ELSEVIER

● *Original Contribution*

REAL-TIME ECHOCARDIOGRAPHY GUIDANCE FOR OPTIMIZED APICAL STANDARD VIEWS

DAVID PASDELOUP,* SINDRE H. OLAISEN,* ANDREAS ØSTVIK,*,† SIGBJORN SABO,* HÅKON N. PETTERSEN,*
ESPEN HOLTE,*,‡ BJØRNAR GRENNE,*,‡ STIAN B. STØLEN,‡ ERIK SMISTAD,*,† SVEIN ARNE AASE,§
HÅVARD DALEN,*,‡ and LASSE LØVSTAKKEN*

* Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and
Technology, Trondheim, Norway; † SINTEF Medical Technology, Trondheim, Norway; ‡ Clinic of Cardiology, St. Olav's Hospital,
Trondheim, Norway; and § GE Vingmed Ultrasound AS, Horten, Norway

**Abstract—Measurements of cardiac function such as left ventricular ejection fraction and myocardial strain are
typically based on 2-D ultrasound imaging. The reliability of these measurements depends on the correct pose of
the transducer such that the 2-D imaging plane properly aligns with the heart for standard measurement views
and is thus dependent on the operator's skills. We propose a deep learning tool that suggests transducer move-
ments to help users navigate toward the required standard views while scanning. The tool can simplify echocardi-
ography for less experienced users and improve image standardization for more experienced users. Training data
were generated by slicing 3-D ultrasound volumes, which permits simulation of the movements of a 2-D trans-
ducer. Neural networks were further trained to calculate the transducer position in a regression fashion. The
method was validated and tested on 2-D images from several data sets representative of a prospective clinical set-
ting. The method proposed the adequate transducer movement 75% of the time when averaging over all degrees
of freedom and 95% of the time when considering transducer rotation solely. Real-time application examples
illustrate the direct relation between the transducer movements, the ultrasound image and the provided feed-
back. (E-mail: david.pasdeloup@ntnu.no) © 2022 The Author(s). Published by Elsevier Inc. on behalf of World
Federation for Ultrasound in Medicine & Biology. This is an open access article under the CC BY license
(http://creativecommons.org/licenses/by/4.0/).**

Key Words: Echocardiography, Navigation, Deep learning, Non-expert, Portable ultrasound.

INTRODUCTION

Echocardiography is the cornerstone image modality for
measuring and evaluating cardiac function, today based
primarily on 2-D image acquisition following a defined
protocol of standard views and measurements used for
subsequent diagnostics (Lang et al. 2015). The acquisi-
tion of these standard views is challenging for inexperi-
enced users, and is prone to substantial variation even
among experienced users. This limits the availability of
echocardiography for the patients and also decreases
measurement reproducibility (Morbach et al. 2018).

The acquisition of the three apical standard views—
apical four-chamber (A4C), two-chamber (A2C) and
long-axis (ALAX)—can be broken down into three
steps, as illustrated in Figure 1. First, the users need to
find the correct intercostal acoustic window where the
anatomical apex of the heart is closest to the transducer.
In the second step, users need to rotate and tilt the trans-
ducer around the left ventricular centerline (between the
apex and the center of the mitral valve) to produce the
three standard apical imaging planes with minimal fore-
shortening. The last step consists of optimizing the
images for both anatomical correctness and image qual-
ity to establish a more detailed cardiac examination. We
here define *anatomical correctness* (AC) to depend only
on the location of the 2-D imaging plane. AC is not
related to the quality of the ultrasound (US) signal.

The starting point of our work is that the standard
apical views acquired by both inexperienced and experi-
enced users have an operator variability in AC, influenc-
ing the measurements that follow. Two examples of

Address correspondence to: David Pasdeloup, Department of
Circulation and Medical Imaging, Faculty of Medicine and Health Sci-
ences, Norwegian University of Science and Technology (NTNU),
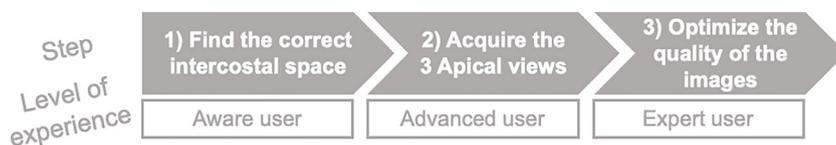7491 Trondheim, Norway. E-mail: david.pasdeloup@ntnu.no

Fig. 1. Procedure steps and experience required for acquisition of the apical standard views.

standard views of different quality are provided in Figure 2. These are recordings from the echo lab and thus represent clinical practice, where anatomical landmarks reveal differences in AC.

Three-dimensional US imaging can be part of a future solution, as volume measurements can be more directly extracted, and standard planes automatically extracted (Chykeyuk et al. 2014; Domingos et al. 2014). However, 3-D and multi-plane imaging still suffers from limitations such as a lower frame rate and suboptimal image quality compared with 2-D imaging, and optimization of each view is often needed to obtain good image quality. Further, 3-D US is not readily available for hand-held systems. We thus believe that 2-D echocardiography will remain important for the foreseeable future.

Some research and development have previously been done to facilitate the echocardiography procedure described in Figure 1. In step 1, finding a suitable intercostal space can be seen as minimizing apical foreshortening, which was previously approached using image segmentation

(a) A2C view – The suboptimal recording (left) has the right ventricle partly visible

(b) ALAX view – The aortic valve is not clearly visible in the sub-optimal recording (left)
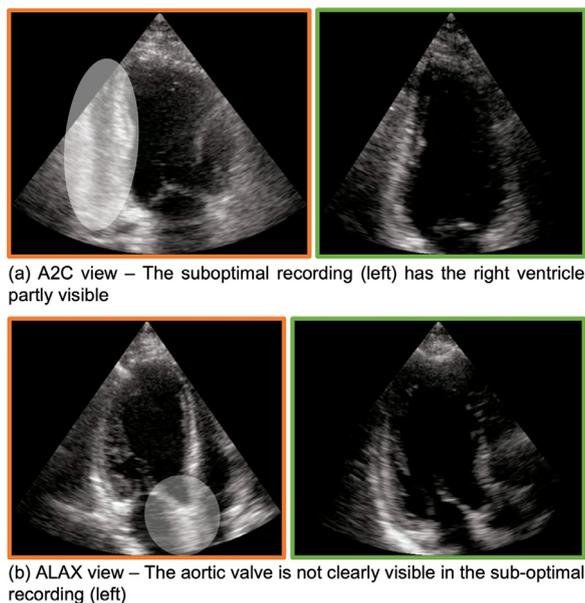
Fig. 2. Example of the inter-operator variability in the imaging plane of clinical recordings. For a given target view on the same patient, the recording on the left acquired is a sub-optimal standard view, whereas the recording on the right (acquired by another operator) complies with the guidelines recommendations. Full cine loops are available in Video S3 (online only).

A2C = apical two-chamber; ALAX = apical long-axis.

(Smistad et al. 2020). Finding the imaging plane in step 2 has been addressed before by both classic and deep learning (DL)−based approaches (Snare et al. 2012; Østvik et al. 2019). However, these methods were only able to grade the AC, and did not give user feedback on the required transducer movements to acquire the standard views. In Østvik et al. (2019), we proposed an exploratory extension of the view classification method with 3-D data for training to enable feedback to the user on rotational movement of the transducer. However, the solution was preliminary, limited to the rotational movement, and struggled to generalize sufficiently for real-time 2-D imaging. To our best knowledge, Toporek et al. (2019) is the only technical description of a DL-based echocardiography guidance system with training data generated with a 2-D transducer and an external positional sensor. Their approach is, however, limited by a low number of patients used in the training data set, an accuracy not quantified for the A2C and ALAX views and a lack of prospective testing. Navigation using an inertial measurement unit could be used to provide feedback and thus guide the user. A solution would, however, still be needed to recognize the target view and the direction to it. Li et al. (2018) used a neural network to calculate the required geometric transformation to obtain a standard plane within a 3-D fetal US volume, and Droste et al. (2020) proposed a neural network that predicts an angle-to-target value given a 2-D fetal image. However, echocardiography introduces additional challenges such as the ribs limiting the acoustic window or anatomical variability. It is therefore not straightforward to apply methods from fetal US to heart US.

Machine learning echocardiography assistance is also a topic of interest for several companies that have developed proprietary solutions. Narang et al. (2021) (Caption Health Inc., Brisbane, CA, USA) reported that their algorithm can help nurses to acquire recordings of diagnostic quality, but did not report results on the correctness of each individual standard view. The same algorithm was tested on medical students by Schneider et al. (2021), who reported the algorithm was helpful in acquiring the A4C and A2C views. However, these two studies lacked a control group of users who scan without the algorithm assistance. The technical challenge of echocardiography guidance is therefore not solved in its entirety.

In this work, we propose a real-time guidance method for 2-D echocardiography that estimates the transducer rotation and tilt in relation to the cardiac anatomy based

only on the 2-D image input. The method provides feedback on how to adjust the transducer position to reach the target standard apical views. This tool can be valuable (i) to improve image view consistency and help limit subsequent measurement variability for experienced users in the echo lab, and (ii) to unleash the potential of echocardiography for non-expert users using hand-held devices, thus unlocking a potential tool for detecting heart disease prior to hospital referral. Finally, the tool can (iii) be used to train novice users in echocardiography.

Main contributions

We have developed and evaluated a software-based echocardiography guidance system based on positional regression directly from 2-D images. The main contributions of the study described here with respect to previous work from Østvik et al. (2019) are as follows:

- An improved protocol for defining echocardiographic anatomical views, including non-standard apical views used for image guiding
- A semi-automated approach for extracting 2-D training data from 3-D US recordings with an accurate reference position
- An more robust neural network for estimating the transducer rotational position
- New neural networks for estimating the transducer tilt position
- Extensive validation and testing on several representative 2-D data sets from the clinic
- A real-time prototype application that shows evidence of the method validity

METHODS

Our main approach was to use deep convolutional neural networks to predict the transducer position relative to the heart in the form of a regression task based on

the 2-D US image input, in which training data were generated from 3-D US volumes, as detailed in the following sections.

Degrees of freedom and problem formulation

Two-dimensional echocardiography involves multiple degrees of freedom (DOFs), with two translation DOFs in the skin plane and three rotation DOFs (transducer rotation, in-plane tilt, out-of-plane tilt). Additional challenges can be attributed to the motion of the heart, patient breathing and the restricted acoustic window between the ribs, limiting the possibility of obtaining good image quality and aligned cardiac views.

To reduce the complexity of the problem, we initially assumed that the heart apex is located at a shallow depth and regarded it as an anchor, meaning that the correct intercostal space is used and that no significant translations at the skin surface are required. Considering that the out-of-plane transducer rotations are difficult for users who lack spatial representation of the heart, we address in this work the rotation and out-of-plane tilt DOFs (referred to further as rotation and tilt DOFs) and leave the in-plane tilt DOFs to separate work.

We considered rotation and tilt separately and trained four individual models, one for the transducer rotation and three for the transducer tilt in the different apical views. The approach of using multiple models mimics the work flow of sonographers who iteratively adjust the transducer tilt and rotation. It is also convenient to isolate method failures and improves explainability.

A particular challenge for position regression is the intrinsic variation of the heart's shape. This is exemplified in Figure 3, which illustrates the patient variability in the amount of rotation from the A4C to the A2C and ALAX views. Because of this variation, we did not design a neural network that outputs an angle-to-target
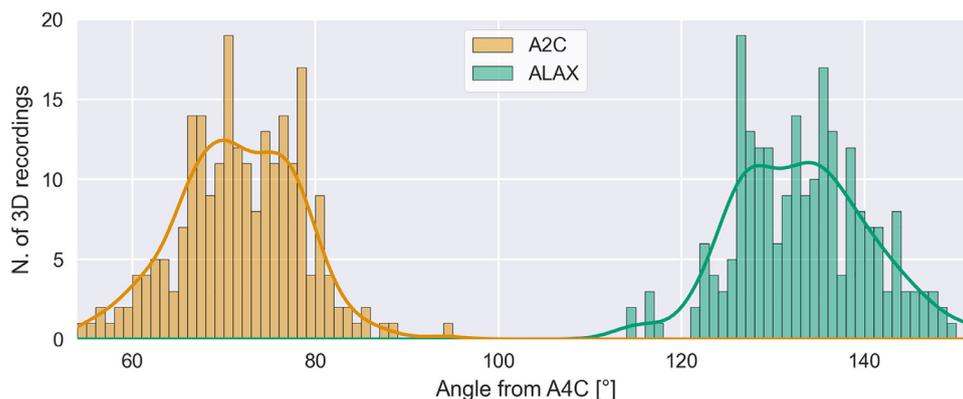

Fig. 3. The angle between A4C−A2C and A4C−ALAX standard views obtained from the manually annotated 3-D data. A2C = apical two-chamber; A4C = apical four-chamber; ALAX = apical long-axis.

value as in Droste et al. (2020), but rather a position relative to the characteristic heart structures defined below.

Training data generation

The 2-D training images were generated from 3-D US recordings according to the procedure illustrated in Figure 4. Starting from a 3-D US recording, we automatically extracted the apex and base landmarks and the maximum left ventricle (LV) radius (steps 1 and 2). The landmark and radius estimates were obtained from the segmentation mask of our 3-D UNet (Smistad et al. 2021), and were quality controlled and corrected by a clinical expert if needed.

Once the LV axis (going through apex and base landmarks) was defined, 360 slices were automatically generated (step 3) at end-diastole (ED) for each patient to simulate a transducer rotation movement around the heart's long axis. A clinical expert annotated 6 of the 360 slices as characteristic cross-sections (CCSs) (step 4). Three were the apical standard views (Fig. 5b), while three were non-standard (Fig. 5c), inserting additional clinical knowledge in the rotation position regression. The characteristic features of the additional views are termed (i) the A42C view (between A4C and A2C), which reveals the presence of a thin right ventricle (RV) but no tricuspid valve is visible, or a thickened inferoseptal wall; (ii) the Annulus Start (ANS) view; and (iii) annulus end (ANE) view. These three are recognizable by the left ventricular outflow tract (LVOT) being slightly visible, while the aorta valve leaflets are not visible. ANS further typically has a thick anterior wall, whereas in the ANE view the RV is visible.

For the tilt DOF, the corresponding CCSs were extracted automatically at the rotational position of the standard views (either A4C, A2C or ALAX). With knowledge of the LV axis length l_{LV} and LV radius r_{LV}, as illustrated in Figure 6a, the tilt span S_w can be expressed as

$$S_w = 2 \cdot \tan^{-1} \frac{\alpha \cdot r_{LV}}{l_{LV}}. \tag{1}$$

α is a coefficient chosen larger than 1 so that the slices at the outer edges of the span include neither the mitral valve (MV) nor the left atrium (LA). The tilt slices were evenly distributed over the tilt span, which should correspond to anatomical features derived from any LV size rather than an absolute tilt angle, thus accounting for patient variability.

Once annotation at ED is done, the training slices are generated from several frames assuming the LV axis and CCSs annotated at ED are stable throughout a single cardiac cycle.
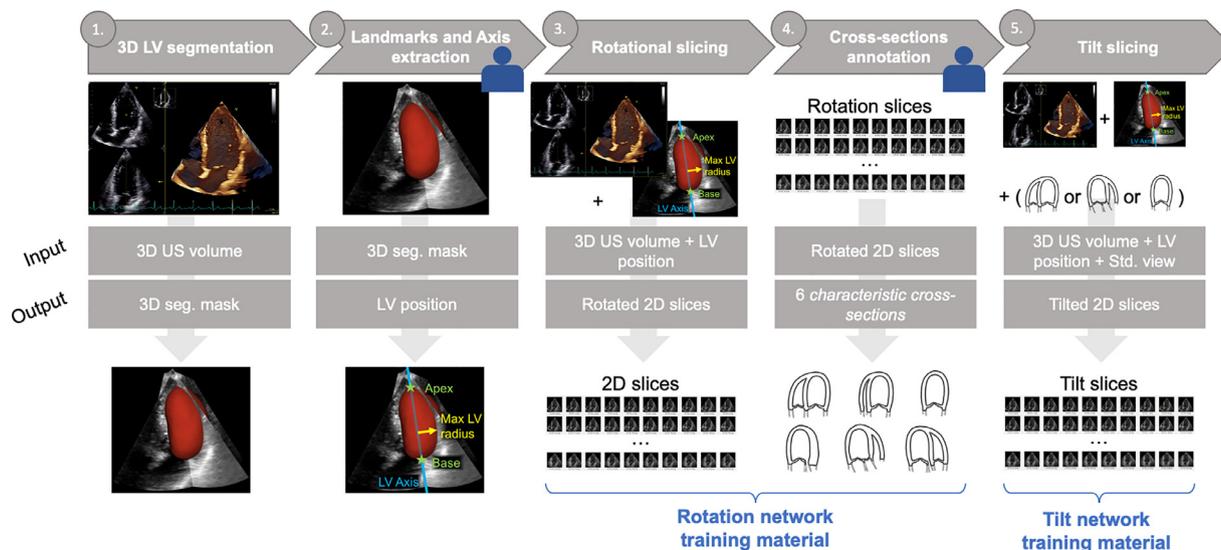

Fig. 4. Training data generation pipeline. Step 1: Takes a 3-D US volume at ED as input and outputs a 3-D mask of the LV. Step 2: Generates, from the mask, the apex and base landmarks, which together form the LV geometric rotational axis. The maximum LV radius is also extracted. The axis and the radius were manually corrected by a clinical expert. Step 3: Automatically generates 360 slices of the volume around the defined rotational axis. Step 4: A clinical expert annotated six slices as CCSs. Step 5: Automatically generates the slices to train the tilt models based on the ED rotational position (either A4C, A2C or ALAX) and the LV axis and radius. The annotations (LV axis and CCSs) were made at the ED frame, whereas the training slices were generated from multiple frames throughout the cardiac cycle. A2C = apical two-chamber; A4C = apical four-chamber; ALAX = apical long-axis; CCS = characteristic cross-section; ED = end-diastole LV = left ventricle; US = ultrasound.
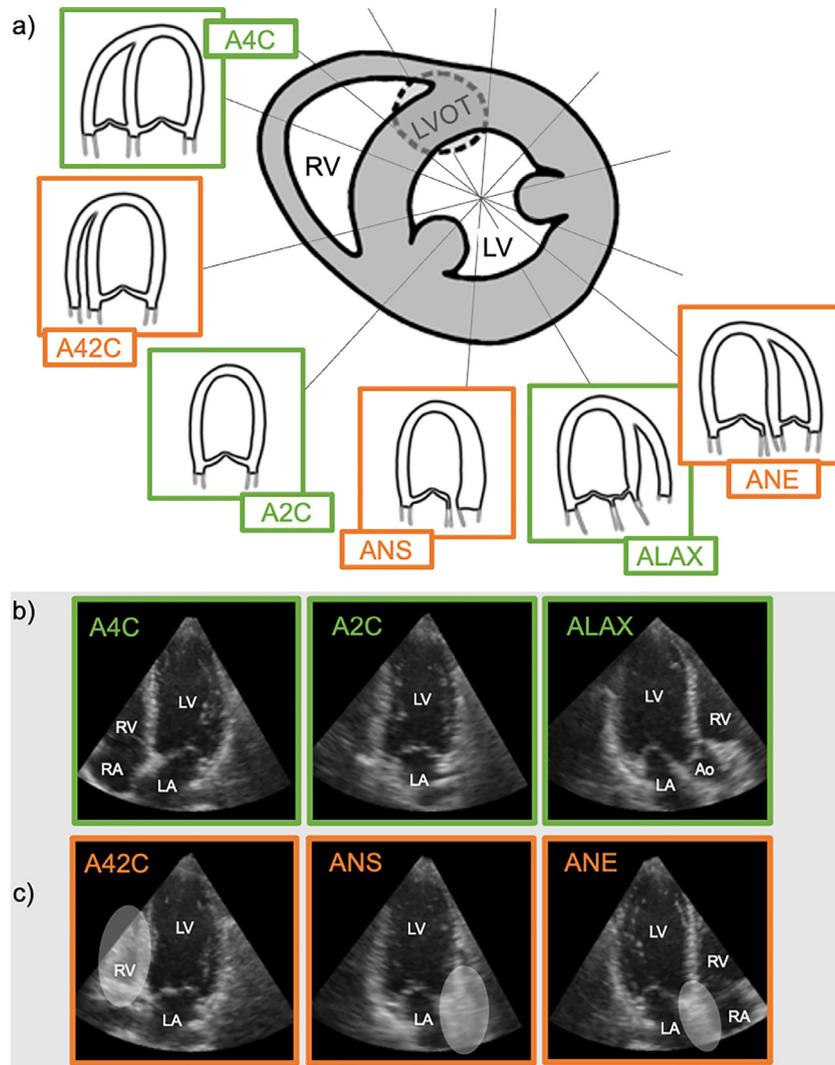
Fig. 5. (a) Positions of the six rotational CSSs in the parasternal short-axis plane. (b) The three CSSs corresponding to the apical standard views and (c) those for non-standard views. Characteristic features of the A42C view is the presence of a thin RV without visible tricuspid valve or a thick inferior septal wall. ANS and ANE are recognized by LVOT being slightly visible, but the aorta valve leaflets not visible. ANS typically has a thick anterior wall, whereas ANE has the RV visible. A2C = apical two-chamber; A4C = apical four-chamber; A42C = view between A4C and A2C views; ALAX = apical long-axis; ANE = annulus end; ANS = annulus start; CCS = characteristic cross-section; LA = left atrium; LV = left ventricle; LVOT = left ventricular outflow tract; RA = right atrium; RV = right ventricle; Ao = aorta.

Image position regression with deep learning

Considering the continuity along the slices for each DOF, the position can be posed as a regression problem and thus addressed with DL-based architectures. Continuity was included in the training procedure using the Earth mover's distance as the loss function (Hou et al. 2016), which measures the distance between the true and predicted positions and penalizes predictions with larger distance error. To account for patient variability, the predicted position $\hat{p}$ is given relative to the labeled CCSs and, thus, to the characteristic heart structures in Figures 5 and 6.

Separate rotation and tilt networks were trained. The rotation network has 12 output categories, corresponding to the 6 rotational CCSs and their flipped counterparts. As the different CCSs are not evenly spaced along the rotational DOF as illustrated in Figure 3, the number of rotational slices is sampled to balance the categories.

The tilt network has 11 output categories sampled equidistantly by the automated tilt annotation procedure described in Figure 4. As there are more slices than labeled CCSs for both rotational and tilt DOFs, the slices inbetween are assigned non-binary labels obtained through linear interpolation of the two
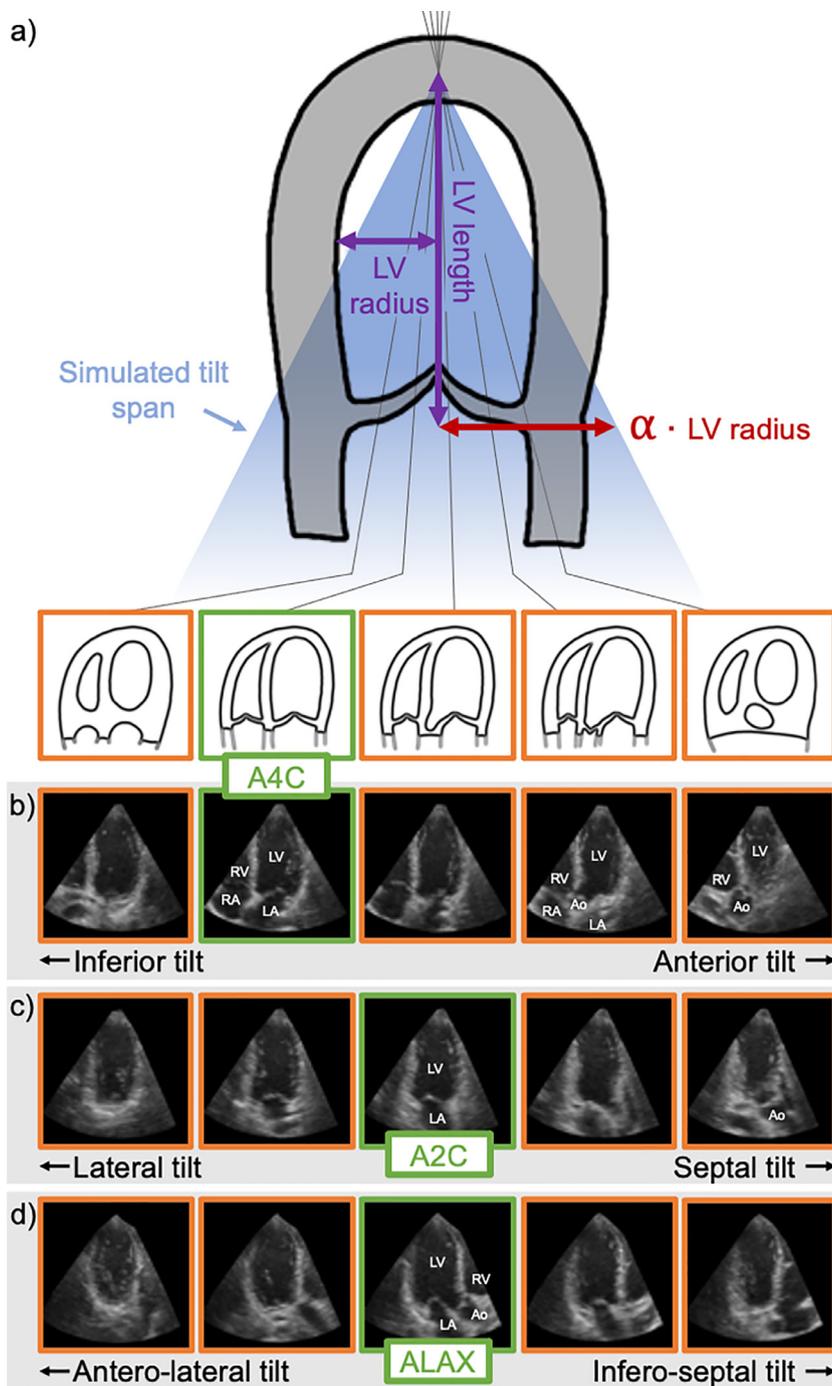
Fig. 6. (a) Description of cross-sections for the A4C tilting. (b) A4C tilt samples. (c) A2C tilt samples. (d) ALAX tilt samples. A2C = apical two-chamber; A4C = apical four-chamber; ALAX = apical long-axis; LA = left atrium; LV = left ventricle; RA = right atrium; RV = right ventricle; Ao = aorta.

closest CCSs so that the models learn the continuous nature of the task.

The predicted relative position $\hat{p}$ for any input frame X is obtained by the dot product of the network output vector C and class index vector as

$$\hat{p}(X) = \sum_{n=0}^{N-1} C^X(n) \cdot n. \tag{2}$$

We can then define the categorical distance d of any frame X to the target CCS as

$$d(X, CCS) = \hat{p}(X) - p(CCS), \qquad (3)$$

where $p(CCS)$ is the index of the target cross-sectional view.

We used our CVCNet topology (Østvik et al. 2019) which is optimized for US image classification and real-time performance. Other model architectures are benchmarked in the supplementary material.

In the real-time application, the target CCS corresponds to the three standard views A4C, A2C and ALAX. The sign of d then indicates the direction of the required transducer movement along the corresponding DOF, and low values of d indicate a correct view.

Data augmentation and pre-processing

Rotation, translation, scaling and gamma intensity transformation augmentations form our baseline set of augmentations used in all experiments. Additionally, we used several US specific augmentations, termed *US augmentations*:

- Gaussian shadows to mimic signal dropouts or acoustic wave propagation artifacts
- Depth attenuation to add varying depth-dependent dampening of the acoustic waves
- Haze artifacts to mimic acoustic haze typically originating from reverberations
- Non-linear color maps to make the models robust to different color transformations
- Contrast scaling and brightness transform to increase robustness to gain and dynamic range
- Field-of-view (FOV) masking to make the models disregard the FOV shape and avoid unintended effects related to the sector width and depth settings

Specifically for our image guiding task, we finally introduced and evaluated two pre-processing techniques related to the use of 3-D volumes as the training data source:

- Spatial reference noise: In the data generation process, we established a reference LV axis position (Fig. 4, step 2). To increase the robustness of the networks to minor errors in transducer tilt and translation, we introduced stochastic deviations to the actual LV axis and rotational position before generating the slices. The amount of deviation remains small compared with the LV dimensions so that the annotations can still be considered valid.
- Gaussian blurring: This pre-processing step aims to improve generalization from 3-D slices to 2-D US images at inference. Blurring is implemented as a Gaussian filter with standard deviation two orders of magnitude lower than the image size. The goal is to

make images more similar by smoothing details such as speckle while preserving the heart structures. Blurring is applied upstream of the convolutional layers for both training and inference.

Data sets

We used data sets of 2-D and 3-D recordings to train, validate and test our models. Regional ethics board approval and written consent were obtained for all studies and patients involved.

Training and internal validation. Three-dimensional recordings were used for training the models and were also used as part of the validation procedure. The FOV of 3-D recordings is often reduced to achieve higher frame rates in clinical practice. Therefore, slices from such volumes are not necessarily representative of what is acquired with 2-D echocardiography. Consequently, we included only 3-D volumes that cover the whole LV and its surrounding structures (RV and LA). This ensures that we have sufficient context to generate slices with structures similar to those visible in 2-D echocardiography which are the data used during inference. The 3-D recordings acquired using ECG gating with visible stitching artifacts were excluded.

We included 3-D US recordings from three different data sets, in total N = 162 patients:

- CETUS: An open data set of n = 45 patients (Bernard et al. 2014), both pathological and healthy. The recordings were acquired with scanners from three different vendors.
- NTNU 3-D A: Recordings from n = 36 patients acquired at St. Olav's Hospital, Trondheim, Norway, using a GE E95 scanner and 4Vc transducer (GE Vingmed Ultrasound, Horten, Norway). The population is representative of the daily routine at the echo lab, with both healthy and diseased patients.
- NTNU 3-D B: n = 81 recordings acquired with a GE E95 scanner at institutions spread over six different countries.

The 3-D recordings from 20% of the patients were set aside for internal validation and are not used for training. Slices from 3-D US volumes have generally lower image quality, especially in the near field. To ensure that the trained models do not overfit the training data, we also validated the method on an external data set made of 2-D US images representative of clinical use as described in the following.

External 2-D validation data set. To validate our method against 2-D images, we specifically acquired the NTNU 2-D Guiding data set, composed of n = 47

patients where each include 15 recordings. For each standard view, the following were acquired by a clinical expert:

- The standard view
- Two views with negative and positive rotation angles compared with the standard one
- Two views with negative and positive tilt angles compared with the standard one

The non-standard recordings have a position error relative to the standard recording that is qualitatively known, allowing evaluation of the ordering metrics given by eqns (4) and (5) given later.

External test data sets. We tested our method on two larger retrospective data sets composed of standard view recordings, without information on the AC: the CAMUS data set (500 patients) (Leclerc et al. 2019), which features A4C and A2C views, and the NTNU-LVD (left ventricular disease) data set (168 patients) (Grue et al. 2018), which features in addition the ALAX view. Both data sets are representative of daily clinical practice in terms of image quality and pathological cases.

Finally, we tested our method on a repeatability study data set of 88 patients who underwent three consecutive exams containing each of the three apical views, carried out by a panel of three sonographers and three cardiologists. We call this data set NTNU 2-D Repeatability.

Experimental setup

Training and validation. Our method gives feedback to help users adjust the transducer position toward a more optimized view. At the end of each training epoch, we ran two validation procedures:

1. For the first procedure the metric of interest is the validation loss, calculated on the internal validation subset of the annotated slices from 3-D. This validation loss is calculated similarly to the training loss. Neither spatial reference noise nor US augmentations are applied when calculating this validation loss.
2. For the second procedure we introduce the ordering success rate (OSR). This metric is more appropriate for evaluating our models from the user perspective of 2-D image guiding. The procedure uses views from the NTNU 2-D Guiding external validation data set. For each combination of patient, standard view, rotation or tilt, we use a set of three recordings corresponding to three different orientations:
   - Recording A: Anatomically correct standard view

- Recording B: Slight positive transducer angulation along the DOF from recording A
- Recording C: Slight negative transducer angulation along the DOF from recording A

The true ordering is then expressed as

$$p(C) < p(A) < p(B), \qquad (4)$$

and the ordering success is defined by

$$\text{Outcome} = \{ \text{Success,} \quad \text{if} \quad \hat{p}(C) < \hat{p}(A) < \hat{p}(B) \text{Fail, otherwise} \qquad (5)$$

where p and $\hat{p}$ are the ground truth and predicted relative positions of the 2-D imaging plane, respectively. The OSR is defined as the ratio of success cases over all cases.

We ran several experiments using the US augmentations, the spatial reference noise and the Gaussian blurring separately and together. The performance is compared with that obtained by using only the baseline set of augmentations. For each DOF, the best model is chosen considering the OSR metric, which reflects the guiding performance on 2-D images. The validation loss remains informative to avoid overfitting to the training data when selecting the model for further testing.

Measure of image standardization. To test our models at a larger scale, we retrospectively run them on the CAMUS and NTNU-LVD external test data sets made of 2-D standard views representative of clinical practice. These data sets contain data labeled as standard views. Although the labels do not contain information on the true AC, this experiment allows us to qualitatively investigate the degree of standardization of recordings acquired in a clinical setup.

Additionally, we study the inter-operator AC variability in the image acquisition by quantifying the AC on recordings from the NTNU 2-D Repeatability data set.

Real-time image-guiding application. We designed and developed a prototype application for image guidance using the FAST framework (Smistad et al. 2019), which combines image streaming from a GE E95 clinical scanner (GE Vingmed Ultrasound) and real-time DL inference and visualization. In the application, as illustrated in Figure 11 and Video S1 (online only), the user chooses a standard view to acquire and a target box corresponding to this view is displayed on the short-axis schematic of the heart. The current position of the transducer relative to the heart is updated in real time based on the models and represented as a *blue line*. When the current view is predicted as anatomically correct by both the rotation and tilt networks,

the *blue line* is inside the target box, which changes color from *red* to *green* to indicate a valid view.

RESULTS

Patient variability

From the output of step 4 in the annotation procedure in Figure 4, we could investigate the variability in rotation angle from A4C to A2C and ALAX views. The results illustrated in Figure 3 reveal variability in heart shapes. This justifies the need for a method that calculates the transducer position relative to some characteristic heart features rather than an angle-to-target value.

Method validation

Transducer rotation estimation. Figure 7a illustrates the predictions through the cardiac cycle for the six rotational CCSs and suggests good convergence of the training procedure on the internal validation set of 3-D slices when using only the baseline augmentations. These predictions are made on 2-D slices from 3-D volumes, as for training, and are expected to yield the best case results. As observed, the rotational model separates the different classes throughout the cardiac cycle although the LV axis and the rotational positions are annotated at ED only. Figure 8a is composed of the same images, but using a model trained with additional US augmentations and specific pre-processing. To evaluate performance on 2-D data during training, the OSR was calculated on recordings from the NTNU 2-D Guiding data set at the end of each epoch. Results are presented in the *Rotation* column of Table 1 and correspond to the average OSR for the three groups of standard views. The best model from the baseline experiment ordered the standard and non-standard 2-D views correctly in 92% of the cases, while an improved score of 95% was

obtained when using our additional augmentations and pre-processing steps. This improvement in the OSR is also visible in Figure 8b, where the predictions are more accurate and precise through the heart cycle than in Figure 7b.

Transducer tilt estimation. On the sliced 3-D data (Fig. 9, left column), the A4C tilt model predictions are consistent with the ground truth CCSs, with stable predictions along the cardiac cycle. Although all three models converge, the variability and stability are slightly worse for the A2C and ALAX tilt models, with more classes overlapping. On the 2-D validation data (center column), the A4C tilt model has the best OSR of 91% whereas the OSR is 67% (resp. 47%) for the A2C (resp. ALAX) view. These scores can be compared against a random choice scenario leading to a 16.6% correct ordering, meaning that all the models perform significantly better than a random baseline. One can also note from Table 1 that the improvements obtained with each augmentation or pre-processing taken separately are not additive when combined.

Real-time image guiding application

In addition to the quantitative offline assessment, we assessed the method qualitatively using our real-time application. Video S1 illustrates only the real-time application being used to acquire the three apical standard views. To the best of our knowledge, this video is the first providing evidence of a correlation between the transducer movements, the ultrasound image and the given feedback for all three apical views throughout a large number of heart cycles. For all standard and non-standard views, the position of the *blue feedback line* on the short axis schematic was responsive to the transducer movements and consistent with the structures visible on
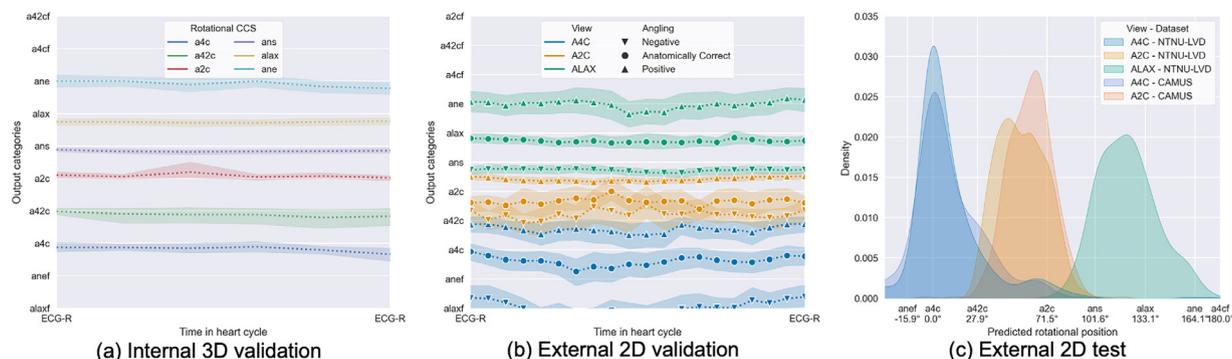

Fig. 7. Rotational results. Model with the best validation loss from the baseline experiment. (a) Model predictions for the annotated CCSs. (b) Model predictions for the NTNU 2-D Guiding recordings. (c) Density plot of the AC for recordings from CAMUS and NTNU-LVD. AC = anatomical correctness; A2C = apical two-chamber; A4C = apical four-chamber; A42C = view between A4C and A2C views; ALAX = apical long-axis; ANE = annulus end; ANS = annulus start; CCS = characteristic cross-section.
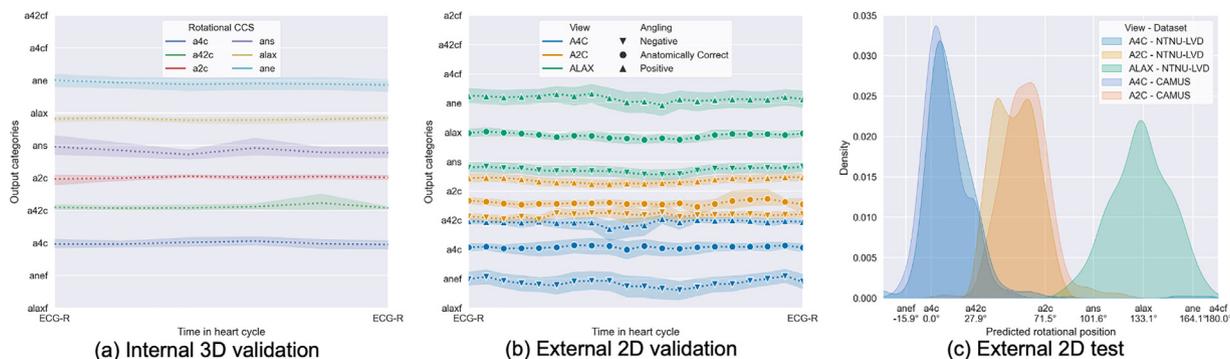
Fig. 8. Rotational results. Model with the best OSR on 2-D non-standard view validation data from the experiment featuring all the pre-processing and augmentations. (a) Model predictions for the annotated CCSs. (b) model predictions for the NTNU 2-D Guiding recordings. (c) Density plot of the AC for recordings from CAMUS and NTNU-LVD. AC = anatomical correctness; A2C = apical two-chamber; A4C = apical four-chamber; A42C = view between A4C and A2C views; ALAX = apical long-axis; ANE = annulus end; ANS = annulus start; CCS = characteristic cross-section; OSR = ordering success rate.

the US image. Figure 11a is a suboptimal A4C view with a visible LVOT. The *blue feedback line* is consequently moved away from the target toward the anterior wall. In Figure 11b, the operator acquired a good A4C view. This is confirmed by the application displaying the feedback line into the target box, which consequently switched color to *green*.

Measure of image standardization

　Retrospective assessment of recordings. The predicted view rotation and tilt for recordings from the CAMUS and the NTNU-LVD data sets are provided in the right columns of Figures 8 and 9. The corresponding visual inspection is available in Video S2 (online only), which illustrates that for each standard view, suboptimal recordings are located at the tails of the distributions whereas standard recordings are located at the center of the distributions. This suggests that our method is robust enough to predict the correct transducer position over a large number of patients.

　Additionally, it can be observed that the positional biases for the external test data sets are consistent with the positional biases on the 2-D validation data from the NTNU 2-D Guiding data set, with, for example, the A2C

rotational position being shifted toward the A4C rotational position.

　Inter-operator variability. We evaluated the AC of the assumed optimal apical standard views from the NTNU 2-D Repeatability data set with our method, with predictions indicating significant differences among operators (Fig. 10). For the A2C rotational position, the method suggested that operator 3 tends to acquire the A2C views closer to the A4C view. This was confirmed by the visual inspection available in Video S3 (online only), in which the RV was typically partly visible. For the ALAX rotational position, the results suggested that the ALAX views of operators 1 and 3 are similar to the A4C view with a vertical flip. Visual inspection of Video S3 revealed that these views included four chambers (the RA is not expected in ALAX) and the LVOT (the aorta valve leaflets are expected to be visible in addition to the LVOT).

DISCUSSION

　We developed and evaluated a deep learning−based method capable of estimating the anatomical orientation of apical views in echocardiography. Multiple deep neural networks were trained to regress

Table 1. Success rate for ordering non-standard views

	Rotation (all views)	Tilt A4C	Tilt A2C	Tilt ALAX
Baseline	0.92	0.81	0.53	0.44
Ultrasound augmentations	**0.93**	**0.95**	0.53	0.33
Gaussian blurring	**0.93**	0.81	**0.60**	0.37
Spatial reference noise	0.90	**0.84**	**0.60**	0.42
All	**0.95**	**0.91**	**0.67**	**0.47**

A2C = apical two-chamber; A4C = apical four-chamber; ALAX = apical long-axis.
Values in boldface show improvement using specific augmentations compared with the baseline training procedure.
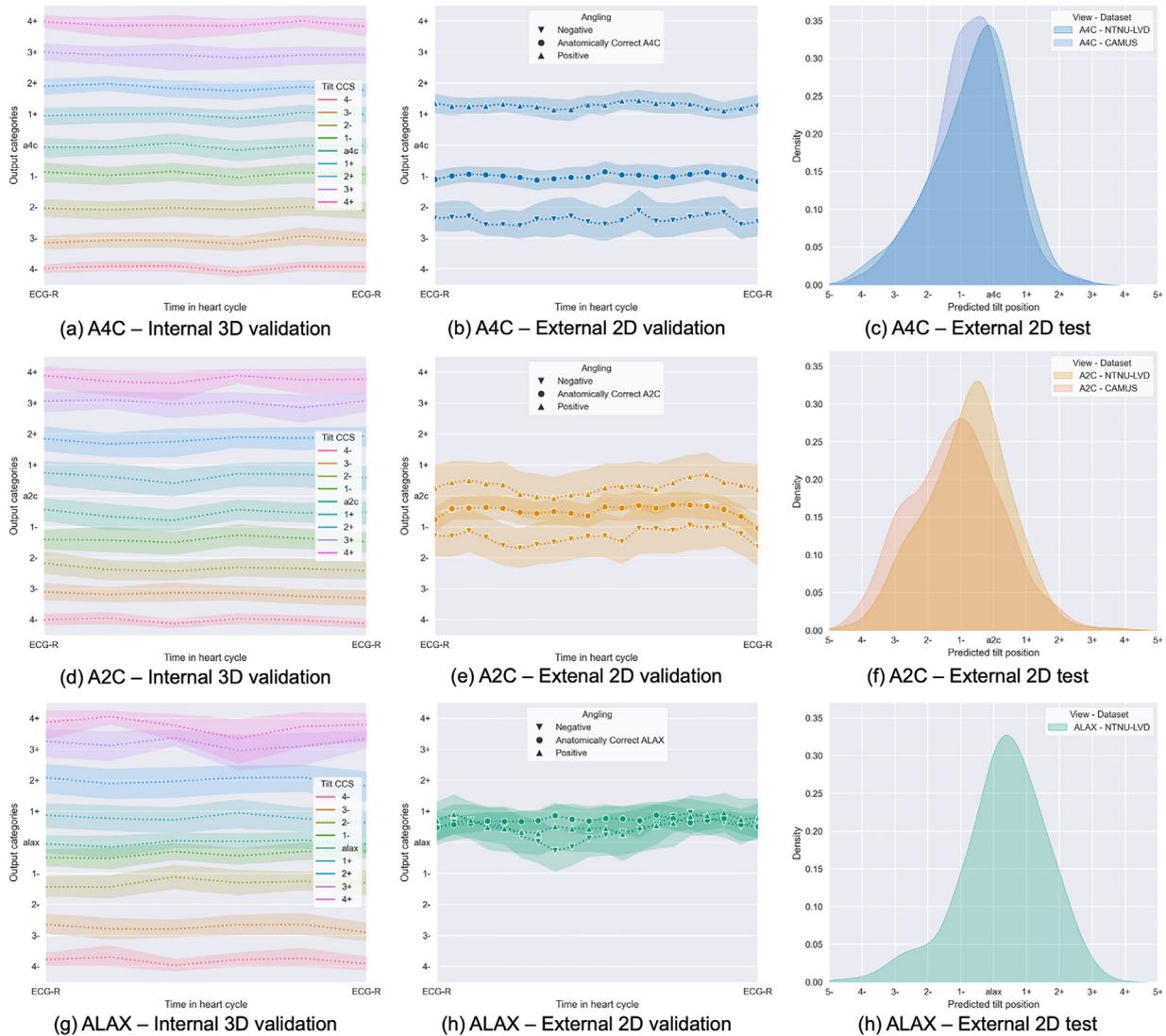
Fig. 9. Tilt results. Model with the best OSR on 2-D non-standard view validation data from the experiment featuring all the pre-processing and augmentations. (left column) Model predictions for the annotated CCSs, (center column) model predictions for the NTNU 2-D Guiding recordings, (right column) density plot of the AC for recordings from CAMUS and NTNU-LVD. AC = anatomical correctness; A2C = apical two-chamber; A4C = apical four-chamber; ALAX = apical long-axis; CCS = characteristic cross-section; OSR = ordering success rate.

the position of the transducer using 2-D slices generated from 3-D US volumes as the training material. So far we considered the transducer rotation and out-of-plane tilt movements, assuming the user has already positioned the transducer such that the image plane passes approximately through the apex of the heart. The method was evaluated using 2-D ultrasound recordings and exhibited promising results in terms of robustness and accuracy. The real-time implementation of the method indicated the ability to provide the user correct feedback on the required movements needed to acquire the valid standard views.

The method is suitable for real-time inference both on off-the-shelf GPUs and on less powerful hand-held

devices. Thus, it may benefit less experienced users to obtain anatomically correct standard views for diagnostics, and can contribute to standardize image views acquired in the echo lab. The method can also be used for quality control, for instance, to help clinicians automatically select the recording and cardiac cycle most suitable for a given measurement within an exam consisting of many recordings. This can also be highly useful in a research setting, for instance, when data mining large amounts of patient examinations.

For both online and offline use cases, the method can reduce the measurement variability introduced during image acquisition and thereby contribute to more accurate patient management. Used retrospectively, the
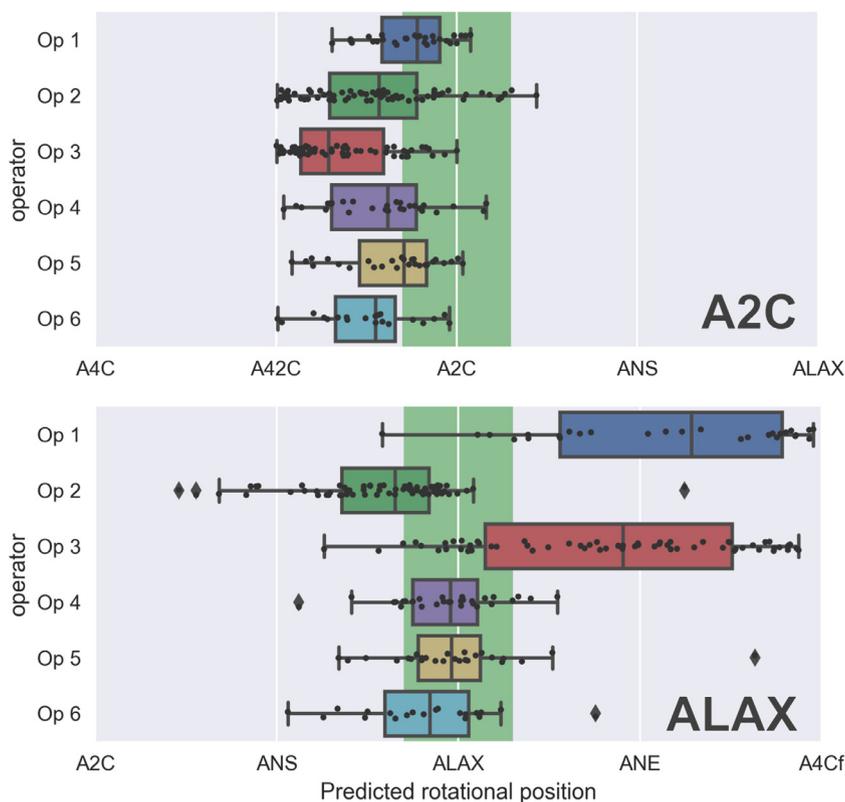
Fig. 10. Predicted rotational position by operator for 2-D recordings for the NTNU 2-D Repeatability data set. A2C = apical two-chamber; A4C = apical four-chamber; A42C = view midway between A4C and A2C views; ALAX = apical long-axis; ANE = annulus end; ANS = annulus start.

method can be used as a tool to analyze and improve scanning habits in individual operators.

An important aspect of our work is the use of 3-D US recordings as the primary data source, which combined with our semi-automatic annotation procedure allowed us to use data from 162 patients from several echo labs. With use of 3-D data, the labeled reference position is given relative to the heart anatomy. Our method therefore compensates for the heart movements inside the chest that occur while breathing, contrarily to Toporek et al. (2019), who used an external sensor to acquire the reference position.

Using 3-D US volumes as the primary data source comes at the price of a discrepancy between the training data made of 3-D slices and the inference data consisting of 2-D recordings. We therefore introduced an additional validation metric, the OSR, which is directly related to the ability to detect small transducer movements on 2-D data and thus has more clinical relevance than the traditional validation loss.

The present work divided the method into simpler sub-problems, which allowed us to better identify some of the challenges related to the apical standard views AC assessment with DL networks. For instance, we found that

the estimated rotational position was more accurate than the estimated tilt position. The average predictions of the rotational model over the cardiac cycle on the NTNU 2-D Guiding in Figure 8b revealed a bias of the A2C rotation toward the A4C. This bias is also present in the CAMUS and NTNU-LVD data sets, as illustrated in Figure 8c. Visual inspection available in Video S2 revealed that many recordings labeled as standard A2C views partly include the RV or the coronary sinus vein. This suggests that the rotational position reference is correct and that the A2C 2-D recordings from the NTNU 2-D Guiding, CAMUS and NTNU-LVD data sets do not fully comply with the expected A2C standard view. Our method thus has the potential to improve AC for experienced users.

Visual inspection of the tilt recordings from the NTNU 2-D Guiding data set in the A2C and ALAX positions revealed that the three recordings are similar, explaining the lower OSR results for the A2C and ALAX tilt models compared with the rotational and A4C tilt models. The lower OSR on the 2-D data is also consistent with the results on the 3-D slices (Fig. 9d, 9g) which have a slightly higher variability than the A4C tilt model (Fig. 9a). This suggests that tilt regression is more difficult to learn in the A2C and ALAX views than in

(a) Incorrect A4C view with visible LVOT – Target box is red

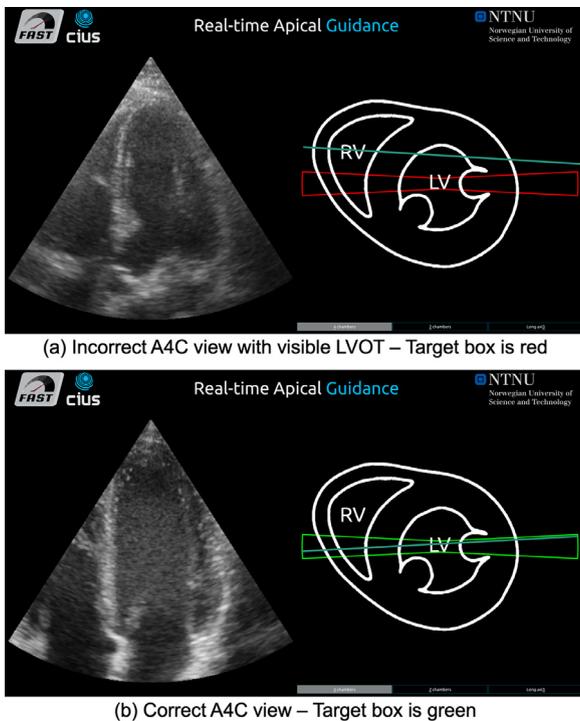(b) Correct A4C view – Target box is green

Fig. 11. Screen captures of the real-time application prototype. Left side of the screen shows the ultrasound image streamed from a GE E95 scanner. Right side of the screen shows the calculated feedback based on the neural network output. A4C = apical four-chamber; LV = left ventricle; LVOT = left ventricular outflow tract; RV = right ventricle.

A4C views, although the data generation and training procedures are identical. Visual inspection of the tilt training data revealed that the slices in the A4C direction are potentially richer in features than in the A2C and ALAX directions.

For the A4C tilt model, one can note, from Figure 9b and 9c, a consistent bias for the 2-D A4C view compared with the A4C sliced in a 3-D volume. Visual inspection of the training material suggested a bias in the reference caused by the automatically generated LV axis. Indeed, we used the geometrical LV axis for the rotational axis during training data generation, whereas the axis used in practice by clinicians crosses the mitral plane slightly closer to the anterior wall. Nevertheless, such reference biases are not an issue for our method as they can be quantified from the validation data set and compensated in post-processing for further inference.

For all DOFs, the position predictions on both validation slices and the NTNU 2-D Guiding data set did not significantly vary throughout the cardiac cycle, suggesting that the annotations made at the ED frame are valid for a complete cycle.

Using the US-specific augmentations improved or maintained the accuracy on the OSR. This supports the hypothesis that adding domain knowledge from US

through data augmentation improves the robustness of the trained networks. Further, the blurring step applied at both training and inference seems to be beneficial for performance, suggesting that local image features are less relevant for the present task.

The CVCNet neural network topology was benchmarked against other topologies (benchmark available in the Supplementary Data, online only). Neither smaller networks (MobileNet V2) nor larger networks (Inveption V3, ResNet50) provided significantly different results. As carrying out a clinical superiority test as proposed by Varoquaux and Cheplygina (2022) is not realistic during early technical development, we assumed our method was topology agnostic and focused our efforts on careful pre-processing of the data and an in-depth evaluation of several external data sets to demonstrate robustness.

Although designed to quantify the guiding abilities of our models, the OSR metric is limited by the fact that we could not control the amount of positive and negative rotation or tilt introduced by the clinicians who acquired the NTNU 2-D Guiding data set. This, in addition to the small size of the NTNU 2-D Guiding data set, makes conclusions on improvements obtained with data augmentation and pre-processing difficult.

Despite limited quantitative results for our method caused by the aforementioned, Videos S1−S3 provide evidence of the accuracy and robustness of our method by showcasing the association between the transducer movements, the US images and the predicted position.

In use, the main limitation of our guidance tool is the need for the correct intercostal point as a starting point. Although this is achievable for experienced users, positioning the transducer over the apex can be challenging for nonexperts. Further clinical testing will identify the improvements required to make the application usable by most users. Future work will address the explainability of the method as this is required for the method to be adopted by the medical community. The same approach should be applicable to make a guiding tool for the parasternal long- and short-axis views. However, as 3-D US is more often performed in the apical window, 3-D data might be less available.

CONCLUSIONS

We have described a method to help ultrasound users acquire apical standard views of the heart. The backbone of the method is based on deep neural networks performing regression of the transducer position relative to the heart. The networks were trained on 2-D slices from 3-D US volumes, where reference positions were obtained using a semi-automated approach. Testing on multiple external data sets of 2-D recordings revealed that the method could detect suboptimal image planes and unveil individual operators' scanning habits. This

suggests that our method is sufficiently robust and accurate to be of clinical value. A real-time application that supports inference of images streamed from a clinical US scanner and displays an intuitive feedback on view position was developed. Examples illustrate the expected relation between the transducer movements, the ultrasound image and the machine learning calculated feedback. Further work will quantify the clinical value of the method inside and outside the echo lab, and map the potential benefit for expert and non-expert users.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ultrasmedbio.2022.09.006.

REFERENCES

Bernard O, Heyde B, Alessandrini M, Barbosa D, Camarasu-Pop, Cervenansky F, Valette S, Mirea O, Galli E, Geleijnse M, Papachristides A, Bosch JG, D'hooge J. Challenge on endocardial three-dimensional ultrasound segmentation (CETUS). Midas J 2014;1–8.
Chykeyuk K, Yaqub M, Alison Noble J. Class-specific regression random forest for accurate extraction of standard planes from 3D echocardiography. In: Menze B, Langs G, Montillo A, Kelm M, Müller H, Tu Z, eds. Medical computer vision. Large data in medical imaging. Lecture Notes in Computer Science MCV 2013. Lecture Notes Comput Sci 2014;8331:53−62.
Domingos JS, Lima E, Leeson P, Noble JA. Local phase-based fast ray features for automatic left ventricle apical view detection in 3D echocardiography. Lecture Notes Comput Sci 2014;8331:119–129.
Droste R, Drukker L, Papageorghiou AT, Noble JA. Automatic probe movement guidance for freehand obstetric ultrasound. Lecture Notes Comput Sci 2020;12263:583–592.
Grue JF, Storve S, Dalen H, Salvesen Ø, Mjølstad OC, Samstad SO, Torp H, Haugen BO. Automatic measurements of mitral annular plane systolic excursion and velocities to detect left ventricular dysfunction. Ultrasound Med Biol 2018;44:168–176.
Hou L, Yu CP, Samaras D. Squared earth mover's distance-based loss for training deep neural networks. 2016;Available at: http://arxiv.org/abs/1611.05916.
Lang RM, Badano LP, Victor MA, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, Lancellotti P, Muraru D, Picard MH, Retzschel ER, Rudski L, Spencer KT, Tsang W, Voigt JU. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015;28:1–39 e14.
Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin PM, Grenier T, Lartizien C, Dhooge J, Lovstakken L, Bernard O. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging 2019;38:2198–2210.
Li Y, Khanal B, Hou B, Alansary A, Cerrolaza JJ, Sinclair M, Matthew J, Gupta C, Knight C, Kainz B, Rueckert D. Standard plane detection in 3D fetal ultrasound using an iterative transformation network. Lecture Notes Comput Sci 2018;11070:392–400.
Morbach C, Gelbrich G, Breunig M, Tiffe T, Wagner M, Heuschmann PU, Störk S. Impact of acquisition and interpretation on total interobserver variability in echocardiography: Results from the quality assurance program of the STAAB cohort study. Int J Cardiovasc Imaging 2018;34:1057–1065.
Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, Chaudhry A, Martin RP, McCarthy PM, Rubenson DS, Goldstein S, Little SH, Lang RM, Weissman NJ, Thomas JD. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. JAMA Cardiol 2021;6:624–632.
Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L. Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. Ultrasound Med Biol 2019;45:374–384.
Schneider M, Bartko P, Geller W, Dannenberg V, König A, Binder C, Goliasch G, Hengstenberg C, Binder T. A machine learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. Int J Cardiovasc Imaging 2021;37:577–586.
Smistad E, Ostvik A, Pedersen A. High performance neural network inference, streaming, and visualization of medical images using FAST. IEEE Access 2019;7 136,310−136,321.
Smistad E, Østvik A, Salte IM, Melichova D, Nguyen TM, Haugaa K, Brunvand H, Edvardsen T, Leclerc S, Bernard O, Grenne B, Løvstakken L. Real-time automatic ejection fraction and foreshortening detection using deep learning. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2595–2604.
Smistad E, Steinsland EN, Lovstakken L. Real-time 3D left ventricle segmentation and ejection fraction using deep learning. Proc IEEE Int Ultrason Symp 2021;2021–2023.
Snare SR, Torp H, Orderud F, Haugen BO. Real-time scan assistant for echocardiography. IEEE Trans Ultrason Ferroelectr Freq Control 2012;59:583–589.
Toporek G, Naidu R, Xie H, Simicich A, Gades T, Raju B. User guidance for point-of-care echocardiography using multi-task deep neural network. Med Image Comput Comput Assist Interv 2019;1:309–317.
Varoquaux G, Cheplygina V. Machine learning for medical imaging: Methodological failures and recommendations for the future. NPJ Digital Med 2022;5:48.