardiologist. This way, a higher-order shape-wise constrain is implied on the segmentation network's output, which can be difficult to express in a standard per-pixel loss function [19]. The critic model can verify the shape integrity of the predicted LV masks and the localization accuracy of the LV landmarks.

### Adversarial training

Given the set of predictions  $\{S_{LV}(f; \theta_s), S_{LM}(f; \theta_s)\}$  and  $C(m; \theta_c)$ , the segmentation model is trained to minimize:

$$\begin{aligned} \mathcal{L}(\theta_s) = & \lambda_1 \mathcal{L}_{LV}(S_{LV}(f; \theta_s), p) + \lambda_2 \mathcal{L}_{LM}(S_{LM}(f; \theta_s), q) \\ & + \lambda_3 \mathcal{L}_C(C(m; \theta_c), \text{True}), \end{aligned} \quad (6)$$

where  $p$  and  $q$  are the respective ground truth for LV segmentation and LV landmark locations;  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are weighting parameters of respective loss terms;  $m = \text{Merge}(S_{LV}(f; \theta_s), S_{LM}(f; \theta_s))$  sums and re-normalizes  $S_{LV}(f; \theta_s)$  and  $S_{LM}(f; \theta_s)$ ; and  $\mathcal{L}_C$  encourages  $S$  to produce segmentation maps that could fool the discriminator  $C$  to recognize the maps as *True*. Throughout the learning phase, the segmentation network and the critic network are alternatively trained together in an adversarial framework. In each learning iteration, the segmentation network is trained to minimize Eq. 6, and the model parameters of the critic network,  $\theta_c$ , are kept unchanged during minimization of the loss in Eq. 6. The critic is also kept trained with  $\mathcal{L}(\theta_c) = \mathcal{L}_C(C(\text{Merge}(p, q); \theta_c), \text{True}) + \mathcal{L}_C(C(m; \theta_c), \text{False})$  to classify between the distribution of ground truth annotations

and the distribution of the predicted masks made by  $S$ . This in turn pushes the segmentation model  $S$  toward generating masks that are similar to the cardiologist's marks and hence, an implicit shape prior is enforced on the joint space of the predicted LV segmentation and landmark locations.

## Network's architecture

The multi-task segmentation model  $S$  is based on the U-net model.  $S$  has four down-sampling and four inverse up-sampling steps with concatenating skip connections. All max-pooling layers have a size of  $2 \times 2$  with a stride of one. All convolutional layers have kernel size of  $3 \times 3$  with the stride of one, followed by a batch normalization layer and *Relu* activation function. The activation function in the last layer is selected to be *sigmoid*. The base number of filters is set to four, which is doubled after each down-sampling step, resulting in a small lightweight network with about 123k trainable parameters suitable to run smoothly on a mobile device.

The critic network  $C$  is a CNN with three convolutional layers followed by two layers of fully connected neurons. The first two convolutional layers are down-sampled using average pooling. Convolutional kernels in  $C$  have the size of  $3 \times 3$ , pooling layers have the size of  $2 \times 2$ , all with the stride of one. The number of filters in the first convolutional layer is set to 16, and doubled after each down-sampling. The network is terminated with a two-layer fully connected network with 64 and one neurons, respectively, the latter of which outputs *True* or *Fake* classifications. All intermediate layers in  $C$  are followed by batch normalization, *Leaky Relu*, and *dropout* with the ratio of 0.25. The activation function in the last layer is *sigmoid*.

## Experiments

### Dataset and implementation details

The proposed application is evaluated over a dataset of 854 echo studies, collected from the Picture Archiving and Communication System at Vancouver General Hospital, with ethics approval of the Clinical Medical Research Ethics Board, in consultation with the Information Privacy Office. The data includes pairs of AP2 and AP4 echo views from 427 patients. For all echo studies, the segmentation and location of landmarks of the LV are annotated by an expert cardiologist at the ED, ES, and a random middle frame between ED and ES phases. The cardiologist's annotations are regarded as the ground truth. For each patient, the ground truth for LVEF is provided using cardiologist's annotations.

Echo cines are loaded onto the mobile application to obtain the AP2 and AP4 LV segmentations, landmarks, and the

biplane ejection fraction. The dataset is randomly split into five non-overlapping groups based on the patients. To obtain results on the entire dataset, the experiment is done five times, where in each run, one group is set aside unseen as the test and the training is done with the other four groups. Therefore, the training to test ratio is 80% to 20%, respectively. Also in each run, 10% of the data in the training is used as validation to search for the optimal hyper-parameters.

The network is implemented in Keras with Tensorflow trained on a PC system. The weights of the network are then frozen and transferred to the mobile application. The mobile device used is a Samsung S8+, with 6 GB of RAM, running a Snapdragon Octa-core processor ( $4 \times 2.45$  GHz and  $4 \times 1.9$  GHz CPUs). *Adam* optimizer is used to train the network.  $\lambda_1$  to  $\lambda_3$  in Eq. 6 are set to 1, 1 and 0.1, respectively. The circles around landmark points in training have the radius of  $R = 7$  pixels in echo images of size  $128 \times 128$ . Two separate networks with similar architecture are trained for AP2 and AP4 views. Lambdas and  $R$  are the hyper-parameters of the model optimized using the validation set. The network's training is done on ED, ES, and the random middle frame (RF) of echo cines, where a ground truth by the cardiologist is available.

### Quantitative evaluation

Here, we evaluate the results in each of the steps of the proposed pipeline. The steps include AP2 LV segmentation, AP2 LV landmark detection, AP4 LV segmentation, AP4 LV Landmark detection, and finally using the segmentation masks to obtain a biplane LVEF estimation.

U-net is considered as a standard state-of-the-art model for medical image segmentation tasks. Works of [35] and [33] propose variations of U-net for echocardiogram segmentation. We compare the performance of the U-net with and without using the proposed training method. The applied U-net has four base filters with about 122 k training parameters. The per-frame run time of the mobile framework with respect to the size of the network and the justification of the choice of four base filters are discussed in Sect. 2.1. In Table 2, we also compare our method to another widely used

**Table 2** Evaluation of LV segmentation performance in AP2 and AP4 echo views

Method	Mean dice score (%)					
	AP2			AP4		
	ES	RF	ED	ES	RF	ED
DeconvNet [25]	86.4	89.7	90.0	85.6	89.5	91.0
U-net [33]	87.9	90.1	91.2	87.6	91.1	92.5
Proposed method	<b>90.3</b>	<b>92.4</b>	<b>92.5</b>	<b>90.3</b>	<b>92.7</b>	<b>93.6</b>

The comparison is done at end-systole (ES), end-diastole (ED), and a random middle frame (RF) between ED and ES. Best results are in bold

**Table 3** Evaluation of LV landmark detection in AP2 and AP4 views

Method	Mean distance (px)			
	AP2		AP4	
	LV apex	Mitral valve	LV apex	Mitral valve
U-net	4.4	4.5	4.8	3.4
Proposed method	<b>3.4</b>	<b>2.0</b>	<b>3.7</b>	<b>2.3</b>

Best results are shown in bold

Detected landmark points by our method are compared to the cardiologist's annotations, using mean Euclidean distance error in pixel space for echo images of size  $128 \times 128$

**Table 4** Evaluation of LVEF estimation in AP2 single view, AP4 single view, and AP2–AP4 biplane views

Method	LVEF estimation error (%)					
	AP2 single		AP4 single		Biplane	
	Median	Mean	Median	Mean	Median	Mean
U-net	8.6	10.1	9.2	10.1	7.1	8.9
Proposed method	<b>7.2</b>	<b>9.4</b>	<b>7.5</b>	<b>9.6</b>	<b>6.2</b>	<b>7.9</b>

Best results are shown in bold

Median and mean of absolute differences between estimated percentage for LVEF by the proposed method compared to the cardiologist's assessment are presented

segmentation model, namely the DeconvNet [25], with four base filters and 149 k training parameters. The comparison is done on ED, ES, and a random frame from systole or diastole phases of the heart. The random middle frame gives an estimation of the segmentation performance over the whole cine clip. The cardiologist's segmentation masks are referenced as ground truth. Results of LV segmentation in AP2 and AP4 echo views are presented in Table 2. Our proposed multi-task network also automatically detects the location of the landmarks of the LV. The Euclidean distance between detected landmark points by our method compared to the cardiologist annotations is shown in Table 3. The distance is presented in pixel (px) space for echo images of size  $128 \times 128$ . The LV segmentation and landmark points are used in the pipeline to automatically calculate biplane LV ejection fraction. LVEF estimation errors are presented in Table 4. Evaluated over 427 patients, automatically estimated biplane LVEF percentage by our method has a median absolute error of 6.2%, and a mean absolute error of 7.8%, compared to the cardiologist's opinion. The importance of the results could be further remarked noticing the reported [8] inter-observer variability of 17.8%, and intra-observer variability of 13.4%, existing for echo biplane LVEF estimation in clinician's examinations.

## Conclusion and discussion

In this paper, we presented a pipeline using mobile POCUS for biplane LVEF estimation. We proposed a lightweight multi-task segmentation framework, based on fully convolutional networks and adversarial training, for simultaneous LV segmentation and LV landmark detection. The software evaluated on pairs of AP2 and AP4 echocardiograms from 427 patients could reach a high correlation compared to the car-

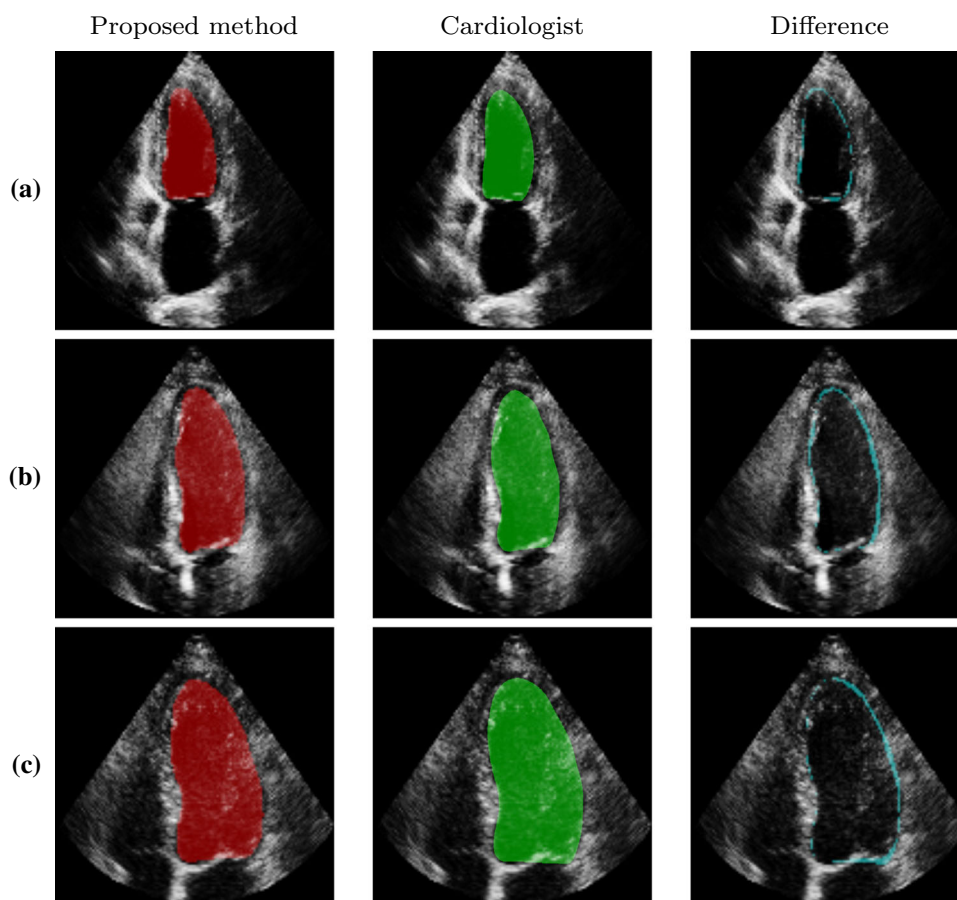
diologist's assessments. Experiments presented show a mean dice score of 92% for LV segmentation, superior to existing comparable methods, and a mean Euclidean distance of 2.85 pixels for LV landmark detection. The predicted annotation set is used in the proposed pipeline to calculate biplane LVEF. The automatically estimated biplane LVEF by the proposed method has a mean absolute error of less than 8% compared to the cardiologist's estimations.

Prognosis and therapeutic cardiac decisions are often based on LVEF measurement. LVEF estimation is one of the key cardiac measurements derived from echo studies. Manual quantification of the LVEF needs cardiologist or sonographer tracing of LV is time-consuming and labor intensive with relatively high inter-observer and intra-observer variability. In recent years, the ultrasound imaging has become widely accessible due to advances in development of cheap portable POCUS devices. This paper is a step toward automatic LVEF estimation on readily available android mobile devices, compatible with POCUS. The mobile cardiac POCUS has advantages of portability, low cost, accessibility, and immediacy of results, vital in applications such as emergency scenarios and anesthesia management [14].

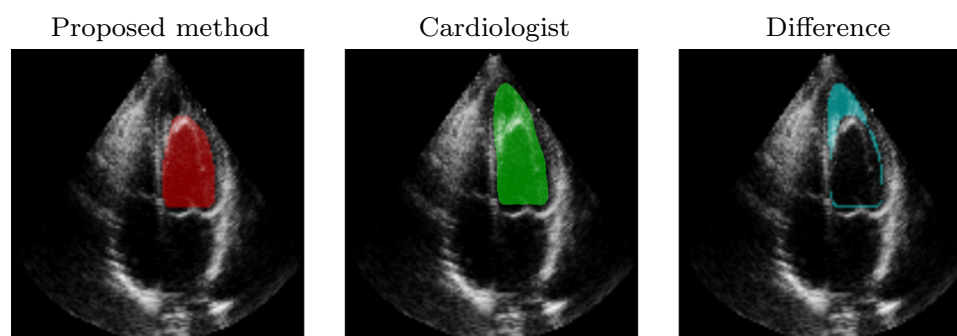
Sample visual results of the LV segmentation by the proposed method compared to the manual annotation by the cardiologist are presented in Fig. 4, where (a) is a AP2 and (b) and (c) are AP4 views. Also, Fig. 5 presents a sample failed case. Low quality of the captured echo, foreshortening, and fuzzy borders could be mentioned as the reasons for the method failure. The LVEF estimation is directly dependent on the accuracy of the LV segmentation. The ratio of maximum to minimum LV volume in a heart cycle is used to estimate LVEF. The LV segmentation error might result in missing a part of LV, or on the contrary, labeling the sur-



**Fig. 4** Sample LV segmentation by our method compared to manual annotation by the cardiologist



**Fig. 5** Sample case of segmentation failure



rounding muscle tissue area of the ventricle as a part of the LV. In each case, the error might directly affect the minimum or maximum measured LV volume, and in turn causes the LVEF estimation error. This source of error leads to an observed bias toward overestimation of the LVEF in our current result set. Investigation of machine learning solutions to guide the operator through acquisition of accurate high-quality echo views could be an improvement which we consider as a future work. Future work also includes the extension of the proposed multi-task segmentation model to other echo views and multiple heart chambers. The model could be extended to derive various clinically demanded cardiac metrics, such as LV wall motion abnormalities.

**Acknowledgements** This work was supported in part by the Natural Sciences and Engineering Research and Council of Canada (NSERC) and in part by the Canadian Institutes of Health Research (CIHR).

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J, Hawley D, Fleming S, Gin K, Swift J (2017) Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. *IEEE Trans Med Imaging* 36(6):1221–1230
- Aveni M, Kheradvar A, Jafarkhani H (2016) A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med Image Anal* 30:108–119
- Carneiro G, Nascimento JC (2013) Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Trans Pattern Anal Mach Intell* 99(1):2592–2607
- Carneiro G, Nascimento JC, Freitas A (2012) The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. *IEEE Trans Image Process* 21(3):968–982
- Chen H, Dou Q, Ni D, Cheng JZ, Qin J, Li S, Heng PA (2015) Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 507–514
- Chen H, Ni D, Qin J, Li S, Yang X, Wang T, Heng PA (2015) Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Health Inform* 19(5):1627–1636
- Chen H, Zheng Y, Park JH, Heng PA, Zhou SK (2016) Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 487–495
- Chuang ML, Hibberd MG, Salton CJ, Beaudin RA, Riley MF, Parker RA, Douglas PS, Manning WJ (2000) Importance of imaging method over imaging modality in noninvasive determination of left ventricular volumes and ejection fraction: assessment by two- and three-dimensional echocardiography and magnetic resonance imaging. *J Am Coll Cardiol* 35(2):477–484
- Fagley RE, Haney MF, Beraud AS, Comfere T, Kohl BA, Merkel MJ, Pustavoitau A, Von Homeyer P, Wagner CE, Wall MH (2015) Critical care basic ultrasound learning goals for American anesthesiology critical care trainees: recommendations from an expert group. *Anesthesia Analgesia* 120(5):1041–1053
- Ghesu FC, Krubasik E, Georgescu B, Singh V, Zheng Y, Hornegger J, Comaniciu D (2016) Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE Trans Med Imaging* 35(5):1217–1228
- Girdhar R, Fouhey DF, Rodriguez M, Gupta A (2016) Learning a predictable and generative vector representation for objects. In: European conference on computer vision. Springer, pp 484–499
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680
- Grossgasteiger M, Hien MD, Graser B, Rauch H, Gondan M, Motsch J, Rosendal C (2013) Assessment of left ventricular size and function during cardiac surgery. An intraoperative evaluation of six two-dimensional echocardiographic methods with real time three-dimensional echocardiography as a reference. *Echocardiography* 30(6):672–681
- Johri AM, Durbin J, Newbigging J, Tanzola R, Chow R, De S, Tam J (2018) Cardiac point-of-care ultrasound: state-of-the-art in medical school education. *J Am Soc Echocardiogr* 31(7):749–760
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
- Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T (2015) Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 16(3):233–271
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
- Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
- Luc P, Couprie C, Chintala S, Verbeek J (2016) Semantic segmentation using adversarial networks. *arXiv preprint. arXiv:1611.08408*
- Mahmood F, Matyal R, Skubas N, Montealegre-Gallegos M, Swaminathan M, Denault A, Snieciński R, Mitchell JD, Taylor M, Haskins S (2016) Perioperative ultrasound training in anesthesiology: a call to action. *Anesthesia Analgesia* 122(6):1794–1804
- McCormick TJ, Miller EC, Chen R, Naik VN (2018) Acquiring and maintaining point-of-care ultrasound (POCUS) competence for anesthesiologists. *Can J Anesth/Journal canadien d'anesthésie* 65(4):427–436
- Moradi M, Guo Y, Gur Y, Negahdar M, Syeda-Mahmood T (2016) A cross-modality neural network transform for semi-automatic medical image annotation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 300–307
- Nascimento JC, Carneiro G (2016) Multi-atlas segmentation using manifold learning with deep belief networks. In: Biomedical imaging (ISBI), 2016 IEEE 13th international symposium on. IEEE, pp 867–871
- Ngo TA, Lu Z, Carneiro G (2017) Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Med Image Anal* 35:159–171
- Noh H, Hong S, Han B (2015) Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 1520–1528
- Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, Cook SA, de Marvao A, Dawes T, O'Regan DP (2018) Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans Med Imaging* 37(2):384–395
- Poudel RP, Lamata P, Montana G (2016) Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: Reconstruction, segmentation, and analysis of medical images. Springer, pp 83–94
- Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint. arXiv:1511.06434*
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
- Rupprecht C, Huaroc E, Baust M, Navab N (2016) Deep active contours. *arXiv preprint. arXiv:1607.05074*
- Schiller NB, Shah PM, Crawford M, DeMaria A, Devereux R, Feigenbaum H, Gutgesell H, Reichek N, Sahn D, Schnittger I

- (1989) Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. *J Am Soc Echocardiogr* 2(5):358–367
32. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
33. Smistad E, ostvik A, Haugen BO, Lovstakken L (2017) 2D left ventricle segmentation using deep learning. In: 2017 IEEE international ultrasonics symposium (IUS), pp 1–4
34. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
35. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Fan E, Aras MA, Jordan C, Fleischmann KE (2017) A computer vision pipeline for automated determination of cardiac structure and function and detection of disease by two-dimensional echocardiography. arXiv preprint. [arXiv:1706.07342](https://arxiv.org/abs/1706.07342)
36. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C (2018) Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation* 138(16):1623–1635
37. Zreik M, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I (2016) Automatic segmentation of the left ventricle in cardiac ct angiography using convolutional neural networks. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). IEEE, pp 40–43

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.