Quantifying Valve Regurgitation using 3-D Doppler Ultrasound Images and Deep Learning

Sigurd Vangen Wifstad*, Lasse Lovstakken*, Jørgen Avdal*,
Erik Andreas Rye Berg*†, Hans Torp*, Bjørnar Grenne*† and Stefano Fiorentini*
*Norwegian University of Science and Technology & Centre for Innovative Ultrasound Solutions, Trondheim, Norway.
†Clinic of Cardiology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway.
Email: sigurd.v.wifstad@ntnu.no

Abstract*—Accurate quantification of cardiac valve regurgitation jets is fundamental for guiding treatment. Cardiac ultrasound is the preferred diagnostic tool, but current methods for measuring the regurgitant volume are limited by low accuracy and high interobserver variability. Following recent research, quantitative estimators of orifice size and regurgitant volume based on high frame rate 3-D ultrasound have been proposed, but measurement accuracy limited by the wide point spread function relative to the orifice size. The aim of this paper was to investigate the use of deep learning to estimate both the orifice size and the regurgitant volume. A simulation model was developed to simulate the power-Doppler images of blood flow through orifices with different geometries. A convolutional neural network was trained on 30000 image pairs. The network was used to reconstruct orifices from power-Doppler data, which facilitated estimators for regurgitant orifice areas and flow volumes. We demonstrate that the network improves orifice shape reconstruction, as well as the accuracy of orifice area and flow volume estimation, compared to a previous approach based on thresholding of the power-Doppler signal (THD), and compared to spatially invariant deconvolution (DC). Our approach reduces area estimation error on simulations: (THD: 13.2 $\pm$ 9.9 mm^2, DC: 12.8 $\pm$ 15.8 mm^2, ours: 3.5 $\pm$ 3.2 mm^2). In a phantom experiment, our approach reduces both area estimation error (THD: 10.4 $\pm$ 8.4 mm^2, DC: 10.98 $\pm$ 8.17, ours: 9.9 $\pm$ 6.0 mm^2), and flow rate estimation error (THD: 20.3 $\pm$ 9.9 ml/s, DC: 18.14 $\pm$ 13.01 ml/s, ours: 7.1 $\pm$ 10.6 ml/s). We also demonstrate *in vivo* feasibility for six patients with aortic insufficiency, compared with standard echocardiography and magnetic resonance references.

***Index Terms*—blood flow measurement, valve regurgitation, 3-D Doppler, deep learning**

I. Introduction

Heart valve regurgitation is a condition where backwards flow of blood due to leaky valves may cause volume overloading and compromised net forward stroke volume, and is associated with a poor prognosis for the patient. The prevalence of valve regurgitation has been estimated to 18–19% in middle aged adults [1], and is projected to increase overall due to an ageing population [2]. Patients with mild or moderate regurgitations undergo regular follow-up, but do not benefit from routine surgery. On the other hand, patients with severe regurgitation generally require surgical intervention to improve symptoms and prevent heart failure. Therefore, it is essential to accurately separate severe from mild and moderate cases of valve regurgitation. In this way, patients with severe regurgitation can receive appropriate treatment while patients with mild/moderate cases can avoid the unnecessary risks associated with surgery.

Transthoracic echocardiography (TTE) is the most common non-invasive tool for assessing the severity of valve regurgitation. Current recommendations [3], [4] recommend a comprehensive evaluation of the severity based on integrating multiple quantitative and qualitative metrics. As a consequence, the grading of valve regurgitation using TTE is a time-consuming procedure and is subject to high inter- and intra-observer variability. According to the recommendations, the main method for quantitative evaluation of valve regurgitation is the two-dimensional proximal isovelocity surface area method (2-D PISA), which provides the effective regurgitant orifice area (EROA), instantaneous flow rates and total regurgitant volume (RVol) from a combination of Color Flow and Continuous Wave (CW) Doppler recordings [5]. However, 2-D PISA is highly user dependent [6], [7] because several important steps must be performed manually, such as selecting the imaging plane, time frame, Color Flow gain and measuring the radius of the flow convergence region. Moreover, accuracy is limited by dynamic changes in regurgitation flow rates during systole and deviations from the assumption of hemispheric convergence zones [8].

Recent research based on 3-D Doppler ultrasound in invasive transesophageal echocardiography (TEE), and have shown promise for accurate and less user-dependent assessment [9]–[13]. However, TTE acquisition is characterized by larger imaging depths and reduced transmit frequencies, which limits both spatial resolution and pulse repetition frequency. Due to this, current TTE 3-D Doppler methods have limited diagnostic value alone [3].

Avdal et al. previously proposed a quantitative estimator for the cross-sectional area (CSA) of the regurgitant jet and RVol based on TTE high frame rate 3-D Doppler ultrasound [14]. A high PRF, weakly focused acquisition was used to acquire the entire region of interest continuously, rather than using packet acquisitions as in Color Flow imaging. The use of continuous acquisitions enabled the estimation of PW Doppler spectra in each voxel. The spectra could be used for maximum velocity envelope estimation, which is more robust to the presence of clutter compared to the autocorrelation estimator used in Color Flow. This method can therefore achieve quantitative flow measures efficiently and in fewer steps than both 2-D/3-D

PISA and current 3-D Doppler based approaches. Using this approach accurate estimates of the flow volume through a circular orifice phantom was achieved. However, the area estimator is highly dependent on the choice of power threshold for detecting voxels which contain flow. Moreover, due to the large point spread function (PSF), it is difficult to accurately depict irregular orifice geometries [15]. This is a challenge when trying to distinguish small orifices from larger ones, limiting the clinical value of the method.

In ultrasound imaging, blurring of the imaging object due to the PSF is a common problem which limits resolution and image quality. Techniques for restoring such images are typically based on decovolution [16]–[18]. Deconvolution aims to restore the object f from the image s, given a model of the imaging system which is commonly described as $s = f * h + \epsilon$, where h is the PSF, and ϵ is noise.

In recent years we have seen an increasing use of convolutional neural networks (CNNs) in ultrasound research. CNNs have been applied for deconvolution and image enhancement [19], [20], as well as a vast amount of structure segmentation from B-Mode images [21]–[24].

Inspired by recent developments in deep learning, we investigate how CNNs can be trained to perform joint deconvolution and segmentation on highly blurred power Doppler images of regurgitant cardiac jets acquired using TTE. We develop a CNN based approach for segmenting the CSA of jets from leaky valves using cross-sectional images extracted from 3-D Doppler recordings. We apply transposed convolutional layers in our CNN to deconvolve and upscale the poor resolution image and segment the jet CSA. Our CNN is trained on simulated power Doppler data, as the amount of real data examples is limited and the target labels are unavailable. The data is simulated using ultrasound simulation software and procedurally generated orifice masks of arbitrary shape and size. Hence training is completely unsupervised, removing the need for manually labeled training data. Finally, we combine velocity estimates from a conventional spectral velocity estimator with the segmented CSA to quantify the instantaneous flow rate.

II. Methods

A. Orifice generation

We simulated pairs of power-Doppler images and binary target label maps which mimic regurgitant orifices. The orifices were generated using Bezier polygons

$$\boldsymbol{B}(t) = \sum_{i=0}^{p} \binom{p}{i} (1-t)^{p-i} t^i \boldsymbol{C}_i, \tag{1}$$

where $\boldsymbol{B}(t)$ is the closed boundary of the Bezier polygon parameterized by t. $\boldsymbol{C}_i$ are the Bezier curve control points which satisfy

$$\boldsymbol{C}_i \begin{cases} \sim \mathcal{U}(x,y) & \text{if } i < p \\ = \boldsymbol{C}_0 & \text{if } i = p, \end{cases} \tag{2}$$

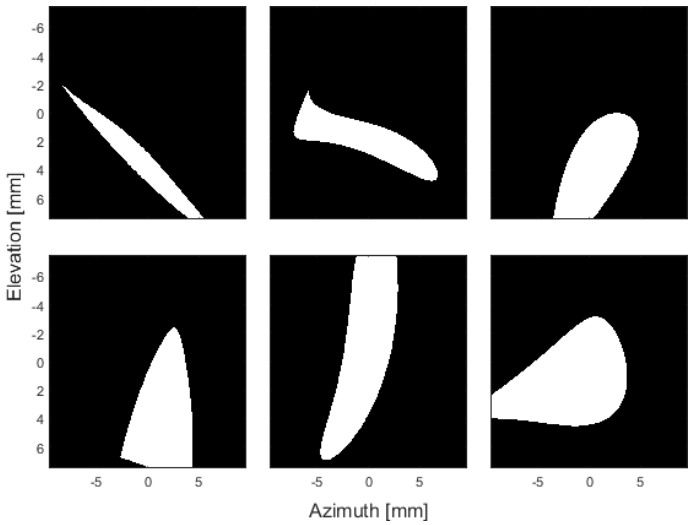


Fig. 1: Example orifices generated using (1) and (2).

where $\mathcal{U}(x,y)$ is a uniform distribution of the image pixel positions (x,y), with a domain equal to the imaging region. $\boldsymbol{C}_0$ is the initial control point drawn from $\mathcal{U}(x,y)$. We define the object function as a binary image $I(x,y)$ where $I(x,y) = 1$ within the region (x,y) enclosed by $\boldsymbol{B}(t)$, and $I(x,y) = 0$ elsewhere. The generation parameters are listed in Table I.

Figure 1 shows six example orifices generated procedurally, showing a variety of possible shapes. In patients with valvular regurgitation, the orifice shape can indeed vary from case to case, based on the cause and position of the regurgitation [25], [26]. Using our approach we can generate a large variety of shapes for the training set, which prevents the model from overfitting on certain geometries.

B. Ultrasound simulation

The power-Doppler images were generated in a two-step process. First, the pulse-echo field h_{pe} is computed assuming Dynamic Receive Focusing, using the Field II ultrasound simulation software [27], [28] and the parameters listed in Table II. Finally, the power-Doppler realizations $R_0(x,y)$ were generated by integrating the pulse-echo contributions from the pixels that belong to the orifice $I(x,y)$.

$$R_0(x,y) = \iint_I P_h(x,y,x_h,y_h) dx_h dy_h, \tag{3}$$

where x_h and y_h denote the coordinates in which the pulse-echo response is calculated, and P_h is the energy of h_{pe} calculated as

$$P_h(x,y,x_h,y_h) = \int_{-\infty}^{+\infty} |h_{pe}(x,y,x_h,y_h,t)|^2 dt. \tag{4}$$

The same pulse-echo field can be used to generate different power-Doppler realizations by changing the orifice map, which allows for quick generation of training data. By integrating the pulse echo signals as in (3) we can achieve simulations

Parameter	Value	Description
p	7	No. of Bezier control points
α	0.01	Augmentation intensity
$N_{\max}$	3	Max no. of superimposed Gaussians
μ_x	(-9.6 mm, 9.6 mm)	Azimuth Gaussian position range
μ_y	(-7.5 mm, 7.5 mm)	Elevation Gaussian position range
S_{xx}	(-0.096 mm, 0.096 mm)	Azimuth covariance range
S_{yy}	(-0.075 mm, 0.075 mm)	Elevation covariance range
S_{xy}	0	Covariance cross-term

TABLE I: Data generation parameters for Bezier polygons and Gaussian blooming augmentation.

Parameter	Value
Demodulation frequency f_c	1–2 MHz
Imaging depth z	8–12 cm
Tx Focus (AZ, EL, RAD)	0, 0, -40 cm
Grid size (AZ×EL)	19.2 mm×15.0 mm
Apodization	None
AZ Elements	100
AZ Pitch	0.215 mm
AZ Kerf	0.215 μm
EL Elements	60
EL Pitch	0.26 mm
EL Kerf	0.26 μm

TABLE II: Field II parameters

efficiently without loss of integrity. This approach was preferred over averaging the backscattered signal from randomly distributed scatterers as this would require too much time considering the amount of training examples needed.

C. Data augmentation

The training set consisted of 30000 power-Doppler and orifice pairs with varying imaging depths and center frequencies. To account for local a reduction in contrast observed in our experimental setup, we superimposed N_i bivariate Gaussian functions to each simulated power-Doppler image $R_0^{(i)}$, creating the augmented image $\tilde{R}_0^{(i)}$

$$\tilde{R}_0^{(i)} = R_0^{(i)} + \alpha \sum_{j=1}^{N_i} \mathcal{N}(\boldsymbol{\mu}_j, S_j), \tag{5}$$

where N_i is uniformly sampled from $\{0, ..., N_{\max}\}$, $\boldsymbol{\mu}_j$ and S_j are uniformly distributed random variables deciding the position and covariance of the jth Gaussian, and α is the augmentation intensity. The data generation parameters are summarized in Table I.

D. Model training

The simulation and training phases are depicted in Figure 2. The GE 4Vc-D geometry was used. The simulation parameters are shown in Table II. N=30000 orifices were procedurally generated with areas uniformly distributed between 0 and 75 mm^2. Power-Doppler images were computed using h_{PE}, with varying transmit configurations, i.e. varying values of z and f_c.

We normalized the images using Z-score standardization and superimposed random Gaussian blooming to each image, as described by (5). Depth and frequency information was added as separate input channels as images with all points

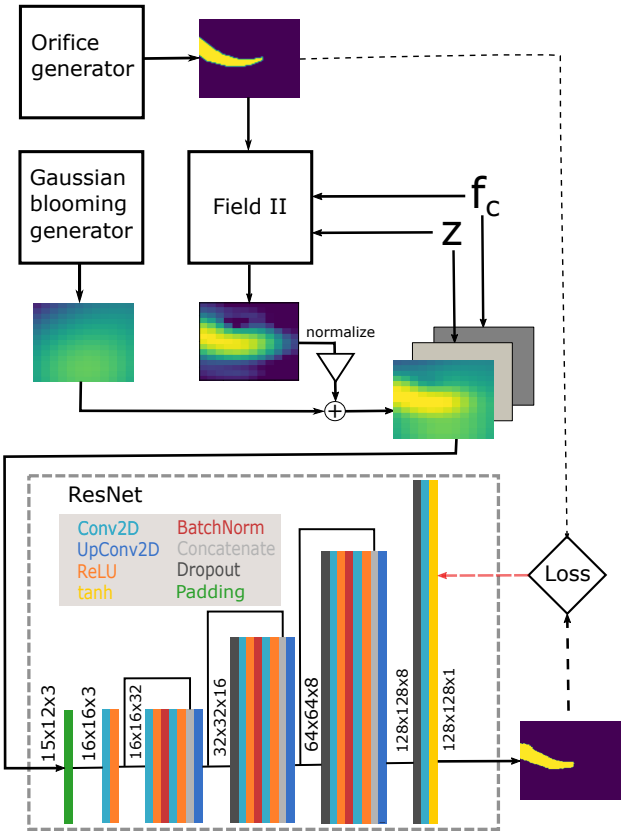


Fig. 2: Schematic representation of the network training. Binary maps representing regurgitant orifices are procedurally generated. Their power-Doppler responses are simulated using Field-II, and Gaussian blooming artifacts are superimposed to the normalized images. The resulting images, along with depth z and frequency f_c parameters, are used to train a CNN (ResNet) to perform segmentation from the power-Doppler images.

having the value of z in meters, and f_c in MHz. For training we generated 1000 images for each transmit configuration, amounting to 30000 images in total. The CNN was trained on the augmented power-Doppler images with the orifice binary maps as target labels, using the Adam optimizer [29] with a learning rate of 0.001 and a binary cross entropy loss function. The augmentation intensity α was chosen by training models with varying augmentation intensities and choosing the one which achieved the highest flow rate estimation accuracy in an experimental flow setup.

The model was implemented using the Keras Python deep learning API and trained on a NVIDIA Quadro RTX 3000. Data generation and training took about two hours. For validation we generated 600 images for each transmit configuration. The model was validated on the *in silico* test dataset. The area estimates from the model predictions were compared with the ground truth and reference segmentation approaches described in Section II-F.

E. Network architecture

The model architecture was a lightweight (∼20000 parameters) network using transposed convolutional layers (UpConv2D) at the end of each convolutional block, such that the number of image pixels at the output of each block is doubled. The input images are hence transformed from 15×12 to 128×128 pixels. The transposed convolutional layers learn to upscale the image directly from the training data as opposed to using simple interpolation techniques. Each convolutional block consists of two convolutional layers with ReLU activations and batch normalization layers to stabilize training. Dropout layers with a dropout rate of 0.25 are added to the end of each block for regularization. The output activation function enforces a binary output. We chose the hyperbolic tangent for this purpose, although a sigmoid would perform equally well. Residual connections connect the first and last layers of each convolutional block to allow for low resolution features to flow through the network. This is a common technique used in deep residual networks (ResNets) [30], which has been shown to improve training stability. We herby refer to our network as "ResNet".

F. Reference segmentation methods

For comparison with the proposed deep learning based segmentation we used two reference approaches. One approach is a conventional -3 dB thresholding of the power-Doppler image, which was used by Avdal et al. [14]. The other approach was spatially invariant non-blind deconvolution, similar to [16]. We used the Richardson-Lucy deconvolution algorithm [31] with 20 iterations to deconvolve the power-Doppler images using the analytical PSF

$$\mathrm{PSF} = \left[\mathrm{sinc}\left(\frac{f_c L_{\mathrm{AZ}}}{cz}x, \frac{f_c L_{\mathrm{EL}}}{cz}y\right)\right]^2, \tag{6}$$

where L_{AZ} and L_{EL} are the aperture dimensions calculated using Table II, c is the speed of sound, z is the imaging depth and x and y are the azimuth and elevation positions. The deconvolved image was segmented by thresholding at 50% pixel intensity.

G. Phantom experiments

We validated the method using the custom-made flow phantom shown in Figure 3. The acquisition and signal processing parameters are summarized in Table III. The phantom was filled with a mixture of water and corn starch to mimic the scattering properties of blood. Channel data were acquired using a GE 4Vc-D probe and a GE E95 scanner operating in high PRF mode. The scanner was locally modified to enable diverging wave acquisitions with a focal point 40 cm behind the transducer. We performed the measurement for insonation angles at 0°, 30°, 40° and 50°, and for three different flow rates. The flow rate was varied by adjusting the height of the upper fluid reservoir. An ultrasonic flow meter (Cynergy3 UF25) was used as a reference. We performed the experiment for circular orifices with sizes 15 mm^2, 25 mm^2, 35 mm^2 and 45 mm^2. We performed a similar experiment for three orifices with noncircular shapes, namely an equilateral triangle (35 mm^2), a half circle (35 mm^2) and a bifurcation of two circular orifices (15 mm^2 and 25 mm^2 respectively).

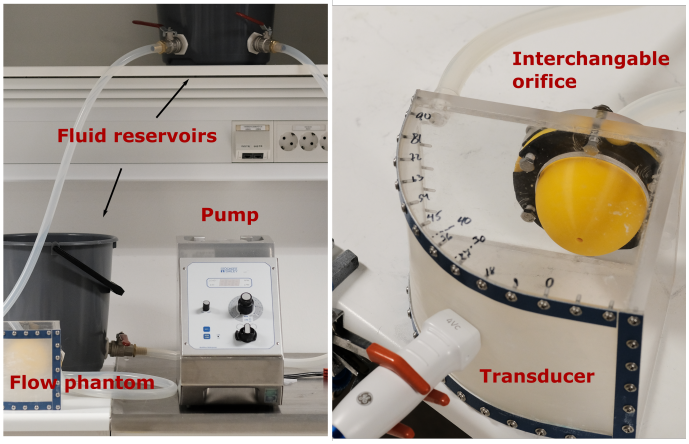


Fig. 3: Experimental setup. Overview (left) and flow phantom closeup (right).

IQ channel data were recorded for offline processing. Channel data were beamformed using the MATLAB UltraSound ToolBox (USTB), and clutter filtered using a FIR filter with an asymmetric frequency response to remove clutter from recruited flow. The passband of the filter was adjusted in each recording to match the observed PW spectrum. We estimated the power-Doppler signal $R_0(x, y, z, t)$ from the filtered signal $s(x, y, z, t)$ by calculating the energy $|s|^2$ with an observation window of 10 ms and an overlap of 50%. The power-Doppler signal was smoothed temporally using moving average filter with length $N_{\mathrm{smooth}} = 11$ (1 ms), and radially using a filter with length $N_z = 3$ (0.2 mm).

The mean velocity $\hat{v}_{\mathrm{mean}}$ was estimated using a spectral envelope estimator.

$$\hat{v}_{\mathrm{mean}} = \hat{v}_{\mathrm{max}} - \frac{\hat{B}}{2}, \tag{7}$$

where $\hat{v}_{\mathrm{max}}$ is the maximum velocity envelope estimated from the PW Dopppler spectrum, and $\hat{B}$ is the estimated bandwidth. We estimated the PW Doppler spectrum using a discrete Fourier transform applied to the same temporal window used to generate the power-Doppler images. The spectrum was smoothed along the temporal dimension similarly to R_0. Before envelope detection the spectrum was binarized automatically using Otsu adaptive thresholding [32]. Estimating the mean velocity using the spectral envelope was preferred over autocorrelation as mean velocity estimators are biased towards by low frequency clutter, which was present in cases of suboptimal clutter filtering. The maximum estimator was shown to be more robust in [14].

Finally, the quantitative metrics were estimated according to

Parameter	Value	Description
f_c	2.05 MHz	Demodulation frequency
c_0	1540	Speed of sound
PRF	10–19 kHz	Pulse repetition frequency
N_c	3.5	Number of cycles
f_s	3.5 MHz	Sampling frequency
Focus depth	-40 cm	Transmit focus depth
Window	Kaiser	For FFT computation
Window size	10 ms	–
Window overlap	50 %	–
Clutter filter	FIR (220)	Filter type and order
N_{FFT}	128	FFT length
N_{smooth}	11	Temporal smoothing filter length
N_z	3	Radial averaging number
Δx	19.2/128 mm	Azimuth voxel length
Δy	15.0/128 mm	Elevation voxel length

TABLE III: Acquisition and signal processing parameters

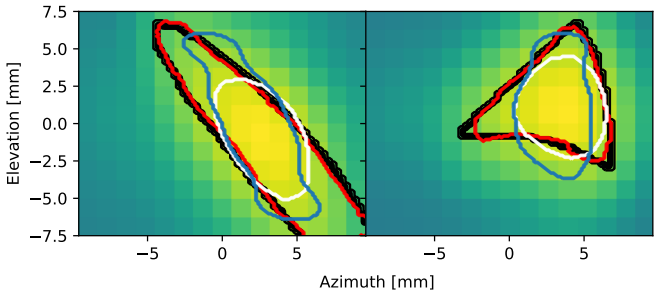


Fig. 4: *In silico* power-Doppler segmentation examples (Red: ResNet; Blue: Deconvolution; White: power thresholding; Black: ground truth).

$$\text{CSA}(t) = \iint_{x,y} g(R_0(x, y, z_{\text{vc}}, t))dxdy, \tag{8}$$

$$Q(t) = \iint_{x,y} \hat{v}_{\text{mean}}(x, y, z_{\text{vc}}, t)g(R_0(x, y, z_{\text{vc}}, t))dxdy, \tag{9}$$

where $\text{CSA}(t)$ and $Q(t)$ is the cross-sectional area and flow rate respectively. The segmentation operator is denoted by $g(\cdot)$. We acquired the segmented orifice image sequence by segmenting the power-Doppler cross-sections $R_0(x, y, z_{\text{vc}}, t)$. The parameter z_{vc} is the vena contracta depth, which was selected manually in the phantom experiment for each recording. For clinical use, z_{vc} needed to be estimated automatically, as described in Section II-H.

H. In vivo feasibility analysis

We acquired 3-D channel data from six patients with aortic valve regurgitation. All patients provided written consent, and approval was given by the regional committee for medical and health research. We used a GE Vivid E95 scanner with a 4V-D probe in high PRF mode, using the same parameters as in section II-G. These recordings were made with a setup using a focal point 30 cm away from the transducer. At the time of the recordings we did not had approval from our industry partners for our improved setup using a -40 cm focal point, but approval was granted at a later time. Comprehensive echocardiograpic examinations were performed to provide reference values for EROA and RVol using 2-D PISA. Magnetic Resonance Imaging (MRI) was also performed to provide RVol. The reference measurements were performed by a cardiologist which was blinded to the results provided with the 3-D Doppler method. We applied our method to the 3-D channel data to estimate RVol and CSA.

The processing chain for the *in vivo* data was the same as in the phantom experiment. However, to account for valve and vena contracta motion in the clinical recordings, we estimated the vena contracta depth z_{vc} at each time t as

$$z_{\text{vc}}(t) = \underset{z}{\text{argmax}} \iint_{x,y} \hat{v}_{\text{max}}(x, y, z, t), \tag{10}$$

where $\hat{v}_{\text{max}}$ is the same maximum velocity envelope as in (7). Here we used the assumption that the maximum velocity occurs at the vena contracta. RVol was estimated by integrating the flow rate $Q(t)$ over the regurgitation time.

III. Results

A. Model training and in-silico validation

We evaluated ResNet segmentation accuracy on a test set consisting of 600 simulated power-Doppler images. The test images were generated similarly to the training images, as explained in Section II-D. The mean area estimation errors were 3.5 ± 2.2 mm^2 for ResNet, 13.2 ± 9.9 mm^2 for power thresholding and 12.8 ± 15.8 mm^2 for deconvolution. Two examples from the test set are shown in Figure 4. ResNet accurately reconstructs the underlying orifice. The deconvolution method is less able to restore the original shapes accurately, likely due to its assumption of spatial invariant PSFs. Thresholding is limited to only providing near elliptical predictions in the object centers, since the PSF severely blurs any sharp edges. The results shown in Figure 5 indicate that ResNet achieves improved segmentation accuracy compared to the references and differentiates better between small and large orifices.

Figure 6 shows the performance of models with different training schemes when subjected to test data with varying imaging depths and transmit frequencies. Results indicate that providing explicit knowledge about depth and frequency during training is beneficial. This was expected, as there will be ambiguities in the relation between the PSF and object size when these parameters are changed. In addition, estimation accuracy decreases with increasing depth, and also decreases for frequencies outside of the training domain. This was also expected, since the transmit frequency and imaging depth affects both the axial and the lateral resolution.

B. Phantom experiments

Figure 7 shows results from models trained with different augmentation intensities α. Following a grid search, a value of $\alpha = 10^{-2}$ gave the best quantitative accuracy while visibly mitigating the effects of local reductions in contrast. The search was performed by training models with α values

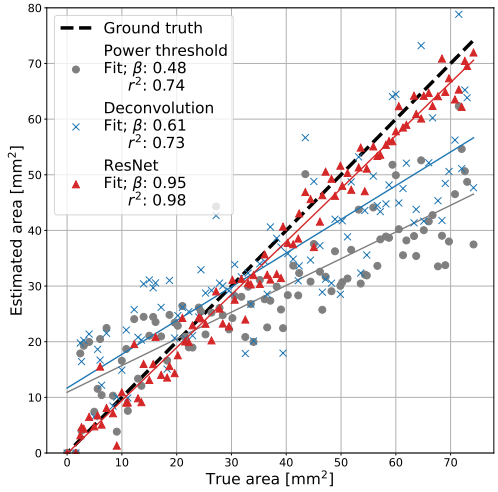
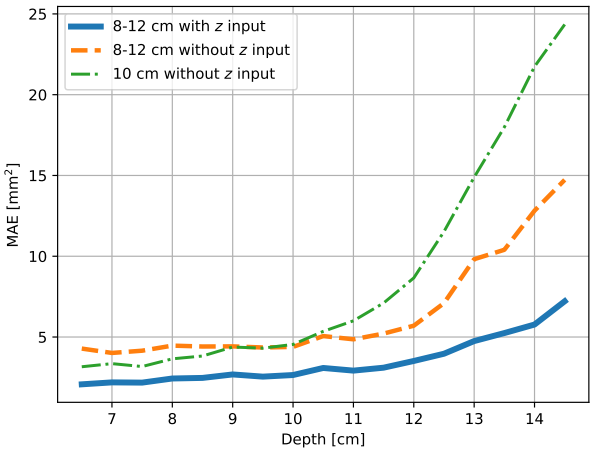
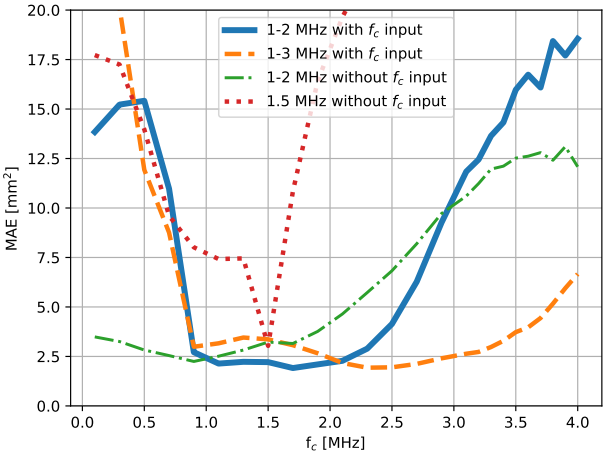


Fig. 5: *In silico* area estimation results. The plots show area estimates using ResNet, deconvolution and power thresholding on a test dataset simulated similarly to the training dataset. Linear regression slopes and coefficients of determination are denoted by β and r^2. Ground truth refers to the areas of the ground truth masks.



(a) Model performances on data with varying imaging depths.



(b) Model performances on data with varying center frequencies.

Fig. 6: Analysis of *in silico* test results for ResNet with different training schemes: (a) on datasets with center frequency $f_c = 1.5$ MHz and varying imaging depth z, and (b) with $z = 10$ cm and varying f_c. The graphs show the mean absolute error (MAE) on test datasets with simulation parameters denoted on the x-axis. The legend describes the training domain for each model and whether the model was given f_c or z as input parameters during training.

between 0 and 1 and computing their average flow rate estimation error for the flow phantom study with circular orifices, as well as monitoring the segmentation qualitatively.

Figures 8 and 9 show the flow rate and area estimates from the phantom setup with four circular orifices, described in Section II-G. Results indicate that ResNet achieves slightly less biased estimates compared to power thresholding and deconvolution, while inter frame variability is similar.

Figure 10 shows flow phantoms with different orifice shapes along with experimental power-Doppler images. The jet cross sections are segmented using both ResNet, power thresholding and deconvolution. Results indicate that ResNet better reconstructs the shape of the orifice than the references.

C. *In vivo feasibility analysis*

Figure 11 shows the results from six patients with aortic regurgitation. The plots compare the RVol and CSA estimated using ResNet, deconvolution and thresholding. The results are compared to Magnetic Resonance Imaging (MRI) and 2-D PISA. In Figure 11b, 2-D PISA EROA estimates are plotted along with CSA as estimated by our method. Note that since 2-D PISA EROA is estimated indirectly using the peak velocity from the CW spectrum, it is not directly comparable to our method, which estimates the CSA directly. Figure 12 shows PW spectra, power-Doppler with jet segmentation and velocity estimates from the six patients. Figure 13 shows a summary of flow rate results, comparing the accuracy from simulations, the experimental validation and the patient data.

We can observe that ResNet is more robust than the other segmentation methods. Power thresholding and deconvolution are more prone to overestimation, most notably in patients 2 and 3. This is attributed to ResNet's ability to infer smaller areas from the highly blurred power-Doppler images, as can be seen in Figure 12. We can see in Figure 11a that this ability has a big impact on the RVol estimates, in which ResNet has a better agreement with the 2-D PISA and MRI references.

IV. Discussion

In this work, we combined deep learning and high frame rate 3-D ultrasound to quantify regurgitant jets in heart valves. This was done by using a neural network trained on simulated data to segment the regurgitant orifice from poor resolution power-Doppler images, which facilitated estimators for the

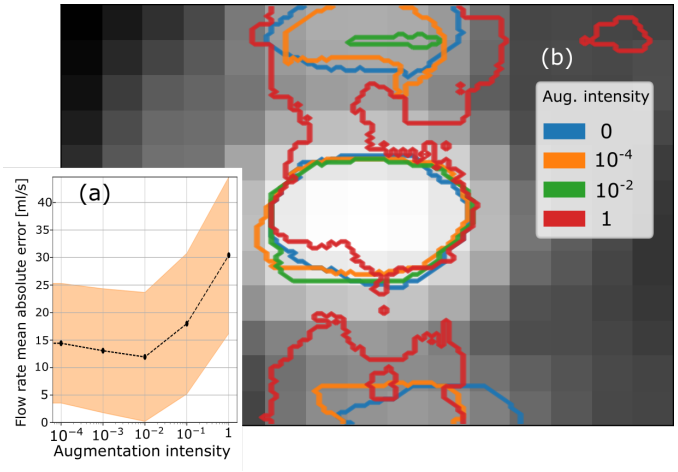
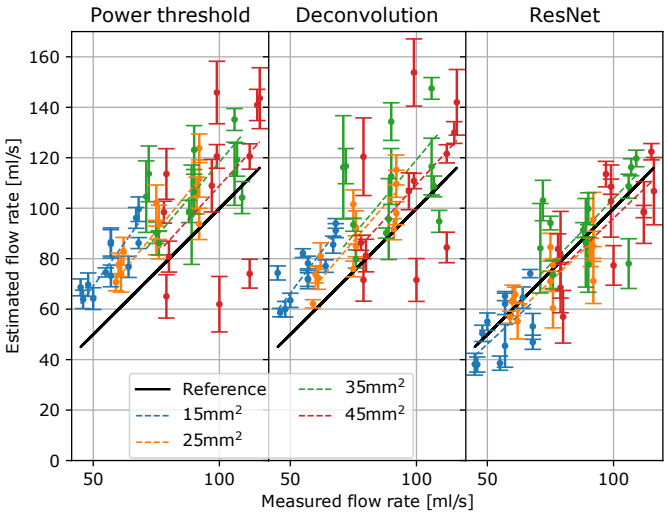


Fig. 7: Results from using models trained with various augmentation intensities α. In (a), a grid search for the best value of α is depicted, showing flow rate estimation mean errors and standard deviation for models with varying α values. The search resulted in $\alpha = 10^{-2}$. In (b), we see a power-Doppler elliptical cross section of a jet from the experimental setup described in II-G. Using $\alpha << 10^{-2}$ results in incorrect segmentation of local areas with reduced contrast. Using $\alpha >> 10^{-2}$ results in a poor training phase and therefore inadequate predictions with high errors. Using $\alpha = 10^{-2}$ we mostly avoid incorrectly segmenting low contrast regions.



	Power threshold	Deconvolution	ResNet
15 mm^2	$\beta: 1.38, r^2: 0.71$	$\beta: 1.33, r^2: 0.69$	$\beta: 0.92, r^2: 0.28$
25 mm^2	$\beta: 1.21, r^2: 0.73$	$\beta: 1.16, r^2: 0.79$	$\beta: 0.95, r^2: 0.54$
35 mm^2	$\beta: 1.18, r^2: -0.16$	$\beta: 1.19, r^2: -0.28$	$\beta: 1.04, r^2: -0.01$
45 mm^2	$\beta: 1.09, r^2: 0.19$	$\beta: 1.09, r^2: 0.22$	$\beta: 0.96, r^2: 0.55$

Fig. 8: Flow rate estimates of circular orifices (15, 25, 35, 45 mm^2) using segmentation from power thresholding, deconvolution, and using ResNet. Each separate measurement is from a recording for a certain orifice size, flow rate and angle. Error bars signify the standard deviation between frames in each recording (10-15 frames per recording). Each recording had a duration of about 60 ms, which is close to the regurgitation durations we observed clinically. The black line shows the flowmeter reference. The estimated velocity field was the same for all the methods. Linear regression slopes and coefficients of determination for each orifice size are denoted by β and r^2.

orifice area and regurgitant volume. Experimental and simulation results shown in Figure 5, 4, 8 and 10 suggest that deep learning based segmentation achieves higher accuracy than power thresholding and spatially invariant deconvolution, and is able to reconstruct the orifice shapes from the low quality images. We also demonstrated feasibility for six *in vivo* cases of aortic regurgitation, as shown in Figure 11 and 12.

The neural network can be trained entirely on simulated data and inference time is short due to the lightweight architecture. Experimental validation showed that our approach is transferable from the simulated domain to real acquisition data, even though the ResNet has been trained solely on simulated data. One challenge we encountered was the difference in signal to noise ratio (SNR) between simulated and observed power-Doppler images. Moreover, we noticed a presence of a diffuse signal surrounding the jet in the observed data, which causes a further reduction in contrast. We believe that the cause for this signal component could be a combination of recruited flow, defocusing due to phase aberration and side lobes. We could account for these blooming effects by using augmentation with Gaussian bivariate functions, a stategy which previously has been applied to account for shadowing artefacts in B-Mode images [33], [34].

To illustrate how increasing realism increases the problem complexity, flow rates from the different test environments are compared in Figure 13. We can observe that when moving from simulations to experimental data, and then to patient data, the accuracy decreases at each step. This trend shows

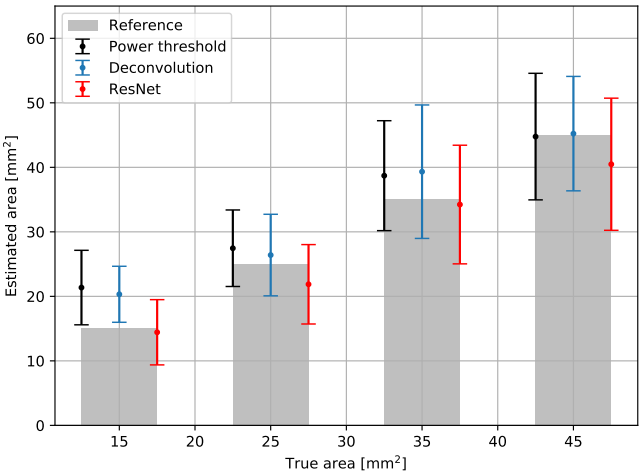


Fig. 9: Angle corrected cross sectional area estimates of circular orifices (15, 25, 35, 45 mm^2) using segmentation from power thresholding, deconvolution, and using ResNet. Error bars signify the standard deviation for all angles and flow rates for a given orifice size.

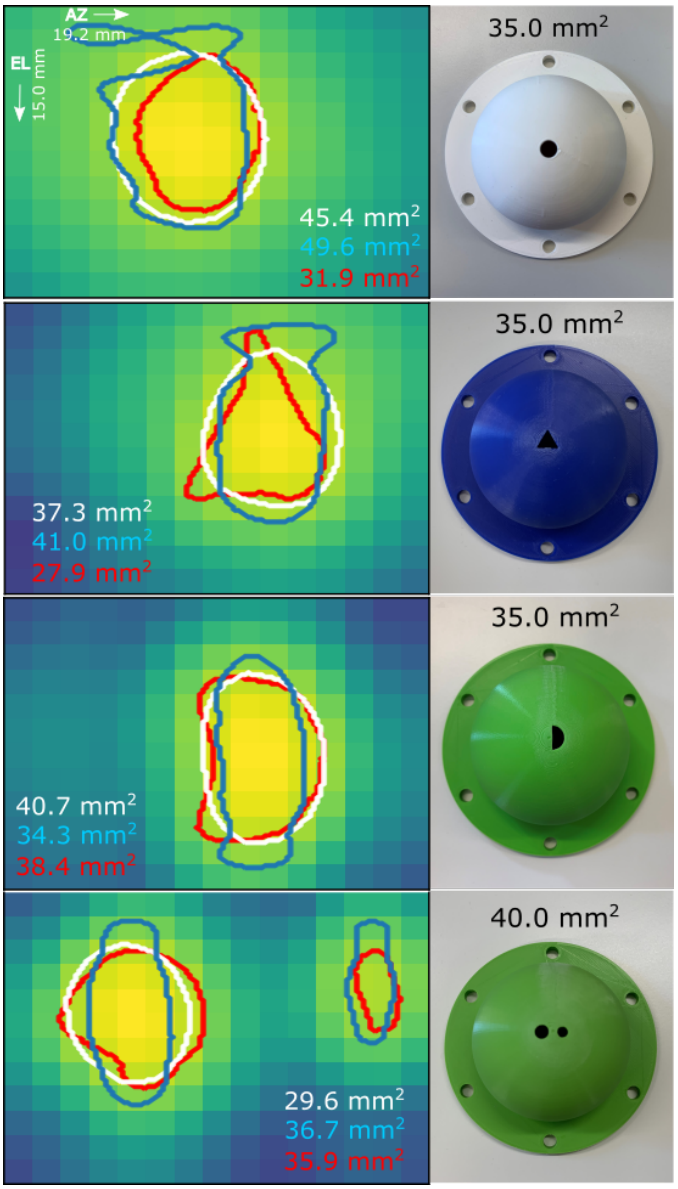
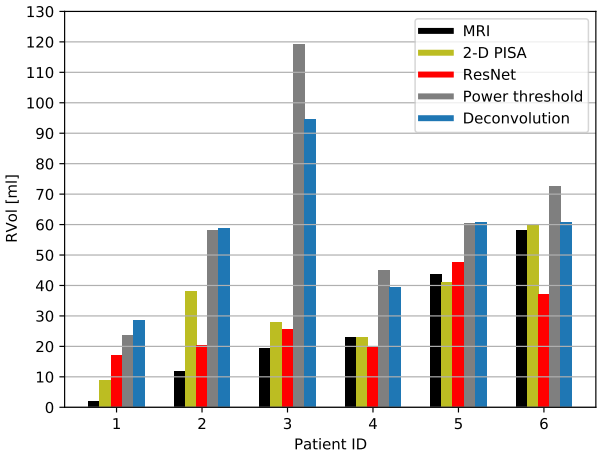
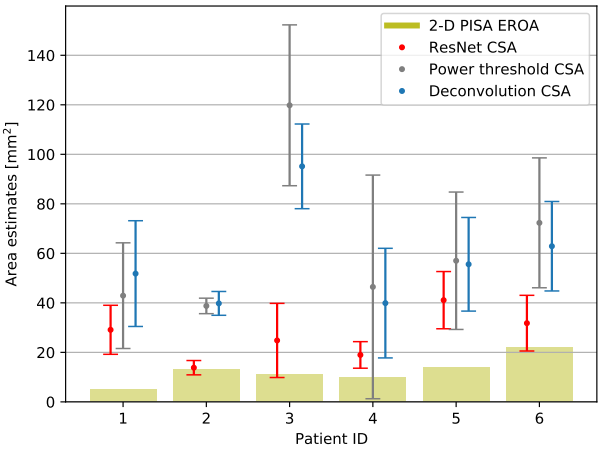


Fig. 10: Recording examples of orifices with different shapes. The left panels show instantaneous power-Doppler images with thresholding segmentation (white), deconvolution (blue), and ResNet segmentation (red). The segmented areas are noted in white, blue and red text respectively. The right panels show the corresponding flow phantoms used in the experiment, with the orifice area noted in black text.



(a) RVol estimates



(b) Area estimates

Fig. 11: RVol and area estimates of six patients with aortic regurgitation from MRI, 2-D PISA, and 3-D Doppler with segmentation using ResNet, deconvolution or power thresholding. MRI measurements were feasible for RVol, but not for area measurements. CSA error bars signify the mean and standard deviation of the area estimates in all frames in the recording. Note that EROA is not directly comparable to CSA since EROA is computed indirectly using a maximum velocity estimate, while CSA is estimated directly. The underestimation of ResNet RVol in patient 6 is likely caused by insufficient PRF in the acquisition.

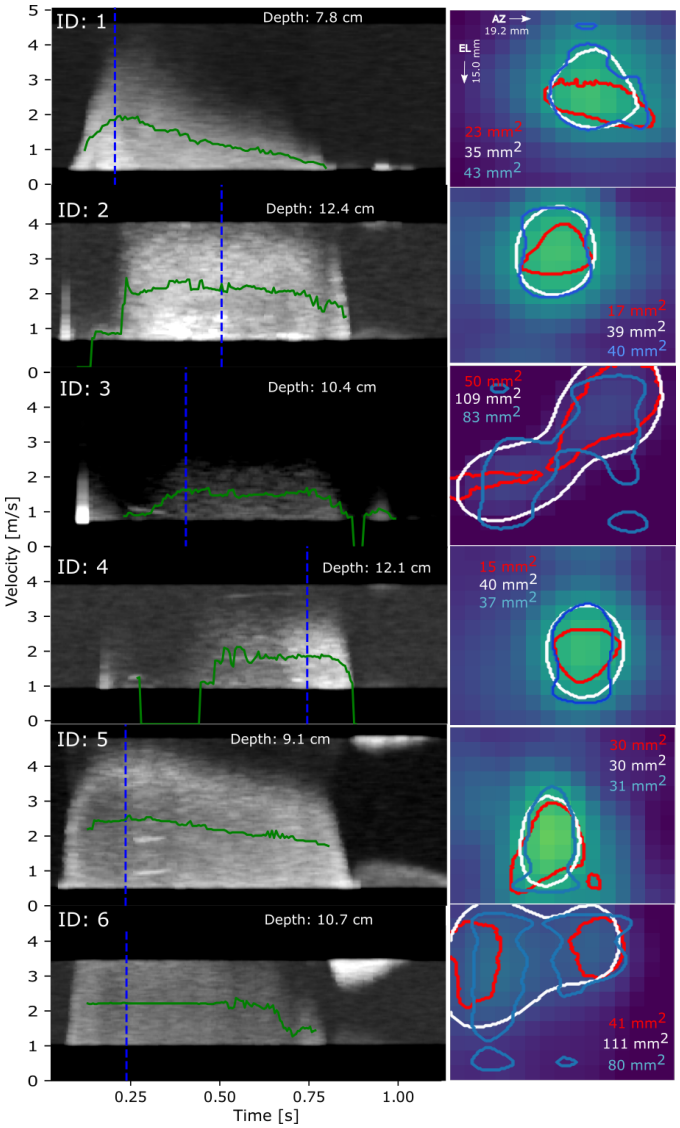


Fig. 12: *In vivo* recordings of six patients with aortic regurgitation. The left panels show the PW spectra, with mean velocity traces plotted in green, and current time frames marked by blue dashed lines. Right panels show power-Doppler cross sections with the CSA segmented using thresholding (white), deconvolution (blue) and ResNet (red). In patient 6 we can observe from the spectrum that the PRF is insufficient to capture the entire velocity envelope. This likely explains the underestimation of RVol for this patient, as can be observed in Figure 11.

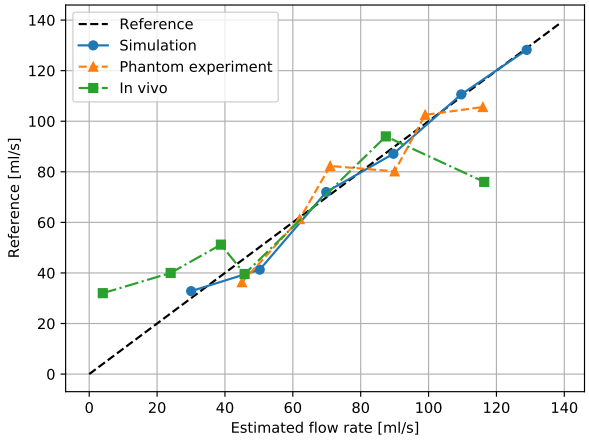


Fig. 13: Comparison of simulated, experimental and *in vivo* flow rate estimation results using ResNet. The graphs are displaying all *in vivo* results with the MRI reference. Simulated and experimental examples are arbitrarily chosen from the results previously shown, while ensuring a representative spread in the reference values. We assume a constant velocity of 2 m/s for the simulated case. *In vivo* flow rates are computed by dividing RVol by the regurgitation duration.

that even though our deep learning model performs well within the simulator conditions, the simulator is limited in providing sufficiently realistic training examples which cover the challenging clinical conditions. The aforementioned effects which are present in a clinical environment, but not in the simulated or experimental environments, have a big impact on the overall signal quality. Valve motion introduces high intensity clutter which is difficult to effectively filter away, and also affects vena contracta depth estimation. We also suspect that aberration and small intercostal windows causes additional deterioration of the image quality.

The experimental setup facilitated validation of our method in a controlled environment where the true orifice geometries and flow rates were known. However, our setup was not intended to accurately mimic the clincal case. The experimental environment has limited realism compared to the clinical cases of TTE for aortic regurgitation. Notable limitations of the experimental setup include the lack of fatty tissue aberrators and ribs, the lack of a moving valve apparatus which may cause shadowing and clutter noise. Future work should aim at creating experimental environments closer to the clinical case. This would facilitate better analysis of the method's limitations in a controlled environment.

Deep learning was only used to segment the regurgitant orifice from power-Doppler images, while velocity was estimated using a conventional PW Doppler estimator. Moreover, the vena contracta depth needed to be estimated prior to segmentation. Future work could expand the method to infer

both area and velocity from 3-D plus time volumes of IQ data, alleviating the need for handcrafted estimators. The neural network architecture would need to be changed in order to incorporate learning correlations in three dimensions and across time. A combination of 3-D convolutional layers, and temporal units such as recurrent layers or attention based mechanisms could be used.

To achieve a model capable of inference directly from 3-D plus time volumes we would need an abundance of such training volumes. This creates the need for a simulator which is fast enough to generate enough training examples in reasonable time, while still accounting for spatially variant PSFs. Field-II would not be fast enough for this purpose, however FLUST [35] is a viable alternative. In the future we are planning to make a 3-D plus time model based on a FLUST simulator, as well as improve key steps in the high PRF acquisition, and processing steps such as adaptive clutter filtering. We believe that these improvements would provide a method capable of producing more robust and accurate results in a clinical environment.

V. Conclusion

In this paper we presented a method that combines deep learning segmentation and 3-D high frame rate ultrasound for the quantification of flow rates, flow volumes, jet areas and shapes for heart valve regurgitation. We showed that our approach better distinguishes between different regurgitation sizes and reconstructs the orifice shape better than a previous approach using thresholding and an approach using spatially invariant deconvolution. *In vivo* feasibility was demonstrated for six patients with aortic regurgitation. Challenges in the acquisition and image formation need to be solved in order to ensure sufficient *in vivo* image quality prior to segmentation. We believe our method could be valuable in clinical assessment in the future, as it could provide higher accuracy results with less user dependency than current recommended methods.

Acknowledgment

This research was made possible due to funding through the Centre for Innovative Ultrasound Solutions (CIUS), appointed by the Norwegian Research Council.

References

[1] J. P. Singh *et al.*, "Prevalence and clinical determinants of mitral, tricuspid, and aortic regurgitation (the framingham heart study)," *American Journal of Cardiology*, vol. 83, no. 6, pp. 897–902, 1999.

[2] 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," *The Lancet*, vol. 368, no. 9540, pp. 1005–1011, 2006.

[3] W. A. Zoghbi *et al.*, "Recommendations for noninvasive evaluation of native valvular regurgitation: A report from the american society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance," *Journal of the American Society of Echocardiography*, vol. 30, no. 4, pp. 303–371, 2017.

[4] P. Lancellotti *et al.*, "Recommendations for the echocardiographic assessment of native valvular regurgitation: an executive summary from the European Association of Cardiovascular Imaging," *European Heart Journal - Cardiovascular Imaging*, vol. 14, no. 7, pp. 611–644, 2013.

[5] G. S. Bargiggia *et al.*, "A new method for quantitation of mitral regurgitation based on color flow doppler imaging of flow convergence proximal to regurgitant orifice," *Circulation*, vol. 84, no. 4, pp. 1481–9, 1991.

[6] N. Thomas, B. Unsworth, E. Ferenczi, J. Davies, J. Mayet, and D. Francis, "Intraobserver variability in grading severity of repeated identical cases of mitral regurgitation," *American Heart Journal*, vol. 156, no. 6, pp. 1089–1094, 2008.

[7] S. Biner *et al.*, "Reproducibility of proximal isovelocity surface area, vena contracta, and regurgitant jet area for assessment of mitral regurgitation severity," *JACC: Cardiovascular Imaging*, vol. 3, no. 3, pp. 235–243, 2010.

[8] T. Buck, B. Plicht, P. Kahlert, I. Schenk, P. Hunold, and R. Erbel, "Effect of dynamic flow rate and orifice area on mitral regurgitant stroke volume quantification using the proximal isovelocity surface area method," *Journal of the American College of Cardiology*, vol. 52, no. 9, pp. 767–778, 2008.

[9] T. C. Tan *et al.*, "Three-dimensional field optimization method: Clinical validation of a novel color doppler method for quantifying mitral regurgitation," *Journal of the American Society of Echocardiography*, vol. 29, no. 10, pp. 926–934, 2016.

[10] M. M. Abudiab, C.-J. Chao, S. Liu, and T. Z. Naqvi, "Quantitation of valve regurgitation severity by three-dimensional vena contracta area is superior to flow convergence method of quantitation on transesophageal echocardiography," *Echocardiography: a Journal of Cardiovascular Ultrasound and Allied Techniques*, vol. 34, no. 7, pp. 992–1001, 2017.

[11] B. Goebel *et al.*, "Vena contracta area for severity grading in functional and degenerative mitral regurgitation: a transoesophageal 3D colour Doppler analysis in 500 patients," *European Heart Journal - Cardiovascular Imaging*, vol. 19, no. 6, pp. 639–646, 2017.

[12] S. Militaru *et al.*, "Validation of semiautomated quantification of mitral valve regurgitation by three-dimensional color doppler transesophageal echocardiography," *Journal of the American Society of Echocardiography*, vol. 33, no. 3, pp. 342–354, 2020.

[13] A. Singh *et al.*, "A novel approach for semiautomated three-dimensional quantification of mitral regurgitant volume reflects a more physiologic approach to mitral regurgitation," *Journal of the American Society of Echocardiography*, 2022.

[14] J. Avdal, A. Rodriguez-Molares, E. Berg, and H. Torp, "Volume flow estimation in valvular jets using 3d high frame rate ultrasound," in *2018 IEEE International Ultrasonics Symposium (IUS)*, 2018.

[15] E. Berg, J. Avdal, T. Espeland, H. Torp, and S. Aakhus, "New 3d doppler flow method for quantitation of aortic regurgitation," *European Heart Journal - Cardiovascular Imaging*, vol. 21, 2020.

[16] C. Dalitz, R. Pohle-Frohlich, and T. Michalk, "Point spread functions and deconvolution of ultrasonic images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 3, pp. 531–544, 2015.

[17] N. Zhao, A. Basarab, D. Kouame, and J.-Y. Tourneret, "Joint segmentation and deconvolution of ultrasound images using a hierarchical bayesian model based on generalized gaussian priors," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3736–3750, 2016.

[18] D. Chen, H. F. Xiao, and J. W. Xu, "An improved richardson-lucy iterative algorithm for c-scan image restoration and inclusion size measurement," *Ultrasonics*, vol. 91, pp. 103–113, 2019.

[19] S. Khan, J. Huh, and J. C. Ye, "Variational formulation of unsupervised deep learning for ultrasound image artifact removal," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 6, pp. 2086–2100, 2021.

[20] A. Makra, W. Bost, I. Kallo, A. Horvath, M. Fournelle, and M. Gyongy, "Enhancement of acoustic microscopy lateral resolution: A comparison between deep learning and two deconvolution methods," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 1, pp. 136–145, 2020.

[21] S. Leclerc *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2d echocardiography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2198–2210, 2019.

[22] S. Liu *et al.*, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.

[23] J. M. Webb, D. D. Meixner, S. A. Adusei, E. C. Polley, M. Fatemi, and A. Alizad, "Automatic deep learning semantic segmentation of ultrasound thyroid cineclips using recurrent fully convolutional networks," *IEEE Access*, vol. 9, pp. 5119–5127, 2021.

[24] Y. Li, Y. Liu, L. Huang, Z. Wang, and J. Luo, "Deep weakly-supervised breast tumor segmentation in ultrasound images with explicit anatomical constraints," *Medical Image Analysis*, vol. 76, 2022.
[25] Y. Otsuji *et al.*, "Insights from three-dimensional echocardiography into the mechanism of functional mitral regurgitation," *Circulation*, vol. 96, no. 6, pp. 1999–2008, 1997.
[26] P. Kahlert, B. Plicht, I. M. Schenk, R. A. Janosi, R. Erbel, and T. Buck, "Direct assessment of size and shape of noncircular vena contracta area in functional versus organic mitral regurgitation using real-time three-dimensional echocardiography," *Journal of the American Society of Echocardiography*, vol. 21, no. 8, pp. 912–21, 2008.
[27] J. Jensen and N. Svendsen, "Calculation of pressure fields from arbitratily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 2, pp. 262–267, 1992.
[28] J. Jensen, "Field: A program for simulating ultrasound systems," *Medical and Biological Engineering and Computing*, vol. 34, pp. 351–352, 1996.
[29] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2014.
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[31] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, p. 745, 1974.
[32] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
[33] E. Smistad, K. Johansen, D. Iversen, and I. Reinertsen, "Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks," *Journal of Medical Imaging*, vol. 5, p. 1, 2018.
[34] A. Østvik *et al.*, "Myocardial function imaging in echocardiography using deep learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1340–1351, 2021.
[35] J. Avdal, I. K. Ekroll, and H. Torp, "Fast flow-line-based analysis of ultrasound spectral and vector velocity estimators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 2, pp. 372–381, 2019.