Doctoral theses at NTNU, 2024:198

Sigurd Vangen Wifstad

Deep Learning Applications for the Assessment of Valvular Heart Disease using Transthoracic Echocardiography

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of
Science and Technology

Sigurd Vangen Wifstad

Deep Learning Applications for the Assessment of Valvular Heart Disease using Transthoracic Echocardiography

Thesis for the Degree of Philosophiae Doctor

Trondheim, May 2024

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

ISBN 978-82-326-7988-1 (printed ver.)
ISBN 978-82-326-7987-4 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:198

Printed by NTNU Grafisk senter

Anvendelser av dyp læring for måling av hjerteklaffsykdom innen transtorakal ekkokardiografi

Hjerteklaffsykdom omfatter sykdommer som påvirker hjerteklaffene og er assosiert med redusert slagvolum og dårlig prognose for pasienten. Transtorakal ekkokardiografi (TTE) er det viktigste verktøyet for vurdering av hjerteklaffsykdom. Imidlertid finnes det for øyeblikket ingen gullstandard for gradering av hjerteklaffsykdom basert på TTE. Derfor er kardiologer avhenginge av en integrert tilnærming med bruk av flere målinger, noe som er både tidkrevende og assosiert med stor grad av målevarabilitet mellom observatører. I lys av nylige fremskritt innen ultralydteknologi og fagfeltet dyp læring (DL) kan noen av de vikigste utfordringene knyttet til arbeidsbelastning og måleusikkerhet bli håndtert. Målet med denne avhandlingen var å utforske hvordan DL kan benyttes i sammenheng med både ny og eksisterende metodikk for kvantisering av hjerteklaffsykdom basert på TTE.

Tre studier blir presentert. Først presenterer vi en DL-tilnærming for automatisk måling av mitralklaffapperatet fra TTE B-Mode-bilder, som ble vist å være muliggjørende for påvisning og kvantisering av mitralprolaps og -stenose. Deretter presenterer vi en DL-tilnærming for automatisk kvantifisering av mitralinsuffisiens ved segmentering og tidsintegrering av konvergensregioner fra 2-D farge-Doppler-sekvenser. Metoden var i stand til å skille mellom milde, moderate og alvorlige tilfeller, og målingene korrelerte godt med referansemålinger fra TTE og magnetisk ressonanstomografi. Til slutt presenterer vi en simuleringsbasert DL-tilnærming for segmentering av dårlig oppløste bilder av klaffeinsuffisiens, i sammenheng med et rammeverk basert på 3-D Doppler ultralydavbildning med høy bildefrekvens. Metoden ble vist å være anvendelig på kliniske data, og sentrale utfordringer ble belyst.

Abstract

Valvular heart disease (VHD) is a class of diseases that affect the heart valves, and are associated with a reduced cardiac output and a poor prognosis for the afflicted patients. Transthoracic echocardiography (TTE) is the cornerstone for the assessment of VHD. However, currently there does not exist any gold standard TTE method for the assessment of the VHD, hence cardiologists rely on an integrative approach of multiple measurements, which is both time-consuming and associated with high inter-observer variability. In light of recent advancements in ultrasound technology and within the field of deep learning (DL), some of the key challenges related to workload and inter-observer variability for the assessment of VHD could be addressed. The goal of this thesis was to explore the integration of DL-based approaches for both new and existing TTE methods for quantifying VHD severity.

Three studies are presented. First, we present a DL approach for automatic monitoring of the mitral valve apparatus from TTE B-Mode images, which was demonstrated feasible for detecting and quantifying mitral valve prolapse and stenosis. Secondly, we present a DL framework for the automatic quantification of mitral regurgitation from segmentation and time integration of regurgitant flow convergence zones from 2-D Color Doppler sequences. This framework was able to distinguish between mild, moderate, and severe cases, and measurements had a good correlation with TTE and magnetic resonance imaging references. Finally, we present a simulation-based DL approach for sub-pixel segmentation of poor resolution images of regurgitant valves, in the context of a high frame rate 3-D Doppler ultrasound framework. Clinical feasibility was demonstrated, and central challenges were highlighted.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of *Philosophiae Doctor* (Ph.D.) at the Faculty of Medicine of the Norwegian University of Science and Technology (NTNU). The research was funded by the Centre for Innovative Ultrasound Solutions (CIUS) and was carried out at the Department of Circulation and Medical Imaging (ISB), NTNU. The main supervisor has been Professor Lasse Løvstakken, and the co-supervisors have been Professor Håvard Dalen, associate professor Jørgen Avdal, associate professor Bjørnar Grenne and Stefano Fiorentini, all from ISB, NTNU.

Acknowledgement

This thesis could not have been completed without the aid of everyone who has supported me throughout these years. I extend my sincere gratitude to my supervisor, Lasse, for his guidance and remarkable expertise in the field of ultrasound. Moreover, I want to express my appreciation to my co-supervisor, Stefano, whose patience and knowledge were essential for keeping me afloat during the initial phases of my PhD studies.

I also want to express my gratitude to the knowledgeable and enthusiastic clinicians Henrik, Erik Andreas, Espen, Bjørnar, and Håvard. They generously took time from their busy schedules to provide me with invaluable data annotations and their profound insights into the field of cardiology.

Furthermore, I would like to thank all my colleagues at ISB who have consistently been available for discussions and have assisted me whenever I got stuck. Special thanks go to Jieyu and David, who helped me numerous times with programming. Moreover, my colleagues have played a crucial role in my overall well-being. We have shared many laughs during coffee breaks, waffle Fridays, travels, and ski trips, and I consider myself fortunate to have you as friends.

Finally, I would like to express my deepest gratitude to my family who has always supported and encouraged me, and to Ingvild, for your love and countless hugs.

Table of Contents

Abbreviations

2-D	Two-dimensional
3-D	Three-dimensional
A2C	Apical two-chamber
A4C	Apical four-chamber view
AI	Artificial intelligence
ALAX	Apical long-axis view
AUNet	Attention UNet
cMRI	Cardiac magnetic resonance imaging
CNN	Convolutional neural network
CSA	Cross sectional area
CT	Computed tomography
CVD	Cardiovascular disease
CW	Continous wave
DL	Deep learning
DS	Deep supervision
DSC	Dice similarity coefficient
EROA	Effective regurgitant orifice area
FIR	Finite impulse response
HPRF	High pulse repetition frequency

ICC	Intraclass correlation coefficient
IQ	In-phase quadrature
LPF	Low pass filter
LV	Left ventricle
LVOT	Left ventricular outflow tract
ML	Machine Learning
MR	Mitral regurgitation
MV	Mitral valve
PISA	Proximal isovelocity surface area
PLAX	Parasternal long-axis view
PRF	Pulse repetition frequency
PSF	Point spread function
PW	Pulsed wave
ResNet	Residual Network
RVol	Regurgitant volume
SNR	Signal to noise ratio
SVD	Singular value decomposition
TCS	Temporal Consistency Score
TEE	Transesophageal echocardiography
TTE	Transthoracic echocardiography
VHD	Valvular heart disease
VTI	Velocity time integral

Chapter 1

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally, and prevalence is increasing with the aging population [1]. As the prevalence increases, so does the workload for clinicians in hospitals around the world, which ultimately may lead to a decline in the quality of patient care. To handle the increase of CVD patients in the future, hospitals need to innovate to spend less time and resources on diagnosis, while still providing accurate disease assessment.

Currently, echocardiography is at the core of CVD assessment. Echocardiography refers to the process of imaging the heart using reflections of high frequency sound waves, i.e. *ultrasound*, transmitted into the tissue. Ultrasonic waves are inaudible to humans, yet highly effective for imaging soft tissues. The earliest experiments successfully demonstrating heart imaging using ultrasound were conducted in the 1950s, and were limited to one-dimensional depictions of the moving tissue [2], [3] and blood flow [4], [5]. As the technology has progressively improved over the years, echocardiography today enables detailed two-dimensional (2-D) and three-dimensional (3-D) imaging of the beating heart, both the tissues of its various structures in motion and blood flow dynamics.

Following the successful development of ultrasound technology, transthoracic echocardiography (TTE), where the ultrasound probe is placed on the skin of the patient's chest, has become the primary imaging modality for assessment of CVD [6]. While enabling detailed imaging of the heart, TTE is non-invasive and non-ionizing and is also attributed to higher accessibility, portability, and inexpensiveness compared to other imaging modalities, such as cardiac magnetic resonance imaging (cMRI) and computed tomography (CT) [7].

A common kind of CVD is *valvular heart disease* (VHD) [8], which is

a class of diseases that affect the heart valves. Although VHD is highly prevalent in the population [9], most cases are mild and do not have a significant impact on the individual's health. However, moderate and severe cases of VHD are associated with increased mortality and morbidity [10], [11]. Consequently, distinguishing between mild, moderate, and severe cases of VHD is imperative to ensure that appropriate treatment is given to those who might be at increased risk while avoiding burdening others with unnecessary treatment.

As with most types of CVD, patients with VHD are initially assessed using TTE [6]. Although abnormalities in the valve apparatus can be identified by experienced operators, accurately assessing the severity of the condition is a significant challenge even for cardiologists with many years of experience. Currently, there are no gold standards for assessment of VHD. Instead, the guidelines [6], [12], [13] recommend an integrative approach of several quantitative and qualitative measurements using TTE. Moreover, patients might additionally be assessed with other imaging modalities, such as transesophageal echocardiography (TEE) and cMRI, as a part of the full examination. Consequently, assessment of VHD is labor intensive and is associated with high inter-observer variability [14]. Thus, methods for severity assessment need to improve in both efficiency and accuracy to ensure sufficient patient care for the aging population.

In recent years we have seen massive advancements within the field of *deep learning* (DL). DL is a sub-category of *machine learning* (ML), which describes systems capable of performing a wide range of sophisticated tasks without explicit programming. DL systems learn symbolic representations of data by adapting thousands to billions of interconnected parameters [15]. The fundamentals of this technology were developed throughout the second half of the 20^{th} century [16]. However, their practical use has only been proven effective in recent years, which is mainly attributed to the availability of computing resources and large datasets [16], [17]. DL systems are rapidly becoming an essential part of society, and most people interact with such systems every day, including facial recognition software [18], media recommender systems [19], and very recently, large language models [15] and other publicly available deep generative models [20]. We also see similar trends in echocardiography, where the amount of research related to DL increases rapidly every year, showing great promise for automation of measurements [21], image enhancement [22], [23], and diagnosis and phenotyping [24]. Given the automation capabilities of DL, it is natural to believe that DL techniques can be applied to the assessment of VHD to alleviate both workload and inter-observer variability.

1.1 Aims of the study

The aims of this study were two-fold. Firstly, we aimed to assess whether automation of existing quantitative metrics associated with VHD was feasible using DL. Secondly, we sought to investigate whether DL could be used to improve the accuracy of both new and existing quantitative metrics for VHD. The aims can be summarized in the following two research questions.

Can deep learning automate the measurement of clinical markers of VHD?

According to the recommendations [6], [12], [13], there exists several quantitative metrics based on TTE which should be integrated as a part of a full assessment of VHD. Some of these include cardiac stroke volumes, maximum velocity and duration of Doppler spectrograms, and caliper measurements of regurgitant jets apparent from color-Doppler imaging. However, as most of these measurements need to be performed manually, they are both tedious and subjective. Moreover, there exists additional metrics that have been proven beneficial for VHD assessment, such as quantitative valve morphology parameters [25], [26] and integrated flow convergence of regurgitant jets from multiple color-Doppler frames (integrated PISA) [27]. However, these metrics are rarely used in practice as it would require considerable additional time per examination. Consequently, many of these metrics are not a part of current recommendations. However, enabling automatic measurements could render these methods viable for clinical practice. Moreover, enabling automatic measurements would facilitate further clinical research on VHD, as gathering quantitative becomes less demanding.

Can deep learning improve the measurement accuracy of current and novel quantitative markers of VHD?

Currently, there are no gold standards for VHD assessment using TTE, and the accuracy of current methods is limited. Current methods are limited by simplifications of underlying physics, and by requiring manual measurements which are prone to human error and practical differences [6]. DL could alleviate subjectivity and hence improve the accuracy and reproducibility of clinical measurements. Moreover, novel methods based on advanced imaging such as 3-D Doppler is limited by image resolution and frame rate [28], [29]. DL systems are capable of approximating any function [30], including inverse mappings from poor resolution images [31], [32]. Hence, DL could be leveraged to mitigate quantitative inaccuracies associated with limited image quality.

1.2 Summary of presented work

Contribution 1: Mitral Valve Segmentation and Tracking from Transthoracic Echocardiography using Deep Learning

The underlying mechanisms leading to VHD are either caused by direct changes in valve morphology and function, or by secondary dysfunction, in which remodeling of the adjacent cavities can cause dilatation of the valvular orifice (annulus) or tethering of the leaflets. Deviations from normal function could lead to improper coaptation of the valve leaflets, effectively creating regurgitant orifices [33]. Moreover, restricted leaflet motion could lead to stenosis in the valve [6]. Assessment of the valve's condition is therefore essential for an accurate description of disease mechanisms and for monitoring the progression of the disease. However, valve function is typically only assessed qualitatively [6]. Quantitative assessment would require tedious manual measurements, which is rarely considered feasible.

In this work, we developed a framework for automatic segmentation of the mitral valve (MV) from multiple standard views in TTE imaging. We trained a convolutional neural network to segment the different anatomical parts of the MV. We compared different network architectures and found that attention gates [34] and deep supervision could be leveraged to increase emphasis on relevant regions in the image, thus mitigating issues of false positives in irrelevant regions. We showed that this improved the robustness and quality of automatically derived biomarkers for MV function.

The biomarkers derived using this framework were valve opening angles, velocities, and leaflet tenting area for all scallops in both leaflets of the MV, similar to VHD biomarkers which have been described in the literature [25], [26], [35]. We applied the framework to exams from patients with VHD and healthy references. We demonstrated that the automatically derived valve opening angles and tenting areas were feasible for detecting and quantifying MV prolapse, and could be associated with the qualitative description from the clinical exam. Similarly, we demonstrated that valve angular displacement and velocity could be associated with stenosis. These initial results hold promise for automatic assessment of VHD in the future, motivating further research on the topic.

Contribution 2: EasyPISA - Automatic Integrated PISA Measurements of Mitral Regurgitation from 2-D Color-Doppler using Deep Learning

One of the most common types of VHD is valve regurgitation, where improper closure of the valve causes leakage, hence reducing net forward stroke

volume. Current guidelines recommend the proximal isovelocity surface area (PISA) method for quantitative assessment of valve regurgitation. However, PISA has several limitations due to physical simplifications and due to requiring accurate manual tracing of the echocardiograms [6], [14].

PISA is associated with high inter-observer variability due to practical differences in measuring the flow convergence region and interpreting the guidelines [14], [36], [37]. Moreover, standard PISA assessment only measures the regurgitant jet in one time instance and therefore needs to be integrated with continuous Doppler measurements to account for temporal variation of the regurgitation [27]. However, with automated measurement it is possible to measure in every frame of the recording, hence accounting for temporal variation directly. Furthermore, automated measurements allow for a more flexible surface area estimator in comparison to the manual approach, for which hemispherical convergence of flow is usually assumed for simplicity of measurement [6], [13], [38].

In this work, we developed a framework for the assessment of mitral regurgitation from 2-D color-Doppler echocardiograms by automating the PISA measurement with DL. The framework, named EasyPISA, automatically segments flow convergence regions from 2-D color-Doppler sequences. The isovelocity surface area was computed without assumptions of hemispherical convergence using a surface of revolution estimator, and regurgitant volumes were derived with time integration of consecutive measurements.

EasyPISA was applied retrospectively to exams of patients with mitral regurgitation, with known reference values from standard examination using TTE and cardiac magnetic resonance imaging (cMRI). We found that EasyPISA estimates had a good correlation with both PISA and cMRI references. Analysis showed that some of the discrepancy could be explained by insufficient frame rate, angle dependency, and over-segmentation, but also by differences in geometry assumptions.

Contribution 3: Quantifying Valve Regurgitation Using 3-D Doppler Ultrasound Images and Deep Learning

One of the fundamental limitations of the 2-D PISA method is that it only captures a cross-section of the regurgitant orifice. Due to this limitation, symmetry needs to be assumed when estimating the out-of-plane component of the lesion. However, the symmetry assumption is not always valid. Depending on the placement and cause of the regurgitation, the orifice could be elliptical or crescent-shaped instead of the assumed circular shape [6], [33].

The 3-D nature of the problem motivates the use of 3-D ultrasound.

While 3-D PISA is an option [39], traditional 3-D Doppler ultrasound is severely limited by frame rate and thus struggles to capture regurgitation dynamics. Because of this, previous work from our group [40] described a method using High Pulse Repetition Frequency (HPRF) 3-D ultrasound, which could achieve images of an *en face* view of the orifice at high frame rate.

However, one of the fundamental challenges with the HPRF 3-D ultrasound technique was the spatial resolution of the images, which made it difficult to accurately delineate the contour of the orifice. This was a major limitation, as accurately estimating the size of the regurgitant orifice is essential for accurately separating between mild, moderate, and severe cases.

In this work, we sought to improve the contour estimation by using DL sub-pixel segmentation for estimating the orifice area with better resolution than what was initially given by the image. Because of limited data availability, we resorted to training the model on simulated data.

Using a lightweight architecture trained on thousands of simulated regurgitant orifices with various shapes and sizes, we achieved a model with high validation accuracy. We also found that applying the model to real data was feasible by using data augmentation to account for noise that could not be directly simulated using the ultrasound simulator.

Experimental analysis showed that the model could delineate the underlying orifice shapes from the poor resolution images better than the original method, and consequently could estimate flow volumes with higher accuracy. Using this model on patient data we achieved comparable results to both 2-D PISA and cMRI. However, other components of the HPRF data acquisition pipeline still pose a challenge to the pipeline's viability for clinical use.

1.3 Discussion of results

1.3.1 Mitral Valve Segmentation and Tracking from Transthoracic Echocardiography using Deep Learning

Challenges in the dataset

Segmentation of the MV is challenging for several reasons. The MV is typically not clearly visible in poor-quality images. Manually annotating such images was therefore a significant challenge. This caused worse model performance on average for poor-quality images, which are common for difficult-to-image patients.

Moreover, the MV is a dynamic, three-dimensional structure. Imaging with 2-D ultrasound is therefore challenging as leaflets can move out of the imaging plane, causing inconsistent segmentation results. To fully assess MV

function it was necessary for the model to segment the MV from multiple standard views. However, as both the MV and the surrounding tissue display different features in the different views, generalizing the segmentation problem across all views was a considerable challenge. Training different models for each view was not feasible at the time due to limitations of the dataset size, hence one model was trained on the full dataset. Consequently, we could observe a reduction in performance for the parasternal long-axis view, and for the anterior leaflet in the apical two-chamber view.

Generalizability was a challenge not only in terms of image quality and view but also pathology. To account for a wide range of pathologies, we trained the model on images from patients with degenerative or rheumatic MV disease with varying severity. However, we could observe that the model would fail for some cases with severe pathology or otherwise substantial deviation from the normal cases. This is a common problem for ML algorithms, as they are prone to overfitting the model parameters to the training dataset [16]. Expanding the training dataset would be beneficial for increasing generalizability. Much more data is currently available but needs to be annotated in order to be used for training. However, ensuring sufficient diversity of pathologies is a challenge.

Model selection

Due to the challenges in the dataset and concerns about generalizability, model design was important to mitigate the specific challenges in the dataset. We found that using attention gates to enforce saliency around the MV could mitigate issues with false positives of surrounding tissue, leading to overall better estimates of the extracted metrics for valve function. However, improvements to the model could be made in the future to further increase the robustness. The MV leaflets are thin structures, that always connect to the annulus points, hence the annulus and leaflet structures might be better modeled as graphs. Followingly, the model could be optimized to minimize a regression loss function of the graph point coordinates. Adapting the segmentation problem to a point coordinate regression problem could be achievable using graph convolutional networks [41], or by combining the proposed model with a differentiable spatial to numerical transform [42].

Moreover, due to the dynamic nature of the MV, leveraging temporal information in the model would likely further increase robustness and temporal consistency. However, this would require a training dataset densely annotated in time, which substantially increases the workload for annotators and was therefore not considered feasible. We experimented with overcoming this issue by using spatiotemporal pseudo-labels generated with a model

trained on individual frames, similar to other research [43]–[45]. However, we did not observe any improved results in our experiments using this approach. Leveraging temporal information could also be possible by using tracking algorithms. Tracking-based algorithms would need to account for large leaflet displacement between imaging frames during the MV opening phase, as well as out-of-plane movement. Various algorithms for tracking MV motion have previously been proposed [46]–[48]. Such algorithms might be combined with supervised DL segmentation methods to improve temporal consistency while still maintaining segmentation accuracy.

Future work

With the knowledge gained from the analysis of the model's performance, the model should be reiterated with some improvements. The training dataset should be expanded to include more examples with various pathologies and image quality by annotating more exams. The current training set size was limited, and we expect an increased performance if the training set were to be expanded. Moreover, the training dataset could be artificially balanced in terms of imaging views, and potentially different pathologies, in order to improve generalization and reduce overfitting. Furthermore, updating the model architecture to enable point coordinate regression, i.e. modeling the MV as a graph, would likely be beneficial.

We are currently running larger retrospective and prospective clinical studies to assess both model robustness and feasibility of the clinical biomarkers. In these studies, we plan to assess the predictive quality of the automatic VHD biomarkers on a larger population, consisting of about 100 exams for retrospective analysis with various pathologies, and about 20 MV patients acquired prospectively using a real-time version of the framework. Initial results indicate good predictive quality for MV prolapse and stenosis. Moreover, results indicate that segmentation performance is better prospectively than retrospectively, as the operator can adjust the imaging view according to the segmentation observed in real-time.

These studies will likely unveil additional unforeseen challenges and potential flaws or limitations of the method, which could be improved in future iterations of the project. Furthermore, the true clinical value of the derived MV biomarkers is hypothesized, but not currently known to a full extent. This is because deriving such measurements in large populations has not been feasible without automated measurement. We expect the ongoing clinical studies will provide more knowledge about the clinical importance of automated MV measurements.

1.3.2 EasyPISA - Automatic Integrated PISA Measurements of Mitral Regurgitation from 2-D Color-Doppler using Deep Learning

Automatic segmentation of flow convergence zones

Similar to the MV segmentation problem, segmentation of flow convergence regions from 2-D color-Doppler images presented some challenges. Firstly, since flow convergence regions were not always visible in the images it was important that the model did not output large false positive predictions for images with no apparent flow convergence. To account for this, about half of the total images used for training and validation were negative images, i.e. did not contain flow convergence annotations. Moreover, the flow convergence labels were small compared to the image dimensions, which, in combination with the large share of negative images, made the dataset subject to a high class imbalance. Furthermore, the image resolution was decreased as a preprocessing step, which could further challenge the segmentation of small flow convergence regions.

To account for class imbalance and to maintain sufficient recall, the loss function used for training the convolutional neural network could be adapted to balance the emphasis on positive and negative examples. However, the trade-off with maintaining a high recall was that the model was more prone to false positives. False positives would typically occur where there was a strong flow signal in irrelevant regions, such as the inlet of the aortic outflow tract. Moreover, false positive segmentation on the atrial side of the mitral valve was a significant challenge as pixels in the regurgitant jet efflux could falsely be segmented along with the flow convergence zone. Similar to the mitral valve segmentation project, as discussed in Section 1.3.1, we attempted to utilize attention gates to mitigate false positives from irrelevant regions. However, we did not observe any clear benefits of this approach at this time. Nonetheless, we believe that attention might prove beneficial in future iterations of the project, but further work is required to adjust the methodology to the problem. As an alternative solution, we could mitigate false positive segmentation on the atrial side of the mitral valve by applying the mitral valve segmentation model to mask out false positive pixels. This conceptually simple, yet effective solution might be essential to ensure a robust application in the future.

Feasibility of flow volume estimation

We demonstrated that EasyPISA had good agreement with the PISA reference values (Intraclass correlation coefficient, ICC: 0.83), but there was a considerable spread in the estimates. The variability was caused by multiple

factors. Some cases were overestimated due to false positive segmentation, as described above. Other cases were underestimated, either due to loss of Doppler signal in regions of high beam-to-flow angles, or due to insufficient frame rate in cases of highly dynamic flow patterns. The PISA method is also known to be limited by high inter-observer variability [36], [37], which also explains some spread in the reference values. Moreover, the EasyPISA method, which is based on the time integration of consecutive frames, is fundamentally different from standard PISA, which relies on the integration of continuous wave Doppler spectra. Furthermore, since EasyPISA does not assume hemispherical convergence, some additional discrepancy between EasyPISA and the PISA reference could be expected in cases of non-hemispherical flow convergence.

In addition to TTE references, cMRI references were also used to better assess the predictive quality of EasyPISA. Although EasyPISA achieved acceptable correlation with cMRI (ICC: 0.66), the most prominent outliers from the comparative analysis to PISA (including severe dynamic, and non-hemispheric cases) did unfortunately not have cMRI reference values. Therefore it was difficult to properly assess the effect of multi-frame integration and eased geometrical assumptions.

The PISA method is challenged by the fact that the method is not necessarily well understood by the operators [14]. Therefore, the operators are prone to making mistakes if they misinterpret the steps of the methodology. This is especially true for challenging cases, such as non-hemispherical convergence [49] and dynamic jets [50]. In our study, the reference PISA and cMRI measurements had an excellent agreement, which was surprising, as other research has demonstrated significant discordance between the methods [37]. We believe that this can be explained by the fact that the cMRI and PISA exams were performed by the same clinician, with knowledge about the patient's condition. Therefore the assessment was not fully blinded, unlike EasyPISA, which was entirely automatic and blind to the patient's condition. Moreover, the clinician was highly experienced with using the PISA method and thus may have performed the measurements more thoroughly than the average clinician.

Future work

We are planning to further assess the clinical feasibility of EasyPISA in the future. Robustness should be verified in datasets with a wide range of pathologies. A larger feasibility study would likely uncover additional challenges that might result in further refinement of parts of the methods, especially the model architecture and training. Moreover, the training

dataset size was limited to 54 patients in this study, and the application would undoubtedly benefit from increasing the size of the training dataset, especially in terms of accounting for a wider range of pathologies. Future studies should also compare EasyPISA with references from multiple clinicians, blinded from the patient's condition and each other's measurements. This way, EasyPISA could more accurately be assessed in comparison to inter-observer variability. Finally, a real-time EasyPISA application is currently under development, which could be used in prospective studies. We believe that a real-time application that indicates to the operator the presence of flow convergence zones and estimates of severity could be useful in initial screenings of patients.

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

Learning from simulations

Unlike the other two projects described in this thesis, we did not have access to annotated HPRF 3-D Doppler data which could be curated for DL. The amount of in vivo recordings was limited, and would furthermore be very challenging to accurately annotate manually due to the poor image quality. However, since the image features were relatively simple and regular, they could effectively be simulated using an ultrasound simulator.

Consequently, the main challenge was related to the domain gap between simulated and real data. The model performance was excellent on the simulated validation data, obtaining an average orifice area error of 3.5 mm^2, which is small compared to the thresholds for severe aortic regurgitation (30 mm^2) and mitral regurgitation (40 mm^2) [6]. However, applying the model to real acquired data unveiled some important challenges.

In an experimental study, we observed that the acquired data was prone to some blooming artifacts which were not observed in the simulations. We believe that this was due to a combination of multiple effects that could not easily be simulated with our simulator, such as sidelobes and recruited flow. The model was affected negatively by the blooming artifacts as it falsely interpreted the artifacts as regions containing flow, causing an overestimation of flow volumes. The gap between the simulated and real domain could be mitigated by using data augmentation, where blooming artifacts were simulated and superimposed on the images. This was shown to mitigate false positive segmentation of blooming artifacts.

The augmented model achieved good results in the experimental tests. The model achieved better flow volume estimates than the references, and could also accurately differentiate between various irregular shapes of the

orifice phantom from highly blurred images. These results demonstrated the capabilities of DL to solve visual recognition problems which can be difficult for humans. Despite promising results in experiments, generalizing the model to clinical data remained a challenge.

Since the model was based on DL from simulations, there is still uncertainty about whether the model could generalize sufficiently well to the clinical data. The clinical data was challenged by additional artifacts, caused by obstruction of view from ribs, reverberation, attenuation, and velocity clutter from rapidly moving tissue. Consequently, this limited the validity of the experimental results, as we could not account for all of the aforementioned effects.

Using the proposed approach we could obtain results comparable to cMRI and PISA. Moreover, we showed that for certain cases with poor image quality, the application greatly improved when using DL segmentation of the orifice, rather than conventional segmentation methods which had been proposed previously. These results were encouraging for prospects of clinical feasibility, however, many challenges still need to be addressed in order for the method to be sufficiently reliable for clinical use.

Challenges with the acquisition and processing pipeline

There were many sources of error in the acquisition and processing pipeline which might have influenced the results. Clutter filtering was a particularly difficult challenge. Heart valve movement and recruited flow are sources of high-velocity clutter which would cause severe reduction in the signal-to-noise ratio if not effectively suppressed using the clutter filter. Moreover, due to the HPRF acquisition, secondary sample volumes at smaller depths could contribute to strong clutter because reflections were less attenuated than from the deeper primary sample volume. We used finite impulse response (FIR) filters for clutter filtering [51], but these needed to be manually adapted to the recordings, and their effectiveness was limited in the presence of high-velocity clutter. Singular value decomposition (SVD) filters, which are the current state of the art in clutter filtering [52], could be worth investigating in the future.

Flow volume estimates depend not only on orifice area estimates but also on velocity estimates. We used a PW Doppler spectrum estimator to estimate the mean velocity in every image pixel. Adaptive thresholding [53] was used to segment the spectral envelope and trace the maximum velocity. However, the thresholding could lead to erroneous estimates in recordings with low signal-to-noise ratios. A DL based estimator could be more robust, but we did not have sufficient training examples to develop a model for spec-

tral envelope segmentation. Moreover, aliasing of high-velocity jets would cause underestimation of velocity, and consequently, flow volumes. Aliasing is caused by a limited PRF, which is challenging for 3-D Doppler ultrasound, albeit the HPRF application does increase the achievable PRF (10-20 kHz) from the baseline (50-200 Hz) [28], [54].

Estimating the vena contracta depth, where the jet cross-section is the most narrow, was important for estimating the jet cross-sectional area. Depth estimation was performed automatically, largely based on the velocity estimator. If the depth was inaccurately estimated, the jet cross-section would appear larger, causing overestimation. Moreover, the image quality would be deteriorated, further challenging the segmentation.

Future work

We are planning a reiteration of this project with an updated pipeline, where many of the issues related to acquisition and processing will be improved. Moreover, the ultrasound simulator could be replaced by a fast 3-D flow-line simulator previously developed in our group [55] to enable the generation of 3-D flow fields for training data, as opposed to static 2-D cross sections. Hence a model could be developed to infer flow volumes directly from acquired 3-D data, which could remove the need for velocity and depth estimation and reduce the sensitivity to data preprocessing. For validation, the realism of the experimental flow phantom could be improved by adding aberrators and rib obstructions. Finally, a larger validation study, preferably with cMRI should be performed using the updated method to assess its clinical feasibility.

1.4 Concluding remarks

In this thesis, we have explored DL methods applied to TTE assessment of VHD. Three projects have been described: automatic mitral valve segmentation and tracking, automatic regurgitant flow volume estimation from 2-D Doppler ultrasound based on the integrated PISA method, and regurgitant flow volume estimation from HPRF 3-D Doppler ultrasound, using DL subpixel segmentation of jet cross sections. The two first projects demonstrated the potential for automatic extraction of existing metrics for VHD, which are currently underused in the everyday clinic. DL enables monitoring metrics that have been recommended by the literature, but which have been challenging to apply in practice due to substantial workload requirements. Furthermore, the third aforementioned project showed how DL could be applied not only for automation but for improving the accuracy of quantitative metrics which were fundamentally limited by image quality.

In current practice, the assessment of VHD is limited by workload and inter-observer variability. The variability is caused by differences in skills and experience, practice differences between institutions, and varying interpretations of the guidelines [14], [56]. As demonstrated in this thesis, DL can be a valuable tool for streamlining workflows and facilitating standardization of measurements. Assessment of valve prolapse and stenosis can be quantified according to automatic measurements following predefined rules, which alleviates subjectivity associated with the assessment. Moreover, automatic measurements of valve regurgitation can alleviate variability associated with improper interpretation of the guidelines, which is especially common for dynamic jets [50], [57], for eccentric jets [58] and non-circular orifice geometries [38], [58].

DL has the potential to solve many problems which have proved to be challenging using classical methods. However, supervised DL techniques are limited by requiring large annotated datasets. In the projects described in this thesis, we found that using annotated data from around 50-100 patients was sufficient for achieving satisfactory results, although the model robustness was challenged by difficult examples. We believe that segmentation robustness can be further increased by annotating more data for training, but it is difficult to know a priori which quantities would be needed. Data annotation is associated with some costs, as it requires experienced clinicians to manually annotate the images. For our datasets annotation time was about 45 minutes per patient. The costs associated with expanding the dataset are manageable, but increasing the dataset size without specific considerations about data distribution would likely result in diminishing returns. Future efforts to expand the datasets should pay attention to the current model's specific failure modes and annotate the images for which the model fails. Moreover, datasets need to be sampled from diverse populations in terms of pathology to ensure generalizability, which otherwise could lead to unwanted model biases.

All medical technology needs to be validated in clinical trials in order to be considered feasible for clinical use. However, since DL systems are data dependent, clinical trials cannot cover all possible cases that could be encountered in practice. For DL systems to be viable for clinical use it is important to ensure they can be trusted by the user, which requires a certain level of explainability. Segmentation-based approaches mimic the cardiologist's workflow and are more explainable than end-to-end learning systems [59] which predict diagnosis directly from the image data. However, segmentation models can fail if presented with low-quality images or data that is outside of the training distribution. Therefore it is important

to incorporate safety measures for practical use which detects segmentation failure and unphysical behaviour. When the DL system is equipped with appropriate safety measures it can be a valuable tool for cardiologists. As demonstrated in this thesis, DL based approaches have the potential for improving VHD assessment by increasing automation and accuracy of existing and novel methods. DL based measurements have already become available in commercial ultrasound scanners [60], [61]. It is therefore credible that DL based methods for VHD disease assessment will be available for clinical use in the future.

1.5 Publication list

In addition to the works enclosed in this thesis, the author has contributed to other research contributions and international conferences. A complete list of contributions follows.

Contributions included in the thesis

- **Sigurd Vangen Wifstad**, Henrik Agerup Kildahl, Bjørnar Grenne, Espen Holte, Ståle Wågen Hauge, Sigbjørn Sæbø, Desalew Mekonnen, Berhanu Nega, Rune Haaverstad, Mette-Elise Estensen, Håvard Dalen and Lasse Lovstakken, "Mitral Valve Segmentation and Tracking from Transthoracic Echocardiography Using Deep Learning", *Ultrasound in Medicine & Biology*, Volume 50, Issue 5, pages 661-670, 2024.

- **Sigurd Vangen Wifstad**, Henrik Agerup Kildahl, Espen Holte, Erik Andreas Rye Berg, Bjørnar Grenne, Øyvind Salvesen, Håvard Dalen and Lasse Lovstakken, "EasyPISA: Automatic Integrated PISA Measurements of Mitral Regurgitation from 2D Color Doppler using Deep Learning", submitted to *Ultrasound in Medicine & Biology*, 2024.

- **Sigurd Vangen Wifstad**, Lasse Lovstakken, Jørgen Avdal, Erik Andreas Rye Berg, Hans Torp, Bjørnar Grenne and Stefano Fiorentini, "Quantifying Valve Regurgitation Using 3-D Doppler Ultrasound Images and Deep Learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Volume 69, Issue 12, pages 3317-3326, December 2022.

Other contributions

- Shivanandan Indimath, **Sigurd Vangen Wifstad**, Vincent Bryon, Bjarne Rosvoll Bøklepp, Lasse Lovstakken, Jørgen Avdal, Stefano Fiorentini and Svein-Erik Måsøy, "Subpixel segmentation of borehole fractures from low resolution Doppler ultrasound images using machine learning", *Geoenergy Science and Engineering*, Volume 235, pages 212703, 2024.

Conference proceedings

- Vincent Bryon, **Sigurd Vangen Wifstad**, Thomas Grønli, Jieyu Hu and Lasse Løvstakken, "Automated Patient-Specific Left Ventricular Simulations for Cardiac Function Evaluation Using Image-Based Computational Fluid Dynamics and Color Flow Imaging," *2023 IEEE International Ultrasonics Symposium (IUS)*, Montreal, Canada, 2023, pages 1-4,

Presentations and abstracts

- **Sigurd Vangen Wifstad**, Henrik Agerup Kildahl, Espen Holte, Erik Andreas Rye Berg, Bjørnar Grenne, Håvard Dalen and Lasse Lovstakken, "Automatic valve quantification using deep learning", *Meeting on Myocardial Function Imaging*, Leuven, Belgium, 2024.

- **Sigurd Vangen Wifstad**, Henrik Agerup Kildahl, Espen Holte, Bjørnar Grenne, Håvard Dalen and Lasse Lovstakken, " Automatic Jet Flow Convergence Detection and Quantification from 2D Color Doppler Flow Using Deep Learning", *IEEE International Ultrasonics Symposium (IUS)*, Montreal, Canada, 2023.

- **Sigurd Vangen Wifstad**, Shivanandan Indimath, Vincent Bryon, Jørgen Avdal, Lasse Lovstakken and Svein-Erik Måsøy, "Can we use Deep Learning to extract important information from poor resolution Doppler images?", *The Artimino conference on Medical Ultrasound Technology*, Artimino, Italy, 2023.

- **Sigurd Vangen Wifstad**, Ståle Wågen Hauge, Sigvard Johansen Seljelv, Sigbjørn Sæbø, Desalew Mekonnen, Berhanu Nega, Sintayehu Abebe, Mette-Elise Estensen, Rune Haaverstad, Håvard Dalen, Lasse Lovstakken, "Tracking heart valve motion from transthoracic echocardiography using deep learning", *IEEE International Ultrasonics Symposium (IUS)*, Venice, Italy, 2022.

- **Sigurd Vangen Wifstad**, Lasse Løvstakken, Jørgen Avdal, Hans Torp and Stefano Fiorentini, "Automatic quantification of valvular insufficiency using 3D Doppler imaging and Deep Learning", *IEEE International Ultrasonics Symposium (IUS)*, Virtual, 2021.

1.6 Thesis outline

The technical and clinical background concepts needed to understand the work discussed in this thesis are presented in Chapter 2. Two contributions are presented in their published form in Chapters 3 and 5. The contribution in Chapter 4 is presented in its submitted form prior to the review of the scientific journal.

References

[1] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, *et al.*, "Heart disease and stroke statistics-2023 update: A report from the american heart association," *Circulation*, vol. 147, no. 8, E93–E621, 2023.

[2] I. Edler and C. H. Hertz, "The use of ultrasonic reflectoscope for the continuous recording of the movements of heart walls," *Clinical Physiology and Functional Imaging*, vol. 24, no. 3, pp. 118–136, 1954.

[3] J. Wild and D. Neal, "Use of high-frequency ultrasonic waves for detecting changes of texture in living tissues," *The Lancet*, vol. 257, no. 6656, pp. 655–657, 1951.

[4] K. Tanaka, T. Wagai, Y. Kikuchi, and R. Uchida, "Studies on detection of intracranial anatomical pathology by ultrasound," *The Journal of the Acoustical Society of Japan*, vol. 8, 1952.

[5] S. Satomura, "Ultrasonic Doppler method for the inspection of cardiac functions," *The Journal of the Acoustical Society of America*, vol. 29, no. 11, pp. 1181–1185, 1957.

[6] A. Vahanian, F. Beyersdorf, F. Praz, *et al.*, "2021 ESC/EACTS guidelines for the management of valvular heart disease," *European Heart Journal*, vol. 43, no. 7, pp. 561–632, 2022.

[7] M. D. Grant, R. D. Mann, S. D. Kristenson, *et al.*, "Transthoracic echocardiography: Beginner's guide with emphasis on blind spots as identified with CT and MRI," *RadioGraphics*, vol. 41, no. 4, pp. 1022–1042, 2021.

[8] B. Iung, V. Delgado, R. Rosenhek, *et al.*, "Contemporary presentation and management of valvular heart disease the EURObservational research programme valvular heart disease II survey," *Circulation*, vol. 140, no. 14, pp. 1156–1169, 2019.

[9] 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.

[10] V. T. Nkomo, "Chronic primary mitral regurgitation: General management," *UpToDate*, 2023.

[11] C. J. Taylor, J. M. Ordonez-Mena, N. R. Jones, *et al.*, "Survival of people with valvular heart disease in a large, english community-based cohort study," *Heart*, vol. 107, no. 16, pp. 1336–1343, 2021.

[12] C. M. Otto, R. A. Nishimura, R. O. Bonow, *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," *Circulation*, vol. 143, no. 5, pp. 35–71, 2021.

[13] S. Robinson, L. Ring, D. X. Augustine, *et al.*, "The assessment of mitral valve disease: A guideline from the british society of echocardiography," *Echo Res Pract*, vol. 8, no. 1, G87–G136, 2021.

[14] A. Hagendorff, F. Knebel, A. Helfen, *et al.*, "Echocardiographic assessment of mitral regurgitation: Discussion of practical and methodologic aspects of severity quantification to improve diagnostic conclusiveness," *Clinical Research in Cardiology*, vol. 110, no. 11, pp. 1704–1733, 2021.

[15] T. Brown, B. Mann, N. Ryder, *et al.*, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 1877–1901.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[17] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[18] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electronics*, vol. 9, no. 8, 2020.

[19] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, 2019.

[20] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10 850–10 869, 2023.

[21] M. Tokodi and A. Kovács, "A new hope for deep learning–based echocardiogram interpretation: The droids you were looking for," *Journal of the American College of Cardiology*, vol. 82, no. 20, pp. 1949–1952, 2023.

[22] M. U. Ghani, F. C. Meral, F. Vignon, and J.-l. Robert, "High frame-rate ultrasound imaging using deep learning beamforming," in *IEEE International Ultrasonics Symposium (IUS)*, ser. IEEE International Ultrasonics Symposium, IEEE, 2019, pp. 295–298.

[23] T. S. W. Stevens, F. C. Meral, J. Yu, I. Z. Apostolakis, J.-L. Robert, and R. J. G. van Sloun, "Dehazing ultrasound using diffusion models," *arXiv*, 2023.

[24] B. Liu, H. Chang, D. Yang, *et al.*, "A deep learning framework assisted echocardiography with diagnosis, lesion localization, phenogrouping heterogeneous disease, and anomaly detection," *Scientific Reports*, vol. 13, p. 3, 2023.

[25] J. Magne, P. Pibarot, F. Dagenais, Z. Hachicha, J. Dumesnil, and M. Sénéchal, "Preoperative posterior leaflet angle accurately predicts outcome after restrictive mitral valve annuloplasty for ischemic mitral regurgitation," *Circulation*, vol. 115, no. 6, pp. 782–791, 2007.

[26] J. J. Silbiger, "Mechanistic insights into ischemic mitral regurgitation: Echocardiographic and surgical implications," *Journal of the American Society of Echocardiography*, vol. 24, no. 7, pp. 707–719, 2011.

[27] 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.

[28] S. Fiorentini, E. A. R. Berg, H. Torp, S. Aakhus, and J. Avdal, "Quantification of flow rates and flow volumes in valve regurgitation using 3-d high frame-rate ultrasound," *IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 3, pp. 29–40, 2023.

[29] 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.

[30] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[31] 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.

[32] 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.

[33] A. Carpentier, "Cardiac valve surgery—the "french correction"," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 86, no. 3, pp. 323–337, 1983.

[34] J. Schlemper, O. Oktay, M. Schaap, *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[35] A. M. Calafiore, S. Gallina, M. Di Mauro, *et al.*, "Mitral valve procedure in dilated cardiomyopathy: Repair or replacement?" *Circulation*, vol. 71, no. 4, pp. 1146–52, 2001.

[36] S. Biner, A. Rafique, F. Rafii, *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.

[37] S. Uretsky, L. Gillam, R. Lang, *et al.*, "Discordance between echocardiography and MRI in the assessment of mitral regurgitation severity: A prospective multicenter trial," *Journal of the American College of Cardiology*, vol. 65, no. 11, pp. 1078–1088, 2015.

[38] A. B. Ziani, D. G. Latcu, S. Abadir, *et al.*, "Assessment of proximal isovelocity surface area (PISA) shape using three-dimensional echocardiography in a paediatric population with mitral regurgitation or ventricular shunt," *Archives of Cardiovascular Diseases*, vol. 102, no. 3, pp. 185–191, 2009.

[39] F. P. Schmidt, T. Gniewosz, A. Jabs, *et al.*, "Usefulness of 3D-PISA as compared to guideline endorsed parameters for mitral regurgitation quantification," *International Journal of Cardiovascular Imaging*, vol. 30, no. 8, pp. 1501–1508, 2014.

[40] 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.

[41] S. Thomas, A. Gilbert, and G. Ben-Yosef, "Light-weight spatio-temporal graphs for segmentation and ejection fraction prediction in cardiac ultrasound," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Cham: Springer Nature Switzerland, 2022, pp. 380–390.

[42] A. Nibali, Z. He, S. Morgan, and L. Prendergast, "Numerical coordinate regression with convolutional neural networks," *arXiv*, 2018.

[43] H. Wei, H. Cao, Y. Cao, *et al.*, "Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Springer International Publishing, 2020, pp. 623–632.

[44] J. Hu, E. Smistad, I. M. Salte, H. Dalen, and L. Lovstakken, "Exploiting temporal information in echocardiography for improved image segmentation," in *2022 IEEE International Ultrasonics Symposium (IUS)*, 2022, pp. 1–4.

[45] H. J. Ling, N. Painchaud, P.-Y. Courand, P.-M. Jodoin, D. Garcia, and O. Bernard, "Extraction of volumetric indices from echocardiography: Which deep learning solution for clinical use?" In *Functional Imaging and Modeling of the Heart*, Springer Nature Switzerland, 2023, pp. 245–254.

[46] S. R. Snare, O. C. Mjølstad, F. Orderud, H. Dalen, and H. Torp, "Automated septum thickness measurement—a kalman filter approach," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 477–486, 2012.

[47] M. Sultan, N. Martins, E. Costa, *et al.*, "Virtual M-mode for echocardiography: A new approach for the segmentation of the anterior mitral leaflet," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 2018.

[48] H. Dröge, B. Yuan, R. Llerena, J. T. Yen, M. Moeller, and A. L. Bertozzi, "Mitral valve segmentation using robust nonnegative matrix factorization," *Journal of Imaging*, vol. 7, no. 10, 2021.

[49] T. Qin, A. Caballero, R. T. Hahn, R. McKay, and W. Sun, "Computational analysis of virtual echocardiographic assessment of functional mitral regurgitation for validation of proximal isovelocity surface area methods," *Journal of the American Society of Echocardiography*, vol. 34, no. 11, pp. 1211–1223, 2021.

[50] Y. Topilsky, H. Michelena, V. Bichara, J. Maalouf, D. W. Mahoney, and M. Enriquez-Sarano, "Mitral valve prolapse with mid-late systolic mitral regurgitation: Pitfalls of evaluation and clinical outcome compared with holosystolic regurgitation," *Circulation*, vol. 125, no. 13, pp. 1643–1651, 2012.

[51] S. Bjaerum, H. Torp, and K. Kristoffersen, "Clutter filters adapted to tissue motion in ultrasound color flow imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, no. 6, pp. 693–704, 2002.

[52] Y. Xu, K.-H. Yiu, and W.-N. Lee, "Fast and robust clutter filtering in ultrafast echocardiography," *Ultrasound in Medicine & Biology*, vol. 49, no. 2, pp. 441–453, 2023.

[53] 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.

[54] J. Provost, C. Papadacci, J. Arango, *et al.*, "3d ultrafast ultrasound imaging in vivo," *Physics in Medicine and Biology*, vol. 59, p. L1, 2014.

[55] 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.

[56] A. Wang, P. Grayburn, J. A. Foster, *et al.*, "Practice gaps in the care of mitral valve regurgitation: Insights from the american college of cardiology mitral regurgitation gap analysis and advisory panel," *American Heart Journal*, vol. 172, pp. 70–79, 2016.

[57] N. Watanabe, "Acute mitral regurgitation," *Heart*, vol. 105, no. 9, pp. 671–677, 2019.

[58] A. Vahanian, F. Beyersdorf, F. Praz, *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)," *European Heart Journal*, vol. 43, no. 7, pp. 561–632, 2021.

[59] M. Vafaeezadeh, H. Behnam, A. Hosseinsabet, and P. Gifani, "Carpnet: Transformer for mitral valve disease classification in echocardiographic videos," *International Journal of Imaging Systems and Technology*, 2023.

[60] V. Roediger, *Autostrain LV/RV/LA - automated strain measurements*, Koninklijke Philips N.V., 2019.

[61] K. McLeod and J. Sprem, *The role of artificial intelligence in streamlining echocardiography quantification*, GE Healthcare, 2022.

Chapter 2

Background

This chapter describes essential background knowledge that is needed to understand the context, the methods and the results presented in this thesis. The chapter begins with a basic introduction to medical ultrasound, including important physical principles and different imaging modalities. The content is largely based on the book *Ultrasound Imaging: Waves, Signals, and Signal Processing* by Anglesen [1], which contains a more comprehensive description of ultrasound technology. Secondly, a description of the heart and heart valve anatomies and valvular heart disease is presented. Thirdly, an excerpt of the recommended integrative approach for the assessment of mitral regurgitation using TTE imaging, according to the American Society of Echocardiography 2017 guidelines [2], is presented. This section aims to give the reader an impression of the extensive workload that is required, and the limitations met when assessing valvular heart disease, hence motivating innovations for automation and improved accuracy. A comprehensive description of the assessment of different types of VHD for all valves can be found in [2]–[5]. Lastly, a brief introduction to the field of deep learning for medical image segmentation is presented. This section is largely based on the book *Deep Learning* by Goodfellow, Bengio, and Corville [6], which provides further details on the topic.

2.1 Echocardiography

Basic principles of medical ultrasound

Ultrasound refers to high frequency sound waves and is widely applied in medicine for imaging of soft tissues. Imaging is possible due to ultrasonic transitivity and reflectivity at the borders between various tissue types, causing reflections at the borders between tissues with differences in *acoustical*

impedance. We define the impedance Z of some material as the product of the speed of sound c in the material, and the mass density ρ:

$$Z = c\rho. \tag{2.1}$$

The acoustic impedance of some different materials [7] are listed in Table 2.1. Figure 2.1 illustrates *specular reflection* of an acoustic wave. At the border between two different materials with impedances Z_1 and Z_2, the acoustic wave will be reflected with magnitude Γ, and further transmitted with magnitude $1 - \Gamma$, where Γ is the reflection coefficient, defined as:

$$\Gamma = \frac{Z_1 - Z_2}{Z_1 + Z_2}. \tag{2.2}$$

Consequently, Γ is large at the borders between materials with large differences in impedance, such as muscle and blood, which for instance results in strong signals from cardiac walls. Conversely, regions where the impedance is homogeneous, like within blood-filled cavities, instead cause weak reflections, and more of the wave energy is propagated further into the medium. Materials with substantial differences in impedance from soft tissues will cause very strong reflections, limiting the use of ultrasound from effectively propagating through such boundaries. According to Table 2.1 and (2.2), Γ is about 0.1 between muscle and fat, meaning that 90% of the energy is transmitted further into the tissue. In comparison, Γ is about 0.64 between bone and muscle, and 0.999 between air and muscle, meaning most of the energy would be reflected at the boundary.

When the ultrasonic wave hits objects that are smaller than its wavelength *scattering*, also known as *diffuse reflection*, will occur. Scattering is the phenomenon where the object absorbs the energy of the incident wave and emits it back in all directions. Such objects are hence referred to as

Material	Acoustic impedance [kg/s/m^2]$\times 10^6$
Air	0.0004
Bone	7.75
Water	1.48
Blood	1.66
Fat	1.38
Muscle	1.68

Table 2.1: Acoustic impedance for different materials.

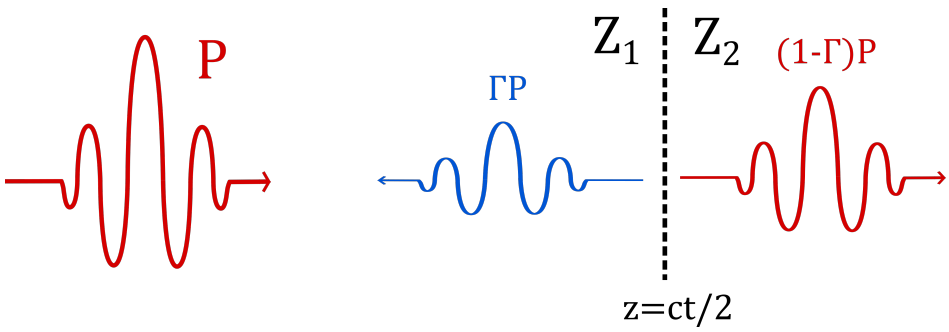


Figure 2.1: Illustration of reflection and transmission of an ultrasonic pulse with amplitude P at the boundary between two impedances Z_1 and Z_2, located at depth z. The amplitudes of the reflected and transmitted waves are reduced by factors of Γ and $1 - \Gamma$ respectively.

scatterers. In medical ultrasound, most tissues are irregular at the surface level and therefore cause scattering from microscopic particles in addition to specular reflection. Specular reflection is unidirectional based on the angle between the incident wave and the surface and can only be received from surfaces perpendicular to the ultrasound beam. This means that most of the received signal in fact originates from scattering.

Moreover, scattering occurs not only on tissue surfaces but also in blood due to the scattering of red blood cells, which constitute the majority of the blood volume. The red blood cell scatterers are densely packed in the blood and therefore interact with each other and can be seen as a continuum, where scattering occurs from fluctuations in mass density and compressability. However, since the mass density and compressibility are similar to those of the surrounding blood plasma the scattering signal is relatively weak compared to scattering from tissue surfaces [8]. The scattering strength of tissue can be as much as 100 dB higher than scattering from blood, which challenges blood flow imaging [9].

By transmitting an ultrasound pulse and recording the received echos from the transmission it is possible to estimate the depth z from which the echos were received according to

$$z = \frac{ct}{2}, \tag{2.3}$$

where t is the time between transmission and the reception of the echo, and c is the speed of sound, which is often assumed to have a constant value of 1540 m/s for soft tissues. Using (2.3), the recorded signal can then be

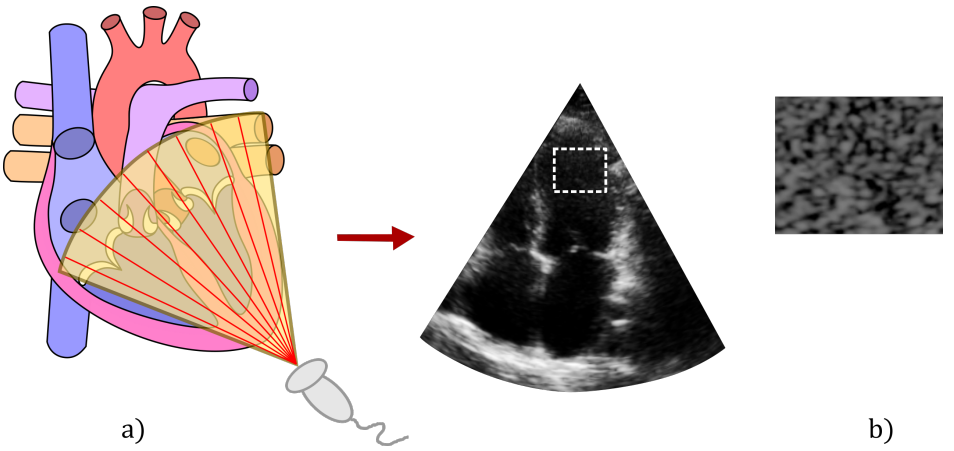


Figure 2.2: a): Formation of a two-dimensional B-mode image. Ultrasonic pulses are transmitted and received in several scan lines. The received signals are mapped to spatial coordinates, forming an image depicting scatterer intensities in a two-dimensional cross-section of the tissue. b): Closeup showing the *speckle pattern* caused by interference of scattering from particles.

visualized as an image where one axis represents depth, and the intensity of the signal represents the material properties at a given depth.

Two-dimensional images depicting scatterer intensity at given depths and lateral positions can be acquired by scanning a sector of an area with repeated ultrasonic transmissions. The image resulting from this process is known as a B-mode image, as depicted in Figure 2.2. Similarly, three-dimensional images can be obtained by additional scanning along a second axis.

The time required to form one B-mode image is fundamentally limited by the propagation time of the transmitted and reflected waves in the medium. To fully form one image the required time T is given by

$$T = \frac{Nz}{2c}, \tag{2.4}$$

where N is the number of scan lines. Thus the maximum achievable frame rate $1/T$ is limited by both depth and the amount of scan lines used. The frame rate limits the detail at which tissue movement can be registered. Moreover, for 3-D ultrasound many more scan lines are required, which limits the achievable frame rate of 3-D imaging.

The detail at which the tissue can be visualized is limited in the radial direction by the length of the ultrasound pulse, given by the center frequency

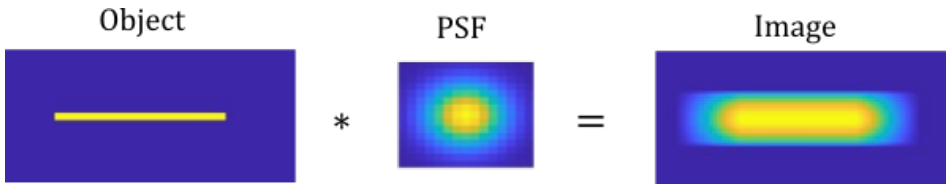


Figure 2.3: Image formation as the convolution (denoted by the operator $*$) between the object and the PSF. The resulting image is a blurred rendition of the underlying object.

and number of cycles of the pulse. In theory, any object cannot be accurately resolved at a better spatial resolution than half the pulse length of the transmitted ultrasound pulse.

To be able to resolve small features radially it is beneficial to use transmit frequencies as high as possible. However, increasing frequency decreases the penetration depth of the transmitted signal. This is due to attenuation μ of the signal amplitude, which follows the relation

$$\mu = \alpha f z, \tag{2.5}$$

where f is the frequency, z is the depth and α is the attenuation coefficient for the material. It follows from (2.5) that high-frequency signals will be heavily attenuated at larger depths, ultimately limiting the achievable radial resolution for imaging deeper tissues.

In the lateral direction, the resolution R is limited by the width of the focused ultrasound beam, conventionally defined as

$$R = f_{\#}\lambda, \tag{2.6}$$

where λ is the transmit wavelength and $f_{\#} = F/D$ is the f-number, defined as the ratio between the focal depth F and the aperture size D. According to (2.6) increasing the aperture size improves lateral resolution. However, aperture size is limited in practice due to technological constraints, as well as physical constraints. The latter is especially true for TTE imaging, as the entire aperture needs to fit in the intercostal space of the patient.

When imaging close to the resolution limits objects become blurred, making it difficult to accurately discern the source of the received signal. The degree to which the ultrasound system blurs the imaging object can be described using the point spread function (PSF) of the imaging system. The PSF describes the equivalent received image of a single point scatterer

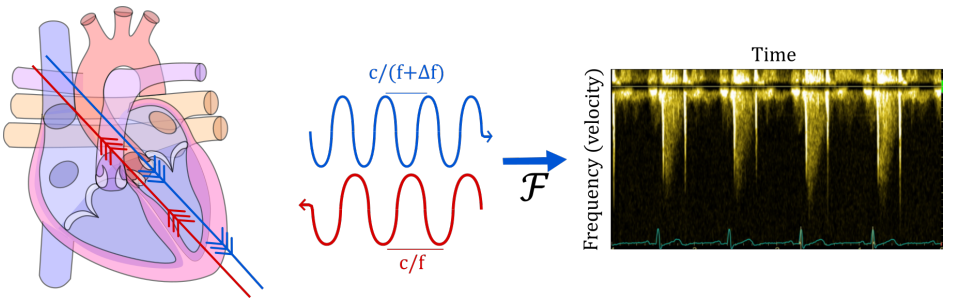


Figure 2.4: Continuous Wave Doppler imaging. A single-frequency continuous wave is transmitted in one direction. The received signal is frequency-shifted according to the Doppler effect. The velocity corresponding to the Doppler shift can be represented using a spectrogram computed via a Fourier transform ($\mathcal{F}$) of the demodulated received signal, showing the frequency content of the received signal at a given time point.

and defines the spatial resolution limits of the system. Consequently, the features of any object that are smaller than the resolution limits given by the PSF will be subject to blurring, as illustrated by Figure 2.3.

Doppler imaging

In addition to imaging tissue structures, ultrasound can also be used for measuring the velocity of moving objects and materials, such as blood. One of the ways this can be achieved is via the Doppler effect. The ultrasound wave transmitted with frequency f hitting an object moving with velocity v will cause a pulse-echo frequency shift Δf in the received signal according to

$$\Delta f = \frac{2f}{c} v \cos\theta, \tag{2.7}$$

where c is the speed of sound in the material and θ is the angle between the velocity vector and the incident ultrasonic wave. The frequency shift Δf is hence linearly related to the radial velocity component of the object. The Doppler effect is utilized in Continuous Wave (CW) Doppler imaging, as illustrated in Figure 2.4. With CW Doppler imaging it is possible to measure the velocity distribution over time in the transmit direction. However, since waves are continuously transmitted and received, it is not possible to determine a time delay between transmit and receive, and hence there is no notion of depth/radial resolution.

Velocity estimation is possible while maintaining radial resolution by using pulsed wave (PW) Doppler imaging. Unlike what the name suggests,

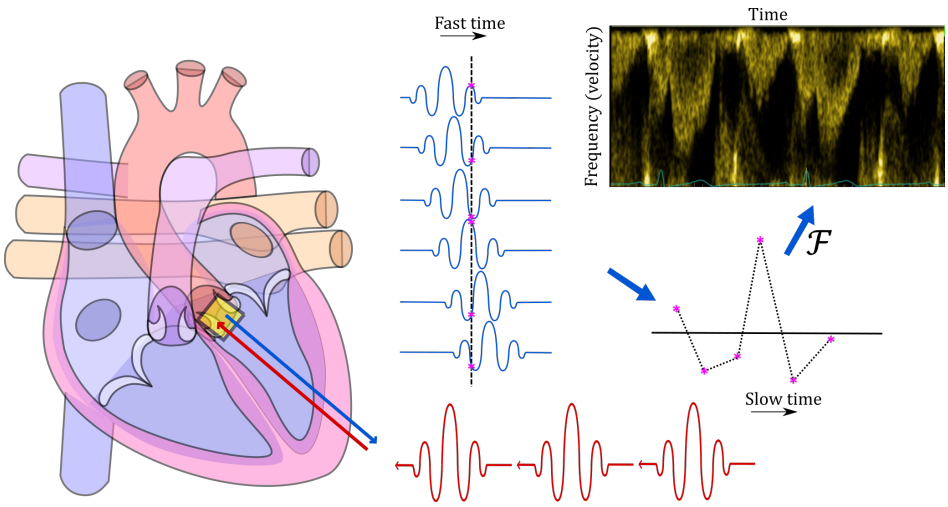


Figure 2.5: Pulsed Wave Doppler imaging. Consecutive ultrasonic pulses are transmitted and received. Each received pulse is sampled at the depth of interest, corresponding to a point in *fast time*. Consecutive fast time samples form a phase-shift signal in *slow time*, from which a velocity spectrogram can be derived via the Fourier transform ($\mathcal{F}$).

PW Doppler imaging does not directly rely on the Doppler effect for velocity estimation. Instead, PW Doppler imaging is based on phase-shift estimation of consecutive pulses, as illustrated in Figure 2.5. By transmitting consecutive pulses at a moving scatterer, the phase shift of the received pulses will be proportional to the scatterer displacement between pulses. Assuming $v << c$, it can be shown that the phase shift $\Delta\phi$ is given by

$$\Delta\phi = \frac{4\pi v f \cos\theta}{c \cdot \mathrm{PRF}}, \tag{2.8}$$

where PRF is the pulse repetition frequency. Combining (2.7) and (2.8) we can express the equivalent Doppler shift as

$$\Delta f = \frac{\mathrm{PRF}}{2\pi}\Delta\phi. \tag{2.9}$$

PW Doppler enables velocity imaging with a notion of spatial locality. However, the phase-estimation technique introduces an upper limit for measurable velocity. This limit, known as the Nyquist limit, is given by the aliasing velocity v_{nq}, which can be expressed as

$$v_{nq} = \frac{c \cdot \text{PRF}}{4f}, \tag{2.10}$$

Velocities that exceed the Nyquist limit will be subject to phase-wrapping, where the apparent velocity wraps around the available velocity range. This can be seen in the Doppler spectrum in Figure 2.5. On the other hand, aliasing does not occur with CW Doppler, hence there is no phase-unwrapping of the velocity spectrum in Figure 2.4. Therefore CW Doppler is more suitable than PW Doppler for measuring very high velocities, for instance, heart valve regurgitation jets, which typically have a peak velocity of 5-6 m/s [10].

Following (2.10), the velocity range with measurable PW Doppler can be extended by increasing the PRF, however, due to transmit and receive times the maximal achievable PRF is given by

$$\text{PRF} < \frac{c}{2z}. \tag{2.11}$$

Increasing the PRF above this limit would result in spatial ambiguity, as a given pulse would be transmitted before the previous one is received. This is known as high PRF (HPRF) mode [11].

Color-Doppler imaging

The phase-estimation technique applied in PW Doppler can also be extended in multiple spatial dimensions. In color-Doppler, also known as Color Flow Imaging, the mean velocities from sample volumes are displayed at their respective spatial locations using a color map representing mean velocity magnitude and direction. Color-Doppler signals are often displayed on top of B-Mode images, resulting in a *duplex image*, displaying both tissue and velocity information. Color-Doppler imaging is illustrated in Figure 2.6.

Color-Doppler imaging enables velocity estimation in two or three spatial dimensions, but practical trade-offs limit both the resolution and frame rate compared to B-mode imaging. As explained earlier, higher transmit frequency results in improved radial resolution, but also leads to increased attenuation. Pulse echo signals from blood flow are typically weak compared to signals from tissue imaging. Therefore, color-Doppler transmit frequencies are usually lower than for B-Mode imaging in order to maintain sufficient penetration depth, and consequently leads to reduced spatial resolution. Moreover, as each duplex image is formed from multiple pulses of both color-Doppler and B-Mode transmissions in multiple scan lines, the high total transmit and receive time limits the frame rate.

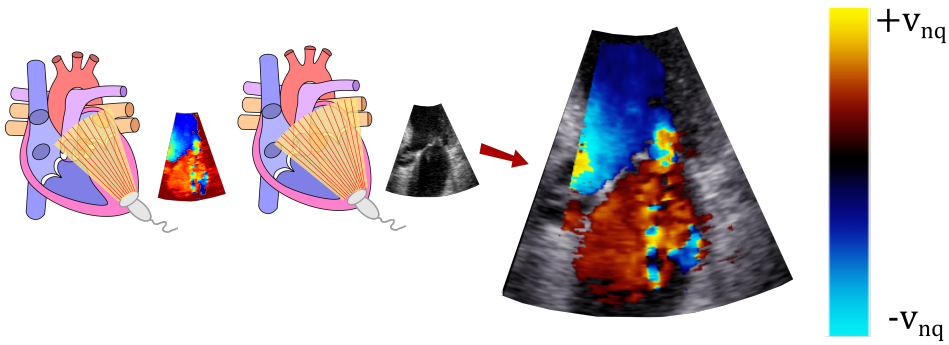


Figure 2.6: Color-Doppler imaging. PW phase estimation is performed in multiple depths for multiple scan lines. The mean velocity in each sample volume is estimated and displayed using a red-blue color map, indicating the velocity magnitude and direction within the available velocity range, given by the aliasing velocity v_{nq}. A B-mode scan is performed and is combined with the color-Doppler image to form a duplex image, where a signal strength arbitration scheme is used in each pixel to choose from which image information is displayed.

Color-Doppler imaging is also subject to aliasing, limiting the measurable range of velocities. The duplex image shown in Figure 2.6 exhibits aliasing in the regions with the highest velocity. The aliasing artifacts are manifested as sudden transitions between red/yellow and blue, as the apparent velocity exceeds the Nyquist limit and thus wraps around the velocity range.

The Nyquist limit acts as an upper bound for measurable velocities. In practice, there is also a lower bound for measurable velocity determined by the presence of *clutter*, which is velocity noise from moving tissues. Since the tissue signal strength is 40-100 dB higher than the blood signal [9], the clutter needs to be removed using a clutter filter, which is illustrated in Figure 2.7. Clutter filtering can be a significant challenge when imaging low-velocity flows, or imaging regions with a presence of high-velocity clutter. For instance, heart valves open with high velocities, which could result in strong clutter that cannot accurately be separated from the blood signal.

Finally, all the aforementioned Doppler imaging modalities are limited to measuring the *radial* component of the velocity vector. Consequently, there is considerable loss of signal for blood moving at high angles relative to the ultrasound transmit beam. This is a major limitation of Doppler techniques, as intracardiac blood flow in reality is largely three-dimensional. Techniques for estimating the multiple components of blood flow are known as vector flow imaging techniques [12]. However, these techniques are associated with a separate set of challenges and are out of the scope of this thesis.

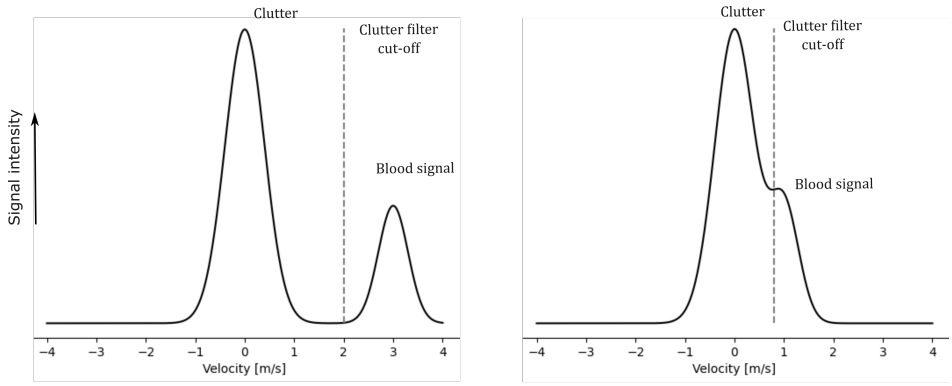


Figure 2.7: Velocity distributions of two different received signals (a and b). In both cases, there is a strong presence of low-velocity clutter. In signal a, the blood component is centered around 3 m/s, and can easily be separated from the clutter using a clutter filter. In signal b, the blood component is centered around 1 m/s. In this case, it is challenging to separate the blood and clutter components without also suppressing parts of the blood component.

2.2 Valvular heart disease

The heart is an organ essential for our survival. Beating about once every second, the heart continuously sends oxygenated blood and nutrients to each cell in the body, while simultaneously pumping deoxygenated blood to the lungs. The heart is illustrated in Figure 2.8 a). The heart consists of four chambers (right and left ventricles and atria), connected by valves, veins, and arteries. Oxygenated blood enters the left atrium from the lungs through the pulmonary veins. During diastole (relaxation) the mitral valve opens and the aortic valve closes, causing blood to flood into the left ventricle from the left atrium. During systole (contraction). The mitral valve closes, the aortic valve opens, and the ventricle contracts, pumping the blood into the aorta, and transporting it out to the body. Similar mechanisms occur simultaneously in the right chambers, moving deoxygenated blood to the lungs via the pulmonary artery.

The valves are imperative for heart function. When closed, they prevent blood from flowing in the wrong direction, hence maintaining the cardiac output. Figure 2.8 b) shows the anatomy of the mitral valve. The mitral valve consists of two leaflets, one on the anterior (A) and one on the posterior (P) side of the left ventricle. Each leaflet is conventionally subdivided into *scallops* 1-3, according to Carpentier's classification [14]. The leaflets are

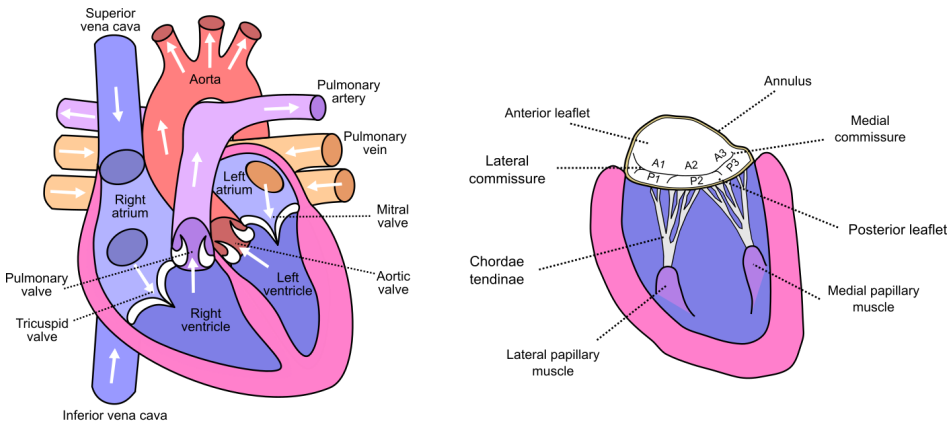


Figure 2.8: a): Illustration of the heart [13], showing the four heart chambers, valves, and interconnecting veins and arteries. b): Illustration of the mitral valve, showing the leaflets, the annulus, the papillary muscles, and the chordae tendinae.

inserted in a fibrous ring along the heart wall known as the *annulus*. In addition, the leaflets are attached to *papillary muscles* with fibrous strings known as *chordae tendinae*, which allow the valve to close instead of swinging back into the left atrium in systole [15].

Valvular heart disease (VHD) is a class of diseases that affect one or several parts of the valve apparatus, causing deviations from normal function. In moderate to severe cases, VHD ultimately leads to reduced cardiac output and is associated with high morbidity and mortality [16], [17]. VHD can be categorized into two main categories: *valve regurgitation*, where improper coaptation of the leaflets causes leakage through the closed valve, and *valve stenosis*, where restricted leaflet motion causes narrowing of the valve orifice when open. Valve disease can be caused either by degenerative or indirect changes to the valve apparatus or by rheumatic heart disease. Rheumatic mitral valve disease, where the mitral valve is scarred following an inflammation, is rare in high-income countries but is still prevalent in low-income countries [16], [18].

Valve regurgitation

Valve regurgitation, also known as valve insufficiency, can be caused directly by changes in the valve apparatus, or indirectly by changes in the surrounding heart tissue [3]. These two cases are respectively categorized into *functional* (also called secondary) insufficiency, and *degenerative* (also called primary) insufficiency [19]. Figure 2.9 shows examples of functional

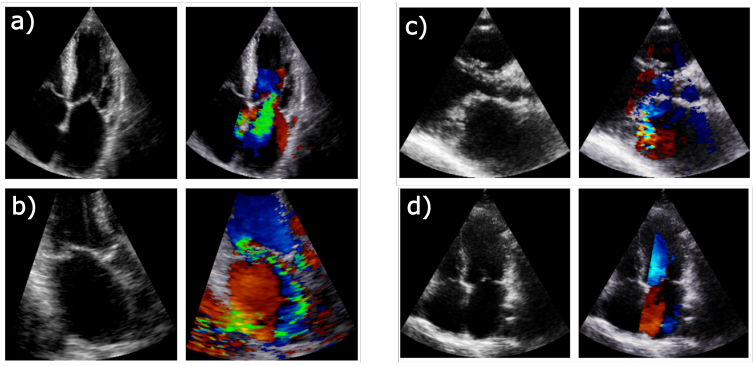


Figure 2.9: a-c): Three cases of mitral regurgitation showing B-mode images on the left and the color-Doppler images on the right. The regurgitant jet is visible as a turbulent signal flowing from the mitral valve into the left atrium. a): Functional insufficiency caused by dilation, which pulls the leaflets apart and creates a regurgitant orifice along the coaptation line. b): Degenerative insufficiency caused by prolapse of the posterior leaflet, causing an orifice at the afflicted scallop (P2), which results in an eccentric jet traveling along the atrial wall. c): Degenerative insufficiency caused by rheumatic valve thickening. d): A healthy reference with no visible regurgitant jet.

and degenerative insufficiency. Degenerative insufficiency can be caused by perforation of the leaflet tissue [14]. However, it is more commonly caused by dysfunction of papillary muscles, or lengthening or rupture of some of the chordae tendinae. This may cause improper coaptation of the leaflets, creating an orifice between the leaflet tips. This is known as leaflet *prolapse* [20], [21]. Rheumatic heart disease can also cause degenerative insufficiency, due to scarring deformations of the leaflet tissue causing improper closure of the valve.

Valve stenosis

Heart valve stenosis refers to the condition where restricted movement of the valve leaflets causes a narrowing of the valve orifice. The narrowing causes a pressure increase in the afflicted heart chambers and may lead to heart failure in severe cases [17], [22]. In high-income countries, stenosis is most common in the elderly population and is typically caused by calcification of the valve. In low-income countries, stenosis is most commonly caused by rheumatic heart disease. Aortic stenosis is the most common type of VHD. Mitral stenosis is less common in high-income countries, however is commonly caused by rheumatic heart disease [3].

2.3 Echocardiographic assessment of mitral valve regurgitation

The integrative approach

Using a combination of B-mode and Doppler imaging, as described in Section 2.1, it is possible to both detect and quantify mitral regurgitation (MR). Experienced operators can in most cases detect abnormalities in valve function, however, accurate disease assessment is difficult since there currently does not exist gold any standard for disease quantification. Consequently, the recommended approach consists of an integrative approach of several quantitative, semi-quantitative, and qualitative measurements.

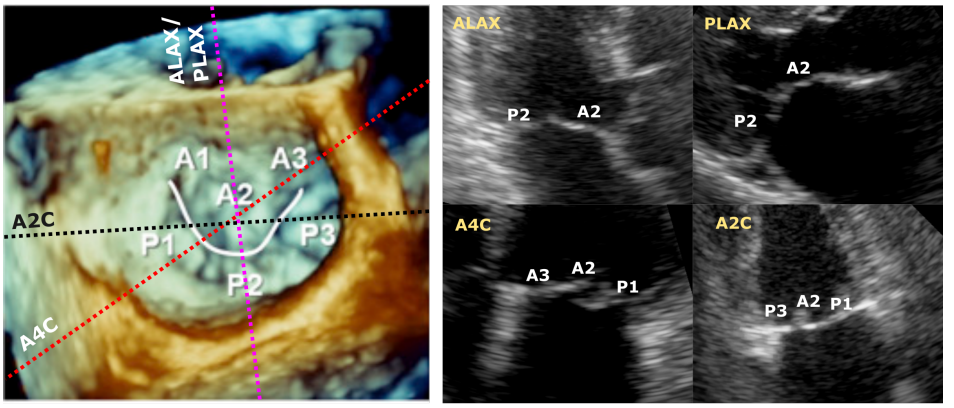
Qualitative and semi-quantitative assessment

The integrative approach is largely based on qualitative and semi-quantitative measurements. In this setting, semi-quantitative refers to quantitative measurements of parameters other than regurgitant flow volumes that are associated with MR.

Qualitative assessment of valve function and morphology is an important step of the integrative approach. Using B-mode imaging, the operator can visualize the morphology of the mitral valve (MV), as shown in Figure 2.10. High-resolution images can be obtained using invasive TEE imaging, as the ultrasound probe can be placed close to the atrium from the esophagus. However, when using TTE imaging the MV morphology is usually better visualized using 2-D B-mode imaging due to limitations of resolution and frame rate, as explained in Section 2.1. The 3-D structure of the MV hence needs to be inferred from multiple standard views.

Visualizing the MV morphology aids the operator in detecting the presence of VHD and describes the underlying mechanisms of the condition. Conditions such as leaflet prolapse, leaflet billowing, restricted leaflet movement, annular dilatation, and chordae rupture can be identified by experienced operators. The morphology and function of the MV are associated with the presence, location, severity, and characteristics of MR jets [14]. Morphological and functional considerations are thus important for correctly interpreting further measurements.

Doppler imaging is suitable for visualizing and measuring the flow dynamics of MR jets. Consequently, many of the integrative measurements are based on both CW, PW, and color-Doppler imaging. The CW Doppler spectral density (Figure 2.11 a)) can be used to qualitatively assess the intensity of the MR jet signal, as higher flow volumes are associated with a higher density of blood cell scatterers. Moreover, measurements of the CW



a) b)

Figure 2.10: B-mode images of the mitral valve. The leaflet scallops (A,P; 1,2,3) are annotated. a): 3-D TEE image showing a top-down view, as seen from the left atrium (adapted from [2]). b): 2-D TTE images showing different cross-sectional standard views of the mitral valve (ALAX: apical long-axis, PLAX: parasternal long-axis, A4C: apical four-chamber, A2C: apical two-chamber). The standard view imaging planes are also displayed in a).

Doppler spectrogram shape, duration, and peak velocity can be important indicators of the severity of the condition.

PW Doppler can be used to measure the diastolic mitral inflow by placing the sample volume at the MV leaflet tips. Although not a direct measurement of MR flow, mitral inflow patterns can be used to distinguish between mild and severe cases of MR. The sample volume can also be placed in the pulmonary vein to measure flow reversal during systole (Figure 2.11 b)). Large MR jets can force blood into the pulmonary veins, hence considerable presence of pulmonary flow reversal can be used to indicate severe MR.

Color-Doppler imaging is useful for visualizing the jet flow dynamics and location. Moreover, the apparent size and shape of the jet can give indications about severity. The vena contracta width (Figure 2.11 c)), which is the width of the narrowest part of the jet at the efflux, is associated with the size of the regurgitant orifice. Moreover, for central jets, the jet area relative to the left atrium area (Figure 2.11 d)) is associated with the regurgitant flow volume.

Finally, 3-D color-Doppler imaging can be utilized to visualize the vena contracta area using a direct planimetry (Figure 2.11 e)). In comparison to the vena contracta width, the vena contracta area provides a more accurate estimate of the regurgitant orifice area in cases of non-circular orifices.

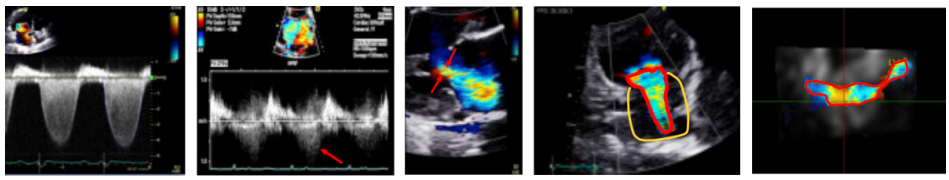


Figure 2.11: Various qualitative and semi-quantitative methods based on Doppler imaging, adapted from [2]. a): CW Doppler of the MR jet. b): PW Doppler showing pulmonary flow reversal. c:) Vena contracta width. d): MR jet and left atrium areas. e): Vena contracta area from 3-D color-Doppler (*en face* view from the left atrium using TEE imaging).

However, accurately tracing the vena contracta contours can be challenging due to the limited quality of 3-D color-Doppler images, as explained in Section 2.1.

Qualitative and semi-quantitative measurements can unveil important aspects of the patient's condition. However, each of the aforementioned Doppler-based methods is limited by several pitfalls, as described in the guidelines [2]–[5]. Consequently, the assessment of MR heavily relies on the experience and skill of the operator. This limits the accuracy of disease severity grading, and results in high inter-observer variability [23], [24].

Quantitative assessment

Quantitative methods aim to measure the total regurgitant volume (RVol) which leaks from the left ventricle to the left atrium during systole. According to the guidelines, there are three recommended quantitative methods for the assessment of MR; the *stroke volume method*, the *volumetric method*, and the *proximal isovelocity surface area method*.

Stroke volume and volumetric methods

The stroke volume method is based on a simple assumption of conservation of mass. The method estimates RVol indirectly by first measuring the mitral inflow stroke volume $\mathrm{SV_{MV}}$ and the aortic outflow stroke volume $\mathrm{SV_{LVOT}}$. Due to the conservation of mass, RVol can be estimated as the difference in stroke volumes:

$$\begin{aligned} \mathrm{RVol} &= \mathrm{SV_{MV}} - \mathrm{SV_{LVOT}} \\ &= \mathrm{CSA_{MV}} \cdot \mathrm{VTI_{MV}} - \mathrm{CSA_{LVOT}} \cdot \mathrm{VTI_{LVOT}}, \end{aligned} \tag{2.12}$$

where CSA and VTI are respectively the the cross sectional area and the velocity time integral of the mitral valve (MV) and the left ventricular outflow

tract (LVOT). The stroke volumes can be estimated using a combination of B-mode and PW Doppler measurements. Firstly, the cross-sectional areas are estimated by measuring the diameters of the MV and LVOT openings, assuming circular geometries. Secondly, velocity time integrals are estimated by tracing PW Doppler spectrograms measuring the flow through the MV and the LVOT.

The volumetric method is similar to the stroke volume method but measures the left ventricular (LV) stroke volume instead of the mitral inflow stroke volume, which in theory are equivalent. The LV stroke volume is estimated by tracing the LV contours in end-diastole and end-systole time points and then estimating the net volume change.

Both the stroke volume and volumetric methods are limited by requiring the integration of multiple measurements, each associated with several sources of error. The measurements of CSA are sensitive to small errors in diameter estimation. These errors are squared in the calculation of the area estimate, assuming a circular geometry. Moreover, diameter and LV estimates can be underestimated in case of imaging view misalignment. Furthermore, the PW Doppler sample volumes can be misplaced, and the spectrograms can be erroneously traced.

Proximal Isovelocity Surface Area (PISA)

Like the stroke volume and volumetric methods, the Proximal Isovelocity Surface Area (PISA) method is also based on the conservation of mass principle. However, the method does not estimate flow volumes from stroke volume differences, but rather directly from the jet influx on the valve surface. The PISA method is depicted in Figure 2.12. As blood converges towards the regurgitant orifice in the MV, velocity increases to above the aliasing velocity, at which the aliasing contours are clearly visible in the color-Doppler image. Assuming hemispherical convergence and conservation of mass, the flow rate Q through the orifice is equal to the hemispherical surface area of the flow convergence region times the velocity at the contour

$$Q = 2\pi r^2 v_{nq}, \tag{2.13}$$

where r is the flow convergence radius and v_{nq} is the aliasing velocity. To integrate temporal information, the maximum velocity envelope is traced in the CW Doppler spectrum which provides the peak velocity v_{max} and the velocity time integral (VTI). Consequently, the regurgitant volume RVol can be computed as

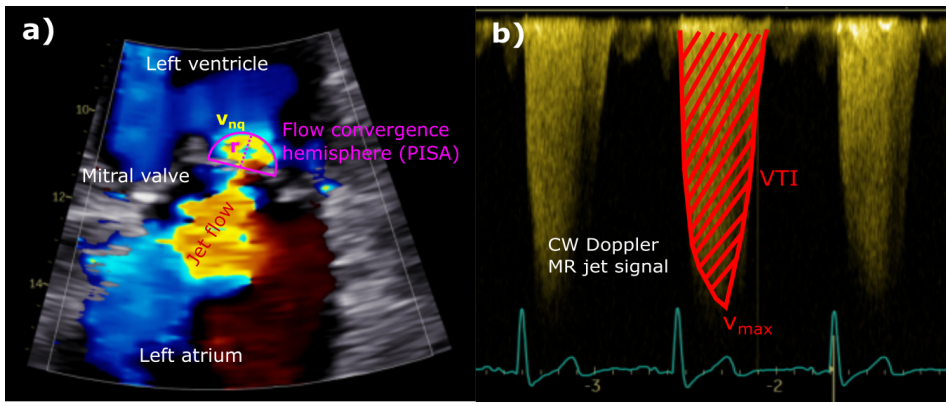


Figure 2.12: The PISA method for quantifying mitral regurgitation (MR). a): Flow convergence surface area estimation. The frame with the largest flow convergence is chosen, and the radius is traced assuming hemispherical convergence. b): CW Doppler of the regurgitant jet. The maximum velocity envelope is traced, leading to estimates of the peak velocity v_{max} and the velocity time integral (VTI).

$$\mathrm{RVol} = Q\frac{\mathrm{VTI}}{v_{\max}}. \tag{2.14}$$

The PISA method is hampered by several practical limitations, one of which is the assumption of hemispherical convergence of blood flow. The hemispherical assumption is inaccurate for cases of functional MR for which the orifice is often crescent-shaped (as can be seen in Figure 2.11 e)). Moreover, PISA is conventionally only computed from a single color-Doppler image, disregarding temporal dynamics of the flow convergence. Overestimation might therefore occur for cases of non-holosystolic jets, especially in case of diffuse CW Doppler spectrograms [25], [26], as illustrated by the weak Doppler spectrogram at end-diastole in Figure 2.12 b). Finally, the PISA method is associated with considerable inter-observer variability [27]. The manual PISA radius r is particularly prone to error, as the resulting RVol estimate is proportional to r^2. Furthermore, both the color-Doppler flow convergence size and the CW Doppler peak velocity are highly sensitive to imaging plane alignment, and insonation angle relative to the jet direction.

The future of valvular heart disease assessment

Echocardiography is a constantly progressing field of research, leading to technical and methodological advancements as a result of the hard labor from researchers and technologists worldwide. Amongst these advancements are

novel quantitative methods for VHD assessment which have been proposed in recent years, with promising results. Recent research [21], [28] has shown that quantifying the average pixel intensity of the CW Doppler spectrogram can provide a robust index for MR severity. Moreover, several methods based on flow line integration from 3-D color-Doppler imaging, as illustrated in Figure 2.13 a), have been proposed [29]–[31]. However, the methods were only demonstrated for TEE imaging. As explained in Section 2.1, TTE is associated with larger imaging depths, limiting the achievable image quality using TTE 3-D color-Doppler. Recent studies from our group demonstrated the possibility of using HPRF 3-D Doppler ultrasound for assessment of valve regurgitation in TTE imaging [11], [32]. However, many technical limitations still need to be addressed in order to achieve sufficiently robust measurements.

Furthermore, recent research has also shown promise for automation of currently recommended methods for VHD assessment, thus aiming to alleviate workload and reduce inter-observer variability. These methods are largely based on deep learning systems applied to image processing of standard echocardiograms. Notable examples include fully automated detection and semi-quantitative measurements of VHD from B-mode and 2-D color-Doppler [33] (Figure 2.13 b)), automated classification of VHD from B-mode video data [34], and automated PISA measurements from one-dimensional color-Doppler imaging [35]. Deep learning-based methods keep gaining attention in the field of echocardiography and hold promise for mitigating issues related to inter-observer variability and manual workload.

2.4 Deep learning for medical image analysis

Automated image analysis lays the foundation for improved workflow for clinicians and radiologists, ultimately improving patient care. In recent years, deep learning has shown to be especially effective for automating tasks, including tasks within the field of medical image analysis. Deep learning has been extensively applied to both disease classification and object detection and segmentation in X-ray images, cMRI, CT and ultrasound [36].

The terms *Artificial Intelligence* (AI), *Machine Learning* (ML) and *Deep Learning* (DL) are often used interchangeably. However, the terms are not strictly equivalent to each other. AI refers to systems that perform one or more tasks that require somewhat intelligent reasoning in order to be completed. This could be, for instance, visual recognition, language processing, or other advanced data analysis tasks. However, discussions about AI today almost exclusively involve systems based on ML.

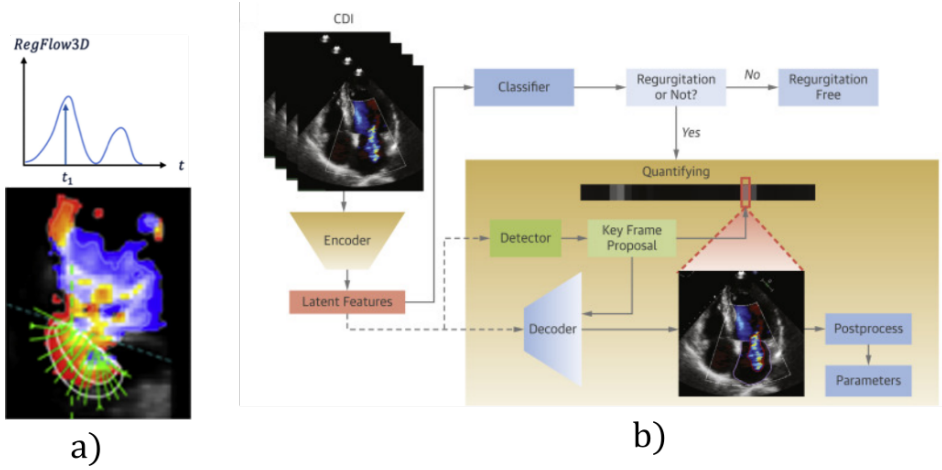


Figure 2.13: a): Semi-automated quantification of MR RVol based on 3-D TEE color-Doppler imaging (adapted from [31]). b): A deep learning-based, automated pipeline for detection and measurements of MR jet area (adapted from [33]).

Machine learning

ML systems are a subclass of AI systems that can perform some task without being explicitly programmed. Instead, they learn to perform the task through *learning* from data. We usually distinguish between *unsupervised learning* and *supervised learning*. Unsupervised learning is when the system learns patterns from data and typically learns to divide instances from the dataset into groups based on the learned patterns. Supervised learning is when the system learns to reproduce some task given an annotated dataset, where for a given data sample x there exists some *ground truth* y that corresponds to the desired result we would like the system to be able to produce from x. In order to achieve this, the system needs to model the relation between x and y. We can define our model in the following way:

$$\hat{y} = f(x|\theta), \tag{2.15}$$

where f is some function and θ are the parameters of the function. Ideally, we would like to choose our function f and its parameters θ such that for any x within the distribution of our dataset, the output $\hat{y}$ of our function is close to the corresponding label y. The way to achieve this is two-fold. Firstly, we need to choose a function f that has sufficient degrees of freedom to be able to model the relation between x and y. Secondly, we need to tune the parameters θ such that f produces the desired output. The parameters θ are

learned through optimization against an objective function that computes a similarity metric between the model output $\hat{y}$ and the ground truth y.

Perhaps the simplest ML model is the linear regression model

$$\hat{y} = wx + b, \tag{2.16}$$

where the function and its parameters w and b model a linear relationship between the data and the labels. To optimize this model, we need to define an objective function subject to optimization. A reasonable choice would be to minimize the squared difference between y and $\hat{y}$. For this simple model, explicit solutions for the optimal choice of w and b given minimal squared distance can be obtained. The solution is dependent on our available training data x and y. How well the model fits the data depends on two factors: how well our sampled data x represent the distribution from which x is sampled, and how well the assumption of a linear relationship between x and y fits the reality.

Although the linear regression model is applicable to many simple machine learning tasks, most modeling problems are much more complex. This is especially true for tasks often attributed with requiring some intelligent behavior, such as visual recognition. Consequently, in order to model such tasks, we require nonlinear functions with many parameters.

Deep Learning

How do we select a model when we cannot cohesively make assumptions about how the desired output y is related to the input x? For the linear regression model, we assumed a priori that x and y were linearly related. However, when we cannot assume simple relations like this, we must resort to functions that can model *any* relationship between x and y. This is possible using a class of functions which are known as *deep neural networks*, as illustrated in Figure 2.14.

A deep neural network is a cascade of linear regression nodes with nonlinear activation functions applied at each node. We can express this recursively using the *forward pass equations*:

$$\begin{aligned} \boldsymbol{h}^{(0)} &= x \\ \boldsymbol{z}^{(l)} &= W^{(l)}\boldsymbol{h}^{(l-1)} + \boldsymbol{b}^{(l)} \\ \boldsymbol{h}^{(l)} &= \sigma^{(l)}(\boldsymbol{z}^{(l)}), \end{aligned} \tag{2.17}$$

where $W^{(l)}$ is the weight matrix containing the weights w_{ij} between node i in layer $l-1$ and node j in layer l, and and $\boldsymbol{b}_l$ is the bias vector containing

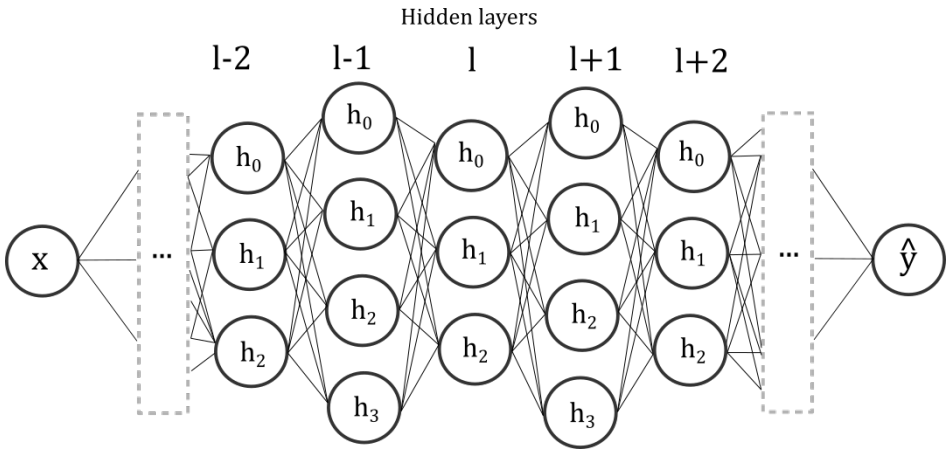


Figure 2.14: Illustration of a fully connected deep neural network. The network consists of multiple interconnected nodes. Each node contains tunable parameter weights w and biases b. The output $h_i^{(l-1)}$ of the node i at *hidden layer* $l-1$ serves as the input to the nodes at layer l.

the biases b_j. The activation function $\sigma^{(l)}$ is some nonlinear function applied to the linear combination at each node, which is needed to assure nonlinear modeling capability.

As the number of parameters in the network increases, the resulting function receives increasing degrees of freedom, hence its modeling capabilities increase. In fact, a sufficiently large network can in theory perfectly approximate *any* function. This statement is known as the Universal Approximation Theorem [37]. Following this statement, we know that regardless of the complexity of a given problem it is possible to effectively approximate the solution using a deep neural network.

Learning with Gradient Descent

Although we know that a sufficiently large deep neural network has the degrees of freedom to model any task, we still need to adjust all the parameters θ according to the desired output of the model. Similar to the linear regression problem, we need to define what the desired output should be, and then fit the parameters according to some objective given the data x and the desired output y. We define a *loss function* $\mathcal{L}(y, \hat{y}|f, \theta)$ which computes the *loss* between y and the model output $\hat{y}$ for each input x. The loss function $\mathcal{L}$ can be any differentiable function, depending on how we wish to define the similarity between the output and the desired output. For instance, we can define a loss function as the mean squared error $\mathcal{L} = \frac{1}{N}\sum ||y - \hat{y}||^2$ between

y and $\hat{y}$, as we did for the linear regression case. However, unlike linear regression, we cannot arrive at analytical solutions for the parameters, because the neural network structure is in general highly complex. Therefore we must resort to loss minimization by numerical optimization.

Optimization of neural networks can be achieved with a process known as *gradient descent*, or more specifically, using the *backpropagation algorithm* [38]. The parameter update step of the backpropagation algorithm is given by

$$\begin{aligned} W &:= W - \mu \nabla_W \mathcal{L} \\ \boldsymbol{b} &:= \boldsymbol{b} - \mu \nabla_b \mathcal{L}, \end{aligned} \tag{2.18}$$

where the operator ∇ denotes the gradient and μ is the *learning rate*. The parameters are updated iteratively until some condition is met, which could be that the predefined number of maximum iterations has been reached, or that the loss has stopped declining. The gradients of the loss with respect to the parameters can be computed at node j in layer l according to

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} &= \frac{\partial \mathcal{L}}{\partial z_j^{(l)}} h_i^{(l-1)} \\ \frac{\partial \mathcal{L}}{\partial b_j^{(l)}} &= \frac{\partial \mathcal{L}}{\partial z_j^{(l)}}. \end{aligned} \tag{2.19}$$

Note how the gradient at layer l is dependent on the values of layer $l-1$. This implies that to update the parameters of all layers, the gradients must be computed from the output and propagated towards the input, hence the name backpropagation.

Convolutional neural networks for image segmentation

Segmentation is a visual recognition task where the goal is to separate image pixels belonging to the foreground, i.e. the objects of interest in the image, from the background. Fully connected deep neural networks can be used for segmentation tasks by vectorizing the image pixels. However, the performance for visual recognition tasks is in general significantly higher for *convolutional neural networks* (CNNs). Instead of applying multiplications with the individual pixels, CNNs apply convolutions with image kernels to the entire image. The forward pass equations can now be expressed as

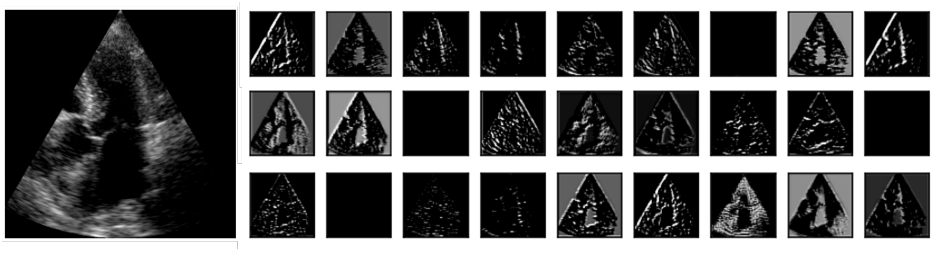


Figure 2.15: An echocardiogram (left) and some learnt feature maps of a CNN trained for segmentation (right).

$$
\begin{aligned}
H^{(0)} &= x \\
Z_i^{(l)} &= W_i^{(l)} * H^{(l-1)} + B_i^{(l)} \\
H_i^{(l)} &= \sigma^{(l)}(Z_i^{(l)}),
\end{aligned}
\tag{2.20}
$$

where $H_i^{(l)}$ is the ith *feature map* at layer l, $W_i^{(l)}$ is the convolutional kernel and $B_i^{(l)}$ are the biases. In this way, the CNN acts like cascaded image filters that decompose the input image into the relevant features for the task, for instance, edges and simple shapes. Figure 2.15 shows some feature maps that are learned by a model trained for the segmentation of echocardiograms. Each feature map highlights different features of the input image. Some highlight variations in image contrast, while others highlight horizontal or vertical lines present in the image. The feature maps are similar to what could be achieved with traditional digital filters from image processing [39], however, the filters are not designed explicitly but learned implicitly from the data through gradient descent.

To make an effective ML system one needs a sufficient amount of representative training data, as well as a suitable function structure, often referred to as the *model architecture*. Deciding which architecture is the most suitable for a given problem is non-trivial as there are infinite ways of combining the different parts of the network, and deciding which performs the best for a given task is only possible by trial and error. However, for medical image segmentation in particular there exists a model architecture that has been proven to provide accurate segmentation results across a wide range of domains, while still being easy to train in the presence of limited training data. This architecture is known as the UNet, first proposed by Ronneberger et al in 2015 [40]. The UNet architecture is illustrated in Figure 2.16

The UNet is based on the *autoencoder*, which progressively compresses

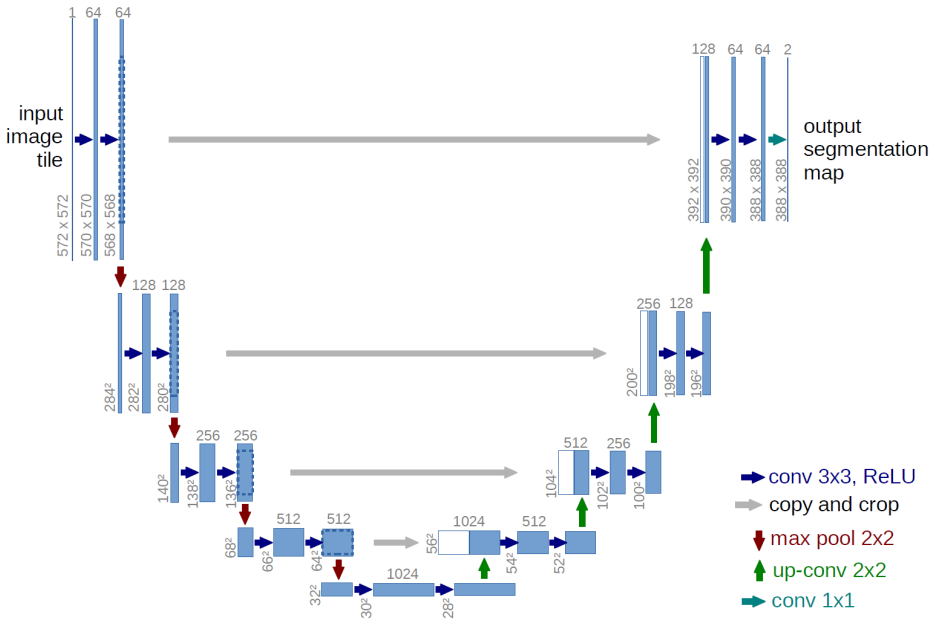


Figure 2.16: Schematic of the UNet architecture (adapted from [40]). The figure illustrates the hidden convolutional layers and their connections.

the input data into latent hidden variables using an *encoder*, and then reconstructs higher dimensional features from the latent variables using a *decoder*. In the UNet, max pooling layers down-sample feature maps, halving the spatial dimensions and compressing information into latent hidden variables. The latent hidden variables are then up-sampled back to the original image size. The innovation of the UNet was to also include *skip connections* at each compression level, where data flows past the compression stages and is combined with the decompressed data. This is believed to give a good balance between feature information and spatial information, causing higher accuracy segmentation results.

Attention Gates

In recent years, the concept of *attention* has garnered growing interest within the DL field. Attention mechanisms lie at the core of models such as generative pre-trained transformers (GPT) for autoregressive language modeling [41]. Moreover, attention mechanisms and transformers have also been applied to visual data [42], holding promise for higher accuracy visual recognition performance compared to CNN architectures.

The idea was originally proposed for machine translation of natural lan-

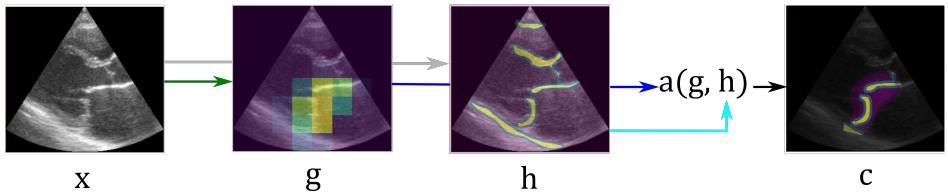


Figure 2.17: Illustration of the attention mechanism for image data, exemplified with a task for localizing the mitral valve from the echocardiogram x. The latent feature representation in g is assumed to contain a coarse representation of the relevant features in the image. The features in h originate from a higher level in the encoder, displaying more global features. The least relevant global features in h are disambiguated based on latent information in g using the approximated function $a(g, h)$.

guage [43]. In this study, the authors used fixed-length masks to search for the most relevant parts of the sentence, without needing to explicitly define which parts were relevant. Instead, the mask was learned from the data using an attention mechanism. The authors showed that by leveraging attention the network achieved better performance at translating long sentences, which was attributed to the attention mechanism's ability to simplify the encoding problem by masking out irrelevant information from the sentences.

The output $c^{(l)}$ of the attention mechanism at encoder layer l can be expressed as

$$\begin{aligned} \psi &= a(g^{(l-1)}, h^{(l)}), \\ \alpha &= \sigma(\psi), \\ c^{(l)} &= \alpha h^{(l)}, \end{aligned} \tag{2.21}$$

where $h^{(l)}$ is the signal from the encoder at compression stage l and $g^{(l-1)}$ is the decoder output at compression stage $l-1$. The function $a(\cdot)$ computes the *alignment scores* and can be approximated with a neural network. σ is a nonlinear activation function mapping the alignment scores to *attention coefficients* α between 0 and 1, consolidating the *context output* $c^{(l)}$, which is the element-wise weighting of the signal from the encoder with the attention coefficients. The context output represents the input signal with a strong emphasis on the salient features. In practice, the alignment score function $a(\cdot)$ can be parameterized along with the rest of the neural network, enabling learning the alignment function through gradient descent of the loss function.

The core idea of the attention mechanism is the assumption that the

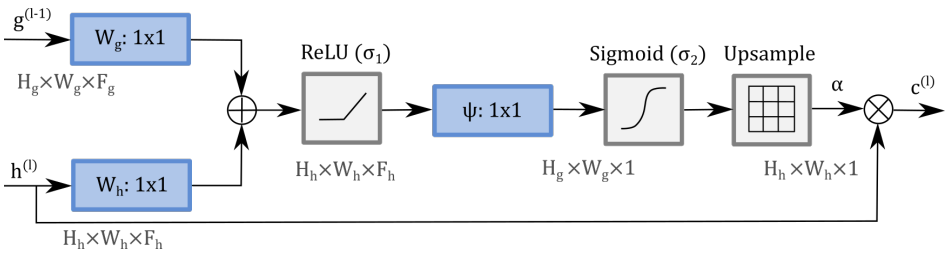


Figure 2.18: Schematic of the attention gate, adapted from [44]. The dimensions of the input signals g^{l-1} and $h^{(l)}$ are $H_g \times W_g \times F_g$ and $H_h \times W_h \times F_h$ respectively, where H, W and F respectively denotes the image height, width and number of features.

alignment score function $a(\cdot)$ can be approximated effectively using the latent feature representation from g^{l-1} in combination with the global features from h^l. It is assumed that when using an encoder-decoder type architecture, the coarse features learned in the encoder relate to the global features from the preceding layers. The idea is illustrated for visual data in Figure 2.17. The signal g is often called the *gating signal* as it serves the purpose of gating the most salient regions of the global features in h.

The attention mechanism was further adapted by Schlemper et al [44], who presented the *attention gate*. The attention gate implements the attention mechanism for CNNs. Consequently, attention gates can be combined with CNN architectures such as the UNet, thus leveraging attention of salient features in image data for visual recognition tasks. The attention gate is illustrated in Figure 2.18.

The attention gate is equivalent to the attention mechanism described by (2.21), but applied to two-dimensional image data. The alignment score function is implemented as a CNN consisting of 1×1 convolutional kernels W_g, W_h and ψ, and a rectified linear unit (ReLU) activation function σ_1. The attention coefficients α are upsampled to enable element-wise multiplication with $h^{(l)}$.

As demonstrated in [44], the attention gate can be inserted into the standard UNet architecture by replacing the concatenation layers with attention gates, resulting in a model architecture commonly referred to as the *Attention UNet*. The Attention UNet was demonstrated to achieve higher precision scores for segmentation tasks, mitigating false positives in irrelevant regions of the image. The authors concluded that Attention UNet can be particularly beneficial for tissue identification and localization of small-size objects.

Loss functions for segmentation

To effectively optimize a CNN to perform image segmentation we need to define an appropriate loss function that can be minimized using gradient descent. This is another non-trivial challenge of DL model design, as the notion of segmentation performance needs to be translated into a comprehensive mathematical function. There are two main categories of loss functions for segmentation, namely distribution-based losses, and region-based losses. For a comprehensive review of loss functions for image segmentation, the reader is referred to [45], [46].

A common choice of a distribution-based loss function is the *binary cross-entropy* function $\mathcal{L}_{\text{CE}}$ defined as

$$\mathcal{L}_{\text{CE}} = -y \log \hat{y} - (1 - y) \log (1 - \hat{y}), \tag{2.22}$$

where y is the ground truth segmentation label and $\hat{y}$ is the prediction confidence score that a given pixel belongs to the foreground class. Conversely, $1 - \hat{y}$ is the prediction confidence score that a given pixel belongs to the background class. The function is computed for every pixel in the image, and the average loss is calculated.

The binary cross entropy loss weights foreground and background predictions equally. However, for unbalanced datasets where one of the classes is over-represented, the term corresponding to this class will dominate the loss. This could result in a model which is poorly aligned to the segmentation task. For instance, in a segmentation task where the foreground object is small compared to the image dimensions, gradient descent could easily lead to a local minimum where the model would only output background labels for all pixels. To mitigate this issue, several balancing strategies have been proposed, one of which is the *balanced cross-entropy loss* $\mathcal{L}_{\text{BCE}}$ defined as

$$\mathcal{L}_{\text{BCE}} = -\beta y \log \hat{y} - (1 - \beta)(1 - y) \log (1 - \hat{y}), \tag{2.23}$$

where β is a weighting parameter between 0 and 1. Setting $\beta > 0.5$ increases the emphasis given to the foreground class during training. Conversely, setting $\beta < 0.5$ increases the emphasis given to the background class. The value of β can either be fixed or adapted to each image. For instance, we can set $\beta = 1 - \bar{y}$, where $\bar{y}$ is the mean pixel value of a given label y. Thus, if y contains many background pixels then β will increase, hence increasing emphasis given to the background, and vice versa.

Another notable distribution based loss function is the *focal loss* [47] defined as

$$\mathcal{L}_{\mathrm{FL}} = -\beta(1-\hat{y})^{\gamma} y \log \hat{y} - (1-\beta) y^{\gamma}(1-y)\log(1-\hat{y}), \tag{2.24}$$

where γ is a focusing parameter. By setting $\gamma = 0$ we arrive at (2.23). Increasing γ modulates the loss function such that less emphasis is given to high-confidence examples. Consequently, the model is encouraged to focus on difficult examples where the output confidence scores are lower.

In addition to distribution-based losses, region-based loss functions are also common. These functions are not computed for every pixel in the image, but rather computed once for the entire image based on some geometrical similarity metric. The majority of these functions aim to maximize some variation of the *dice similarity coefficient* (DSC). Maximizing the DSC is equivalent to minimizing the dice loss function $\mathcal{L}_{\mathrm{DL}}$ which can be implemented as

$$\mathcal{L}_{\mathrm{DL}} = 1 - \frac{1 + 2\sum_n y_n \hat{y}_n}{1 + \sum_n y_n + \hat{y}_n} \tag{2.25}$$

where n denotes the pixel index [46].

As discussed in [46], the best choice of loss function depends on the problem at hand. The loss function typically needs to be adapted to the data and the desired solution of the segmentation problem. Moreover, multiple loss functions can be combined, and regularizers can be added to incorporate task-specific priors in the learning phase.

Evaluation metrics

When assessing the performance of segmentation models it is common to evaluate standard metrics for segmentation performance [48]. One common metric is the aforementioned DSC which measures the degree of overlap between two regions A and B. The DSC is illustrated in Figure 2.19 a) and is defined as

$$\mathrm{DSC} = \frac{2(A \cap B)}{A + B}, \tag{2.26}$$

where $A \cap B$ is the intersection between A and B. A value of DSC $= 1$ signifies perfect overlap, while DSC $= 0$ signifies no overlap. The DSC is the harmonic mean of the *precision* and *recall*, which can be defined as

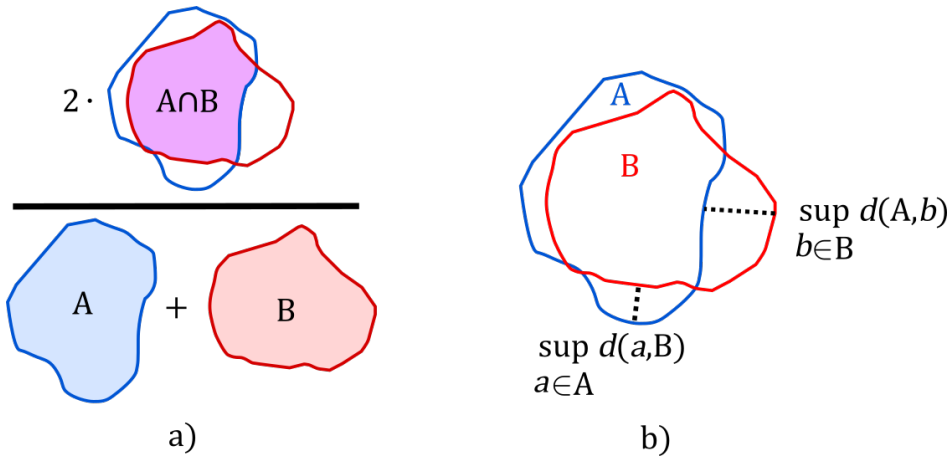


Figure 2.19: a): The dice similarity coefficient, is computed as the ratio between twice the area of the intersection of the regions A and B divided by the area of the individual regions. b): The Hausdorff distance d_{HD}, which is the maximal distance between the subsets A and B

$$\begin{aligned} \text{Precision} &= \frac{A \cap B}{A}, \\ \text{Recall} &= \frac{A \cap B}{B}, \end{aligned} \tag{2.27}$$

assuming that A is the predicted region and B is the true region. Precision is a measure of how many of the predicted foreground pixels are relevant. A value of 1 signifies that all the predicted foreground pixels are within the actual foreground region, ignoring false negatives. Recall measures how many of the relevant pixels are retrieved by the prediction. A value of 1 signifies that all of the true foreground pixels were predicted, ignoring false positives.

In addition to DSC, precision, and recall, another common segmentation metric is the *Hausdorff distance* [49]. The Hausdorff distance d_{HD} measures the maximum distance between the subsets A and B and is formally defined as

$$
\begin{aligned}
d(a, B) &= \inf_{b \in B} d(a, b), \\
d(A, b) &= \inf_{a \in A} d(a, b), \\
d_{\mathrm{HD}} &= \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(A, b) \right\}
\end{aligned}
\tag{2.28}
$$

where $d(a, b)$ is the distance between a point a in A and a point b in B, sup denotes the supremum operator (the least upper bound), and inf denotes the infimum operator (the greatest lower bound). The Hausdorff distance is illustrated in Figure 2.19 b). This metric can be useful to quantify large deviations from the ground truth region, which might indicate that the segmentation model produces false positives far away from the ground truth region. The DSC will not be greatly affected in such cases when the false positive region is small compared to the true positive region.

Note that cross-entropy-based metrics which are typically used as loss functions are not feasible to be used as evaluation metrics. This is because the cross entropy is unbounded, and the magnitude is highly dependent on the specific dataset. Therefore, a stand-alone cross-entropy value does not convey any information about performance. On the other hand, the DSC, precision, and recall are normalized scores, while the Hausdorff distance is bounded by the image dimensions. However, the values of each metric should be interpreted according to the specific segmentation problem. For instance, for tasks where the segmentation area is small relative to its contour, the DSC tends to be lower on average compared to cases where the area is large relative to the contour [50]. Moreover, certain datasets can contain more challenging cases than others (in terms of image quality, variations in anatomy, etc.), which implies that reported metrics should be interpreted relative to the dataset quality and size.

Furthermore, the aforementioned metrics are only proxies for the addressed clinical problem. Therefore, problem-specific quantitative metrics with clinical relevance should ideally be evaluated alongside standard segmentation metrics. Additionally, a detailed qualitative analysis should be performed to address segmentation quality and identify failure modes.

The generalization challenge

One of the key challenges with current DL methods is related to their ability to *generalize* to unseen data. As previously described, we know that for a given task there exists a neural network architecture that can perfectly approximate a solution to the task [37]. However, the solution can only be

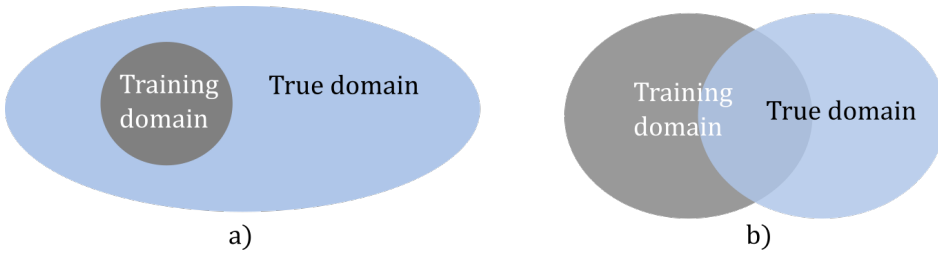


Figure 2.20: Differences in the training domain and the true domain of data. a): The training domain is a subset of the true domain. b): The training domain and the true domain are only partly overlapping, and there is a gap between the domains.

adapted to the available training data. To assure that a DL model is able to sufficiently generalize to new data we often require training and testing data that is sufficiently representative for the task. In medical imaging, acquiring sufficiently representative data can be challenging in several ways.

When curating a dataset of medical images the resulting dataset will be a subset of the set of all possible images, as illustrated in Figure 2.20 a). This can pose a challenge for model generalization if the images are sampled from a population that is limited in terms of anatomy and pathology. For instance, when developing a model for VHD assessment using TTE images it is important to consider differences in VHD ethiology between various countries, as explained in Section 2.2. Moreover, DL models trained solely on medical image data have been demonstrated to be sensitive to factors such as ethnicity [51]. Thus it is important not to assume a given model generalizes to outside of the training domain without properly evaluating the claim, as the model might be biased to the training domain.

Another variation of the generalization challenge is the possibility of a *domain gap* between the training domain and the true domain, as illustrated in Figure 2.20 b). In medical imaging, there will be a domain gap between datasets acquired using different scanners or different imaging protocols which can deteriorate model performance when applying the same model across different domains [52], [53]. Moreover, there will be a considerable domain gap when using simulated data for training, as for the project described in Section 5. In this case, the importance of the domain gap will depend on how realistic the simulations are compared to the real data. Decreasing the domain gap can sometimes be achieved using *data augmentation*, where the goal is to broaden the training domain by transforming the image data, for instance using geometrical transformations, masking, or image quality deterioration [54].

References

[1] B. Angelsen, *Ultrasound Imaging: Waves, Signals, and Signal Processing* (Ultrasound Imaging: Waves, Signals, and Signal Processing). Emantec, 2000.

[2] W. A. Zoghbi, D. Adams, R. O. Bonow, *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.

[3] A. Vahanian, F. Beyersdorf, F. Praz, *et al.*, "2021 ESC/EACTS guidelines for the management of valvular heart disease," *European Heart Journal*, vol. 43, no. 7, pp. 561–632, 2022.

[4] C. M. Otto, R. A. Nishimura, R. O. Bonow, *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," *Circulation*, vol. 143, no. 5, pp. 35–71, 2021.

[5] S. Robinson, L. Ring, D. X. Augustine, *et al.*, "The assessment of mitral valve disease: A guideline from the british society of echocardiography," *Echo Res Pract*, vol. 8, no. 1, G87–G136, 2021.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[7] H. Azhari, *Basics of Biomedical Ultrasound for Engineers*. Wiley - IEEE press, 2010, Appendix A.

[8] B. Angelsen, "A theoretical-study of the scattering of ultrasound from blood," *Modeling, Identification and Control*, vol. 2, 1981.

[9] S. Bjaerum, H. Torp, and K. Kristoffersen, "Clutter filters adapted to tissue motion in ultrasound color flow imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, no. 6, pp. 693–704, 2002.

[10] A. Salustri and A. Almaghrabi, "Mitral valve disease: Correlation between the most important echocardiographic parameters and haemodynamics," *E-Journal of Cardiology Practice*, vol. 16, no. 24-26, 2018.

[11] 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.

[12] J. A. Jensen, S. I. Nikolov, A. C. H. Yu, and D. Garcia, "Ultrasound vector flow imaging-part i: Sequential systems," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency control*, vol. 63, no. 11, pp. 1704–1721, 2016.

[13] E. Pierce, *Diagram of the human heart*, [Online; accessed November 08, 2023], 2003.

[14] A. F. Carpentier, A. Lessana, J. Y. Relland, *et al.*, "The "physio-ring": An advanced concept in mitral valve annuloplasty," *The Annals of Thoracic Surgery*, vol. 60, no. 5, pp. 1177–1186, 1995.

[15] S. Harb and B. Griffin, "Mitral valve disease: A comprehensive review," *Current Cardiology Reports*, vol. 19, 2017.

[16] J. S. Aluru, A. Barsouk, K. Saginala, P. Rawla, and A. Barsouk, "Valvular heart disease epidemiology," *Med Sci (Basel)*, vol. 10, no. 2, p. 32, 2022.

[17] C. J. Taylor, J. M. Ordonez-Mena, N. R. Jones, *et al.*, "Survival of people with valvular heart disease in a large, english community-based cohort study," *Heart*, vol. 107, no. 16, pp. 1336–1343, 2021.

[18] S. Coffey, R. Roberts-Thomson, A. Brown, *et al.*, "Global epidemiology of valvular heart disease," *Nat Rev Cardiol*, vol. 18, no. 12, pp. 853–864, 2021.

[19] V. T. Nkomo, "Chronic primary mitral regurgitation: General management," *UpToDate*, 2023.

[20] S. Gunnal, R. Wabale, and M. Farooqui, "Morphological study of chordae tendinae in human cadaveric hearts," *Heart views : the official journal of the Gulf Heart Association*, vol. 16, pp. 1–12, 2015.

[21] V. Kamoen, M. De Buyzere, M. El Haddad, T. L. M. de Backer, and F. Timmermans, "Average pixel intensity method for prediction of outcome in secondary mitral regurgitation," *Heart*, vol. 106, no. 12, pp. 904–909, 2020.

[22] A. B. Kvaslerud, K. Santic, A. I. Hussain, *et al.*, "Outcomes in asymptomatic, severe aortic stenosis," *PLoS One*, vol. 16, no. 4, e0249610, 2021.

[23] A. Hagendorff, F. Knebel, A. Helfen, *et al.*, "Echocardiographic assessment of mitral regurgitation: Discussion of practical and methodologic aspects of severity quantification to improve diagnostic conclusiveness," *Clinical Research in Cardiology*, vol. 110, no. 11, pp. 1704–1733, 2021.

[24] A. Wang, P. Grayburn, J. A. Foster, *et al.*, "Practice gaps in the care of mitral valve regurgitation: Insights from the american college of cardiology mitral regurgitation gap analysis and advisory panel," *American Heart Journal*, vol. 172, pp. 70–79, 2016.

[25] Y. Topilsky, H. Michelena, V. Bichara, J. Maalouf, D. W. Mahoney, and M. Enriquez-Sarano, "Mitral valve prolapse with mid-late systolic mitral regurgitation: Pitfalls of evaluation and clinical outcome compared with holosystolic regurgitation," *Circulation*, vol. 125, no. 13, pp. 1643–1651, 2012.

[26] N. Watanabe, "Acute mitral regurgitation," *Heart*, vol. 105, no. 9, pp. 671–677, 2019.

[27] S. Biner, A. Rafique, F. Rafii, *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.

[28] J. Verbeke, V. Kamoen, M. De Buyzere, T. Claessens, and F. Timmermans, "The pixel variation score: An echocardiographic index to assess temporal variation of mitral regurgitant flow," *Journal of the American Society of Echocardiography*, 2023.

[29] T. C. Tan, X. Zeng, Y. Jiao, *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.

[30] S. Militaru, O. Bonnefous, K. Hami, *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.

[31] A. Singh, J. Su, A. This, *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.

[32] S. Fiorentini, E. A. R. Berg, H. Torp, S. Aakhus, and J. Avdal, "Quantification of flow rates and flow volumes in valve regurgitation using 3-d high frame-rate ultrasound," *IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 3, pp. 29–40, 2023.

[33] F. Yang, X. Chen, X. Lin, *et al.*, "Automated analysis of doppler echocardiographic videos as a screening tool for valvular heart diseases," *JACC: Cardiovascular Imaging*, vol. 15, no. 4, pp. 551–563, 2022.

[34] M. Vafaeezadeh, H. Behnam, A. Hosseinsabet, and P. Gifani, "Carpnet: Transformer for mitral valve disease classification in echocardiographic videos," *International Journal of Imaging Systems and Technology*, 2023.

[35] K. Tang, Z. Ge, R. Ling, *et al.*, "Mitral regurgitation quantification from multi-channel ultrasound images via deep learning," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Cham: Springer Nature Switzerland, 2023, pp. 223–232.

[36] M. Kim, J. Yun, Y. Cho, *et al.*, "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, pp. 657–668, 2019.

[37] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[38] D. Rumelheart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[39] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Pearson, 2018.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer International Publishing, 2015, pp. 234–241.

[41] T. Brown, B. Mann, N. Ryder, *et al.*, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 1877–1901.

[42] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[43] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Advances in Neural Information Processing Systems (NIPS)*, ser. Advances in Neural Information Processing Systems, vol. 28, 2015.

[44] J. Schlemper, O. Oktay, M. Schaap, *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[45] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artificial Intelligence in Review*, vol. 54, no. 1, pp. 137–178, 2021.

[46] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, 2020, pp. 115–121.

[47] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *arXiv*, 2018.

[48] S. Ostmeier, B. Axelrod, F. Isensee, *et al.*, "Use-evaluator: Performance metrics for medical image segmentation models supervised by uncertain, small or empty reference annotations in neuroimaging," *Medical Image Analysis*, vol. 90, p. 102 927, 2023.

[49] T. Birsan and D. Tiba, "One hundred years since the introduction of the set distance by dimitrie pompeiu," in *System Modeling and Optimization*, F. Ceragioli, A. Dontchev, H. Futura, K. Marti, and L. Pandolfi, Eds., Boston, MA: Springer US, 2006, pp. 35–39.

[50] A. Trutti, L. Fontanesi, M. Mulder, P.-L. Bazin, B. Hommel, and B. Forstmann, "A probabilistic atlas of the human ventral tegmental area (VTA) based on 7 tesla MRI data," *Brain Structure and Function*, vol. 226, 2021.

[51] J. W. Gichoya, I. Banerjee, A. R. Bhimireddy, *et al.*, "AI recognition of patient race in medical imaging: A modelling study," *The Lancet Digital Health*, vol. 4, no. 6, e406–e414, 2022.

[52] R. Kushol, P. Parnianpour, A. Wilman, S. Kalra, and Y.-H. Yang, "Effects of mri scanner manufacturers in classification tasks with deep learning models," *Scientific Reports*, vol. 13, 2023.

[53] C. Li, X. Lin, Y. Mao, *et al.*, "Domain generalization on medical imaging classification using episodic training with task augmentation," *Computers in Biology and Medicine*, vol. 141, p. 105 144, 2022.

[54] F. Garcea, A. Serra, F. Lamberti, and L. Morra, "Data augmentation for medical imaging: A systematic literature review," *Computers in Biology and Medicine*, vol. 152, p. 106 391, 2023.

Chapter 3

Mitral Valve Segmentation and Tracking from Transthoracic Echocardiography using Deep Learning

Valvular heart diseases (VHD) pose a significant public health burden, and deciding the best treatment strategy necessitates an accurate assessment of heart valve function. Transthoracic Echocardiography (TTE) is the key modality to evaluate VHD, but the lack of standardized quantitative measurements leads to subjective and time-consuming assessments. We aimed to use deep learning to automate the extraction of mitral valve (MV) leaflets and annular hinge points from echocardiograms of the MV, improving standardization and reducing workload in quantitative assessment of MV disease. We annotated the MV leaflets and annulus points in 2931 images from 127 patients. We propose an approach for segmenting the annotated features using Attention UNet with deep supervision and weight scheduling of the attention coefficients to enforce saliency surrounding the MV. The derived segmentation masks were used to extract quantitative biomarkers for specific MV leaflet scallops throughout the heart cycle. Evaluation performance was summarized using dice: 0.63 ± 0.14, annulus error: 3.64 ± 2.53, and leaflet angle error: 8.7 ± 8.3 °. Leveraging Attention UNet with deep supervision robustness of clinically relevant metrics was improved compared to UNet, reducing standard deviation by: 2.7 ° (angle error), 0.73 mm (annulus error). We correctly identified cases of MV prolapse, cases of stenosis, and healthy references from clinical material using the derived biomarkers. Robust deep

learning segmentation and tracking of MV morphology and motion is possible by leveraging attention gates and deep supervision, and holds promise for enhancing VHD diagnosis and treatment monitoring.

3.1 Introduction

Valvular heart disease (VHD) is a class of diseases that affect the function of the heart valves, including obstruction of blood flow (valve stenosis) and leakage of blood (valve regurgitation), representing an increasing burden to the public health system [1]. Transthoracic Echocardiography (TTE) is the primary modality for assessing VHD. Current recommendations [2], [3] emphasize an integrative approach of quantitative and indirect measurements limited to blood flow measurements (Doppler) and dimensions and volumes of heart chambers. This approach is time-consuming, dependent on the clinician's experience and skill, and prone to significant interobserver variability [4]–[6].

Quantitative predictors for MV morphology and motion based on leaflet angle and tenting area have previously been proposed [7]–[9]. Currently, direct quantitative measurements require tedious manual annotation of several images for each recording, which is rarely considered feasible. Automating direct quantification of VHD could facilitate standardization and reduce workload in the clinic.

Moreover, standardized measurements of leaflet motion could be beneficial for regular patient follow-up to monitor the disease progression of the valve condition according to a given treatment. Tracking the MV through the heart cycle could also unveil relevant information about cardiac function and filling pressures. By further enabling the extraction of these measurements in large databases, it may be possible to improve the diagnostic and prognostic value of echocardiographic markers of VHD and cardiac function.

Measuring mitral valve (MV) morphology and motion has previously been achieved with various unsupervised approaches [10]–[13]. However, these approaches are susceptible to motion artifacts from surrounding tissue and poor performance for poor-quality images.

In recent years, convolutional neural networks (CNNs) have shown great promise in medical image segmentation. CNNs have also been successfully applied to segmenting the MV leaflets from TTE images. Costa et al. [14] trained a UNet to segment the MV leaflets in children with rheumatic valve disease, with promising results. In addition to segmentation, classification of VHD from TTE images using deep learning has also recently been proposed [15]. Moreover, several studies [16]–[19] have explored 3-D segmentation and tracking of the MV leaflets and annulus in Transesophageal Echocar-

diography (TEE). 3-D TTE is non-invasive but currently hampered by a lower frame rate and poorer image quality, which makes segmentation and tracking of the mitral valve more challenging.

The UNet architecture is often used for medical image segmentation as it has been proven to provide robust segmentation results across a wide range of tasks. However, segmentation of the MV from echocardiograms poses extra challenges for learning spatial features. The MV leaflets are dynamic, thin structures with varying profiles throughout the cardiac cycle, which are influenced by patient-specific image quality, potential obstruction in the view, and the possibility of misidentification with other visible cardiac structures, such as other valves and cardiac walls. Thus, robust tracking of the MV over time is a significant challenge.

We aimed to automate the extraction of key measurements of the MV leaflets and annular hinge points from TTE images in apical two-, four-, and five-chamber (A2C, A4C, A5C) and apical and parasternal long-axis (ALAX, PLAX) views. Secondly, we aimed to evaluate the robustness of MV segmentation for different deep learning architectures. The overarching goal was to enhance standardization and reduce operator bias and workload in the quantitative assessment of MV pathology.

The main contribution of this work includes a deep learning-based approach for segmenting the MV, utilizing attention gates and deep supervision for improved robustness. We also demonstrate the automatic extraction of clinical biomarkers for MV motion and morphology and how they relate to VHD.

3.2 Methods

Application Pipeline

An MV segmentation network was combined with cardiac event timing (systole/diastole) and imaging view classifiers to produce automated quantitative measurements for specific leaflet scallops at relevant time points. The pipeline is illustrated in Figure 3.1. For each of the identified scallops, we extracted quantitative biomarkers (leaflet angle, leaflet tenting area, and leaflet angular velocity), which were used for tracking of MV motion and morphology and identification of MV disease. For imaging view and cardiac timing classification, we used neural networks from previous work in our research group [20], [21].

Datasets

The dataset used for training and evaluation consisted of transthoracic echocardiograms from 78 MV disease patients acquired at St. Olavs Hospi-

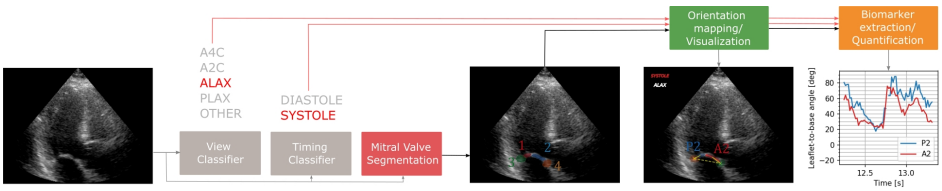


Figure 3.1: Application pipeline overview. MV leaflets and annulus hinge points are segmented from TTE echocardiograms. The relevant leaflet scallop labels (P2; central posterior scallop, A2: central anterior scallop) are assigned to the segmentation classes according to the predicted view. Biomarkers for each leaflet are computed from the segmentation output sequence. The timing prediction extracts biomarkers at relevant phases in the cardiac cycle.

tal in Trondheim, Norway, between 2015 and 2021, and 49 rheumatic MV disease patients acquired at Tinkur Anbessa Specialized Hospital in Addis Abeba, Ethiopia, between 2017 and 2019. The regional medical and health research committee approved the use of all image data in this work. The dataset consisted of B-Mode echocardiograms acquired using GE Vivid E9/E95 scanners (GE Vingmed Ultrasound AS, Norway), with the primary standard views for VHD assessment (A2C, A4C, A5C, ALAX, PLAX). A total of 2931 images from 438 recordings from 127 patients were annotated. The dataset had a representative distribution of image quality and consisted of both high and poor-quality images. About 5-10 image frames were annotated in each recording. The image frames were chosen manually by the annotators to capture the MV range of motion, including end systole and end diastole frames and the MV opening phase. The anterior and posterior valve leaflets and the annulus hinge points were annotated using AnnotationWeb [22], as illustrated in Figure 3.2. Experienced clinicians performed the annotations.

Attention UNet with Deep Supervision

To mitigate false positive segmentation of irrelevant features, focus should be given to the salient region around the MV during training. We used the Attention UNet [23] to enhance focus on the region of interest. This architecture introduces attention gates that leverage the multiplication of saliency maps with the image tensors, encouraging the network to segment only relevant regions. We applied deep supervision (DS) on the attention coefficients to enforce saliency around the MV. We achieved this by computing segmentation loss on the output and attention maps, where the latter was optimized to enforce saliency heatmaps around the MV. The model architecture is illustrated in Figure 3.3.

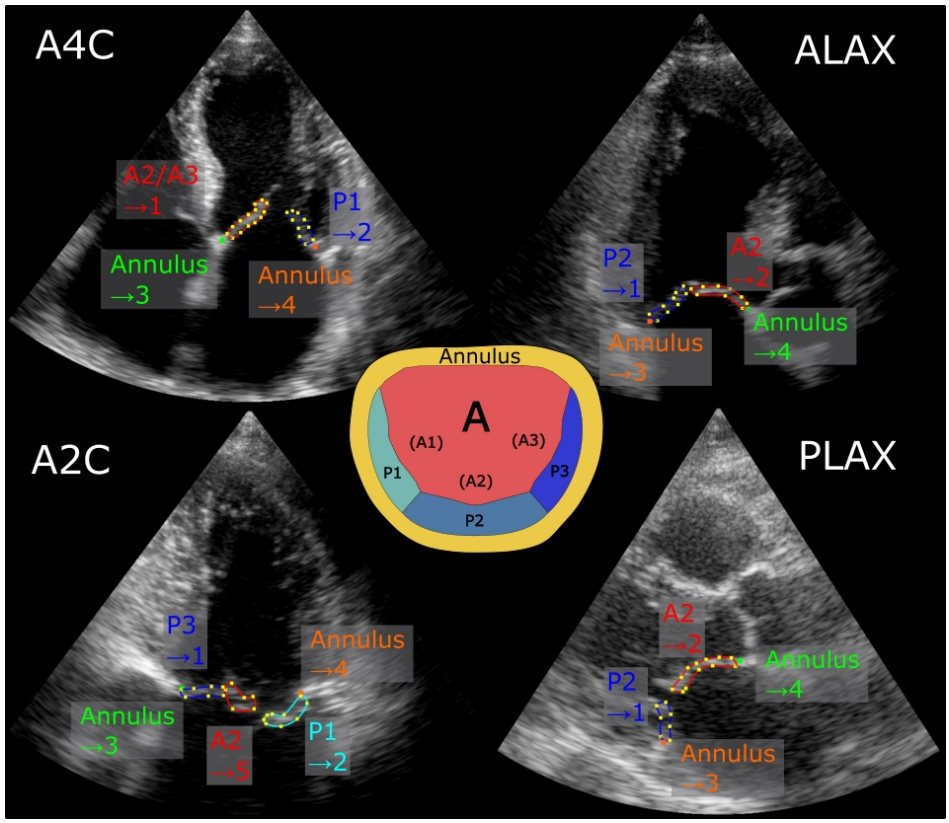


Figure 3.2: Four images with different cardiac views (A4C, A2C, ALAX, PLAX) annotated using AnnotationWeb. The leaflets and annulus points were annotated as separate classes according to the MV anatomy, depicted by the center illustration (A, anterior; P, posterior; 1, anterolateral; 2, central; 3, posteromedial). To simplify the complexity of the learning task across different views, the classes were mapped to values 1-5 according to left-right orientation labeling, as indicated by the arrows. 5-10 keyframes at different points in the cardiac cycle were annotated for each recording. A5C views were grouped under the same category as A4C.

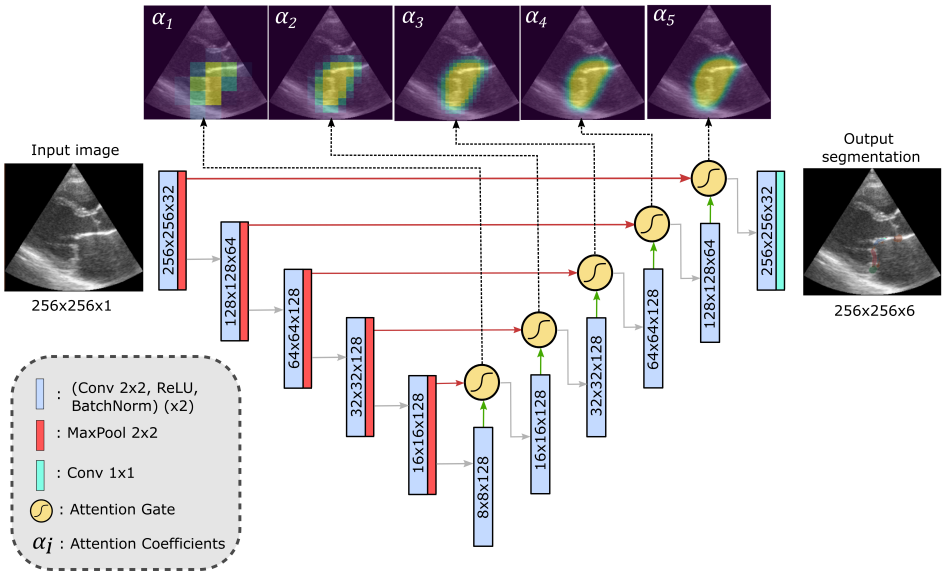


Figure 3.3: Model architecture: Attention UNet with deep supervision. The network outputs six segmentation class labels (background, three leaflets, two annulus hinge points) for a given image. In addition, attention coefficient maps α_i are taken as outputs for early-stage deep supervision. The network is trained to output saliency maps surrounding the MV in this stage. Hence α_i are primed to focus on the correct spatial features in the image before the segmentation task is gradually introduced and eventually dominates the learning task.

Since the MV is small compared to the image dimensions, the training data had a high class imbalance. To mitigate the class imbalance, we used a balanced cross-entropy loss. The segmentation loss function $\mathcal{L}_{seg}$ computed between the ground truth tensor y and the predicted tensor $\hat{y}$ is defined as

$$\begin{aligned}\mathcal{L}_{seg} = &- \beta y \log \hat{y} \\ &- (1-\beta)(1-y) \log (1-\hat{y})\end{aligned} \tag{3.1}$$

where β is the mean value of each channel in the tensor y.

The total loss function $\mathcal{L}$ was computed as a weighted sum of all of the loss functions according to

$$\mathcal{L} = \lambda(n)\mathcal{L}_{seg} + \frac{1-\lambda(n)}{5}\sum_{i=1}^{5}\mathcal{L}_{\alpha}^{(i)}, \tag{3.2}$$

where $\mathcal{L}_{\alpha}^{(i)}$ is the ith attention coefficient map loss, computed as the binary cross entropy between the attention coefficient map α_i and the saliency heatmap label. The saliency heatmaps were created by taking the convex hull of the segmentation labels and smoothing the edges using a Gaussian filter.

A weight scheduler allocated weights to the different loss functions such that more emphasis was given to the saliency maps at the beginning of the learning phase, and more emphasis was given to the segmentation output towards the end. The loss weight scheduler coefficient $\lambda(n)$ is defined as

$$\lambda(n) = \begin{cases} 0, & \text{if } n \leq N_0 \\ \frac{n-N_0}{N_1-N_0}, & \text{if } N_0 < n < N_1 \\ 1, & \text{if } n \geq N_1, \end{cases} \tag{3.3}$$

where n is the epoch number. The constants N_0 and N_1 were set to 40 and 200, respectively.

Biomarkers for Mitral Valve Disease

Following segmentation of the MV, we extracted biomarkers for MV morphology and motion, namely *leaflet angle*, *leaflet tenting area*, and *leaflet angular velocity*. These quantitative predictors correspond to biomarkers described in previous research [7]–[9], but are extracted automatically. The biomarker extraction is illustrated in Figure 3.4.

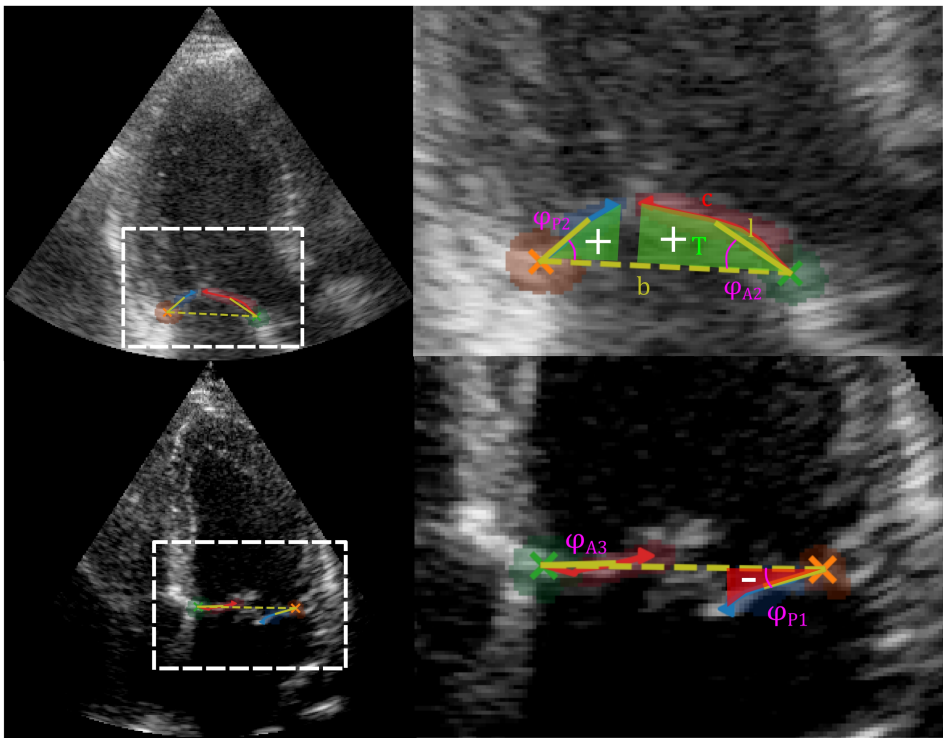


Figure 3.4: Segmentation examples with an illustration of biomarker extraction. Leaflet vectors $\boldsymbol{l}$ (solid yellow lines) are computed from the segmentation output centroid of the leaflet and the corresponding annulus hinge point (green and orange cross). The annular plane vector $\boldsymbol{b}$ (dashed yellow line) is computed between the two hinge points. For each identified leaflet scallop (A, anterior; P, posterior; 1, anterolateral; 2, central; 3, posteromedial) we compute the leaflet opening angle ϕ, the leaflet angular momentum $\mathrm{d}\phi/\mathrm{d}t$ and the signed tenting area T between the leaflet center curves $\boldsymbol{c}$ (red and blue curves) and $\boldsymbol{b}$.

The leaflet angle ϕ is computed as

$$\phi = \arcsin(\boldsymbol{l} \times \boldsymbol{b}), \tag{3.4}$$

where $\boldsymbol{l}$ is the hinge-to-leaflet centroid vector and $\boldsymbol{b}$ is the annular plane vector, defined by the two annulus hinge points. The leaflet tenting area is computed by extracting the center curve spline $\boldsymbol{c}$ of the leaflet segmentation mask and computing the signed area T between $\boldsymbol{c}$ and $\boldsymbol{b}$. The parts where $\boldsymbol{c}$ lies below $\boldsymbol{b}$ negatively affect the total area. The leaflet angular momentum is computed as the time derivative $\mathrm{d}\phi/\mathrm{d}t$ of the leaflet angle. Angular velocity is converted to m/s by multiplying by the length $|\mathbf{l}|$ between the hinge point and the leaflet centroid.

Model Training and Evaluation

We trained and compared UNet, Attention UNet, and Attention UNet + DS on the annotated data. The images were reshaped into 256×256 pixels. Annulus hinge points were converted into Gaussian heatmaps with a full-width half maximum of 10 pixels. The networks were trained for 500 epochs, using the Adam optimizer with an initial learning rate of $5 \cdot 10^{-4}$ and a batch size of 64. We used balanced cross entropy segmentation loss (3.1) for all models. We applied gamma augmentation to the images with random gamma values between 0.4 and 1.7 to handle varying image contrast.

All models were trained on the same 70% of the annotated subjects and tested on the remaining 30%. Moreover, we also performed the same experiment with 30% training data and 70% testing data. During testing, we evaluated standard segmentation metrics and the accuracy of leaflet angle estimation, annulus point placement, and leaflet center curve spline estimation (12-point cubic splines). Additionally, we assessed segmentation quality throughout every frame in each recording by calculating a Temporal Consistency Score $\mathrm{TCS}(n)$ for each frame $\hat{y}(n)$, defined as:

$$\mathrm{TCS}(n) = \begin{cases} 1, & \text{if } \hat{y}(n) \text{ is consistent} \\ 0, & \text{otherwise} \end{cases}$$

where $\hat{y}(n)$ is considered consistent if the following conditions are met:

- Both annulus hinge points are segmented.
- None of the annulus hinge points have moved from the previous frame with a velocity higher than 0.35 m/s.
- None of the leaflets have a velocity higher than 1.5 m/s.

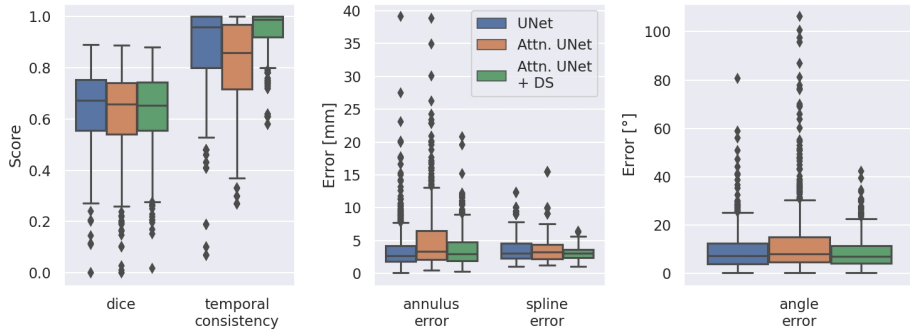


Figure 3.5: Boxplot of central test results for the different models (70/30% train/test split). Metrics were computed as explained in Table 3.1. We can observe notable outlier reduction for Attention UNet with DS compared to UNet and Attention UNet without DS for temporal consistency and annulus, spline, and angle errors.

The annulus velocity threshold was set above twice the highest expected velocity in healthy individuals [24]. Similarly, the leaflet velocity threshold was chosen based on measured velocities in healthy individuals, which never exceeded 0.75 m/s in our datasets.

Clinical feasibility

The proposed method was applied retrospectively to three TTE exams of healthy references with no VHD, three exams of patients with MV prolapse, and three exams of patients with MV stenosis. MV motion and morphology biomarkers for each leaflet scallop were extracted from representative recordings in each exam. The extracted quantitative measurements were compared to the qualitative descriptions of the MV condition in the cardiologist's report. In addition, we applied the proposed leaflet angle tracking to one healthy reference and one patient with restrictive filling of the left ventricle. The tracking results were compared qualitatively to pulsed wave Doppler spectra of MV inflow from the same exams.

3.3 Results

Segmentation validation

Evaluation results are reported in Table 3.1. Key metrics are also shown in Figure 3.5. Attention UNet without DS performed notably worse than the others. Attention UNet with DS performed similarly to UNet on average; however, robustness, temporal consistency, and biomarker accuracy were

Metrics	**70/30% Train/Test Split**		
	UNet	**Attn. UNet**	**Attn. UNet + DS**
Dice Score	**0.64 ± 0.15**	0.63 ± 0.16	0.63 ± 0.14
Precision	**0.69 ± 0.17**	0.64 ± 0.17	0.65 ± 0.17
Recall	0.62 ± 0.17	0.64 ± 0.19	**0.64 ± 0.17**
Hausdorff [mm]	**12.91 ± 12.78**	21.82 ± 19.64	14.17 ± 11.88
Annulus Error [mm]	3.64 ± 3.36	4.85 ± 5.53	3.64 ± 2.53
Spline Error [mm]	3.57 ± 1.92	3.74 ± 2.25	**3.11 ± 1.06**
Angle Error [°]	9.56 ± 10.99	12.67 ± 17.01	**8.67 ± 8.32**
TCS	0.85 ± 0.20	0.79 ± 0.19	**0.93 ± 0.10**
Metrics	**30/70% Train/Test Split**		
	UNet	**Attn. UNet**	**Attn. UNet + DS**
Dice Score	0.36 ± 0.22	0.43 ± 0.23	**0.46 ± 0.19**
Precision	0.35 ± 0.21	**0.42 ± 0.22**	0.39 ± 0.18
Recall	0.41 ± 0.28	0.48 ± 0.28	**0.61 ± 0.26**
Hausdorff [mm]	37.08 ± 20.16	38.26 ± 20.12	**24.34 ± 13.66**
Annulus Error [mm]	8.21 ± 9.73	9.50 ± 10.65	**6.93 ± 7.25**
Spline Error [mm]	7.48 ± 3.63	6.83 ± 4.42	**5.34 ± 3.01**
Angle Error [°]	23.50 ± 24.42	24.65 ± 27.99	**18.35 ± 16.85**
TCS	0.60 ± 0.25	0.59 ± 0.25	**0.92 ± 0.13**

Table 3.1: Performance comparison of different models. Significant improvements ($p < 0.05$) are emphasized. Dice score, precision, and recall are only evaluated for the leaflet classes. The temporal consistency score (TCS) is the fraction of frames in a recording that passes the automatic temporal consistency evaluation. Spline error is computed as the mean absolute error of estimated center curve splines from the ground truth annotations relative to the predictions.

improved. Moreover, Attention UNet with DS significantly improved all metrics when using a 30/70% train/test split. Figure 3.6 demonstrates how UNet was prone to produce significant outliers in edge cases, which were mitigated using attention gates and DS. Figure 3.7 shows how attention improved temporal consistency as more coherent biomarker tracking results with fewer errors were derived.

For the Attention UNets, we can observe significant improvement when applying DS. Without DS, predictions had a relatively high error, especially regarding angle, annulus displacement, and Hausdorff. Figure 3.8 shows that without applying DS, the attention coefficient maps did not necessarily emphasize only the relevant features. Consequently, as illustrated in Figure 3.6, Attention UNet without DS was prone to false positives in irrelevant regions. However, such false positives were mitigated using DS.

Biomarker feasibility

We evaluated the automatically extracted biomarkers for VHD on three patients with MV prolapse, three patients with MV stenosis, and three healthy references. We here used the Attention UNet with DS to perform segmentation. Figure 3.9 shows heatmaps illustrating predicted biomarkers in the different MV scallops. These were created by combining segmentation results with view and timing classification from multiple recordings of the same subject. The results indicated that negative systolic angle and tenting area were associated with MV prolapse. Similarly, reduced leaflet movement was associated with stenosis.

MV motion and Cardiac Function

In addition to aiding the assessment of VHD, MV motion patterns might relate to other aspects of cardiac function, for instance, MV inflow patterns, as illustrated in Figure 3.10. MV motion could provide additional information about flow and pressures, which, unlike Doppler measurements, is independent of insonation angle and could be measured in several views without the need for angle correction. Cardiac function is complicated; hence, the assessment might benefit from detailed measurements, which are currently infeasible but possibly achievable through automation.

3.4 Discussion

We have described a framework for segmenting the MV leaflets and annulus points in five TTE standard views. Based on the segmentation, we proposed automatic spatiotemporal biomarkers for quantifying MV morphology and motion, including leaflet angle, velocity, and systolic tenting area. The advantages of improved robustness by leveraging attention gates and deep su-

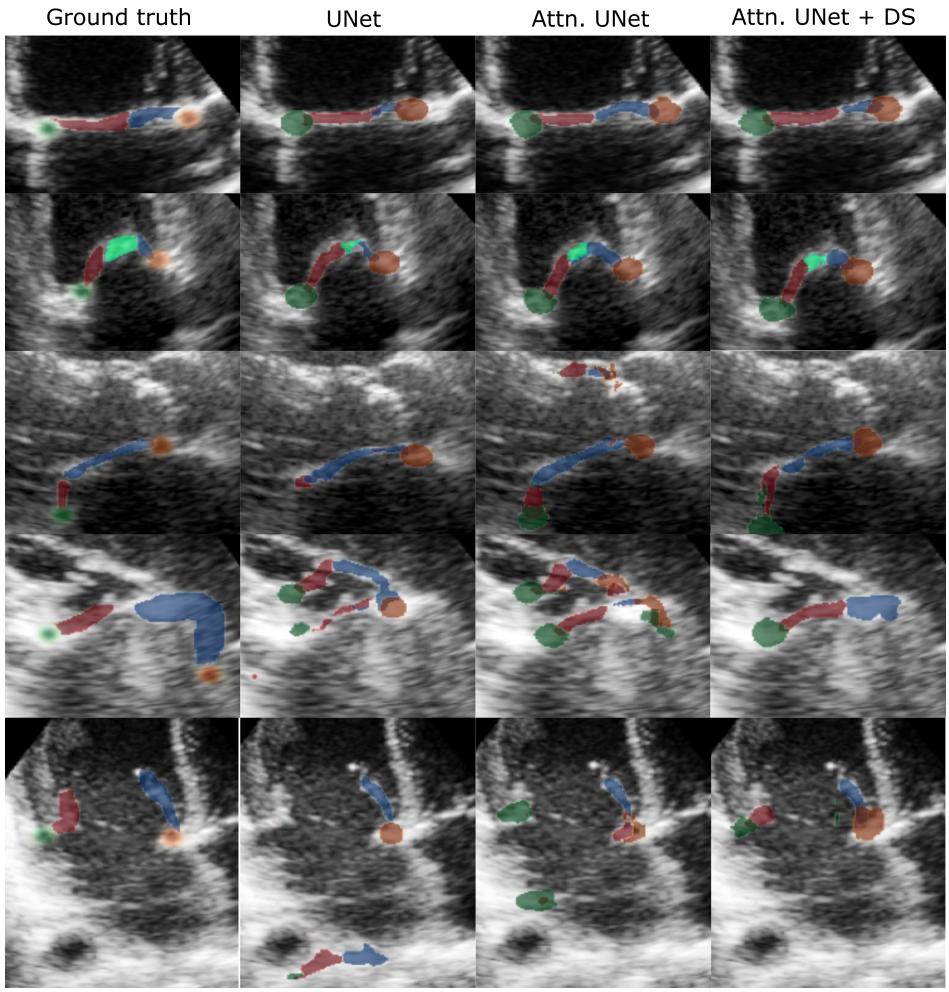


Figure 3.6: Segmentation examples (zoomed) from the test set (70/30% train/test split). Attention UNet + DS mitigates false positives of other valves and tissue in the bottom three examples. Note that the bottommost example is an offset ALAX view where UNet failed considerably, whereas Attention UNet + DS provided reasonable results. This example indicates that our approach might be more robust in the presence of suboptimal views, which is a common occurrence in the clinic.

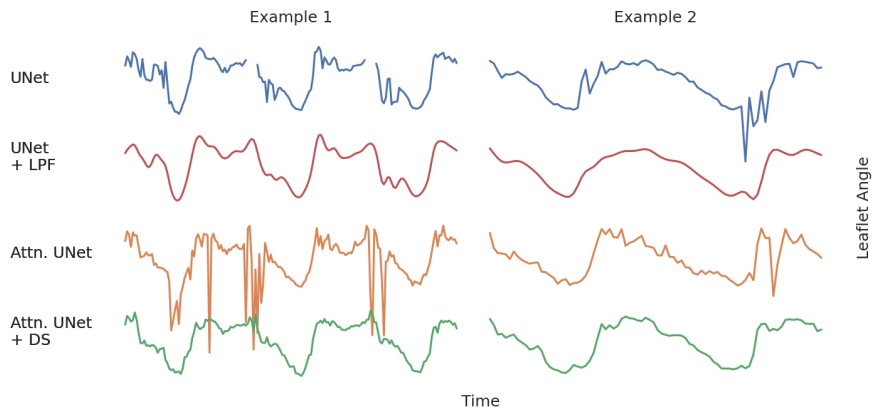


Figure 3.7: Leaflet angle tracking plots of two examples. In these examples, Attention UNet + DS had the highest temporal consistency and provided smoother tracking with fewer errors than the other models. Smoothing could also be achieved by low pass filtering (LPF) of the tracking result; however, this could impact accuracy and is limited to offline post-processing.

pervision were emphasized. Finally, we demonstrated the clinical feasibility of the aforementioned biomarkers.

Attention UNet with DS outperformed other models by focusing on relevant regions and reducing confusion from irrelevant structures with similar features. This improved robustness, biomarker accuracy, and temporal consistency, despite a slight decrease in pixel-wise accuracy and dice score. Moreover, when using a lower training fraction, Attention UNet with DS demonstrated significant improvements in all metrics, implying that Attention UNet with DS was more robust when exposed to a larger test set. This improvement could indicate that it converges faster towards a desirable solution during training when labeled data is scarce. The scarcity of labeled data is the main limitation when implementing deep learning models in practice since annotation is a substantial workload for clinicians.

Traditional metrics for segmentation performance are of limited use for the MV due to its small leaflet area relative to its contour. When evaluating leaflet angles, annulus placement, and center curve splines, our approach significantly improved compared to the baseline. Most notably, we observed a reduction of outliers in these metrics. This reduction can be explained by an increased robustness to out-of-domain cases, for which UNet was prone to significant failure. Such failure cases can be detected by user supervi-

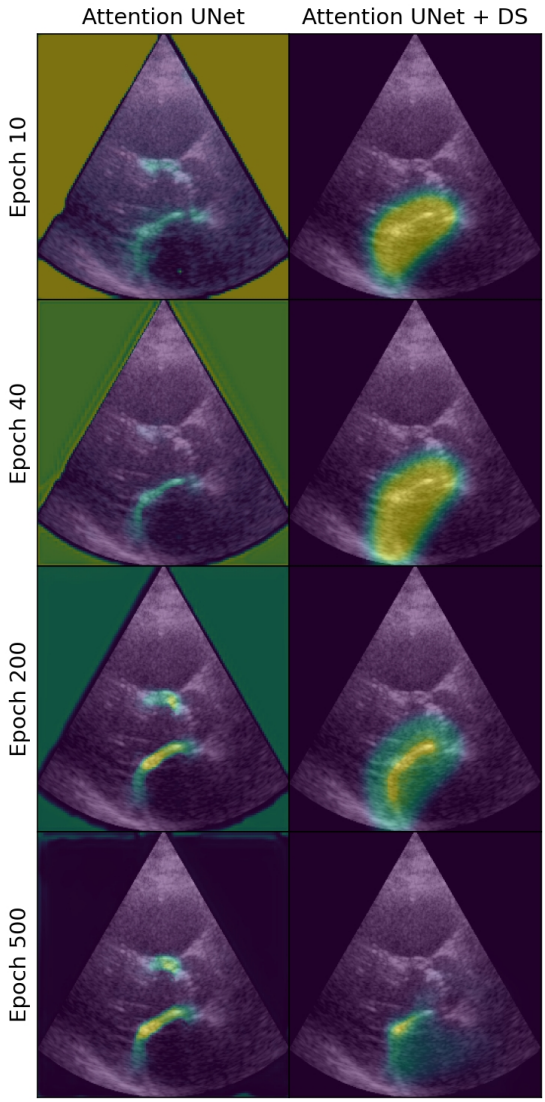


Figure 3.8: Evolution of the attention coefficient map α_5 during training for Attention UNet with and without DS applied. Using DS, the attention coefficient maps quickly converge to emphasize the relevant region in the image. Between epochs 40 and 200, the loss weight scheduler gradually introduced the segmentation loss into the loss function. Beyond epoch 200, only the segmentation loss was optimized. We can observe that the salient region was then narrowed further towards the contour of the MV. By contrast, Attention UNet without DS highlighted both the mitral and aortic valves, proving less aligned with the desired objective.

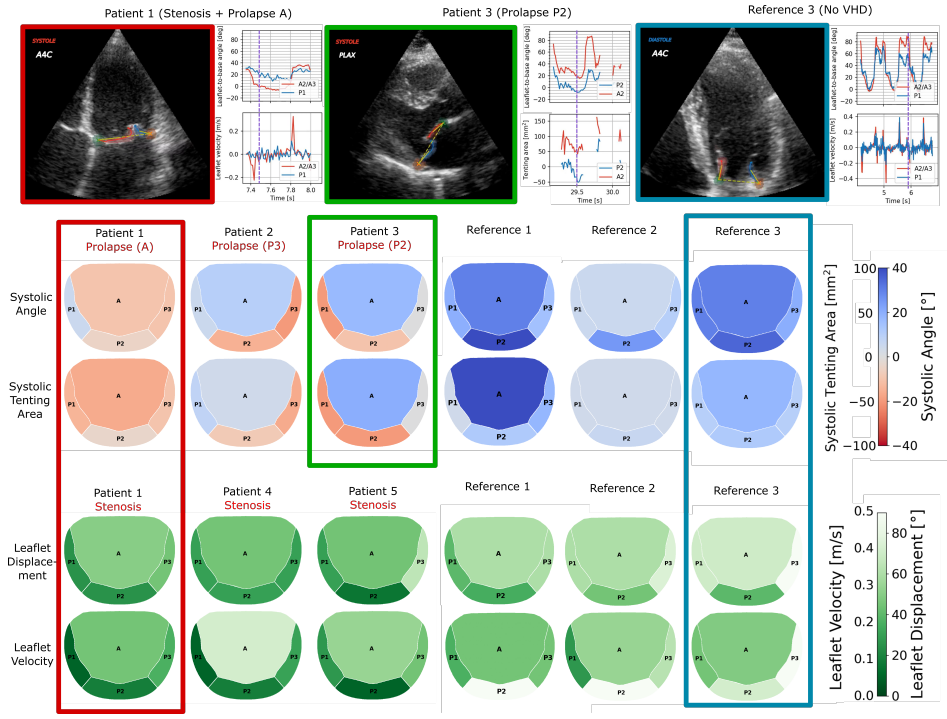


Figure 3.9: Top: Examples of MV segmentation, tracking, and biomarker extraction for cases of stenosis, prolapse, and a healthy reference. Bottom: Heatmaps of the MV (A, anterior; P, posterior; 1, anterolateral; 2, central; 3, posteromedial) showing systolic tenting area, systolic angle, leaflet displacement, and velocity measured by applying our approach on exams consisting of multiple standard echocardiogram recordings. Patients 1–3 had mitral valve prolapse at the time of the recording. According to the cardiologist's report, the scallops subject to prolapse are denoted in red text. Patients 1, 4, and 5 had mitral stenosis. References 1–3 were healthy individuals with no VHD.

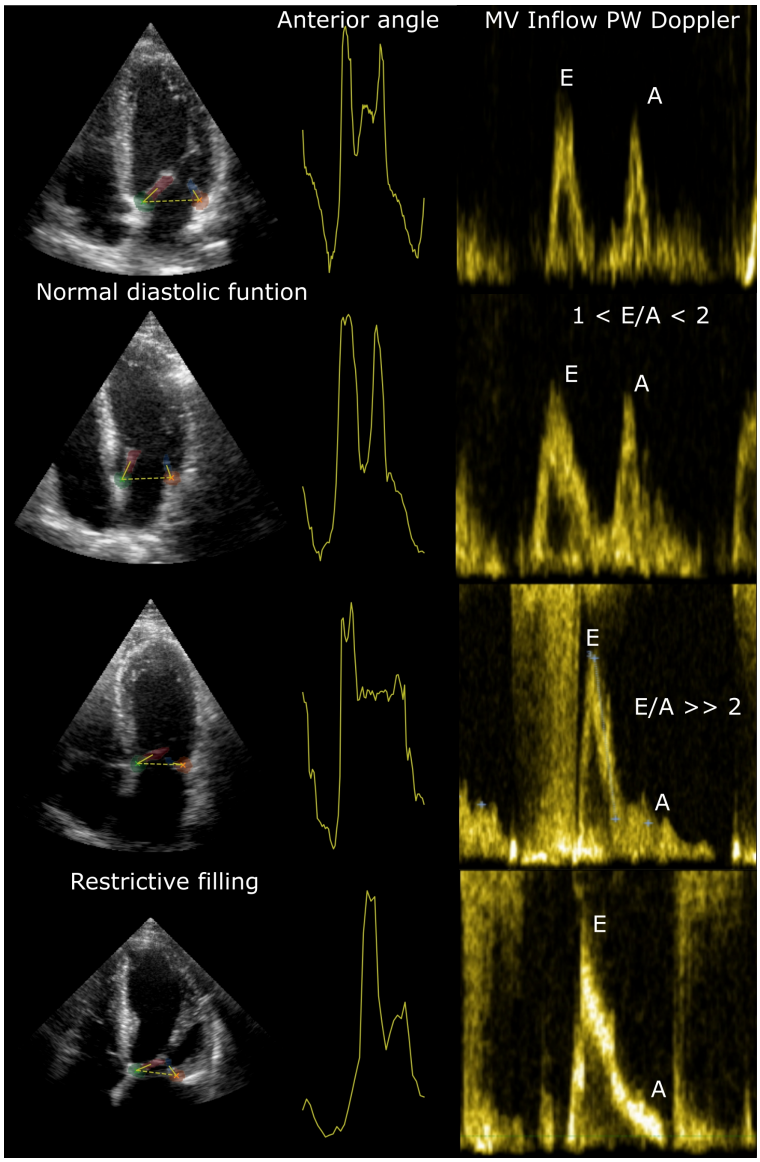


Figure 3.10: Tracking patterns of the anterior leaflet for A4C views in two cases of normal diastolic function (top) and two cases of restrictive filling (bottom). The tracking results (middle) are compared to MV Inflow Doppler spectra (right). We can observe that in cases of restrictive filling, the atrial systole peak angle is notably reduced, similar to the A wave velocity peak in the Doppler spectrum.

sion; however, mitigating failure cases is imperative for building trust among users.

Despite satisfactory results, challenges exist for MV segmentation, especially when subject to poor image quality or when leaflets are partially visible or overlap. In poor-quality images, the MV is sometimes barely visible. These cases pose challenges to the annotators, leading to higher uncertainty in the reference annotations. There was also increased uncertainty in the reference annotations for cases of severe rheumatic valve thickening, which made the model prone to flickering segmentation contours in such cases. Moreover, echocardiography is intrinsically a spatio-temporal modality and MV leaflet motion is highly dynamic. Therefore, we believe leveraging temporal information during training is likely beneficial for improving segmentation accuracy and robustness. Similar to previous research [25]–[27], this might be achievable using semi-supervised learning with spatio-temporal pseudo-labels.

Moreover, the MV structure might be better modeled as a graph using a Graph Convolutional Neural Network (GCN) [28]. We also expect further improvement if separate models are trained for each view, as suggested by previous research [14]. However, this requires an increased training set for all models to learn effectively.

Automatic measurements of the MV from standard echocardiograms offer benefits such as facilitating standardized assessments, reducing clinic workload, and aiding non-expert users with improved visualization and detection of abnormalities. Since manual tracking of MV leaflet motion is infeasible, there is limited knowledge about how MV leaflet motion relates to VHD and heart function. Future studies will assess the clinical viability of our approach and, if possible, determine cutoff values for the proposed biomarkers. Future work should also further investigate the use of MV motion time series concerning cardiac function. We believe that studying MV leaflet motion at the various stages of diastolic dysfunction might uncover important information about interpreting the observed patterns.

3.5 Conclusions

In this work, we presented a deep learning framework for segmentation of the mitral valve leaflets and annulus hinge points from transthoracic echocardiograms in five standard views. Moreover, we derived robust quantitative measures of angular movement and tenting areas for the different leaflet scallops, which were feasible for assessing mitral valve prolapse and stenosis. We showed that Attention UNet with deep supervision of the attention coefficients may improve focus on salient regions in the image, improving ro-

bustness and providing higher temporal consistency. We believe automated measurements of MV morphology and motion could be valuable in the assessment of mitral valve disease and might additionally provide information about cardiac function.

References

[1] 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.

[2] C. M. Otto, R. A. Nishimura, R. O. Bonow, *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," *Circulation*, vol. 143, no. 5, pp. 35–71, 2021.

[3] A. Vahanian, F. Beyersdorf, F. Praz, *et al.*, "2021 ESC/EACTS guidelines for the management of valvular heart disease," *European Heart Journal*, vol. 43, no. 7, pp. 561–632, 2022.

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

[5] A. Wang, P. Grayburn, J. A. Foster, *et al.*, "Practice gaps in the care of mitral valve regurgitation: Insights from the american college of cardiology mitral regurgitation gap analysis and advisory panel," *American Heart Journal*, vol. 172, pp. 70–79, 2016.

[6] A. Hagendorff, F. Knebel, A. Helfen, *et al.*, "Echocardiographic assessment of mitral regurgitation: Discussion of practical and methodologic aspects of severity quantification to improve diagnostic conclusiveness," *Clinical Research in Cardiology*, vol. 110, no. 11, pp. 1704–1733, 2021.

[7] J. J. Silbiger, "Mechanistic insights into ischemic mitral regurgitation: Echocardiographic and surgical implications," *Journal of the American Society of Echocardiography*, vol. 24, no. 7, pp. 707–719, 2011.

[8] J. Magne, P. Pibarot, F. Dagenais, Z. Hachicha, J. Dumesnil, and M. Sénéchal, "Preoperative posterior leaflet angle accurately predicts outcome after restrictive mitral valve annuloplasty for ischemic mitral regurgitation," *Circulation*, vol. 115, no. 6, pp. 782–791, 2007.

[9] A. M. Calafiore, S. Gallina, M. Di Mauro, *et al.*, "Mitral valve procedure in dilated cardiomyopathy: Repair or replacement?" *Circulation*, vol. 71, no. 4, pp. 1146–52, 2001.

[10] S. R. Snare, O. C. Mjølstad, F. Orderud, H. Dalen, and H. Torp, "Automated septum thickness measurement—a kalman filter approach," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 477–486, 2012.

[11] M. Sultan, N. Martins, E. Costa, *et al.*, "Virtual M-mode for echocardiography: A new approach for the segmentation of the anterior mitral leaflet," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 2018.

[12] L. Corinzia, F. Laumer, A. Candreva, M. Taramasso, F. Maisano, and J. M. Buhmann, "Neural collaborative filtering for unsupervised mitral valve segmentation in echocardiography," *Artificial Intelligence in Medicine*, vol. 110, p. 101 975, 2020.

[13] H. Dröge, B. Yuan, R. Llerena, J. T. Yen, M. Moeller, and A. L. Bertozzi, "Mitral valve segmentation using robust nonnegative matrix factorization," *Journal of Imaging*, vol. 7, no. 10, 2021.

[14] E. Costa, N. Martins, M. S. Sultan, *et al.*, "Mitral valve leaflets segmentation in echocardiography using convolutional neural networks," in *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG)*, 2019, pp. 1–4.

[15] M. Vafaeezadeh, H. Behnam, A. Hosseinsabet, and P. Gifani, "Carpnet: Transformer for mitral valve disease classification in echocardiographic videos," *International Journal of Imaging Systems and Technology*, 2023.

[16] J. Chen, H. Li, G. He, *et al.*, "Automatic 3D mitral valve leaflet segmentation and validation of quantitative measurement," *Biomedical Signal Processing and Control*, vol. 79, no. 2, 2023.

[17] B. S. Andreassen, D. Volgyes, E. Samset, and A. H. S. Solberg, "Mitral annulus segmentation and anatomical orientation detection in tee images using periodic 3D cnn," *IEEE Access*, vol. 10, pp. 51 472–51 486, 2022.

[18] M. Ivantsits, B. Pfahringer, M. Huellebrand, *et al.*, "3D mitral valve surface reconstruction from 3d tee via graph neural networks," in *Statistical Atlases and Computational Models of the Heart: Regular and CMRxMotion Challenge Papers*, O. Camara, E. Puyol-Anton, C. Qin, *et al.*, Eds., ser. Lecture Notes in Computer Science, vol. 13593, 2022, pp. 330–339.

[19] 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 Computind and Computer Assised Intervention (MICCAI) 2021, PT V*, M. deBruijne, P. Cattin, S. Cotin, *et al.*, Eds., ser. Lecture Notes in Computer Science, vol. 12905, 2021, pp. 459–468.

[20] A. M. Fiorito, A. Østvik, E. Smistad, S. Leclerc, O. Bernard, and L. Lovstakken, "Detection of cardiac events in echocardiography using 3D convolutional recurrent neural networks," in *2018 IEEE International Ultrasonics Symposium (IUS)*, 2018, pp. 1–4.

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

[22] E. Smistad, A. Østvik, and L. Løvstakken, "Annotation web - an open-source web-based annotation tool for ultrasound images," in *2021 IEEE International Ultrasonics Symposium (IUS)*, 2021, pp. 1–4.

[23] J. Schlemper, O. Oktay, M. Schaap, *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[24] J. F. Grue, S. Storve, A. Støylen, *et al.*, "Normal ranges for automatic measurements of tissue doppler indices of mitral annular motion by echocardiography. Data from the HUNT3 study," *Echocardiography*, vol. 36, no. 9, pp. 1646–1655, Sep. 2019.

[25] H. Wei, H. Cao, Y. Cao, *et al.*, "Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Springer International Publishing, 2020, pp. 623–632.

[26] J. Hu, E. Smistad, I. M. Salte, H. Dalen, and L. Lovstakken, "Exploiting temporal information in echocardiography for improved image segmentation," in *2022 IEEE International Ultrasonics Symposium (IUS)*, 2022, pp. 1–4.

[27] H. J. Ling, N. Painchaud, P.-Y. Courand, P.-M. Jodoin, D. Garcia, and O. Bernard, "Extraction of volumetric indices from echocardiography: Which deep learning solution for clinical use?" In *Functional Imaging and Modeling of the Heart*, Springer Nature Switzerland, 2023, pp. 245–254.

[28] S. Thomas, A. Gilbert, and G. Ben-Yosef, "Light-weight spatio-temporal graphs for segmentation and ejection fraction prediction in cardiac ultrasound," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Cham: Springer Nature Switzerland, 2022, pp. 380–390.

Chapter 4

EasyPISA - Automatic Integrated PISA Measurements of Mitral Regurgitation from 2-D Color-Doppler using Deep Learning

This paper is submitted for publication and is therefore not included.

Chapter 5

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

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 is 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 estima-

tion error (THD: 10.4 ± 8.4 mm^2, DC: 10.98 ± 8.17, ours: 9.9 ± 6.0 mm^2), and flow rate estimation error (THD: 20.3 ± 9.9 ml/s, DC: 18.14 ± 13.01 ml/s, ours: 7.1 ± 10.6 ml/s). We also demonstrate *in vivo* feasibility for six patients with aortic insufficiency, compared with standard echocardiography and magnetic resonance references.

5.1 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 be 18–19% in middle-aged adults [1], and is projected to increase overall due to an aging 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] 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-Doppler and Continuous Wave (CW) Doppler recordings [4]. However, 2-D PISA is highly user-dependent [5], [6] because several important steps must be performed manually, such as selecting the imaging plane, time frame, color-Doppler 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 [7].

Recent research based on 3-D Doppler ultrasound in invasive transesophageal echocardiography (TEE), and have shown promise for accurate and less user-dependent assessment [8]–[12]. 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 [13]. A high PRF, weakly focused acquisition was used to acquire the entire region of interest continuously, rather than using packet acquisitions as in color-Doppler 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-Doppler. 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 were achieved. However, the area estimator is highly dependent on the choice of power threshold for detecting voxels that contain flow. Moreover, due to the large point spread function (PSF), it is difficult to accurately depict irregular orifice geometries [14]. 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 that limits resolution and image quality. Techniques for restoring such images are typically based on deconvolution [15]–[17]. 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 [18], [19], as well as a vast amount of structure segmentation from B-Mode images [20]–[23].

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.

5.2 Methods

Orifice generation

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

$$\boldsymbol{B}(t) = \sum_{i=0}^{p} \binom{p}{i} (1-t)^{p-i} t^i \boldsymbol{C}_i, \tag{5.1}$$

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

$$\boldsymbol{C}_i \begin{cases} \sim \mathcal{U}(x,y) & \text{if } i < p \\ = \boldsymbol{C}_0 & \text{if } i = p, \end{cases} \tag{5.2}$$

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 5.1.

Figure 5.1 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 [24], [25]. Using our approach we can generate a large variety of shapes for the training set, which prevents the model from overfitting on certain geometries.

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 [26], [27] and the parameters listed in Table 5.2. Finally, the power-Doppler realizations $R_0(x,y)$ were

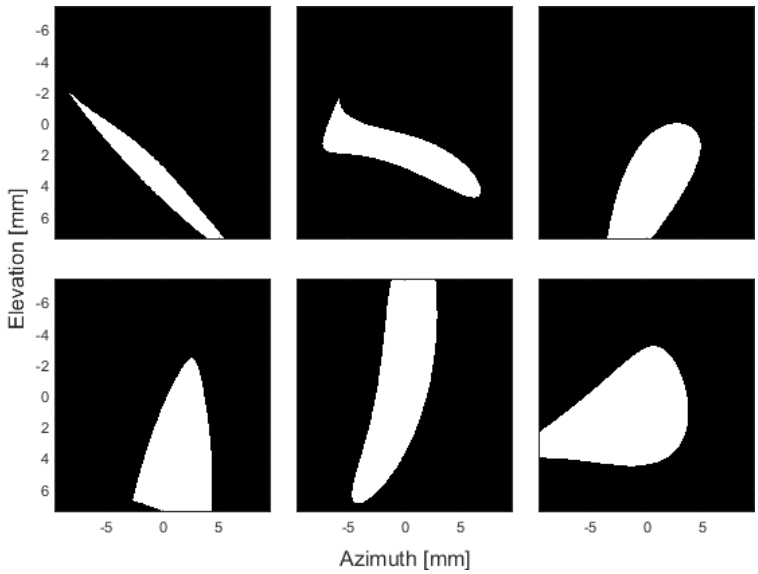


Figure 5.1: Example orifices generated using (5.1) and (5.2).

generated by integrating the pulse-echo contributions from the pixels that belong to the orifice $I(x, y)$.

$$R_0(x, y) = \iint_I P_h(x, y, x_h, y_h) dx_h dy_h, \tag{5.3}$$

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

$$P_h(x, y, x_h, y_h) = \int_{-\infty}^{+\infty} |h_{pe}(x, y, x_h, y_h, t)|^2 dt. \tag{5.4}$$

The same pulse-echo field can be used to generate different power-Doppler realizations by changing the orifice map, which allows for a quick generation of training data. By integrating the pulse echo signals as in (5.3) we can achieve simulations 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.

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

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

bivariate Gaussian functions to each simulated power-Doppler image $R_0^{(i)}$, creating the augmented image $\tilde{R}_0^{(i)}$

$$\tilde{R}_0^{(i)} = R_0^{(i)} + \alpha \sum_{j=1}^{N_i} \mathcal{N}(\boldsymbol{\mu}_j, S_j), \tag{5.5}$$

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 5.1.

Model training

The simulation and training phases are depicted in Figure 5.2. The GE 4Vc-D geometry was used. The simulation parameters are shown in Table 5.2. 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 (5.5). Depth and frequency information was added as separate input channels as images with all points 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 [28] 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 that

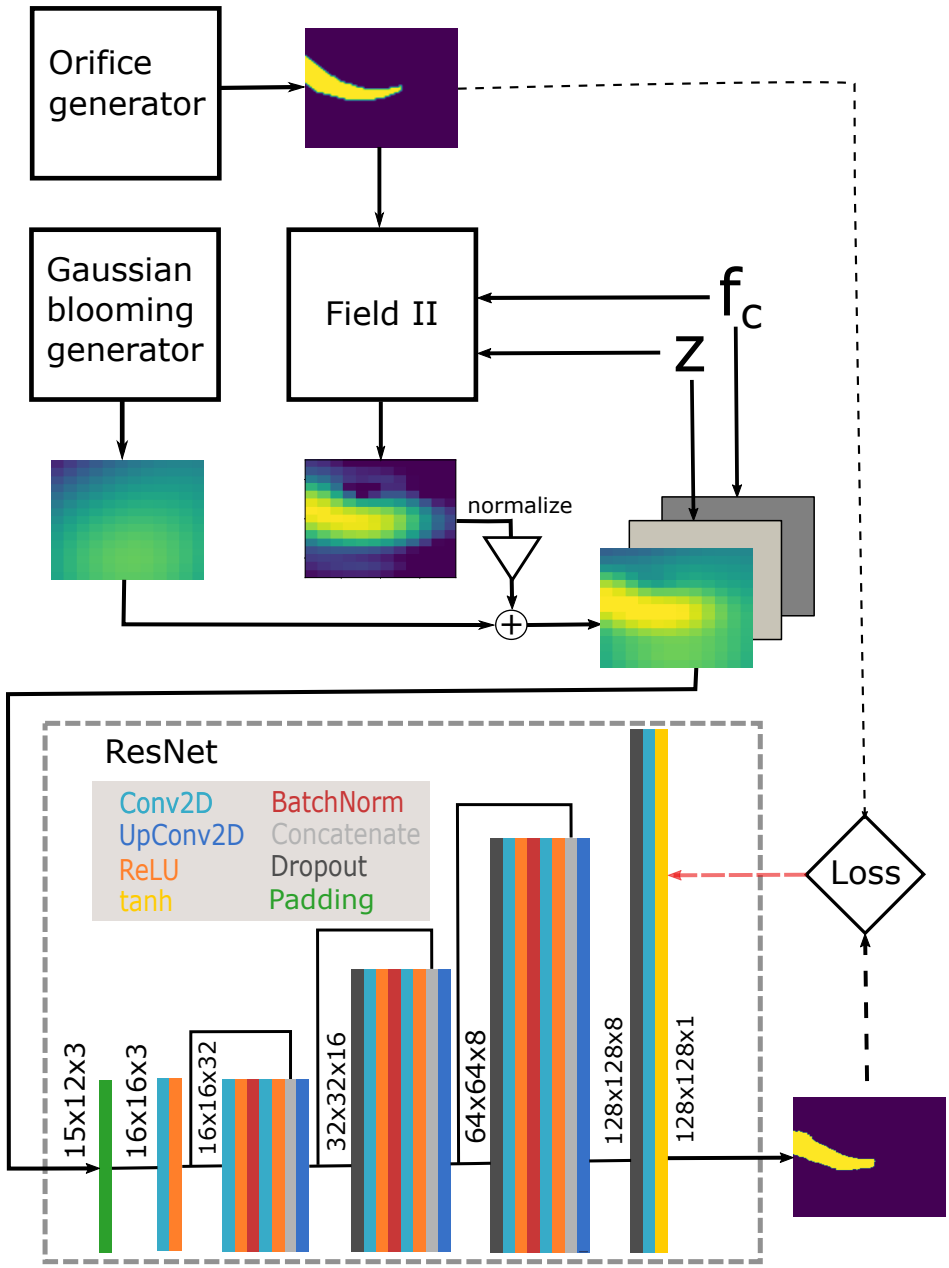


Figure 5.2: Schematic representation of the network training. Binary maps representing regurgitant orifices are procedurally generated. Their power-Doppler responses are simulated using Field-II, and Gaussian blooming artifacts are superimposed to the normalized images. The resulting images, along with depth z and frequency f_c parameters, are used to train a CNN (ResNet) to perform segmentation from the power-Doppler images.

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 5.2: Field II parameters

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 an 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.

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) [29], which has been shown to improve

training stability. We hereby refer to our network as "ResNet".

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. [13]. The other approach was spatially invariant non-blind deconvolution, similar to [15]. We used the Richardson-Lucy deconvolution algorithm [30] with 20 iterations to deconvolve the power-Doppler images using the analytical PSF

$$\mathrm{PSF} = \left[\mathrm{sinc}\left(\frac{f_c L_{\mathrm{AZ}}}{cz}x, \frac{f_c L_{\mathrm{EL}}}{cz}y\right)\right]^2, \tag{5.6}$$

where L_{AZ} and L_{EL} are the aperture dimensions calculated using Table 5.2, 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.

Phantom experiments

We validated the method using the custom-made flow phantom shown in Figure 5.3. The acquisition and signal processing parameters are summarized in Table 5.3. The phantom was filled with a mixture of water and cornstarch 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).

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 the 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

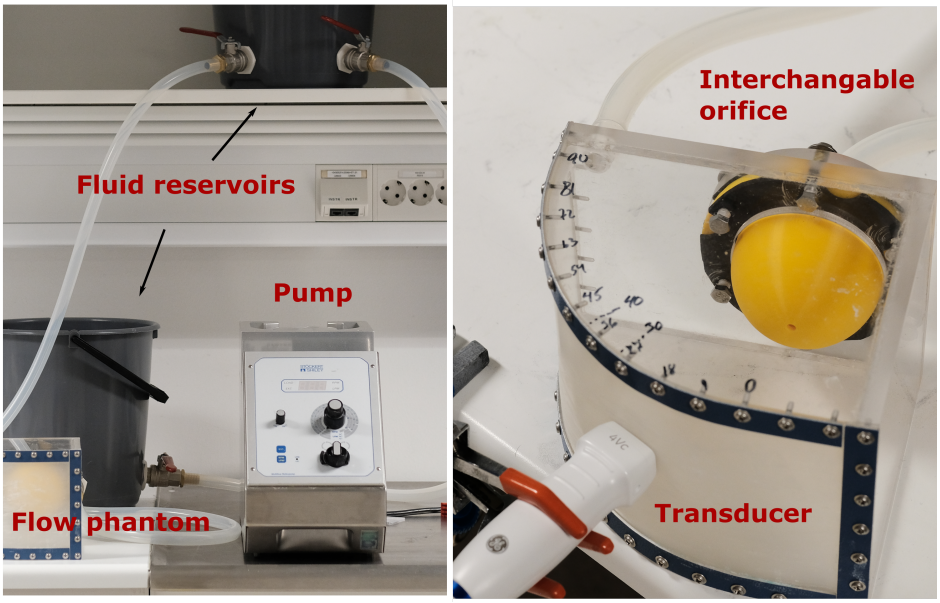


Figure 5.3: Experimental setup. Overview (left) and flow phantom closeup (right).

of 50%. The power-Doppler signal was smoothed temporally using a moving average filter with length $N_{\text{smooth}} = 11$ (1 ms), and radially using a filter with length $N_z = 3$ (0.2 mm).

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

$$\hat{v}_{\text{mean}} = \hat{v}_{\text{max}} - \frac{\hat{B}}{2}, \tag{5.7}$$

where $\hat{v}_{\text{max}}$ is the maximum velocity envelope estimated from the PW Doppler 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 [31]. Estimating the mean velocity using the spectral envelope was preferred over autocorrelation as mean velocity estimators are biased towards low-frequency clutter, which was present in cases of suboptimal clutter filtering. The maximum estimator was shown to be more robust in [13].

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 5.3: Acquisition and signal processing parameters

$$\text{CSA}(t) = \iint_{x,y} g(R_0(x, y, z_{\text{vc}}, t))dxdy, \tag{5.8}$$

$$Q(t) = \iint_{x,y} \hat{v}_{\text{mean}}(x, y, z_{\text{vc}}, t)g(R_0(x, y, z_{\text{vc}}, t))dxdy, \tag{5.9}$$

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.

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 for the experimental setup. 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 have approval from our industry partners for our

improved setup using a -40 cm focal point, but approval was granted at a later time. Comprehensive echocardiographic examinations were performed to provide reference values for EROA and RVol using 2-D PISA. Cardiac Magnetic Resonance Imaging (cMRI) 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

$$z_{\mathrm{vc}}(t) = \underset{z}{\operatorname{argmax}} \iint_{x,y} \hat{v}_{\max}(x, y, z, t) dx dy, \tag{5.10}$$

where $\hat{v}_{\max}$ is the same maximum velocity envelope as in (5.7). 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.

5.3 Results

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. 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 5.4. 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 5.5 indicate that ResNet achieves improved segmentation accuracy compared to the references and differentiates better between small and large orifices.

Figure 5.6 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

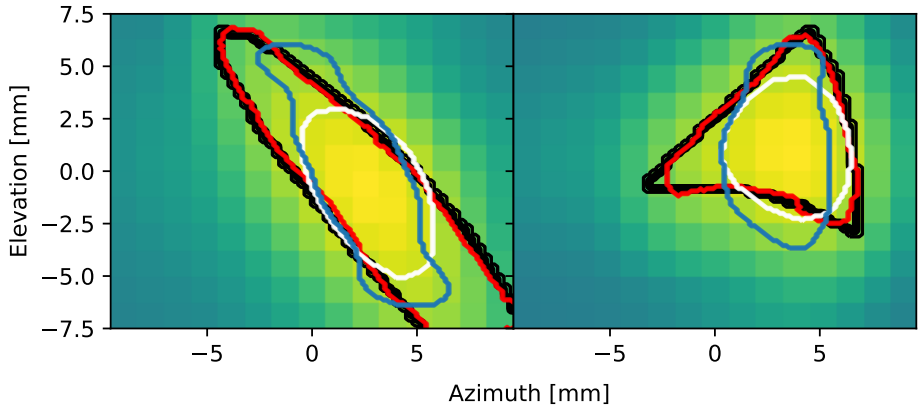


Figure 5.4: *In silico* power-Doppler segmentation examples (Red: ResNet; Blue: Deconvolution; White: power thresholding; Black: ground truth).

increasing depth and also decreases for frequencies outside of the training domain. This was also expected since the transmit frequency and imaging depth affect both the axial and the lateral resolution.

Phantom experiments

Figure 5.7 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 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 5.8 and 5.9 show the flow rate and area estimates from the phantom setup with four circular orifices. Results indicate that ResNet achieves slightly less biased estimates compared to power thresholding and deconvolution, while inter-frame variability is similar.

Figure 5.10 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.

In vivo feasibility analysis

Figure 5.11 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 Cardiac Magnetic

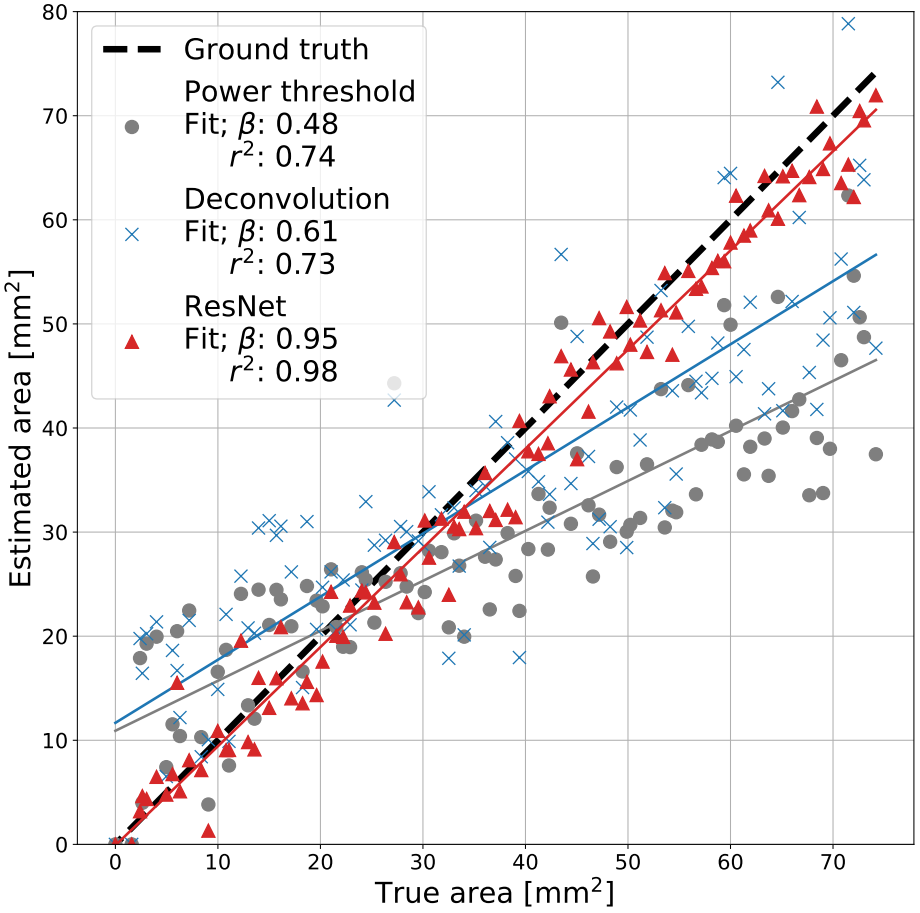
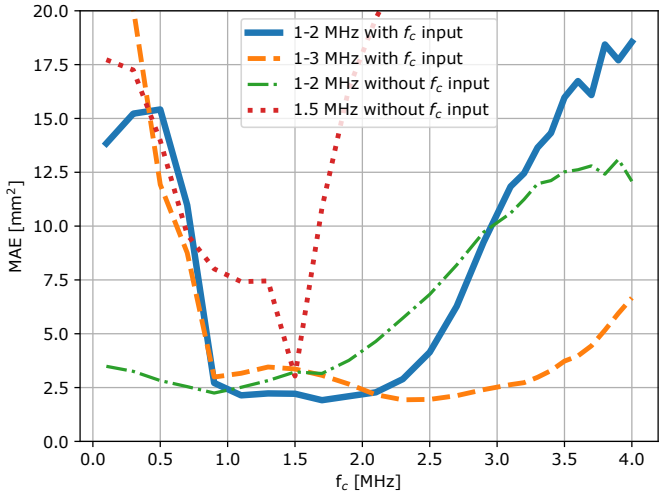
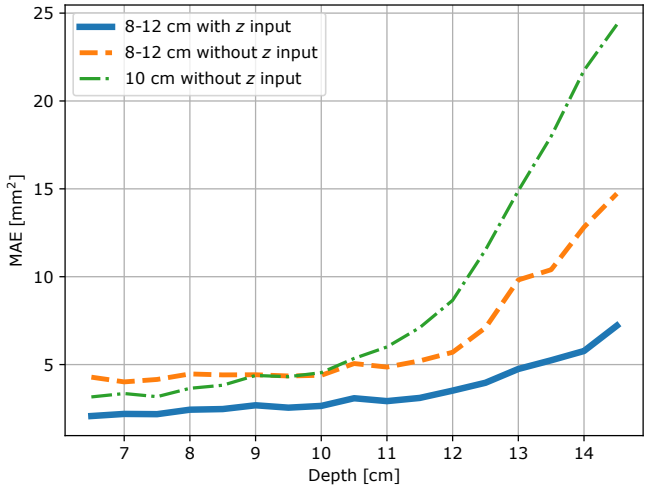


Figure 5.5: *In silico* area estimation results. The plots show area estimates using ResNet, deconvolution, and power thresholding on a test dataset simulated similarly to the training dataset. Linear regression slopes and coefficients of determination are denoted by β and r^2. Ground truth refers to the areas of the ground truth masks.



a)



b)

Figure 5.6: 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.

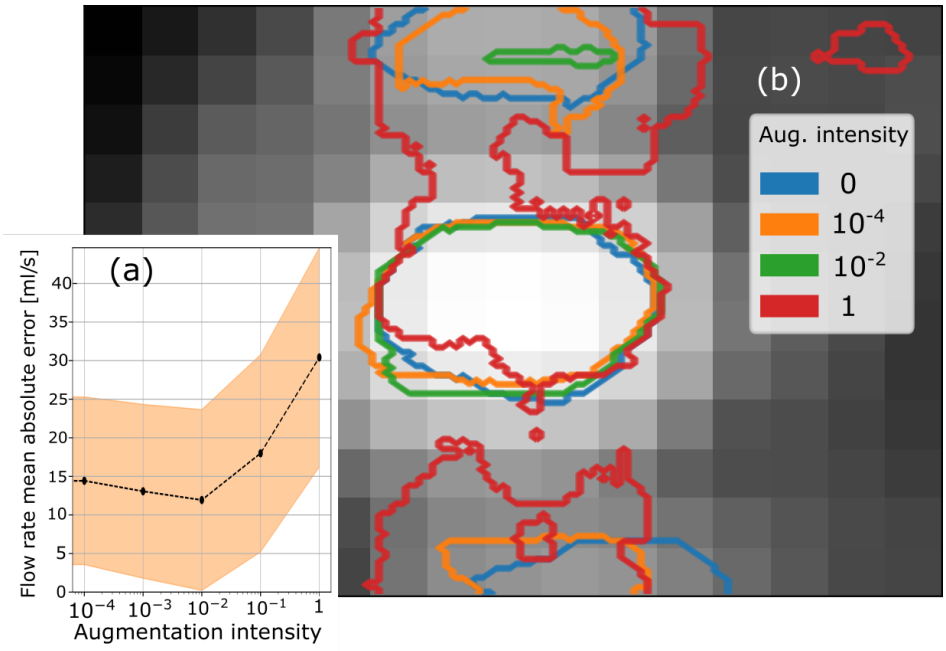
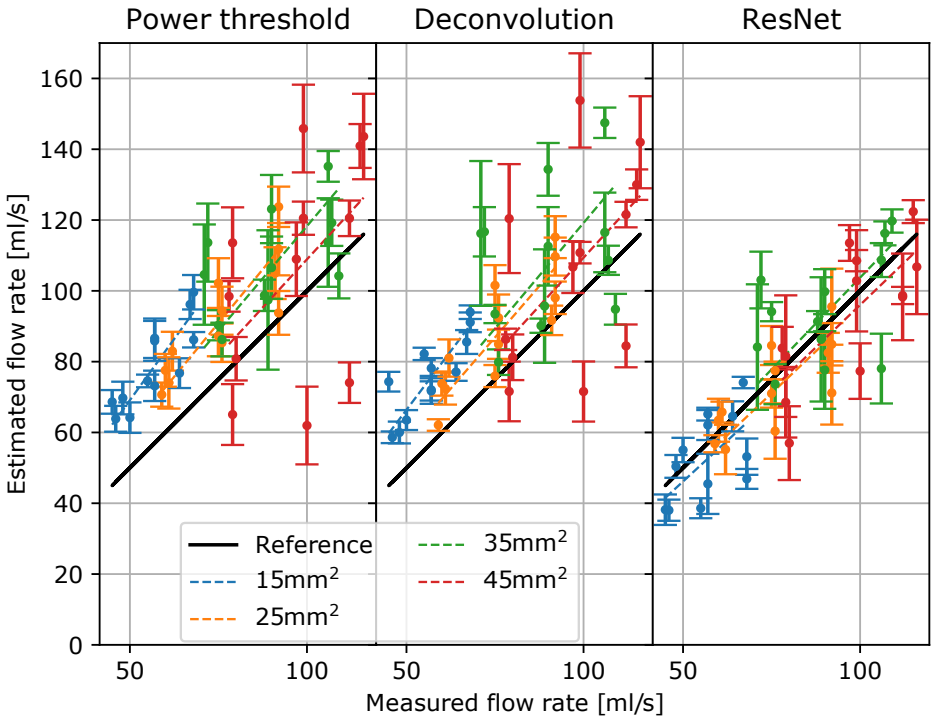


Figure 5.7: Results from using models trained with various augmentation intensities α. In (a), a grid search for the best value of α is depicted, showing flow rate estimation mean errors and standard deviation for models with varying α values. The search resulted in $\alpha = 10^{-2}$. In (b), we see a power-Doppler elliptical cross-section of a jet from the experimental setup. 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 mm²	$\beta: 1.38, r^2: 0.71$	$\beta: 1.33, r^2: 0.69$	$\beta: 0.92, r^2: 0.28$
25 mm²	$\beta: 1.21, r^2: 0.73$	$\beta: 1.16, r^2: 0.79$	$\beta: 0.95, r^2: 0.54$
35 mm²	$\beta: 1.18, r^2: -0.16$	$\beta: 1.19, r^2: -0.28$	$\beta: 1.04, r^2: -0.01$
45 mm²	$\beta: 1.09, r^2: 0.19$	$\beta: 1.09, r^2: 0.22$	$\beta: 0.96, r^2: 0.55$

Figure 5.8: 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.

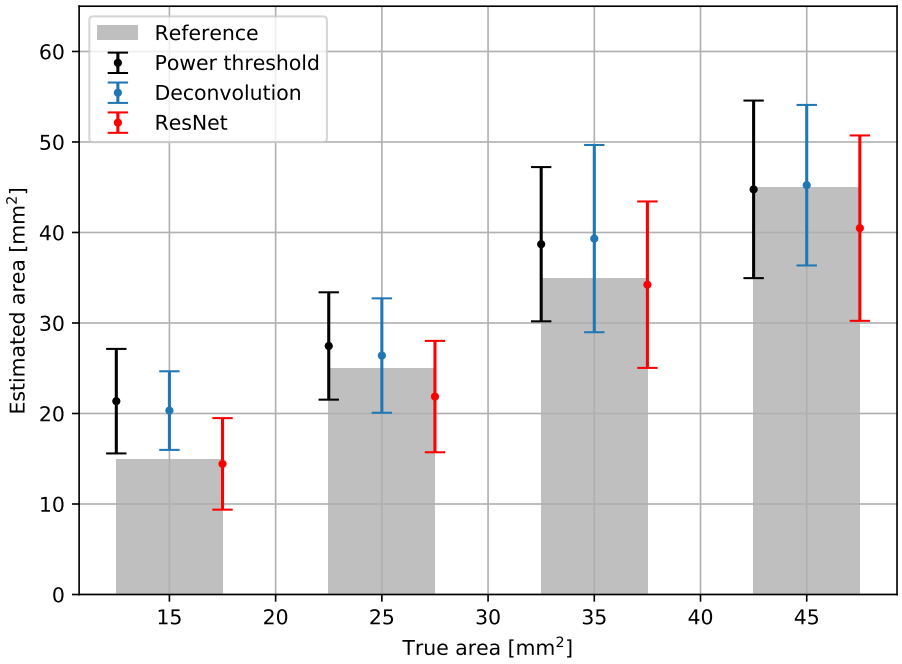


Figure 5.9: Angle corrected cross-sectional area estimates of circular orifices (15, 25, 35, 45 mm^2) using segmentation from power thresholding, deconvolution, and using ResNet. Error bars signify the standard deviation for all angles and flow rates for a given orifice size.

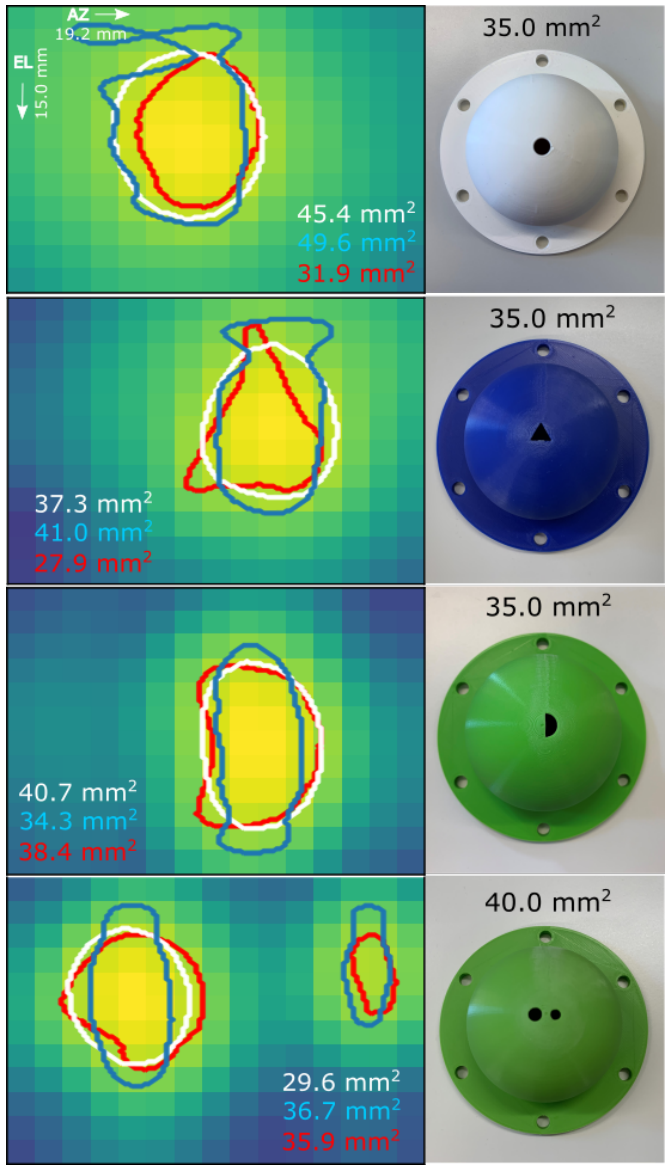


Figure 5.10: Recording examples of orifices with different shapes. The left panels show instantaneous power-Doppler images with thresholding segmentation (white), deconvolution (blue), and ResNet segmentation (red). The segmented areas are noted in white, blue, and red text respectively. The right panels show the corresponding flow phantoms used in the experiment, with the orifice area noted in black text.

Resonance Imaging (cMRI) and 2-D PISA. In Figure 5.11 b), 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 5.12 shows PW spectra, power-Doppler with jet segmentation, and velocity estimates from the six patients. Figure 5.13 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 5.12. We can see in Figure 5.11 a) that this ability has a big impact on the RVol estimates, in which ResNet has a better agreement with the 2-D PISA and cMRI references.

5.4 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 orifice area and regurgitant volume. Experimental and simulation results shown in Figure 5.5, 5.4, 5.8 and 5.10 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 low-quality images. We also demonstrated feasibility for six *in vivo* cases of aortic regurgitation, as shown in Figure 5.11 and 5.12.

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 the presence of a diffuse signal surrounding the jet in the observed data, which caused a further reduction in contrast. We believe that the cause for this signal component could be a combination of recruited flow, and defocusing due to phase aberration and side lobes. We could account for these blooming effects by using augmentation with Gaussian bivariate functions, a strategy which previously has been applied to account for shadowing artifacts in B-Mode images [32], [33].

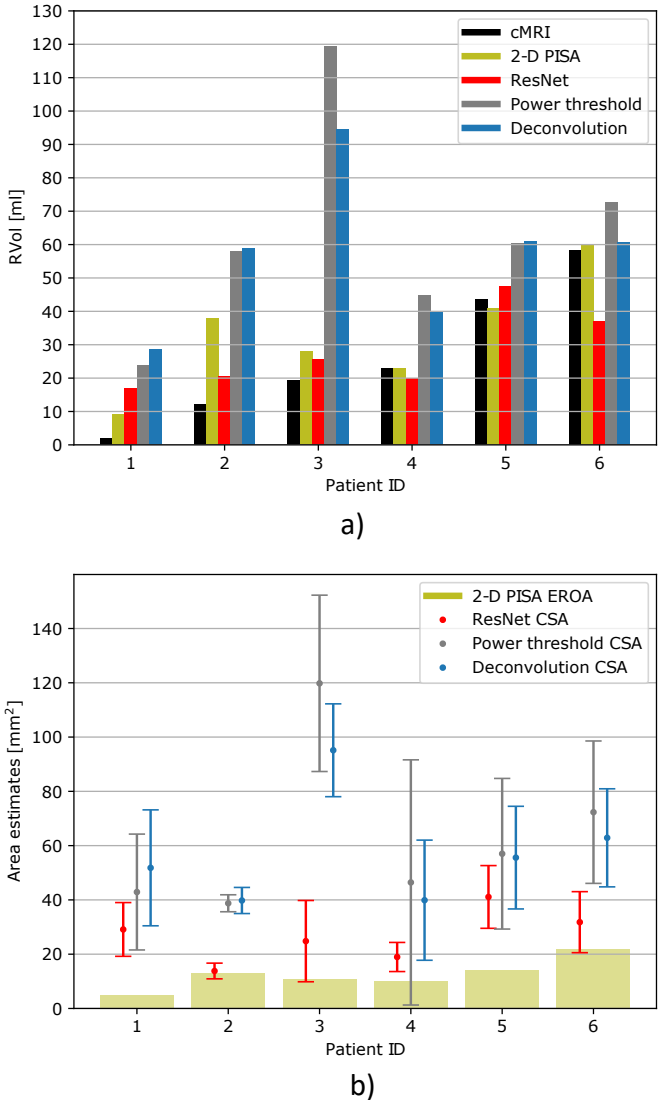


Figure 5.11: RVol (a) and area (b) estimates of six patients with aortic regurgitation from cMRI, 2-D PISA, and 3-D Doppler with segmentation using ResNet, deconvolution or power thresholding. cMRI measurements were feasible for RVol, but not for area measurements. CSA error bars signify the mean and 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.

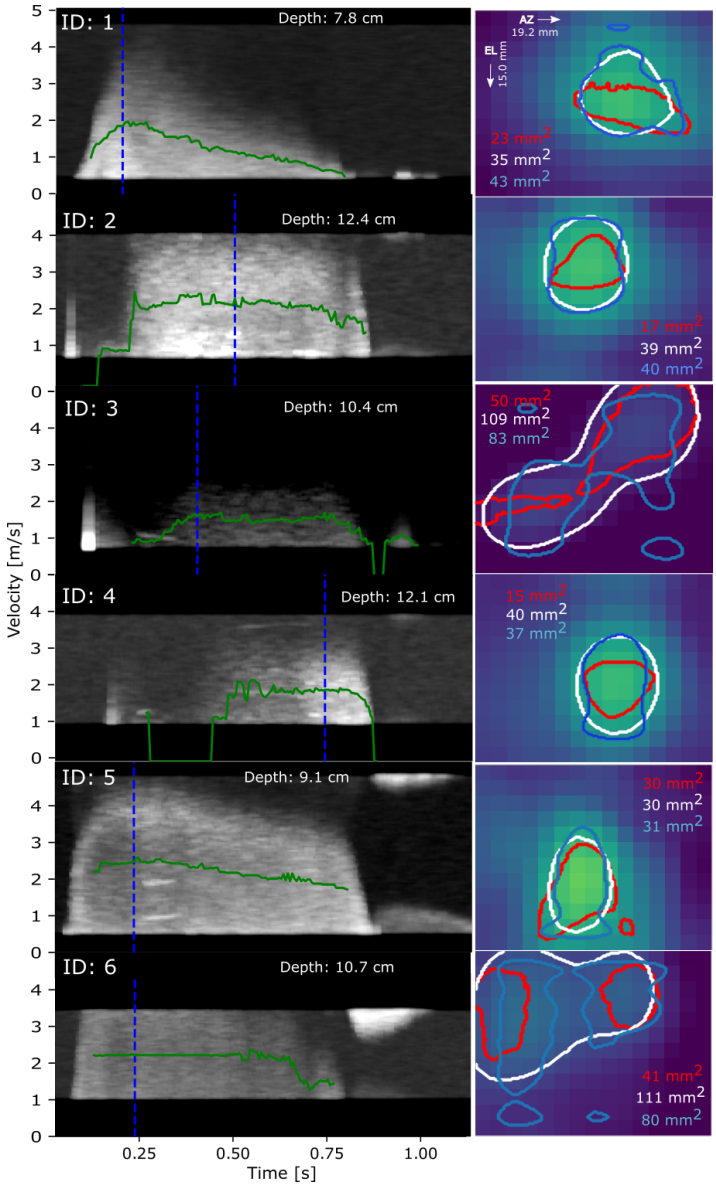


Figure 5.12: *In vivo* recordings of six patients with aortic regurgitation. The left panels show the PW spectra, with mean velocity traces plotted in green, and current time frames marked by blue dashed lines. The right panels show power-Doppler cross sections with the CSA segmented using thresholding (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 5.11.

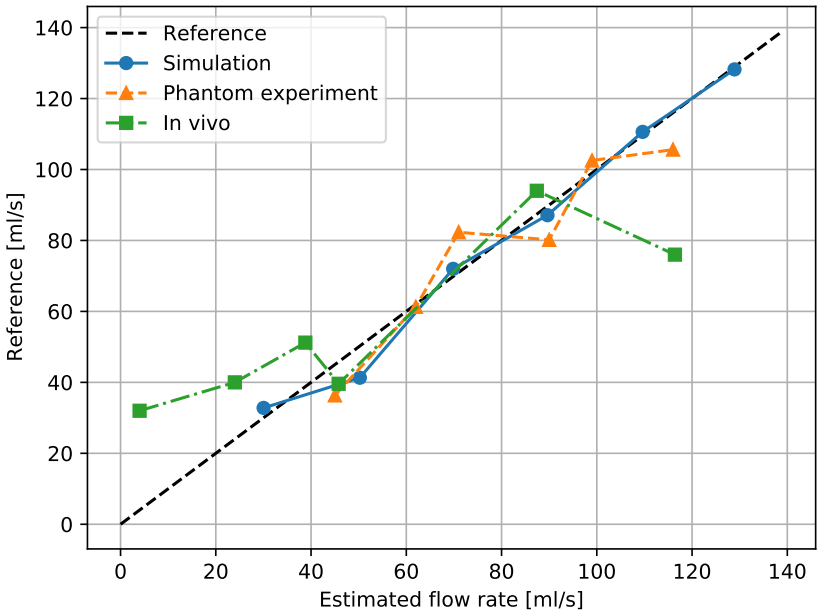


Figure 5.13: Comparison of simulated, experimental and *in vivo* flow rate estimation results using ResNet. The graphs are displaying all *in vivo* results with the cMRI reference. Simulated and experimental examples are arbitrarily chosen from the results previously shown while ensuring a representative 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.

To illustrate how increasing realism increases the problem complexity, flow rates from the different test environments are compared in Figure 5.13. 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 that even though our deep learning model performs well within the simulator conditions, the simulator is limited in providing sufficiently realistic training examples that 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 the 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 clinical 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 and 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 a 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 that is fast enough to generate enough training examples in a reasonable time, while still accounting for spatially variant PSFs. Field-II would not be fast enough for this purpose, however, FLUST [34] 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.

5.5 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.

References

[1] J. P. Singh, J. C. Evans, D. Levy, *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, D. Adams, R. O. Bonow, *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] G. S. Bargiggia, L. Tronconi, D. J. Sahn, *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.

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

[6] S. Biner, A. Rafique, F. Rafii, *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.

[7] 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.

[8] T. C. Tan, X. Zeng, Y. Jiao, *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.

[9] 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.

[10] B. Goebel, R. Heck, A. Hamadanchi, *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.

[11] S. Militaru, O. Bonnefous, K. Hami, *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.

[12] A. Singh, J. Su, A. This, *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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] S. Leclerc, E. Smistad, J. Pedrosa, *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.

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

[22] 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.

[23] C. Li, X. Lin, Y. Mao, *et al.*, "Domain generalization on medical imaging classification using episodic training with task augmentation," *Computers in Biology and Medicine*, vol. 141, p. 105 144, 2022.

[24] Y. Otsuji, M. Handschumaker, E. Schwammenthal, *et al.*, "Insights from three-dimensional echocardiography into the mechanism of functional mitral regurgitation," *Circulation*, vol. 96, no. 6, pp. 1999–2008, 1997.

[25] 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.

[26] 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.

[27] J. Jensen, "Field: A program for simulating ultrasound systems," *Medical and Biological Engineering and Computing*, vol. 34, pp. 351–352, 1996.

[28] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2014.

[29] 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.

[30] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, p. 745, 1974.

[31] 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.

[32] 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.

[33] A. Østvik, I. M. Salte, E. Smistad, *et al.*, "Myocardial function imaging in echocardiography using deep learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1340–1351, 2021.

[34] 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.

ISBN 978-82-326-7988-1 (printed ver.)
ISBN 978-82-326-7987-4 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology