IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received April 21, 2022, accepted May 4, 2022, date of publication May 10, 2022, date of current version May 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174059

Mitral Annulus Segmentation and Anatomical Orientation Detection in TEE Images Using Periodic 3D CNN

BØRGE SOLLI ANDREASSEN[1], DAVID VÖLGYES[1], EIGIL SAMSET[1,2,3], AND ANNE H. SCHISTAD SOLBERG[1]

[1]Department of Informatics, University of Oslo, 0316 Oslo, Norway
[2]GE Healthcare, 3183 Horten, Norway
[3]ProCardio Center for Innovation, 0372 Oslo, Norway

Corresponding author: Børge Solli Andreassen (borgesan@ifi.uio.no)

ABSTRACT Segmentation of the mitral annulus is often an important step in cardiac examinations. We propose a robust 3D method for predicting the anatomical orientation and segmentation of the mitral annulus in 3D transesophageal echocardiography. The method takes advantage of the circular anatomy of the annulus by utilizing cylinder coordinate samples and a 3D convolutional neural network with circular convolutions. Furthermore, the paper proposes new landmark detection loss functions based on the earth mover's distance. The method's effectiveness was demonstrated by training a HighRes3dNet model and evaluating its performance on a separate test set consisting of 135 frames from 19 examinations. The obtained coordinate prediction error was 1.96 ± 1.62 mm, and the anatomical orientation prediction error was $9.7° \pm 15.8°$. The robust and fully automatic mitral annulus segmentation and orientation prediction provided by the method can ease the workload of clinicians and provide time savings in clinics.

INDEX TERMS Deep learning, earth mover's distance, echocardiography, landmark detection, mitral annulus segmentation.

I. INTRODUCTION

The mitral valve is located between the left atrium and left ventricle, allowing blood to flow from the atrium to the ventricle and preventing backflow of blood during the ventricular systole. The mitral valve complex comprises the mitral annulus, two mitral leaflets that attach to the mitral annulus, chordae tendineae, and papillary muscles [1]. The saddle-shaped mitral annulus changes size and shape throughout the cardiac cycle [2]. Papillary muscles, connected to the mitral leaflets through fibrous cords (the chordae tendineae), prevent the mitral leaflets from prolapsing into the atrium during the ventricular systole.

Valvular heart disease is a common health problem in both industrialized and developing countries [3]. A population-based study by Nkomo *et al.* [4] estimates a prevalence of valvular heart disease in the United States of 2.5%. Diseases related to the mitral valve include mitral stenosis and mitral regurgitation. Mitral stenosis is a narrowing of the mitral valve that is most prevalent in developing countries — as it is often associated with rheumatic fever [5] — however, it is also present in industrialized countries [6]. Mitral regurgitation is characterized by the valve leaflets not providing a tight seal, resulting in backflow of blood from the ventricle to the atrium during systole [7]. Nkomo *et al.* [4] found that the prevalence of mitral regurgitation increases with age — with almost 10% of the population older than 75 years of age being affected by moderate to severe regurgitation.

Echocardiography, or cardiac ultrasound, is the recommended image modality for initial evaluation of valvular heart disease, according to the guidelines of the ESC/EACTS [8] and the ACC/AHA [9]. Transesophageal Echocardiography (TEE) is an imaging modality where the ultrasound probe is passed through the esophagus, resulting in high-quality images due to the proximity to the heart. Images of

The associate editor coordinating the review of this manuscript and approving it for publication was Chulhong Kim.

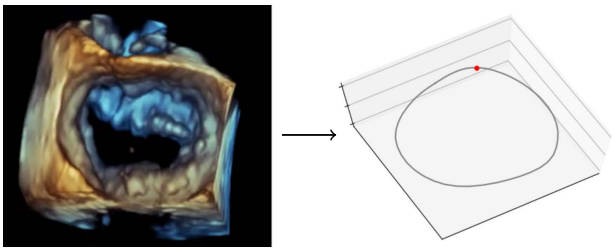

FIGURE 1. Left subplot shows a 3D TEE acquisition of the mitral valve from the view of the probe. Right subplot illustrates the 3D curve of the mitral annulus, where the red point shows the center of the aortic outflow tract. The proposed method predicts both the segmentation of the mitral valve and the anatomical orientation of the valve from 3D TEE images.

the mitral valve can be acquired with the TEE probe at the midesophageal level — a standard view in TEE examinations [1].

In recent years, the mitral annulus anatomy has received increased attention, e.g., related to mitral regurgitation volume assessments and mitral valve repair planning [1], [10]. Precise mitral annulus segmentation can facilitate accurate, quantitative measurements for the above applications. The method proposed in this paper provides automatic segmentation of the mitral annulus from midesophageal-level 3D TEE acquisitions, as illustrated in Figure 1. The method can save valuable time during TEE examinations and subsequent analysis, as immediate evaluation of the automated segmentation would allow the clinician to quickly verify that the acquisition is suitable for the measurements.

A. RELATED WORK

The increasing amount of medical imaging data being generated has manifested a need for automated algorithms that can provide fast, accurate, and reliable information from the data [11]. Segmentation methods for ultrasound images have long been an active research field [12], with multiple recent applications [13], [14].

Deformable models represent a widely used class of methods that has been applied for medical image segmentation for decades [15]. These methods span from curvature and image gradient methods to statistical shape models that take advantage of 3D anatomical shape models as a priori knowledge [16]. However, a common limitation of deformable models is that they often require manual initialization.

Methods based on machine learning have become increasingly popular in medical image segmentation [17]. A wide range of machine learning methods exists where feature representations are learned instead of applying handcrafted features. Recent years have seen a growing amount of research applying artificial intelligence to cardiovascular imaging, particularly deep learning methods using Convolutional Neural Networks (CNN) [14]. While CNN-based models can learn features and properties from the training data and do not require initialization, these methods rely on the availability of labeled data.

While we limit the majority of this overview to pertain to mitral valve segmentation, examples of other applications in echocardiography are 2D and 3D left ventricle segmentation [18], [19], which for instance can be applied for ejection fraction estimation and foreshortening detection [20].

Mitral annulus segmentation is closely related to segmentation of the mitral valve leaflets — as the outer borders of the mitral leaflets define the mitral annulus — however, the tasks are often complementary since mitral leaflet segmentation generally does not ensure a smooth, continuous delineation of the mitral annulus morphology. Several previous works on mitral leaflet segmentation illustrate this point by highlighting the prediction accuracy close to the annulus as a challenge [21]–[23].

Previous work on delineating the mitral valve annulus in TEE imaging covers a range of computer vision methods. Ionasec *et al.* [24] proposed a method that fits a physiological shape model to the aortic and mitral valve in both CT and TEE. Schneider *et al.* [25] proposed applying a 'thin tissue detection' algorithm to find the mitral leaflets and then applying graph cut segmentation to identify the mitral annulus. Later, Schneider *et al.* [26] utilized optical flow to track the mitral annulus through the diastole, starting from a segmentation using the method presented in [25] on a systolic frame. Voigt *et al.* [27] proposed a two-component method, using a probabilistic boosting tree to find an initial prediction, then applying optical flow to track the prediction temporally. Sotaquira *et al.* [28] applied Dijkstra's algorithm to segment the mitral annulus and leaflets. Tiwari and Patwardhan [29] applied the thin tissue detector of [25] with a Naive Bayes classifier for the annulus localization.

Previous methods applying 2D CNNs to the task also exist. Our previous work [30] presented a method for mitral annulus segmentation using a 2D CNN on individual image slices. A limitation of [30] is that it did not use 3D context but obtained 3D predictions using an iterative post-processing algorithm along the 2D planes — where the prediction in a plane relied on the result of the neighboring plane. Zhang *et al.* [31] presented another mitral annulus segmentation pipeline — using a combination of deep reinforcement learning, 2D landmark detection using a 2D CNN, and spline fitting to produce the 3D annulus predictions.

Deep learning has also been proposed for the related task of mitral valve leaflets segmentation, where Carnahan *et al.* [32] in a recent work presented a fully automatic method for segmenting the mitral valve leaflets in 3D TEE, using a 3D CNN.

B. CONTRIBUTIONS

First, the proposed method provides a highly accurate and automatic mitral valve segmentation for 3D TEE images. The approach has similarities to our previous work [30]; however, instead of predicting individual 2D planes, the inference is done on the entire 3D volume. Further, the method proposed in this paper does not include nor require any post-processing

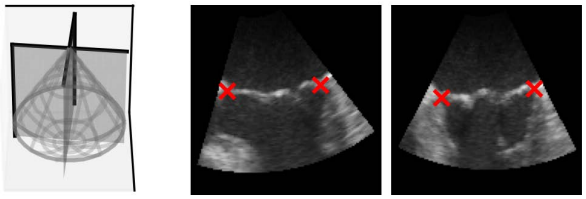

FIGURE 2. Two orthogonal planes from one of the DICOM files. Left: Illustration showing the location of the two planes in Cartesian space. The meshgrid illustrates the field of view of the probe. Center and right: Projected images from the DICOM with the mitral annulus labels marked with a red x. Note that the mitral valve intersects both planes in two points. The aortic outflow tract can be seen on the left side of the center subplot.

of the output, unlike [30]. To the best of our knowledge, our method is the first to apply a 3D CNN to the mitral annulus segmentation task. Our approach simplifies and decomposes the problem by applying cylinder coordinate samples (Section II-A) and a 3D CNN with periodic boundary conditions (Section II-C) — enabling the CNN to learn to exploit the 3D contextual information. The method is adaptable to other applications with a circular nature, also outside medical imaging.

Second, the proposed method predicts the anatomical orientation of the mitral valve, enabling, e.g., automatic view selection and calculation of additional clinical measurements.

Third, we introduce several tools for general CNN landmark detection applications:

- Introducing an earth mover's distance loss function for landmark detection, using the distance from the label coordinate to the 2D prediction heatmap.
- Using the angular moment of prediction heatmaps as a confidence proxy to weigh their contribution and calculate an ensembled heatmap.
- Using the geometric median as coordinate prediction estimator, since it is less susceptible to outliers than the 'center of mass' and less susceptible to noise than using 'arg max'.

II. METHODOLOGY

A. CYLINDER COORDINATE SAMPLES

The method uses cylinder coordinate to take advantage of the midesophageal-level view of the acquisitions. In this view, the mitral valve is *en face* in the acquisition, i.e., the valve is forward-facing when seen from the position of the probe, as shown in Figure 1.

A cylinder coordinate system is created by rotating an initial plane around its depth axis. Denote the coordinate dimensions by α, h, and w — corresponding to the angle of rotation around the centerline, the depth from the probe, and finally, the distance from the centerline. In this coordinate system, each rotational plane intersects the mitral valve in two positions,[1] as illustrated in Figure 2.

[1]Requirements for the planes to intersect cylinder coordinate planes are described in Section III-B.

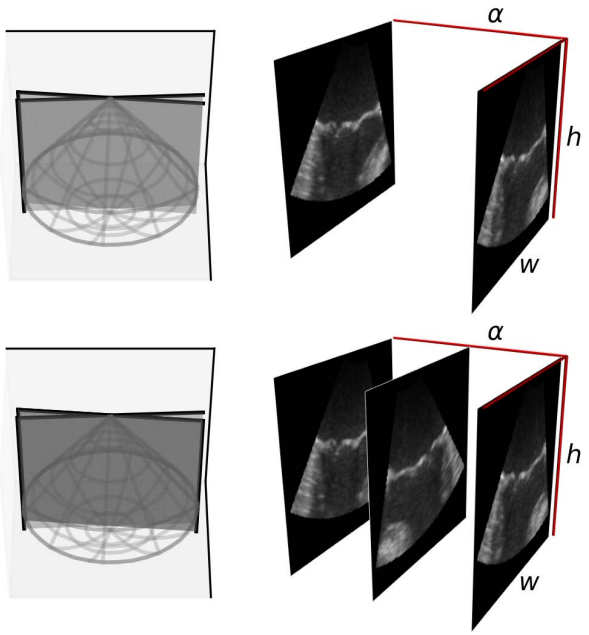

FIGURE 3. Illustration showing a cylinder coordinate sample, as introduced in Section II-A, with meshgrid mock-ups of probe field of view with example planes (left) and corresponding image planes positioned in the cylinder volume (right). The figure only shows a few rotational planes for visualization purposes. An animation showing all the rotational planes is available in the supplementary material (video 1). Top: Two neighboring planes. Note that while the planes are opposite to each other in the cylinder coordinate space (right column), they are neighbors as the rotational dimension is periodic. Bottom: The same planes as in the top row, in addition to a plane 180° (i.e., $n_\alpha/2$ planes) away from the plane in the front, i.e., the front and middle planes are mirror images. Note that the added plane is hard to see in the meshgrid, as it completely overlays its mirror plane.

Cylinder coordinate samples, $\mathbf{V}$, are obtained by projecting the DICOM data to n_α rotational planes. The resulting sample dimensions are (n_α, n_h, n_w), where (n_h, n_w) are the image dimensions of each plane (pixels). The n_α planes are obtained by iteratively rotating the initial coordinate plane by $360/n_\alpha$ degrees around its depth axis. Note that the two orthogonal planes shown in Figure 2 correspond to two planes in this cylinder coordinate system with $\Delta\alpha = 90°$.

Let subscripts a, i, and j denote the discrete indices of the rotational, height, and width dimensions. The a-th rotational plane is denoted $\mathbf{V}_a$ and contains the projected image at the plane rotated by $a/n_\alpha \cdot 360$ degrees around the depth centerline from a *base plane*. Consider the index notation of the rotational dimension to be modulus n_α, i.e., $\mathbf{V}_a = \mathbf{V}_{a+n_\alpha}$ implicitly, as the rotational dimension is periodic.

Figure 3 illustrates how coordinate planes in Cartesian space correspond to the cylinder coordinate samples. As illustrated in the figure, the planes, $\mathbf{V}_a$, are periodic in the rotational dimension and mirror-symmetric every $n_\alpha/2$-th plane — as this corresponds to a 180-degree rotation around the plane centerline. This mirror symmetry is exploited during inference, as described in Section II-E.

B. CYLINDER COORDINATE LABELS

Each sample has two label types: the mitral annulus coordinate labels (in-plane coordinates) and the anatomical orientation label (rotation angle).

By itself, the mitral annulus segmentation leaves one degree of freedom to be determined: the orientation of the valve. A suitable landmark to determine this orientation is the aortic outflow tract — as it is clearly distinguishable in midesophageal-level TEE acquisitions (see Figure 2). The mitral annulus segmentation and orientation enable automating several standard views (e.g., the surgical 3D view and anteroposterior view) and related measurements.

The raw annotations used to calculate the mitral annulus labels are 58 Cartesian coordinate points that delineate the annulus. Further, the anatomical orientation label is calculated from a single Cartesian coordinate point at the location of the aortic outflow tract. Section III-A describes the raw annotation of the data.

As described in Section II-A, the mitral annulus intersects each plane, $\mathbf{V}_a$, in two points. Let the superscripts l and r denote the left and right intersection points, respectively. Normalized label coordinates — $\mathbf{y}_a^l$ and $\mathbf{y}_a^r$ — are calculated for each plane, $\mathbf{V}_a$, by linear interpolation between the closest of the Cartesian coordinate points on each side of the respective plane. Let $\mathbf{y}^l$ and $\mathbf{y}^r$ be the coordinate labels through $\mathbf{V}$ — forming two curves corresponding to the rotational dimension's left and right intersection points.

Let the anatomical orientation label, o, be defined as the normalized plane index where the center of the aortic outflow tract lies on the right side of $\mathbf{V}_a$. The label is calculated using the Cartesian coordinate for the aortic outflow tract introduced above and can have values in the range [0, 1], corresponding to [0°, 360°] rotation.

Each cylinder coordinate sample consists of the volume and labels, i.e., $\{\mathbf{V}, \mathbf{y}^l, \mathbf{y}^r, o\}$.

C. DEEP LEARNING MODEL

The proposed method relies on a 3D fully convolutional neural network to simultaneously predict the mitral annulus and anatomy orientation — making it a multitask learning problem. Multitask learning is generally known to have a regularizing effect, which can reduce the risk of overfitting [33].

The model, $\mathcal{M}$, is a heatmap regression model [34], which takes cylinder coordinate volumes, $\mathbf{V}$ — as described in Section II-A — as its input and yields three output channels with the same dimensions as the input. The two first channels are used for coordinate predictions, the final channel for predicting the anatomical orientation. Denote the raw channel output of the final model layer $\{\check{\mathbf{H}}^l, \check{\mathbf{H}}^r, \check{\mathbf{O}}\}$.

Spatial softmax is applied to each rotational plane of the two first channels, obtaining normalized heatmaps along the rotational dimension. The third channel, $\check{\mathbf{O}}$, is aggregated to a 1D vector by computing the mean value over the spatial dimensions and applying 1D softmax. Specifically, the model output is:

$$\mathcal{M}(\mathbf{V}) \rightarrow \begin{cases} \mathbf{H}^l = \Phi(\check{\mathbf{H}}^l), & (n_\alpha \times n_h \times n_w) \\ \mathbf{H}^r = \Phi(\check{\mathbf{H}}^r), & (n_\alpha \times n_h \times n_w) \\ \mathbf{O} = \mathcal{S}\left(\frac{1}{n_h \cdot n_w} \sum_{i,j} \check{\mathbf{O}}\right), & (n_\alpha) \end{cases} \tag{1}$$

where Φ is the spatial softmax function applied to each plane along the rotational dimension and $\mathcal{S}$ is the 1D softmax function.

The method takes advantage of the cylinder coordinate samples (see Section II-A), as the region around the mitral annulus has similar image-features along the rotational dimension (see Figure 3 and supplementary video 1). This similarity reduces the variability that the 3D CNN feature extractors needs to learn, and thereby simplifies the learning task compared to samples in Cartesian coordinates.

Section III-C details the specific model architecture and model parameters used in the reported experiments of this paper.

D. LOSS FUNCTION

The proposed method has two goals: Segmenting the mitral valve annulus and predicting the anatomical orientation of the volume. The loss function reflects these goals by applying a weighted combination of two loss terms:

$$\mathcal{L} = \lambda \mathcal{L}_{\text{c}} + (1 - \lambda)\mathcal{L}_{\text{ao}}, \quad \lambda \in [0, 1], \tag{2}$$

where $\mathcal{L}_{\text{c}}$ optimizes coordinate predictions, $\mathcal{L}_{\text{ao}}$ optimizes the anatomical orientation prediction, and λ is a hyperparameter.

Both loss functions are based on the earth mover's distance with single point labels, resulting in the two closed-form loss functions introduced below. Further, both losses apply a transportation cost proportional to the square distance to the label, penalizing more significant errors and prioritizing more challenging samples.

1) COORDINATE PREDICTION LOSS

Let the target distribution for each prediction plane, $\mathbf{H}_a^p, p \in \{l, r\}$, be a heatmap with a Kronecker-delta centered at the normalized coordinate $\mathbf{y}_a^p$. Then, each pixel's resulting *transportation cost* contribution is proportional to the squared distance to the label, $\mathbf{y}_a^p$. Specifically, the loss contribution of a single heatmap plane is:

$$E(\mathbf{H}_a^p, \mathbf{y}_a^p) = \sum_{i,j} \mathbf{H}_a^p \circ \mathcal{D}^2_{\mathbf{y}_a^p}, \tag{3}$$

where $\mathcal{D}_{\mathbf{y}_a^p}$ is a distance matrix with dimensions (n_h, n_w) and pixel values contain the normalized distance from each pixel to the normalized label, $\mathbf{y}_a^p$. Equation (3) applies the square of this distance matrix. Figure 4 illustrates the coordinate prediction loss function applied to a synthetic example heatmap.

The coordinate prediction loss for each volume, $\mathbf{V}$, is the average loss across the rotational planes for both the left and

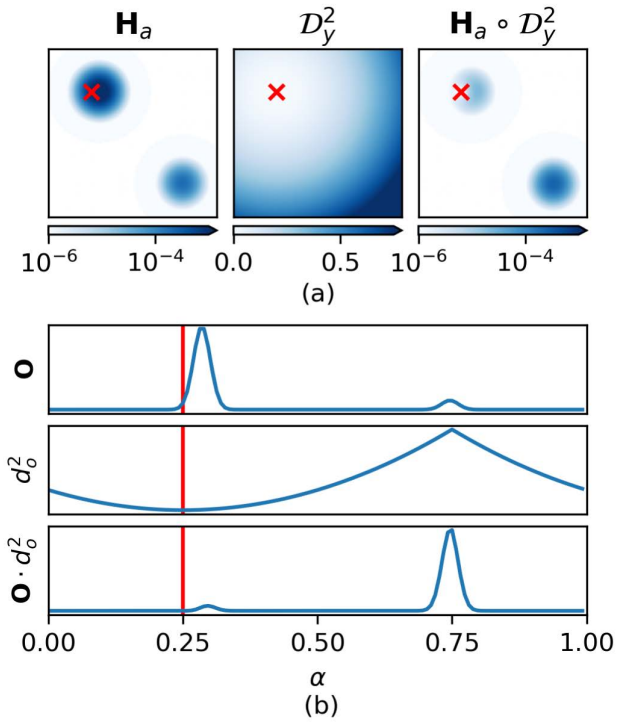

FIGURE 4. Synthetic examples of the two earth mover's distance loss functions introduced in Section II-D. (a) Illustration of the coordinate prediction loss function (3). Left plot: Synthetic prediction heatmap (with sum of one) that has two modes: a strong mode centered close to the label position, $\mathbf{y}$ (red x), and a weaker, erroneous mode in the lower right corner. Center plot: Distance matrix, $\mathcal{D}_{\mathbf{y}}^2$, which has pixel values proportional with the square distance to the label, $\mathbf{y}$. Right plot: The Hadamard product $\mathbf{H} \circ \mathcal{D}_{\mathbf{y}}^2$. The coordinate prediction loss (3) is the sum of the elements of this product. (b) Illustration of the anatomical orientation loss function (5). Top plot: Synthetic anatomical orientation output $\mathbf{O}$ with two modes: a strong mode centered close to the label position, o (red line), and a weaker, erroneous mode. Center plot: Distance vector, $\mathbf{d}_o^2$, with values proportional with the square distance to the label, o. Bottom plot: Elementwise product between $\mathbf{O}$ and $\mathbf{d}_o^2$. The anatomical orientation loss function (5) is the sum of these values.

the right prediction heatmaps:

$$\mathcal{L}_{\mathrm{c}}(\mathcal{M}(\mathbf{V}), \mathbf{y}^l, \mathbf{y}^r) = \frac{1}{2n_\alpha} \sum_{p\in\{l,r\}} \sum_a E(\mathbf{H}_a^p, \mathbf{y}_a^p). \tag{4}$$

2) ANATOMICAL ORIENTATION LOSS

Let the target prediction of the anatomical orientation output, $\mathbf{O}$, be a Kronecker-delta at the label orientation, o. Then, the anatomical orientation loss applies a periodic earth mover's distance between $\mathbf{O}$ and the target orientation, specifically:

$$\mathcal{L}_{\mathrm{ao}}(\mathcal{M}(\mathbf{V}), o) = \mathbf{O} \cdot \mathbf{d}_o^2, \tag{5}$$

where $\mathbf{d}_o$ is a vector with length n_α and the values are the shortest circular distance to the anatomical orientation label, o. Note that $\mathbf{d}_o$ contains distances normalized to $[0, 1]$ and that the loss uses the square distance. The anatomical orientation loss function is illustrated in Figure 4.

E. MODEL INFERENCE

During inference of heatmap regression models, the final coordinate predictions are calculated from the prediction heatmaps. We apply the weighted geometric median to calculate the mitral annulus coordinates and anatomical orientation predictions. The geometric median, also known as the Fermat-Weber point, is the point that minimizes the sum of distances, i.e., the $\mathcal{L}_1$ distance, to the set of sample points. The weighted geometric median assigns a weight contribution to each point.

1) OVERVIEW

The first step towards calculating the coordinate predictions is introducing a combined heatmap, $\mathbf{H}^c$. The combined heatmap takes advantage of the periodic symmetry of the samples (introduced in Section II-A) to weigh the contributions of the left and right heatmaps, as described below. Then, applying the geometric median to each plane yields the coordinate predictions. Next, calculating the anatomical orientation prediction utilizes the geometric median after projecting $\mathbf{O}$ to the unit circle. Figure 5 shows an example of $\mathbf{O}$ and a 3D rendering of $\mathbf{H}^c$ from a single sample, $\mathbf{V}$.

2) COMBINED COORDINATE PREDICTION HEATMAP

Due to the periodic mirror symmetry of $\mathbf{V}$ — described in Section II-A and illustrated in Figure 3 — the image plane $\mathbf{V}_{a+n_\alpha/2}$ is the mirror image of $\mathbf{V}_a$, for each a. This property enables combining the predictive power of the heatmaps $\mathbf{H}^l$ and $\mathbf{H}^r$ by introducing a shifted mirroring of $\mathbf{H}^r$. Let $\tilde{\mathbf{H}}^r$ be the 180° shifted mirroring of $\mathbf{H}^r$, i.e., $\tilde{\mathbf{H}}_a^r = \mathrm{Ref}(\mathbf{H}_{a+n_\alpha/2}^r)$, for each a, with $\mathrm{Ref}(\cdot)$ denoting a reflection across the depth centerline of each plane.

The target coordinate of $\mathbf{H}_a^l$ and $\mathbf{H}_a^r$ is $\mathbf{y}_a^l$ and $\mathbf{y}_a^r$, respectively, see (4). Due to the periodic mirror symmetry, the target coordinate of $\tilde{\mathbf{H}}^r$ is $\mathbf{y}_\alpha^l$. Consequently, the following combined prediction heatmap takes advantage of both the left and right heatmap predictions:

$$\mathbf{H}^c : \mathbf{H}_a^c = \mathbf{w}_a^l \cdot \mathbf{H}_a^l + \mathbf{w}_a^r \cdot \tilde{\mathbf{H}}_a^r, \tag{6}$$

for weights $\mathbf{w}^l$ and $\mathbf{w}^r$, calculated as described below.

3) COMBINED HEATMAP WEIGHTS

The aim of weighting the individual heatmap planes in (6) is to obtain a more robust combined heatmap — as one of the heatmaps may provide a good prediction while the other fails. A proxy for prediction confidence is applied when calculating the relative weighting, resulting in a natural prediction ensemble.

The angular moment of the prediction heatmaps provides one such confidence proxy. In particular, a highly focused heatmap with all energy centered around a point will yield a low angular moment, while a heatmap with more spread or multiple modes will result in a higher angular moment. The following choice of weights takes advantage of this: let $\mathbf{w}_a^l$ and $\mathbf{w}_a^r$ in (6) be inversely proportional to the squared angular

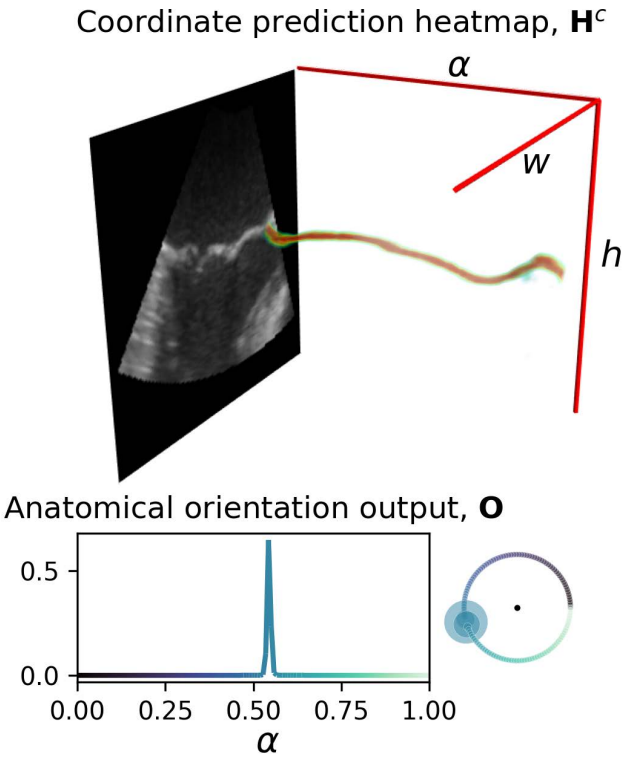

FIGURE 5. Example of model output. Top: Illustration of the heatmap prediction in the 3D cylinder coordinate system. A single plane, $\mathbf{V_0}$, is shown together with $\mathbf{H^c}$ throughout the volume. A logarithmic scale colormap is used for the coordinate prediction heatmap — with low voxel values set to be transparent. Note that the energy of the coordinate predictions are highly focused across the rotational planes in the shown example. An animation showing the predictions in all the rotational planes is available in the suplementary material (video 2). Bottom: Anatomical orientation output, **O**, from the above sample. The bottom right figure shows the values of **O** projected onto the unit circle (used to decode the anatomical orientation prediction, $\hat{o}$, as described in Section II-E) with scatter sizes proportional with the values of **O**. The color gradient (dark to light blue) highlights the point correspondence between the left and right anatomical orientation plots. The anatomical orientation prediction of the shown example is $\hat{o} = \mathcal{G}_{ao}(\mathbf{O}) = 194°$.

moment of the respective heatmap planes, normalized so that $\mathbf{w}_a^l + \mathbf{w}_a^r = 1$, for each a.

4) COORDINATE PREDICTIONS

The final coordinate predictions for a sample, $\mathbf{V}$, are calculated by applying the weighted geometric median to each rotational plane of the combined heatmap, $\mathbf{H}^c$. Each rotational plane's weighted geometric median decoding is:

$$\hat{\mathbf{y}}_a = \mathcal{G}_c(\mathbf{H}_a^c) = \arg\min_{\mathbf{x}} \sum_{i,j} \mathbf{H}_a^c \circ \mathcal{D}_{\mathbf{x}}, \tag{7}$$

using distance matrix $\mathcal{D}_x$ as introduced in Section II-D. Applying (7) to each rotational plane of $\mathbf{H}^c$ results in the coordinate predictions through the entire volume, $\hat{\mathbf{y}}$. Cartesian coordinate predictions are obtained by projecting the normalized cylinder coordinates onto Cartesian coordinates.

5) ANATOMICAL ORIENTATION

The estimate of the anatomical orientation of the heart is calculated from the model output, $\mathbf{O}$. The first decoding step is to project $\mathbf{O}$ to the unit circle, where each point $\mathbf{O}_a$ corresponds to the angle $a/n_\alpha \cdot 360°$ (see Section II-A) as illustrated in Figure 5. Then, the anatomical orientation prediction, $\hat{o}$, is obtained by applying the weighted geometric median to the points on the unit circle and calculating the resulting angle. Let the following notation denote this operation:

$$\hat{o} = \mathcal{G}_{ao}(\mathbf{O}). \tag{8}$$

The above approach for calculating the geometric median in a periodic domain is similar to Bai & Breen [35], except that they applied the center of mass.

F. OVERVIEW

Figure 6 shows an outline of the model inference of a trained model.

III. EXPERIMENT

A. RAW DATA ACQUISITION AND ANNOTATION

This research study was conducted retrospectively using fully anonymized echocardiography images. The DICOM data, annotations, and dataset split in this paper are the same as in [30].[2] Legal agreements with the data providers ensured compliance with local requirements for consent and secondary use.

The dataset covers a range of image quality and includes several mitral valve diseases — including Barlow's disease, fibroelastic deficiency, and functional mitral regurgitation. The exclusion criteria in [30] were acquisitions that generally would not be used for mitral valve analysis in clinical practice due to parts of the valve being outside the acquisition or acquisition with very low image quality.

The dataset consists of 111 midesophageal-level 4D TEE acquisitions of the mitral valve from 89 patient examinations acquired at three clinical sites. The acquisitions consist of a variable number of frames, totaling 700 3D frames. The data split ensured that volumes from the same examination only appear in one of the training, validation, or test sets to avoid bias. Table 1 shows the training, validation, and test sets data split.

Annotations were made using the commercially available medical software EchoPAC (GE Vingmed Ultrasound, Horten, Norway) and are the same as in [30] — where the annotations were generated by the first author, using 4D Auto MVQ, upon instructions from a trained cardiology expert. The annulus segmentation of 4D Auto MVQ uses multiple manually placed landmarks to make an initial segmentation. The annotator then manually edits the segmentation by modifying the points along the contour as needed.

[2]While the raw data used is the same, the generated samples used in this method are 3D cylinder coordinate volumes (Section II-A), in contrast to individual 2D planes used in [30].

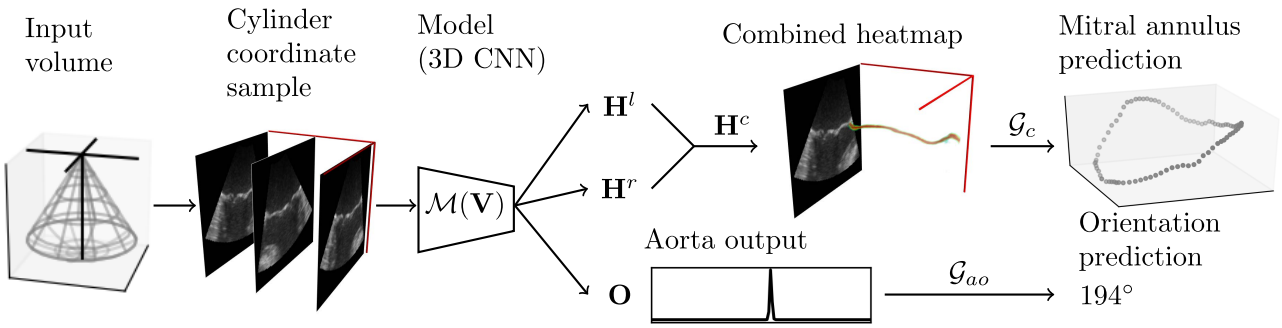

FIGURE 6. Illustration shows the inference pipeline of the method. Given a model, $\mathcal{M}$, trained as described in Section III-C, using the loss functions described in Section II-D. The model, $\mathcal{M}$, accepts a full 3D cylinder coordinate sample, $\mathbf{V}$, yielding $\mathcal{M}(\mathbf{V}) = \{\mathbf{H}^l, \mathbf{H}^r, \mathbf{O}\}$ (Section II-C). A combined heatmap, $\mathbf{H}^c$, is calculated from $\mathbf{H}^l$ and $\mathbf{H}^r$, from which the final coordinate predictions are calculated — using the geometric median (Section II-E). The anatomical orientation prediction is obtained by projecting $\mathbf{O}$ onto the unit circle, applying the geometric median, and calculating the angle of the result (Section II-E).

TABLE 1. Overview of the data used in the experiments. Some examinations consists of multiple DICOM files and each DICOM has several systolic frames. Five cylinder volume samples are created for each included DICOM frame in the training and validation sets — as described in Section III-B.

Dataset	Exams (DICOM files)	Systole 3D frames	Cylinder coordinate samples
Training	55 (74)	459	2295
Validation	15 (18)	106	530
Test	19 (19)	135	135

Using annotations generated with 4D Auto MVQ has a significant benefit compared to manual annotations in individual points along cross-sections of the mitral valve: When editing the initial segmentation, 4D Auto MVQ provides the annotator with full 3D context and the capability to switch between frames for additional temporal context. Furthermore, when correcting the curve segmentation in a point, 4D Auto MVQ updates nearby points on the curve to ensure the segmented mitral annulus is smooth. In contrast, it is challenging to segment a smooth 3D curve of the mitral annulus manually, as it requires annotations in several individual slices through the volume.

The raw annotations consist of 58 Cartesian coordinates for each DICOM time frame — delineating the mitral valve annulus — and a single Cartesian landmark coordinate centered at the aortic valve. [3]

The DICOM images used in this paper were acquired on GE Vivid E9 and GE Vivid E95 scanners (GE Vingmed Ultrasound, Horten, Norway). Only systolic time frames are included in the data sets, as 4D Auto MVQ generates labels for the systolic phase.

B. DATASET GENERATION

The samples for the training, validation, and test sets were created as described in Sections II-A and II-B. Each sample consists of the cylinder coordinate volume, left and right mitral annulus labels, and the anatomical orientation label, i.e., $\{\mathbf{V}, \mathbf{y}^l, \mathbf{y}^r, o\}$.

All planes were cropped to a spatial resolution of 80×80 mm height and width, and the sample dimensions (n_α, n_h, n_w) were set to (128, 128, 128), i.e., 128 rotational planes, each with an image resolution of (128, 128).

1) TRAINING AND VALIDATION SET

Multiple samples are generated for each DICOM frame in the training and validation sets. The purpose of generating multiple samples from each frame is to introduce augmented samples to reduce the likelihood of overfitting and increase the model's robustness to different orientations of the anatomical features. Recall that each cylinder coordinate sample is generated from an initial plane (see Section II-A). This initial plane was obtained by applying a sequence of geometric transforms to the 2D plane at zero degrees elevation in the volume. The geometric transforms and value range are given in Table 2.

Due to the rotations and translations, different samples from the same frame result in different anatomical orientations. Five such *view-augmented* samples are generated for each DICOM in both the training and validation sets, as reflected in the rightmost column of Table 1. Note that each sample is calculated from the respective DICOM acquisition, not by interpolating a Cartesian voxel grid.

2) TEST SET

A single sample was generated for each DICOM frame in the test set. The initial plane for these cylinder coordinate samples was the plane at 0° elevation in the DICOM volume.

3) SAMPLE REQUIREMENTS

Two pre-requisites were made for the samples included in the experiment:

1) the axis of rotation (centerline of the base plane) must be inside the mitral annulus,
2) the mitral annulus coordinates are not closer to the top or bottom of the volumes than 10% (8mm).

[3] Other landmark points are also placed in 4D Auto MVQ, but these are not used by this method.

TABLE 2. Overview of transforms used to create view-augmented samples in the training and validation set, as discussed in Section III-B. The three transforms are applied sequentially, with a random value in the listed range. Cylinder samples are generated from the translated planes, as described in Section II-A.

Transform	Axis/Magnitude	Value range
Rotation	Tilting plane	$[-5°, -5°]$
Rotation	Centerline (depth dimension)	$[0°, 360°]$
Translation	Random direction	$[0\text{mm} - 5\text{mm}]$

TABLE 3. Overview of parameters in the final model runs. The values were selected based on preliminary experiments, as outlined in Section III-C.

Run	lr	lr-decay	Batch size
A	$1e-2$	1/2	4
B	$5e-2$	2/3	3
C	$5e-3$	2/3	3

All test set samples fulfilled these pre-requisites, while a few augmented geometries in the training and validation set were translated before creating the cylinder sample volume to comply.

4) FINAL DATASET

The rightmost column of Table 1 shows the final number of samples in the training, validation, and test set.

C. MODEL ARCHITECTURE AND TRAINING

The HighRes3DNet model architecture proposed by Li *et al.* [36] is used in the experiments of this paper — modified to apply circular convolutions in the rotational dimension. Most parameters used in the experiment were selected based on preliminary model training. The final experiment consisted of training three models using the pre-selected parameters, with only the minor differences reported in Table 3. The choice to train a small number of models was informed by the required machine resources — with approximately 60 hours to train each model. As such, the reported experiment aims to show the feasibility and effectiveness of the proposed method without applying an extensive hyperparameter search.

1) MODEL/TRAINING PARAMETERS

A benefit of the HighRes3DNet architecture is that it requires a comparatively low number of model parameters [36]. Several model sizes were tested in preliminary experiments. The models in the final experiment all use the model configuration as presented in [36], resulting in a model with approximately 0.8M trainable parameters. The constant, λ, for balancing the two losses in (2) was set to $\lambda = 0.7$, based on results in the preliminary experiments.

Each model was trained for 200 epochs with an epoch size of 100. Small batch sizes were used (see Table 3) due to the large memory requirements during model training, and instance normalization [37] was applied. Model training applied a learning rate reduction on validation loss plateaus with a ten epoch patience — with the initial learning rates and decay factors specified in Table 3. The training, validation, and test set described in Section III-B was used for the experiments.

2) TRAINING AND VALIDATION

A random subset of the training set was sampled for each epoch — applying weighted sample frequencies to obtain a balanced sampling from the examinations. After each epoch, weighted validation set statistics were calculated for the entire validation set to monitor the performance and select a final model.

3) TEST SET INFERENCE

The test set was kept separate until a final model had been selected. Inference on the test set was only done using the final model. The criterion to select the final model checkpoint was the lowest in-plane coordinate prediction error on the validation set.

4) IMPLEMENTATION AND HARDWARE

The PyTorch implementation[4] of the HighRes3DNet model was modified to use circular convolutions in the rotational dimension and zero-padding in the remaining dimensions.

Models were trained using three/four (one GPU per sample in the batch, see Table 3) Nvidia Quadro P6000 GPUs on a machine with an Intel Xeon E5-2620 CPU.

D. RESULT METRICS

The coordinate prediction error metrics reported in the paper are:

1) *In-plane error:* prediction error measured for each rotational plane of the cylinder coordinate samples (absolute distance)
2) *Curve-to-curve distance error:* average shortest error between the prediction and the label curves
3) *Surgical view angle*: average error in degrees between a plane fitted to the prediction and the plane fitted to the label curve coordinates
4) *Perimeter:* Relative perimeter error.

Illustrations and further details about the in-plane, curve-to-curve, and surgical view angle error metrics are given in Appendix A.

The anatomical orientation prediction errors are reported in degrees and plane indices, i.e., *n* planes out of 128 planes. Both mean values with standard deviation and median are reported for the anatomical orientation error.

All reported means and standard deviations are weighted per examination to take into account that different examinations have a different number of frames.

IV. RESULTS

The test set results of the selected model obtained a weighted mean error of 1.96 ± 1.62 mm for the coordinate predictions. The weighted curve-to-curve prediction error was

[4]HighRes3DNet implementation: https://github.com/fepegar/highresnet

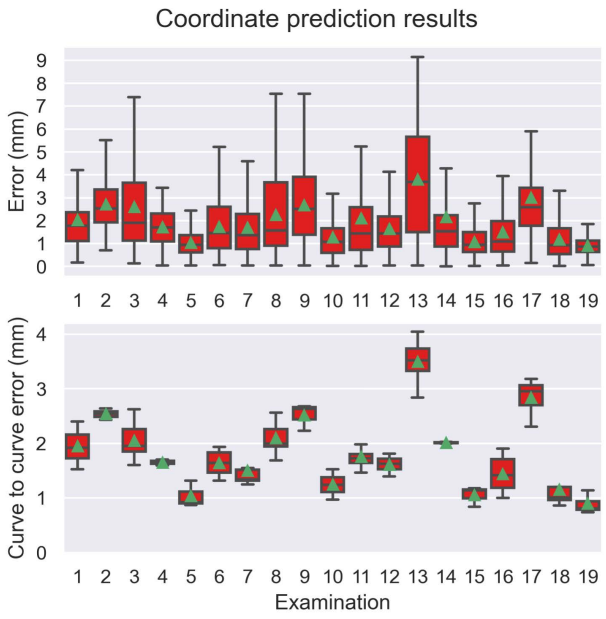

FIGURE 7. Box plots showing the coordinate prediction errors across the 19 examinations in the test set. The error metrics are introduced in Section III-D and the average across the entire test set is presented in Section IV. Top: In-plane errors. Predictions for all individual planes and time frames contribute to the metric. Bottom: Curve-to-curve errors. Each frame yields a single error. Plot settings: Green triangles show mean error for each examination. Boxes show median error and quartiles. Whiskers show the 1.5 times the interquartile range. Outliers are not shown in the plots.

1.82 ± 0.70 mm. Figure 7 shows a box plot of the coordinate errors across the examinations in the test set.

Further, the model obtained a surgical view prediction error of $3.28 \pm 2.92°$ and a relative perimeter error of $5.8 \pm 4.8\%$. Table 4 presents the coordinate prediction metrics introduced in Section III-D with the respective comparable results from [25], [29], [30], and [31].

The anatomical orientation predictions on the test set obtained a weighted mean error of 9.7 ± 15.8 degrees (3.5 ± 5.6 plane indices of the 128 planes) and a median prediction error of 5.6 degrees (2 plane indices). Table 5 presents the anatomical orientation results, and Figure 8 shows the result for each examination as a box plot.

V. DISCUSSION

A. COMPARISON WITH EXISTING METHODS

The results from our experiment indicate that the mitral annulus segmentation results from our proposed method are on par with the existing literature and in the range of interobserver variability.

To estimate the interobserver variability of the mitral annulus annotations, we compared the test set labels against separate annotations — made by an echocardiography expert using 4D Auto MVQ on the mid-systolic frame of the test set examinations. The mean distance between the annotations was 2.20 ± 0.77 mm (in-plane error). This is in line with the results from Schneider *et al.* [25] — who reported an

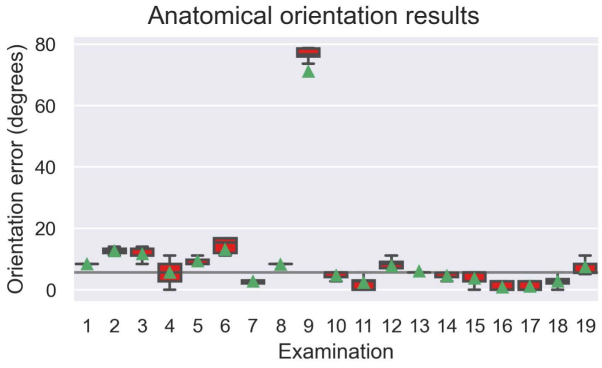

FIGURE 8. Box plot showing the anatomical orientation prediction results for each examination in the test set — as introduced in Section IV. All frames except for examination 9 have an anatomical orientation error of less than 17°. The horizontal line shows the median prediction (see Table 5). The failing anatomical orientation prediction for examination 9 is investigated and discussed further in Section V-B. Plot settings: See Figure 7.

average Mahalanobis distance of 1.63 ± 0.76 mm, measured between ten experts annotators across ten volumes. Our estimate between two annotators seems reasonable since the Mahalanobis distance in [25] measured the normalized distance from each expert to the average of the other nine experts.

As described in Section III-A, 4D Auto MVQ requires manual user input. The same applies to several other tools in clinical use and the methods presented in [25] and [29]. Compared to the above methods, a benefit of our work is that no such manual initialization is required. Another benefit of our method stems from the capacity of the 3D CNN: More data used during model training can generally be expected to result in increased robustness and better generalization capability — as the performance of CNN methods generally improves with more training set samples [38].

The most direct comparison can be made with our previous work [30], as the same DICOM dataset is used. While a paired t-test indicates that the average coordinate predictions per examination are statistically indistinguishable for the two methods, the results are in the range of interobserver variability — as discussed above. Our proposed method has several benefits over [30]: It uses full 3D context during inference (in contrast to a separate post-processing algorithm to combine 2D predictions in [30]), it ensembles coordinate predictions proportional with angular moment (stabilizing model inference), and uses geometric median to calculate the prediction coordinates (more robust to heatmap outliers in $\mathbf{H}^c$). Figure 11 in Section V-C illustrates prediction improvements on two of the worst results in [30]. Further, fewer hyperparameters need to be manually set — related to both the model and to the model training — compared to [30]. Finally, our proposed method has the added benefit of predicting the global anatomical orientation of the valve, which for instance can be used to automate view selection.

Zhang *et al.* [31] report a surgical view prediction error that is significantly larger than our experiment. However,

TABLE 4. Coordinate prediction results, presented with results from [25], [29], [31], and [30]. An overview of the metrics is given in Section III-D. Note: †: Metrics calculated from final results in [30], see Appendix B for details. ‡: In [25], the standard deviation is calculated across the average prediction of each sample (the corresponding result for our experiment is a standard deviation of 0.80 mm).

Measurement error [unit]	Schneider et al. [25]	Tiwari and Patwardhan [29]	Zhang et al. [31]	Andreassen et al. [30]	Proposed (this)
Annulus in-plane [*mm*]	$1.81 \pm 0.78^{\ddagger}$	2.59	1.57	$2.04 \pm 1.87^{\dagger}$	1.96 ± 1.62
Annulus curve-to-curve [*mm*]	–	–	3.49 ± 2.21	$1.94 \pm 0.82^{\dagger}$	1.82 ± 0.70
Surgical view [*degrees*]	–	–	9.62 ± 10.46	$3.26 \pm 2.26^{\dagger}$	3.28 ± 2.92
Perimeter [%]	–	–	10 ± 16	6.1 ± 4.5	5.8 ± 4.8
Number of test set volumes	10	15	432	135	135

TABLE 5. Anatomical orientation prediction results. The results are given both in terms of degrees rotation (first row) and error in terms of number of plane indices, e.g., out of the 128 rotational planes (second row).

	Anatomical orientation errors	
	Weighted mean	Median
Degrees	$9.7° \pm 15.8°$	$5.6°$
Planes	3.5 ± 5.6 planes	2 planes

as they use three points to calculate each plane, their plane estimates are more susceptible to small coordinate prediction errors. The main metric for mitral annulus predictions in [31] is the curve-to-curve metric. In our experiment, only three of 135 volumes (from examination 13) have a larger curve-to-curve error than 3.5 mm — the reported average prediction error in [31]. Zhang *et al.* [31] reports a 1.57 mm in-plane error as a comparable result to the reported in-plane error of [30]; however, it is unclear how the in-plane error was calculated, as this result is only mentioned in a footnote. All volumes in our test set have a lower curve-to-curve error than average in-plane error. Further, [31] fits a spline to seven points predicted by their method. Their reported error metric for five of these points (projection distance from the point to the annotated annulus) corresponds closely to the in-plane error metric applied in this work and [30]. These points (in particular '*P*' and '*Aux Lmks*' in Table 1 of [31]) have an average prediction error higher than 3.5 mm. Our understanding is that the 1.57 mm in-plane error reported in [31] implies that their spline prediction is significantly better than their five explicitly predicted landmark points.

B. ANATOMICAL ORIENTATION PREDICTIONS

There are several benefits to predicting the orientation of the mitral valve. The delineation of the mitral valve and the prediction of the center of the left ventricle outflow tract fully determine the mitral valve's anatomical position in the volume. This can be used for automatic view selection, e.g., the 3D surgical view (see Figure 1) and the midesophageal long-axis view. The combined mitral valve segmentation and orientation prediction could also be used to automate leaflet segmentation methods that require manual interaction when used in isolation — for instance, the mitral leaflet segmentation of 4D Auto MVQ.

The anatomical orientation prediction is successful for most samples in our experiment. Specifically, all time frames for 18 out of the 19 test set examinations have an anatomical orientation prediction error lower than 17° (six plane indices); see Figure 8.

However, for examination 9, the anatomical orientation error is significant for most frames. Investigating the anatomical orientation output, $\mathbf{O}$, reveals that the prediction contains two modes at different angles — one close to the correct orientation and the other approximately 80° away, where the ventricular wall is outside the field of view, as shown in Figure 9.

Detecting predictions with multiple modes — as the example shown in Figure 9 — could be used to flag predictions for manual inspection or post-processing. Figure 10 shows the feasibility of this approach by plotting a measure of the prediction spread of $\mathbf{O}$ against the prediction error. As the domain is periodic, the prediction spread in the figure uses (5); however, with the prediction, $\hat{o}$ from (8) substituted as the orientation label — specifically $\mathbf{O} \cdot d_{\hat{o}}^2$. Figure 10 shows that the samples from examination 9 all have a large prediction spread. Most other samples have a low prediction spread and low prediction error, while a few samples have a larger spread but still make good predictions as the largest mode of $\mathbf{O}$ is close to the correct orientation. A large prediction spread does not imply a prediction error but possibly lower confidence.

C. MITRAL ANNULUS COORDINATE PREDICTIONS

Our presented method performs the mitral annulus coordinate predictions well throughout most samples in the test set. As discussed in Section V-A, the mean results are in the range of interobserver variability. Notably, the method performs well in cases where our previous method [30] failed, e.g., in the area around the aortic outflow tract of examinations 3 and 6, as shown in Figure 11. These improvements are likely due to our proposed method's additional 3D spatial context.

Overall, most test set samples yielded a highly focused prediction heatmap. Using the geometric median (7) also provides stability with respect heatmap noise (see Section V-E). However, a small number of prediction errors were caused by prediction heatmaps with multiple modes. This error type could be detected by evaluating heatmap spread — similar to detecting multiple modes in the orientation predictions (Section V-B).

Examination 13 has the highest average prediction error in the test set. Figure 12 shows the prediction error along the rotational dimension and the prediction in three rotational

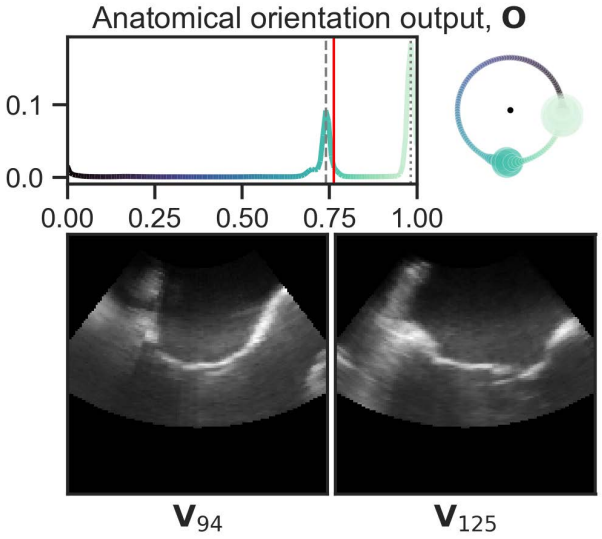

FIGURE 9. Anatomical orientation prediction, **O**, for a frame from examination 9. Top: Model output, **O**, along the normalized rotational axis [0, 1]. As mentioned in Section V-B, the model outputs two prediction modes for this examination. Red vertical line shows the orientation of the label, o. Gray vertical lines correspond to the rotational planes shown below, at the peaks of the two modes. Subfigure to the right shows **O** projected to the unit circle, as introduced in Section II-E and as shown in Figure 5. Bottom left: Rotational plane $\mathbf{V}_{94}$ (dashed gray line), with the aortic outflow tract visible. Bottom right: Rotational plane $\mathbf{V}_{125}$ (dotted gray line), that corresponds to the largest mode of **O**. Note that the ventricular valve is outside the field of view in this plane, making it look very similar to a plane with the aortic outflow tract.

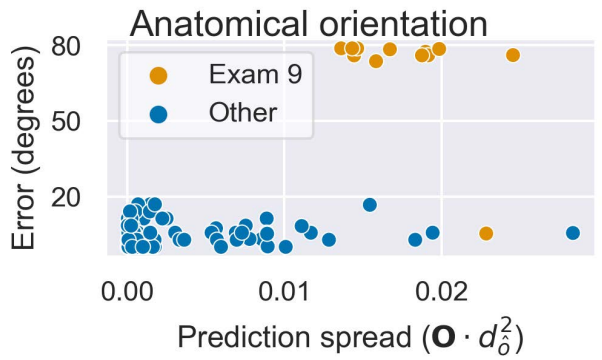

FIGURE 10. Scatter plot showing anatomical orientation prediction spread against anatomical orientation prediction error for the samples in the test set. The prediction spread highlights one way for detecting multiple modes in **O**, that could be used to trigger user interaction or initiate prediction post-processing, discussed in Section V-B. Colors: The orange dots highlight the results from examination 9, which yields the highest prediction errors (see Figure 7). Note that a single frame from examination 9 has a low prediction error, as the mode around the label orientation was slightly larger in this frame; thus, the geometric median decoding results ended up on the correct mode. The blue dots show the results of all frames from the other examinations. Prediction spread: The x-axis applies anatomical orientation loss (5), with the predicted anatomical orientation, $\hat{o} = \mathcal{G}_{ao}(\mathbf{O})$ from (8), in place of the label.

planes from one examination 13 sample. The mitral leaflet tissue is thick in the depicted region, and it is difficult to discern the mitral annulus delineation without temporal context.

In the case of examination 17, the annotations are several millimeters onto the mitral leaflets in a region of the volume.

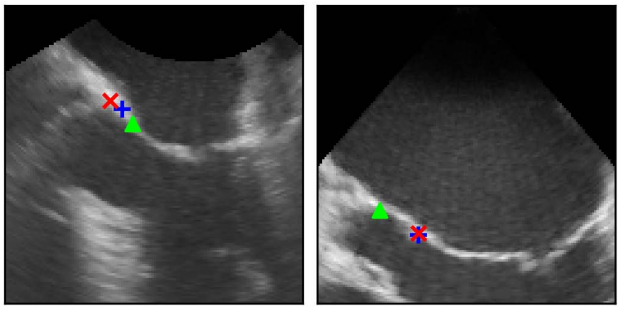

FIGURE 11. Illustration of results in the aortic outflow region of examination 3 and 6. Red x: label coordinate, blue +: prediction (this method), green triangle: prediction from [30]. The examples illustrate the importance of the 3D context of our proposed method, as these planes were highlighted in [30] as challenging, with the largest prediction errors of these examinations. The 2D CNN used in [30] detected the transition to thinner tissue on the mitral leaflet as a point on the annulus (examination 3, left subplot), and predicted the annulus point to be approximately 1 cm beyond the open aortic valve (examination 6, right subplot). The post-processing algorithm in [30] did not correct these predictions. For our new method, the predictions around the depicted area were consistent throughout the time frames of the examinations.

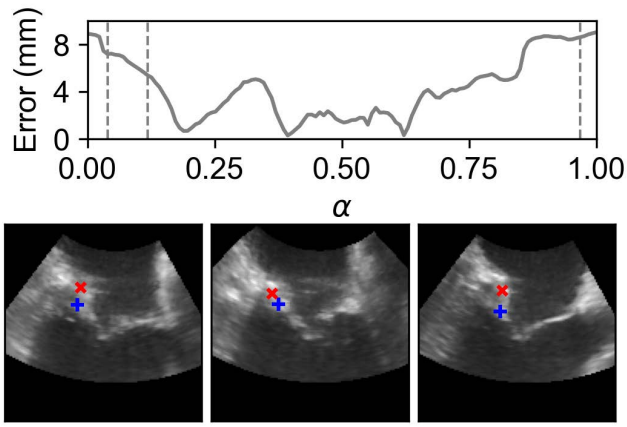

FIGURE 12. Illustration of the in-plane prediction error of a single time frame from examination 13 — the examination with the largest average prediction error, see Figure 7. Top: The error in millimeters, along the normalized rotational axis, α. The three vertical lines correspond to the three image planes visualized in the bottom row, in order from left to right. There is approximately thirty degrees rotation between each depicted plane. Bottom: Three image planes correspond to the highlighted angles, as described above. The red x corresponds to the label location, while the blue + corresponds to the prediction. Segmenting the mitral valve in the shown region of this acquisition can be challenging without temporal information.

The prediction accuracy in this region is good, and the main contribution to the prediction error for this sample stems from the annotation error in this region. We address the limitation related to annotations in Section V-G.

Using the same samples in the test set as [30] enables direct comparison, as illustrated in Figure 11. A possible downside is the potential to introduce adaptive overfitting, i.e., overfitting caused by reusing the test set. However, we do not expect this to be a problem in our study, as all improvements proposed in this paper are generic and sample independent (e.g., introducing 3D context), and we applied a

strict policy of selecting the final model before running any inference on the test set (see Section III-C).

D. LOSS FUNCTION

The proposed earth mover's coordinate loss function (4) conditions the network towards a single label point without the need to tune or select any loss function parameters. In contrast, both the $\mathbf{L}^2$ matrix loss (applied in [31]) and the divergence loss (applied in [30]) use a Gaussian template map with standard deviation, σ, as a hyperparameter. An additional benefit of the proposed earth mover's loss function is that it increases with the distance to the label, as illustrated in Figure 4, penalizing large errors more than small ones. The $\mathbf{L}^2$ and the divergence-based loss functions are pointwise differences that do not encode the distance to the target coordinate. The periodic loss function (5), used to optimize the anatomical orientation predictions, has similar properties as the coordinate loss.

Yan *et al.* [39] recently proposed a landmark detection loss function based on the earth mover's distance. Their proposed loss function is similar to (4); however, they employ a Gaussian distribution around the label coordinate as the target distribution of the earth mover's distance. Consequently, their loss does not simplify into a closed-form equation, and the resulting computational requirement is significantly larger than our proposed loss function. The benefit of (4) is twofold: It removes the σ as a parameter that needs to be selected and simplifies the computation of the earth mover's function to an exact closed-form equation.

E. INFERENCE USING GEOMETRIC MEDIAN

The final inference step applies the geometric median to compute the coordinates and orientation predictions from $\mathbf{H}^c$ and $\mathbf{O}$, respectively. The geometric median is known to be a stable estimator that is robust to outliers [40], with a breakdown point shown to be 0.5 [41].

To the best of our knowledge, the geometric median has not previously been applied to calculate coordinates for heatmap regression models. The common approach is to apply the center of mass or 'arg max' to calculate coordinates in heatmap regression models; however, the center of mass is sensitive to multiple prediction modes, and 'arg max' is sensitive to prediction noise. However, a downside of using the geometric median is that it does not have a closed-form solution.

F. INFERENCE TIME

The inference time of a single 3D volume is 1.8 ± 0.04 seconds (across 100 runs on a machine with an Intel Xeon CPU E5-2620 and a single Nvidia Quadro P6000 GPU).

We carried out a subsequent experiment to get an indication of whether our proposed method can be applied in near real-time applications: A single model training of a small HighRes3dNet model using lower resolution samples. The inference time of the resulting model is 0.13 ± 0.01 seconds — with average coordinate prediction errors within 10%, and average orientation errors within 15% of the main results. See Appendix C for details.

Different applications would give different weights to the trade-off between the accuracy and the speed of the method. On the one hand, the near real-time inference described above could provide the operator with automated, low delay assistance for aligning the probe and centering the mitral valve in the acquisition. On the other hand, measurement accuracy outweighs inference time when doing clinical measurements of the mitral valve.

Note that the code was not optimized with respect to speed and there is likely a potential to optimize model inference time.

G. EXPERIMENT LIMITATIONS AND FUTURE WORK

While the results presented in this paper demonstrate our proposed method's technical merit and feasibility, we acknowledge that the total number of acquisitions is a limitation of the study.

The training set consists of 74 acquisitions from 55 patients; however, by using multiple time frames and applying view augmentation, our test set had more than two thousand volumes. Further, as discussed in Section III-A, the dataset covers a range of image quality and mitral valve diseases. Nevertheless, a larger training set would increase the range of anatomical variation that the model learns from, likely improving its capacity to generalize. While the test set size in our experiment is of the same magnitude as many comparable studies, a larger number of test set acquisitions would be beneficial to provide more information about how well the method generalizes. Another limitation is that the annotations were not created by clinical experts, see Section III-A.

A clinical study would address the above limitations and would be necessary for evaluating our proposed method for clinical application. The clinical study should include several expert annotators and evaluate the performance of the method against interobserver and intraobserver variability on the same test set acquisitions. This study could also include a comparison of other clinically certified methods. We consider such a clinical study as important future work.

After clinical validation, the proposed method could be integrated into medical software. The method could replace semi-automated methods, e.g., the initial segmentation of 4D Auto MVQ. This would still allow the user to modify the resulting segmentation if needed. This workflow would save time, and the automatic predictions can be a step toward standardized segmentation results and hopefully be of use for clinicians in training.

VI. CONCLUSION

We have proposed a robust and automatic 3D method for segmentation of the mitral annulus and predicting the anatomical orientation of the valve from TEE images. The results are state of the art, with an average mitral annulus

prediction error of less than 2 mm, and an average anatomical orientation prediction less than ten degrees.

The proposed earth mover's distance loss functions provide landmark detection losses without the need to tune or select any hyperparameters. Further, the ensemble weighting of multiple prediction heatmaps — using the angular moment as a confidence measure — provides a simple approach for combining multiple prediction heatmaps. Finally, applying the geometric median to calculate coordinate predictions from heatmaps provides an estimator that is robust to noise and secondary modes.

APPENDIX A
ERROR METRICS

In-plane error: The in-plane error metric calculates the prediction error for each plane as the Euclidean distance between the label and prediction, as illustrated in Figure 13 (a). In the reported experiments of this paper, each sample has 128 error measurements, as 128 rotational planes were used.

Curve-to-curve error: The *curve-to-curve* error metric — as applied in Zhang *et al.* [31] — measure the distance between two curves as the average distance from each point on one curve to the closest point on the other curve. Given a prediction curve D (detection) and label curve G (ground truth), the directed distance from the prediction curve to the label curve is:

$$\mathrm{E}_{D,G} = \mathbb{E}_{d \in D} \inf_{g \in G} \|d - g\|_2.$$

Above, the shortest distance to G is calculated for each point d on D, as shown in Figure 13 (b). Note that distance $\mathrm{E}_{D,G}$ and $\mathrm{E}_{G,D}$ generally are not equal.

The *curve-to-curve* metric is defined as the average of the distance $\mathrm{E}_{D,G}$ and $\mathrm{E}_{G,D}$:

$$d(D, G) = \frac{1}{2}\Big(\mathrm{E}_{D,G} + \mathrm{E}_{G,D}\Big), \tag{9}$$

as illustrated in Figure 13 (b).

Surgical view angle: The surgical view angle metric measures how well the surgical view can be set by the mitral annulus prediction. A plane is fitted to the prediction curve and a separate plane to the label curve, using singular value decomposition. Our surgical view angle metric is the angle between these two planes, measured by the angle between their normal vectors. Figure 14 shows an illustration of the metric.

APPENDIX B
RESULT PRECISION AND ADDITIONAL RESULT METRICS FROM PREVIOUS WORK

To enable comparing the results of this paper with the results of our previous work [30], it was required to calculate additional metrics from the final predictions of [30]. Specifically, we calculated the curve-to-curve errror and the surgical view error, as reported in Table 4.

Additionally, in-plane errors were reported using one decimal in [30]. To allow comparison with the results of this paper and with the results of Zhang *et al.* [31], the in-plane errors from [30] were re-calculated and are reported with two decimal precision in Table 4.

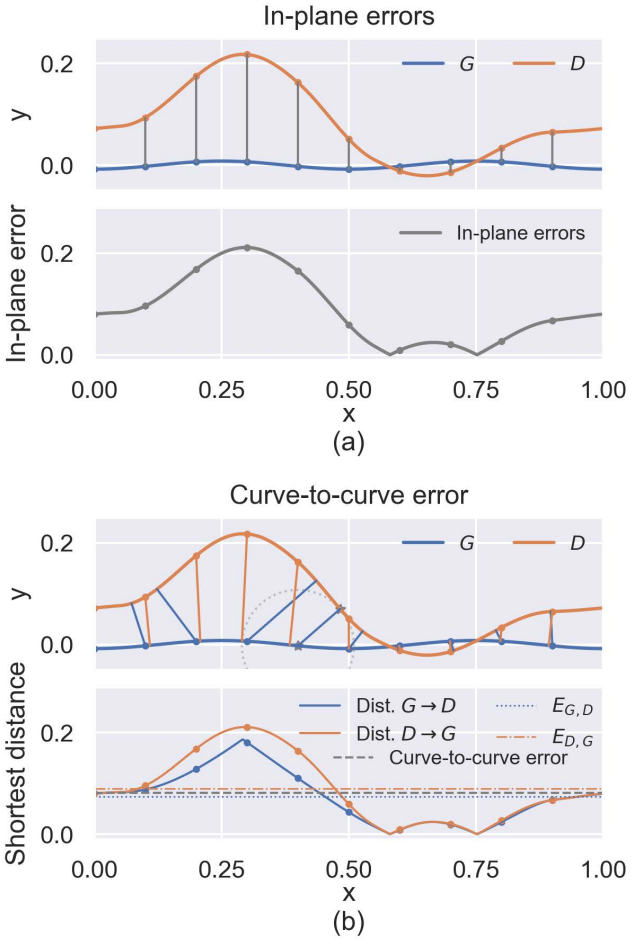

FIGURE 13. Illustration of the coordinate prediction error metrics introduced in Section III-D. In the synthetic two-dimensional example, G is the label (ground truth) and D is the prediction (detection). (a) In-plane error: Top: Label (blue) and prediction (orange) curves. For illustration purpose, the error contributions are visualized as grey lines at evenly distributed intervals on the curves. Bottom: The in-plane error distance visualized for all points along the curve, with the n-th point is the Euclidean distance between the n-th points on the label and prediction curves. The shown example has an in-plan error of 0.10 ± 0.07. (b) Curve-to-curve error: The curve-to-curve error metric (9) yields a single number per 3D sample. Top: The same synthetic label (blue) and prediction (orange) as in (a). The distance to the other curve is highlighted with the same interval as in (a). Straight lines between the curves illustrate the distance from G to D (blue lines) and from D to G (orange lines). The dotted circle (gray) highlights the shortest distance from G to D at $x = 0.4$, with endpoints marked as gray stars. Note that the distance is not symmetric. Bottom: The curve-to-curve error is shown as the gray dashed line, and is the mean value of $\mathrm{E}_{D,G}$ and $\mathrm{E}_{G,D}$. The solid lines shows the directed distance between the curves (orange from D to G, and blue from G to D). The scatter points correspond to the length of the straight lines in the above plot. The shown example has a curve-to-curve error of 0.09.

APPENDIX C
INFERENCE TIME EXPERIMENT

Details about the subsequent inference time experiment described in Section V-F are outlined below.

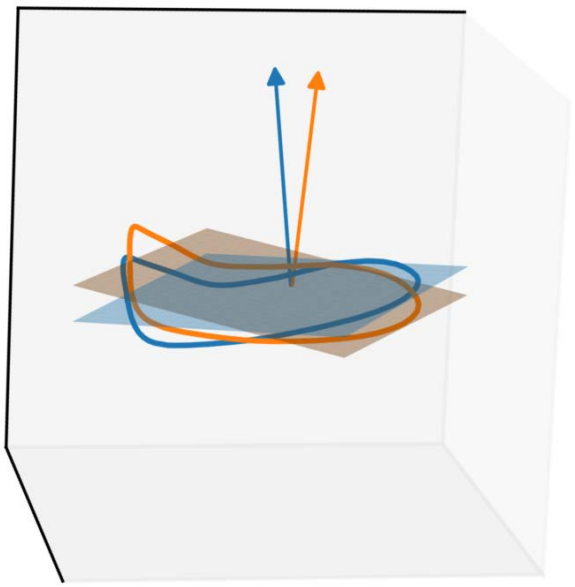

FIGURE 14. **Illustration of the surgical view error metric — introduced in Section III-D. The error metric measures the angle between fitted planes, as the angle between their normal vectors (shown as arrows). The planes are calculated using the singular value decomposition of each curve, (shown as plane segments in the figure). The illustrated curves shows the delineation of a mitral annulus curve together with a rotated copy of the same curve. The surgical view error of the shown example is 11.2°.**

Dataset: Training, validation, and test datasets for the experiment was generated as described in Section III-B, except with sample dimensions (n_α, n_h, n_w) set to (64, 64, 64).

Model and training: The model used for the experiment was a small HighRes3dNet model with approximately thirty-two thousand parameters (four channels in the first block and two residual blocks per dilation block, otherwise the same as the main experiment). A single model was trained and a single model checkpoint with good validation set results was run on the test set.

Results: The result of the single model training was: In-plane error: 2.1 ± 1.6 mm, curve-to-curve error: 2.0 ± 0.7 mm, anatomical orientation error: $11.1° \pm 12.4°$.

Inference time: The full inference time with the above model configuration was 0.13 seconds — from which the 3D CNN inference took 0.05 seconds. The calculations after the CNN (i.e., (6), (7), and (8)) could be optimized to reduce the total inference time.

ACKNOWLEDGMENT

The authors would like to thank Andrew Gilbert and Sarina Thomas for the valuable discussions and input to the article.

REFERENCES

[1] R. T. Hahn, T. Abraham, M. S. Adams, C. J. Bruce, K. E. Glas, R. M. Lang, S. T. Reeves, J. S. Shanewise, S. C. Siu, W. Stewart, and M. H. Picard, "Guidelines for performing a comprehensive transesophageal echocardiographic examination: Recommendations from the American society of echocardiography and the society of cardiovascular anesthesiologists," *J. Amer. Soc. Echocardiography*, vol. 26, no. 9, pp. 921–964, Sep. 2013.

[2] M. Garbi and M. J. Monaghan, "Quantitative mitral valve anatomy and pathology," *Echo Res. Pract.*, vol. 2, no. 3, pp. R63–R72, Sep. 2015.

[3] B. Iung and A. Vahanian, "Epidemiology of acquired valvular heart disease," *Can. J. Cardiol.*, vol. 30, no. 9, pp. 962–970, Sep. 2014.

[4] V. T. Nkomo, J. M. Gardin, T. N. Skelton, J. S. Gottdiener, C. G. Scott, and M. Enriquez-Sarano, "Burden of valvular heart diseases: A population-based study," *Lancet*, vol. 368, pp. 1005–1011, Sep. 2006.

[5] Y. Chandrashekhar, S. Westaby, and J. Narula, "Mitral stenosis," *Lancet*, vol. 374, no. 9697, pp. 1271–1283, 2009.

[6] B. Iung *et al.*, "Contemporary presentation and management of valvular heart disease: The EURObservational research programme valvular heart disease II survey," *Circulation*, vol. 140, pp. 1156–1169, Jan. 2019.

[7] M. Enriquez-Sarano, C. W. Akins, and A. Vahanian, "Mitral regurgitation," *Lancet*, vol. 373, no. 9672, pp. 1382–1394, 2009.

[8] A. Vahanian *et al.*, "2021 ESC/EACTS guidelines for the management of valvular heart disease: Developed by the task force for the management of valvular heart disease of the European society of cardiology (ESC) and the European association for cardio-thoracic surgery (EACTS)," *Eur. Heart J.*, vol. 43, pp. 561–632, Aug. 2021.

[9] C. M. Otto *et al.*, "2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American college of cardiology/American heart association joint committee on clinical practice guidelines," *J. Amer. College Cardiol.*, vol. 77, no. 4, pp. e25–e197, 2021.

[10] S. Robinson, L. Ring, D. X. Augustine, S. Rekhraj, D. Oxborough, A. Harkness, P. Lancellotti, and B. Rana, "The assessment of mitral valve disease: A guideline from the British society of echocardiography," *Echo Res. Pract.*, vol. 8, no. 1, pp. G87–G136, Mar. 2021.

[11] J. Weese and C. Lorenz, "Four challenges in medical image analysis from an industrial perspective," *Med. Image Anal.*, vol. 33, pp. 1339–1351, Oct. 2016.

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

[13] S. Gandhi, W. Mosleh, J. Shen, and C.-M. Chow, "Automation, machine learning, and artificial intelligence in echocardiography: A brave new world," *Echocardiography*, vol. 35, no. 9, pp. 1402–1418, 2018.

[14] M. Sermesant, H. Delingette, H. Cochet, P. Jaïs, and N. Ayache, "Applications of artificial intelligence in cardiovascular imaging," *Nature Rev. Cardiol.*, vol. 18, pp. 600–609, Mar. 2021.

[15] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: A survey," *Med. Image Anal.*, vol. 1, no. 2, pp. 91–108, 1996.

[16] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, Aug. 2009.

[17] B. Erickson, P. Korfiatis, Z. Akkus, and T. Kline, "Machine learning for medical imaging," *RadioGraphics*, vol. 37, no. 2, pp. 505–515, Mar. 2017.

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

[19] S. Leclerc, E. Smistad, J. Pedrosa, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, J. Dhooge, L. Lovstakken, and O. Bernard, "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019.

[20] E. Smistad, A. Ostvik, I. M. Salte, D. Melichova, T. M. Nguyen, K. Haugaa, H. Brunvand, T. Edvardsen, S. Leclerc, O. Bernard, B. Grenne, and L. Lovstakken, "Real-time automatic ejection fraction and foreshortening detection using deep learning," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 12, pp. 2595–2604, Dec. 2020.

[21] A. M. Pouch, H. Wang, M. Takabe, B. M. Jackson, J. H. Gorman, R. C. Gorman, P. A. Yushkevich, and C. M. Sehgal, "Fully automatic segmentation of the mitral leaflets in 3D transesophageal echocardiographic images using multi-atlas joint label fusion and deformable medial modeling," *Med. Image Anal.*, vol. 18, no. 1, pp. 118–129, Jan. 2014.

[22] J. Pedrosa, S. Queiros, J. Vilaca, L. Badano, and J. D'hooge, "Fully automatic assessment of mitral valve morphology from 3D transthoracic echocardiography," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–6.

[23] P. Carnahan, O. Ginty, J. Moore, A. Lasso, M. A. Jolley, C. Herz, M. Eskandari, D. Bainbridge, and T. M. Peters, "Interactive-automatic segmentation and modelling of the mitral valve," in *Functional Imaging and Modeling of the Heart*, Y. Coudière, V. Ozenne, E. Vigmond, and N. Zemzemi, Eds. Cham, Switzerland: Springer, 2019, pp. 397–404.
[24] R. I. Ionasec, I. Voigt, B. Georgescu, Y. Wang, H. Houle, F. Vega-Higuera, N. Navab, and D. Comaniciu, "Patient-specific modeling and quantification of the aortic and mitral valves from 4-D cardiac CT and TEE," *IEEE Trans. Med. Imag.*, vol. 29, no. 9, pp. 51–1636, Sep. 2010.
[25] R. J. Schneider, D. P. Perrin, N. V. Vasilyev, G. R. Marx, P. J. del Nido, and R. D. Howe, "Mitral annulus segmentation from 3D ultrasound using graph cuts," *IEEE Trans. Med. Imag.*, vol. 29, no. 9, pp. 1676–1687, Sep. 2010.
[26] R. J. Schneider, D. P. Perrin, N. V. Vasilyev, G. R. Marx, P. J. del Nido, and R. D. Howe, "Mitral annulus segmentation from four-dimensional ultrasound using a valve state predictor and constrained optical flow," *Med. Image Anal.*, vol. 16, no. 2, pp. 497–504, Feb. 2012.
[27] I. Voigt, M. Scutaru, T. Mansi, B. Georgescu, N. El-Zehiry, H. Houle, and D. Comaniciu, "Robust live tracking of mitral valve annulus for minimally-invasive intervention guidance," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2019, pp. 439–446.
[28] M. Sotaquira, M. Pepi, L. Fusini, F. Maffessanti, R. M. Lang, and E. G. Caiani, "Semi-automated segmentation and quantification of mitral annulus and leaflets from transesophageal 3-D echocardiographic images," *Ultrasound Med. Biol.*, vol. 41, no. 1, pp. 251–267, Jan. 2015.
[29] A. Tiwari and K. A. Patwardhan, "Mitral valve annulus localization in 3D echocardiography," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1087–1090.
[30] B. S. Andreassen, F. Veronesi, O. Gerard, A. H. S. Solberg, and E. Samset, "Mitral annulus segmentation using deep learning in 3-D transesophageal echocardiography," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 4, pp. 994–1003, Apr. 2020.
[31] Y. Zhang, A.-A. Amadou, I. Voigt, V. Mihalef, H. Houle, M. John, T. Mansi, and R. Liao, "A bottom-up approach for real-time mitral valve annulus modeling on 3d echo images," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham, Switzerland: Springer, 2020, pp. 458–467.
[32] P. Carnahan, J. Moore, D. Bainbridge, M. Eskandari, E. C. S. Chen, and T. M. Peters, "DeepMitral: Fully automatic 3D echocardiography segmentation for patient specific mitral valve modelling," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham, Switzerland: Springer, 2021, pp. 459–468.
[33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org
[34] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Regressing heatmaps for multiple landmark localization using CNNs," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016, pp. 230–238.
[35] L. Bai and D. Breen, "Calculating center of mass in an unbounded 2D environment," *J. Graph. Tools*, vol. 13, no. 4, pp. 53–60, Jan. 2008.
[36] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the compactness, efficiency, and representation of 3D convolutional networks: Brain parcellation as a pretext task," in *Information Processing in Medical Imaging*, M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, and D. Shen, Eds. Cham, Switzerland: Springer, 2017, pp. 348–360.
[37] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, pp. 1–6, Jul. 2016.
[38] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 843–852.
[39] Y. Yan, S. Duffner, P. Phutane, A. Berthelier, C. Blanc, C. Garcia, and T. Chateau, "2D Wasserstein loss for robust facial landmark detection," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107945.
[40] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "Robust statistics on Riemannian manifolds via the geometric median," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[41] H. P. Lopuhaa and P. J. Rousseeuw, "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *Ann. Statist.*, vol. 19, no. 1, pp. 229–248, Mar. 1991.

BØRGE SOLLI ANDREASSEN received the M.Sc. degree in industrial mathematics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2012. He is currently pursuing the Ph.D. degree in computer vision for medical imaging with the University of Oslo, Norway, in collaboration with GE Healthcare. He was a Systems Engineer at Aker Solutions AS, in 2017, specializing in front-end engineering for subsea production system projects.

DAVID VÖLGYES received the M.Sc. degree in physics from Eötvös Loránd University (ELTE), Budapest, Hungary, in 2008, and the Ph.D. degree in computer science from NTNU, Trondheim, Norway, in 2018. His Ph.D. research at the University of Oslo, Norway, focused on applying machine learning for CT image analysis. He is currently a Senior Software Engineer with Science and Technology AS and a Guest Researcher with the University of Oslo. His main research interests include image processing and analysis using machine learning.

EIGIL SAMSET received the M.Sc. degree in engineering cybernetics from NTNU, Trondheim, Norway, in 1997, and the Ph.D. degree in MRI-guided therapy from the Faculty of Medicine, University of Oslo, Norway, in 2003. He performed his Postdoctoral Fellowship at Brigham and Women's Hospital, Boston, MA, USA. He is the Global Chief Technology Scientist for cardiology solutions with GE Healthcare and a Professor with the University of Oslo. Currently, he is leading strategy and product development across the cardiology care area for GE Healthcare and is involved in developing AI and data-driven approaches throughout the continuum of care.

ANNE H. SCHISTAD SOLBERG received the M.Sc. degree in computer science and the Ph.D. degree in image analysis from the University of Oslo, Norway, in 1989 and 1995, respectively. She is currently a Professor with the Digital Signal Processing and Image Analysis Group, Department of Informatics, University of Oslo. She has worked on a broad range of machine learning applications within image analysis. Her research interests include medical imaging, remote sensing, sonar imaging, ultrasound imaging, seismic imaging, feature extraction, and machine learning.