# Intraventricular Vector Flow Mapping by Color Doppler using Physics-based Neural Networks

Hang Jung Ling

N° d'ordre NNT : 2024ISAL0087

**THESE de DOCTORAT DE L'INSA LYON,**
**membre de l'Université de Lyon**

**Ecole Doctorale** N° 162
**Mécanique, Énergétique, Génie civil, Acoustique (MEGA)**

**Spécialité/ discipline de doctorat : Acoustique**

Soutenue publiquement le 07/11/2024, par :

**Hang Jung LING**

---

# Écoulement intraventriculaire en échocardiographie Doppler avec réseaux de neurones fondés sur la physique

---

Devant le jury composé de :

| | | | |
|---|---|---|---|
| Vignon-Clementel, Irène | DR | INRIA | Présidente |
| Løvstakken, Lasse | PR | NTNU Norvège | Rapporteur |
| Kachenoura, Nadjia | DR | INSERM | Rapporteure |
| Sermesant, Maxime | DR | INRIA | Examinateur |
| Villemain, Olivier | PH | CHU Bordeaux | Examinateur |
| Garcia, Damien | DR | INSERM | Directeur de thèse |
| Bernard, Olivier | PR | INSA de Lyon | Co-directeur de thèse |
| Courand, Pierre-Yves | PU-PH | Hospices Civils de Lyon | Co-directeur de thèse |

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Département de la Formation par la Recherche et des Études Doctorales (**FEDORA**)

Bâtiment Charlotte Perriand 3e étage
17, av. des Arts
69621 Villeurbanne Cédex
fedora@insa-lyon.fr

Référence : TH1150_LING Hang Jung

L'INSA Lyon a mis en place une procédure de contrôle systématique via un outil de détection de similitudes (logiciel Compilatio). Après le dépôt du manuscrit de thèse, celui-ci est analysé par l'outil. Pour tout taux de similarité supérieur à 10%, le manuscrit est vérifié par l'équipe de FEDORA. Il s'agit notamment d'exclure les auto-citations, à condition qu'elles soient correctement référencées avec citation expresse dans le manuscrit.

Par ce document, il est attesté que ce manuscrit, dans la forme communiquée par la personne doctorante à l'INSA Lyon, satisfait aux exigences de l'Etablissement concernant le taux maximal de similitude admissible.

**INSA LYON**
**Campus LyonTech La Doua**
20, avenue Albert Einstein - 69621 Villeurbanne cedex - France
Tél. +33 [0]4 72 43 83 83 - Fax +33 [0]4 72 43 85 00
**www.insa-lyon.fr**

membre de UNIVERSITÉ DE LYON

# Département FEDORA – INSA Lyon - Ecoles Doctorales

| SIGLE | ECOLE DOCTORALE | NOM ET COORDONNEES DU RESPONSABLE |
|---|---|---|
| ED 206<br>CHIMIE | **CHIMIE DE LYON**<br>https://www.edchimie-lyon.fr<br>Sec. : Renée EL MELHEM<br>Bât. Blaise PASCAL, 3e étage<br>secretariat@edchimie-lyon.fr | **M. Stéphane DANIELE**<br>C2P2-CPE LYON-UMR 5265<br>Bâtiment F308, BP 2077<br>43 Boulevard du 11 novembre 1918<br>69616 Villeurbanne<br>directeur@edchimie-lyon.fr |
| ED 341<br>E2M2 | **ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION**<br>http://e2m2.universite-lyon.fr<br>Sec. : Bénédicte LANZA<br>Bât. Atrium, UCB Lyon 1<br>Tél : 04.72.44.83.62<br>secretariat.e2m2@univ-lyon1.fr | **Mme Sandrine CHARLES**<br>Université Claude Bernard Lyon 1<br>UFR Biosciences<br>Bâtiment Mendel<br>43, boulevard du 11 Novembre 1918<br>69622 Villeurbanne CEDEX<br>e2m2.codir@listes.univ-lyon1.fr |
| ED 205<br>EDISS | **INTERDISCIPLINAIRE SCIENCES-SANTÉ**<br>http://ediss.universite-lyon.fr<br>Sec. : Bénédicte LANZA<br>Bât. Atrium, UCB Lyon 1<br>Tél : 04.72.44.83.62<br>secretariat.ediss@univ-lyon1.fr | **Mme Sylvie RICARD-BLUM**<br>Laboratoire ICBMS - UMR 5246 CNRS - Université Lyon 1<br>Bâtiment Raulin - 2ème étage Nord<br>43 Boulevard du 11 novembre 1918<br>69622 Villeurbanne Cedex<br>Tél : +33(0)4 72 44 82 32<br>sylvie.ricard-blum@univ-lyon1.fr |
| ED 34<br>EDML | **MATÉRIAUX DE LYON**<br>http://ed34.universite-lyon.fr<br>Sec. : Yann DE ORDENANA<br>Tél : 04.72.18.62.44<br>yann.de-ordenana@ec-lyon.fr | **M. Stéphane BENAYOUN**<br>Ecole Centrale de Lyon<br>Laboratoire LTDS<br>36 avenue Guy de Collongue<br>69134 Ecully CEDEX<br>Tél : 04.72.18.64.37<br>stephane.benayoun@ec-lyon.fr |
| ED 160<br>EEA | **ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE**<br>https://edeea.universite-lyon.fr<br>Sec. : Philomène TRECOURT<br>Bâtiment Direction INSA Lyon<br>Tél : 04.72.43.71.70<br>secretariat.edeea@insa-lyon.fr | **M. Philippe DELACHARTRE**<br>INSA LYON<br>Laboratoire CREATIS<br>Bâtiment Blaise Pascal, 7 avenue Jean Capelle<br>69621 Villeurbanne CEDEX<br>Tél : 04.72.43.88.63<br>philippe.delachartre@insa-lyon.fr |
| ED 512<br>INFOMATHS | **INFORMATIQUE ET MATHÉMATIQUES**<br>http://edinfomaths.universite-lyon.fr<br>Sec. : Renée EL MELHEM<br>Bât. Blaise PASCAL, 3e étage<br>Tél : 04.72.43.80.46<br>infomaths@univ-lyon1.fr | **M. Hamamache KHEDDOUCI**<br>Université Claude Bernard Lyon 1<br>Bât. Nautibus<br>43, Boulevard du 11 novembre 1918<br>69 622 Villeurbanne Cedex France<br>Tél : 04.72.44.83.69<br>direction.infomaths@listes.univ-lyon1.fr |
| ED 162<br>MEGA | **MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE**<br>http://edmega.universite-lyon.fr<br>Sec. : Philomène TRECOURT<br>Tél : 04.72.43.71.70<br>Bâtiment Direction INSA Lyon<br>mega@insa-lyon.fr | **M. Etienne PARIZET**<br>INSA Lyon<br>Laboratoire LVA<br>Bâtiment St. Exupéry<br>25 bis av. Jean Capelle<br>69621 Villeurbanne CEDEX<br>etienne.parizet@insa-lyon.fr |
| ED 483<br>ScSo | **ScSo**[1]<br>https://edsciencessociales.universite-lyon.fr<br>Sec. : Mélina FAVETON<br>Tél : 04.78.69.77.79<br>melina.faveton@univ-lyon2.fr | **M. Bruno MILLY** (INSA : J.Y. TOUSSAINT)<br>Univ. Lyon 2 Campus Berges du Rhône<br>18, quai Claude Bernard<br>69365 LYON CEDEX 07<br>Bureau BEL 319<br>bruno.milly@univ-lyon2.fr |

[1] ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Damien Garcia, Olivier Bernard, and Pierre-Yves Courand, for their invaluable guidance, patience, and constant support throughout my PhD journey. Their encouragement and unwavering belief in my potential have been instrumental to the completion of this thesis. I am immensely grateful for the freedom they allowed me to explore and for the rigorous standards they set, which have continuously motivated me to strive for excellence.

Thank you, Damien, for introducing me to the fascinating world of ultrasound and vector flow imaging. Your humor and light-hearted jokes made our early morning 8 a.m. meetings much more engaging and lively (sometimes a little too much :)). Olivier, as a wonderfully attentive and caring mentor, thank you for always being there for me—nearly 24/7—and even answering my messages on Sundays. Your rich knowledge of AI and medical imaging guided me in making crucial decisions for my deep learning methods. Pierre-Yves, thank you for your willingness to acquire clinical data for my projects and for joining my supervision board in the final year of my thesis. I am excited to work with all of you again during my post-doctoral research, and I truly believe our future projects will be extraordinary. Thank you for your confidence in inviting me to join this next adventure, and special thanks to Damien for providing the funding to make it possible.

I would also like to extend my heartfelt thanks to our collaborators from IMAG Montpellier, Franck Nicoud, Simon Mendez, and Salomé Bru, for providing us with their exceptional simulations. I am grateful for the warm welcome you gave me during my visits to Montpellier and for the constructive feedback you offered on my work. My sincere thanks to Salomé for her kind advice and guidance in visualizing and analyzing the simulated data.

I am deeply appreciative of my colleagues and friends in the lab, who have been a constant source of inspiration and joy—especially those from Office No. 19. Nathan, my go-to reference whenever I faced difficulties, thank you for all the valuable ideas, insightful discussions, and our collaborations, which have been truly phenomenal. Marta, I will never forget our chats and the occasional complaints about our supervisor when work became overwhelming. Reviewing your work and slides has expanded my knowledge of MRI, and I'm grateful for that. ;) To Shusong, now a post-doctoral researcher in Reims, thank you for your

care and, of course, for the delicious hotpots. Celia, my beta tester and the new “chouchou” of Olivier’s team—thank you for patiently working with my codes and congratulations on placing third in the Myosaiq challenge. Robin, your endless energy brought life to our office, and your high-standard tutorials at the Myriad journal club were top-notch. Pierre, thank you for the joy and the lively political and philosophical discussions. Maylis, thank you for sharing interesting, useful, and often humorous finds from social media. Pierre-Elliott, I appreciate your thoughtful questions, which always prompted a deeper understanding of our work. Morgane, thank you for your friendly presence in the office. To the newcomers of Office No. 19, thank you for bringing fresh energy and perspectives to our space. I look forward to seeing all the great work you will contribute and to sharing more memorable moments together.

My gratitude also extends to those outside of Office No. 19 who have supported me along the way. To Julia, thank you for your assistance with my simulations and for our wonderful discussions and collaborations on Damien’s projects. Émile, I am grateful for our friendship that began back in TDSI of 5GE. Sophie, thank you for your cheerfulness and the beautiful photos you captured on so many occasions. Raoul, I appreciate your kindness and benevolence in the lab. Matthis, your humor and expressive reactions have always brightened the day. Jean-Baptiste, I will always cherish our shared moments at conferences and our good-natured teasing of one another. Mohamed, you are incredibly hardworking—I still remember your dedication, working late into the night during our conference stay in Venice. Célestine, thank you for the great times we had at the IUS conference. And to Tatiana and Lorena, I have enjoyed our pleasant lunchtime conversations together.

In addition to the support from my colleagues and friends, I would like to acknowledge the Malaysian Public Service Department (JPA) for sponsoring my six-year degree and master’s studies in France. Without their generous support, I would not have had the opportunity to pursue my PhD here at CREATIS.

Lastly, I am deeply indebted to my family, whose steadfast support, understanding, and encouragement sustained me through this journey. My deepest appreciation goes to my beloved parents for their unconditional love and sacrifices, which has made me who I am today. I pray for your good health always. To my elder brother, my role model since childhood, thank you for all the help you have given me, both in my studies and in daily life. To my younger sister, soon to be a doctor, thank you for the joy and care you’ve always brought into my life; I wish you all the best as you begin your housemanship. And to my elder sister, I know you would be proud of your little brother’s achievements. Your smiling face is my antidote in life and my motivation to overcome challenges—I will never forget your words!

To all who have contributed to this thesis in ways both great and small, thank you!

# Abstract

Cardiovascular diseases are the leading causes of death globally, accounting for more than 20 million lives annually, according to the World Health Organization. The heart, as the central organ in the cardiovascular system, is responsible for pumping blood to the body's cells and tissues. Assessing cardiac health is crucial for the early prevention of cardiovascular diseases. For this purpose, echocardiography is the preferred imaging modality in clinical settings due to its bedside portability and affordability. Echocardiographic examinations typically evaluate the heart's efficiency during filling (diastolic function) and ejection (systolic function). While systolic function is measured by parameters like ejection fraction and global longitudinal strain, diastolic function relies on metrics related to mitral valve and annulus velocities, which can sometimes lead to inconsistent diagnoses. Therefore, alternative biomarkers, such as intracardiac blood flow patterns and vortex characteristics, may offer more accurate diastolic function quantification.

Color Doppler echocardiography provides a scalar blood flow field in the heart, representing the projection of actual blood velocities along ultrasound scanlines. Intraventricular vector flow mapping (*i*VFM) is a recent technique that reconstructs vector blood flow from these scalar fields, enhancing intracardiac blood flow quantification. However, this method requires time-consuming preprocessing steps, including segmenting the left ventricular cavity and correcting aliasing artifacts in the color Doppler data.

This thesis leverages deep learning (DL) to automate the preprocessing steps required for vector flow mapping. First, we trained 3D DL models for left ventricular segmentation on echocardiographic sequences, ensuring temporal consistencies by considering the temporal dimension as the third spatial dimension. We then proposed DL-based methods to resolve aliasing artifacts in color Doppler data, involving segmentation and deep unfolding approaches.

Finally, we performed *i*VFM through physics-informed (PINNs) and physics-guided neural networks. For PINNs, we used gradient descent and neural networks to solve the *i*VFM optimization problem, utilizing various techniques for loss balancing and enhancing computational efficiency. For the physics-guided method, we created a physics-constrained dataset to train a neural network for reconstructing intracardiac vector blood flow from color Doppler data in a supervised manner, incorporating a physical regularization term in the

training loss functions to address the dataset size limitation. Neural network-based methods demonstrated performance comparable to the original least-squares-based *i*VFM approach, with the physics-guided method showing independence from explicit boundary conditions. Our results indicated the potential application of PINNs in ultrafast color Doppler imaging and the use of fluid dynamics equations to improve reconstruction accuracy with minimal adjustments to the loss functions.

The automation of the entire *i*VFM pipeline using neural networks, from segmentation to vector flow reconstruction, enhances the reliability of *i*VFM as a tool. The next step would be its application in clinical settings to explore and extract new flow-based biomarkers.

***Keywords*—**Cardiac flow, color Doppler, dealiasing, deep learning (DL), echocardiography, physics-guided neural networks (PGNNs), physics-informed neural networks (PINNs), segmentation, ultrasound, vector flow imaging

# Table of contents

# List of figures

# List of tables

# Nomenclature

### Greek Symbols

| | |
|---|---|
| $\nabla\times$ | Curl |
| $\Delta f$ | Doppler frequency shift |
| $\Delta\varphi$ | Doppler phase shift |
| $f_0$ | Center frequency |
| $f_{\mathrm{RX}}$ | Receiving frequency |
| $f_{\mathrm{TX}}$ | Transmitting frequency |
| $\psi$ | Stream function |
| $u_{\mathrm{D}}$ | Doppler velocity |
| $v_{\mathrm{D}}$ | Sign-inverted Doppler velocity |
| $v_{\mathrm{N}}$ | Nyquist velocity |

### Acronyms / Abbreviations

| | |
|---|---|
| *i*VFM-v1 | Intraventricular vector flow mapping proposed by Garcia et al. |
| *i*VFM-v2 | Intraventricular vector flow mapping proposed by Assi et al. |
| *i*VFM-v3 | Intraventricular vector flow mapping proposed by Vixège et al. |
| A2C | Apical two-chamber |
| A3C | Apical three-chamber |
| A4C | Apical four-chamber |

| | |
|---|---|
| ACDC | Automated Cardiac Diagnosis Challenge |
| AI | Artificial intelligence |
| AUROC | Area under the receiver operating characteristic curve |
| B-mode | Brightness mode |
| BST | Blood speckle tracking |
| CASTOR | Cardiac segmentation with constraints method proposed by Painchaud et al. |
| CDI | Color Doppler Imaging |
| *i*VFM | Intraventricular vector flow mapping |
| CE-US | Contrast-enhanced ultrasound |
| CFD | Computational fluid dynamics |
| CMRI | Cardiac Magnetic Resonance Imaging |
| CNN(s) | Convolutional neural network(s) |
| CV | Computer vision |
| DAS | Delay-and-sum |
| PRF | Pulse repetition frequency |
| DL | Deep learning |
| ECG | Electrocardiogram |
| Echo-PIV | Echocardiographic particle image velocimetry |
| ED | End-diastole/-diastolic |
| EDV | End-diastolic volume |
| EF | Ejection fraction |
| ES | End-systole/-systolic |
| ESV | End-systolic volume |
| fps | Frame per second |

| | |
|---|---|
| GLS | Global Longitudinal Strain |
| GRU(s) | Gated Recurrent Unit(s) |
| I/Q | In-phase and quadrature |
| LLM(s) | Large language model(s) |
| LSTM | Long short-term memory |
| LVEF | Left ventricle ejection fraction |
| LA | Left atrium |
| LV | Left ventricle |
| RA | Right atrium |
| RV | Right ventricle |
| $ML_d$ | Myocardial length at end-diastole |
| MLP(s) | Multi-layer perception(s) |
| $ML_s$ | Myocardial length at end-systole |
| NLP | Natural language processing |
| NN(s) | Artificial neural network(s) |
| nRMSE | Normalized root-mean-square error |
| PGNN(s) | Physics-guided neural network(s) |
| PINN(s) | Physics-informed neural network(s) |
| PISA | Proximal Isovelocity Surface Area |
| PLAX | Parasternal long-axis |
| PW Doppler | Pulsed-wave Doppler |
| ROI(s) | Region(s) of interest |
| SAM | Segment Anything Model |
| SOTA | State-of-the-art |

| | |
|---|---|
| SV | Sample volume |
| TDI | Tissue Doppler Imaging |
| VAE | Variational autoencoder |
| ViT | Vision transformer |

# Part I

# Résumé en Français
# (French Summary)

L'étude présentée dans ce manuscrit porte sur "Écoulement intraventriculaire en échocardiographie Doppler avec réseaux de neurones fondés sur la physique". Cette première partie vise à fournir un résumé en français des points essentiels de chaque chapitre du manuscrit complet, rédigé en anglais.

Le résumé est organisé comme suit :

- un sommaire de la thèse qui synthétise la problématique et les contributions de cette recherche, ainsi qu'une introduction à la problématique clinique de la thèse et aux axes de recherche ciblés pour résoudre ce problème ;
- un résumé de chaque chapitre, incluant les contributions, dans lequel un aperçu du travail réalisé et une conclusion partielle sont fournis ;
- une conclusion avec les perspectives pour les futurs travaux.

# Sommaire

Les maladies cardiovasculaires sont les principales causes de décès dans le monde, causant plus de 20 millions de décès chaque année. L'évaluation de la santé cardiaque est cruciale pour prévenir ces maladies. Pour cela, l'échocardiographie est couramment utilisée en routine clinique à cause de sa portabilité et de son coût abordable. Les examens échocardiographiques évaluent la fonction systolique et diastolique du cœur, mais les mesures de la fonction diastolique peuvent parfois donner des résultats de diagnostic discordants. Explorer des biomarqueurs alternatifs, comme le flux sanguin intracardiaque, pourrait améliorer la précision de la quantification de la fonction diastolique.

La cartographie du flux vectoriel intraventriculaire ou "intraventricular vector flow mapping" (*i*VFM) est une technique qui reconstruit le flux sanguin vectoriel à partir des champs scalaires fournis par l'échocardiographie Doppler couleur, mais elle nécessite des étapes de prétraitement chronophages. Dans cette thèse, nous avons utilisé l'apprentissage profond (DL) pour automatiser ces étapes, y compris la segmentation du ventricule gauche et la correction des artefacts de repliement de phase ou l'aliasing. Nous avons également développé des méthodes basées sur les réseaux de neurones fondés sur la physique pour reconstruire l'écoulement vectoriel intraventriculaire, montrant que ces approches peuvent améliorer l'efficacité et la précision de l'*i*VFM.

L'automatisation complète du pipeline d'*i*VFM à l'aide de réseaux de neurones, de la segmentation à la reconstruction du flux vectoriel, améliore la fiabilité de l'*i*VFM. La prochaine étape serait d'appliquer cet outil en milieu clinique pour explorer et extraire de nouveaux biomarqueurs basés sur le flux, ce qui pourrait bénéficier à la détection précoce des maladies cardiovasculaires.

# Introduction

Le système cardiovasculaire se compose d'un réseau complexe d'organes et de vaisseaux. Au cœur de ce système se trouve le cœur, qui pompe le sang et les nutriments vers chaque cellule et tissu du corps. Il est essentiel de connaître le cycle cardiaque, qui se divise en deux phases : la systole (contraction et éjection du sang) et la diastole (relaxation et remplissage).

Cette thèse se concentre principalement sur le ventricule gauche, qui pompe le sang vers le reste du corps. Quantifier l'efficacité des fonctions systolique et diastolique du ventricule gauche est primordial pour les cliniciens afin d'évaluer avec précision la santé cardiaque. Des mesures telles que la fraction d'éjection et la déformation longitudinale globale du myocarde offrent des évaluations quantitatives de la fonction systolique, reflétant la capacité du cœur à pomper le sang efficacement lors de la contraction. La fonction diastolique, quant à elle, est évaluée par des paramètres tels que les vitesses du flux mitral et de l'anneau mitral, fournissant des informations précieuses sur la capacité du cœur à se détendre et à se remplir adéquatement entre les battements. Cependant, ces paramètres peuvent conduire à des conclusions diagnostiques variées, incitant à explorer des biomarqueurs alternatifs tels que les schémas de flux intracardiaque, la vorticité ou les gradients de pression pour des diagnostics potentiellement plus fiables en pratique clinique.

En diagnostic cardiaque, l'échocardiographie émerge comme la modalité d'imagerie préférée en clinique en raison de son faible coût, de ses capacités en temps réel, de sa portabilité et de son caractère non invasif. Le mode d'imagerie Doppler couleur, en particulier, est utilisé comme outil qualitatif pour visualiser le flux sanguin dans le cœur et les vaisseaux. Une limitation majeure de l'imagerie Doppler couleur est qu'elle ne fournit que des informations scalaires, qui sont la projection des vitesses réelles du flux sanguin le long des lignes de balayage ultrasonores.

Une technique récente de cartographie du flux vectoriel intraventriculaire ou "intraventricular vector flow mapping" (*i*VFM) permet de reconstruire le flux sanguin vectoriel à partir des champs scalaires donné par l'échocardiographie Doppler couleur, facilitant ainsi la quantification du flux sanguin intracardiaque. Cependant, cette technique de reconstruction du flux vectoriel intraventriculaire nécessite plusieurs étapes de prétraitement semi-automatiques et chronophages, telles que la segmentation du ventricule gauche pour extraire les conditions aux limites et la suppression des artefacts de repliement de phase sur les images Doppler couleur, et elle manque également de flexibilité. C'est là que réside la motivation de la présente étude—exploiter la puissance de l'apprentissage profond pour automatiser entièrement les étapes de prétraitement (voir Chapitres 3 et 4) et utiliser des réseaux de neurones basés sur la physique pour améliorer la reconstruction et la caractérisation du flux vectoriel intraventriculaire par échocardiographie Doppler couleur (cf. Chapitre 5).

# Contexte clinique

## Système cardiovasculaire

Le cœur humain, situé légèrement à gauche du centre de la poitrine, est le muscle le plus sollicité du corps, battant environ 100 000 fois et pompant jusqu'à 7 500 litres de sang chaque jour. Maintenir une bonne santé cardiaque est essentiel pour garantir une bonne qualité de vie. Malheureusement, selon l'Organisation mondiale de la santé, les maladies cardiovasculaires sont la principale cause de décès dans le monde, représentant un tiers des décès en 2019 [1].

Le système cardiovasculaire, comprenant le cœur et les vaisseaux sanguins, distribue le sang oxygéné et les nutriments à travers le corps. Le cœur se compose de quatre chambres et fonctionne par des cycles de contraction et de relaxation, assurant une circulation efficace. Deux phases critiques du cycle cardiaque, le remplissage rapide et la diastase, durant lesquelles les vortex intracardiaques se forment, sont particulièrement étudiées dans cette thèse.

Pour mieux détecter et diagnostiquer ces maladies, plusieurs examens, principalement réalisés par échocardiographie, ont été intégrés dans les routines cliniques. Cependant, les métriques actuelles peuvent ne pas capturer les changements subtils dans la dynamique du flux sanguin intracardiaque, typiques des premiers stades des maladies cardiovasculaires. Cette thèse vise donc à estimer les flux sanguins vectoriels dans le ventricule gauche pour dériver de nouveaux biomarqueurs basés sur le flux, dans l'espoir d'une détection précoce de ces maladies.

## Imagerie ultrasonore : échocardiographie

L'échocardiographie, une modalité d'imagerie par ultrasons, est largement utilisée en pratique clinique pour examiner le cœur grâce à ses capacités en temps réel, sa portabilité au chevet du patient et son coût abordable. Bien que l'échocardiographie offre des résultats d'imagerie immédiats, l'imagerie par résonance magnétique cardiaque fournit une qualité d'image supérieure mais nécessite des temps d'acquisition plus longs, généralement d'environ trente minutes.

L'imagerie par ultrasons repose sur la transmission et la réception d'ondes sonores de haute fréquence pour générer des images en temps réel des structures internes du corps. Les ondes acoustiques utilisées dépassent la limite supérieure de l'audition humaine et les fréquences médicales courantes vont de 2 à 18 MHz, les fréquences plus élevées offrant une meilleure résolution spatiale mais une pénétration tissulaire réduite. Les sondes à ultrasons, composées d'éléments piézoélectriques, d'un matériau de soutien, d'une couche d'adaptation

acoustique et d'une lentille acoustique, jouent un rôle crucial dans l'émission et la réception des ondes ultrasonores, optimisant ainsi la qualité des images obtenues.

Différents types de sondes à ultrasons sont utilisés en fonction des applications cliniques spécifiques. Les sondes linéaires, convexes et à réseau phasé, chacune avec ses propres caractéristiques, permettent d'imager des structures superficielles, des zones anatomiques plus grandes ou des espaces restreints comme entre les côtes. En échocardiographie transthoracique, la sonde est placée sur la paroi thoracique pour obtenir diverses vues du cœur, telles que les vues apicales, parasternales et sous-costales. Ces acquisitions fournissent des métriques cliniques essentielles pour évaluer le fonctionnement des différentes structures cardiaques, en se concentrant particulièrement sur les acquisitions apicales et la vue trois-chambres du ventricule gauche, sujet central de cette thèse.

# État de l'art

Dans le chapitre précédent, nous avons présenté le contexte clinique de cette thèse et souligné la nécessité d'explorer de nouveaux biomarqueurs basés sur le flux pour une meilleure évaluation de la fonction diastolique, en raison des limites des métriques cliniques existantes. Pour atteindre cet objectif, il est essentiel d'analyser le flux sanguin vectoriel intracardiaque dans le ventricule gauche (LV). Par conséquent, nous commençons ce chapitre avec un aperçu des différentes techniques d'imagerie du flux vectoriel intraventriculaire, en couvrant leurs principes et leurs limitations.

L'objectif ultime de cette thèse est d'automatiser entièrement le pipeline de cartographie du flux vectoriel intraventriculaire (*i*VFM) en utilisant des réseaux de neurones. Cela inclut à la fois les étapes de prétraitement et de reconstruction du flux, de la segmentation du LV à la correction de l'aliasing Doppler couleur, jusqu'à la reconstruction du flux sanguin vectoriel intracardiaque. Par conséquent, nous fournissons un aperçu des architectures de réseaux de neurones populaires.

## Estimation de champs de vitesse intraventriculaire

Les techniques traditionnelles d'imagerie par ultrasons utilisées en pratique clinique ne permettent pas de reconstruire le champ de vitesse vectoriel au sein du LV. Diverses études ont proposé des solutions à cette limitation, y compris la vélocimétrie par image de particules échocardiographiques (echo-PIV), le suivi de speckles sanguins (BST) et *i*VFM. L'echo-PIV, par exemple, utilise des microbulles remplies de gaz comme agents de contraste injectés dans le sang humain [2]. Ces microbulles peuvent voyager librement à travers les capillaires pulmonaires et systémiques. Cependant, une limitation majeure de l'echo-PIV est la nécessité d'injections de microbulles, ce qui peut être chronophage et nécessiter du personnel supplémentaire, rendant cette méthode peu pratique pour une utilisation clinique de routine.

Le BST est une technique qui suit directement les speckles sanguins en utilisant des séquences d'ultrasons ultrarapides, éliminant ainsi le besoin d'un agent de contraste [3, 4]. Cette méthode peut être combinée avec l'imagerie Doppler couleur (CDI) pour améliorer l'efficacité computationnelle et réduire la variance. Cependant, en raison de la fréquence d'acquisition plus élevée par rapport à l'échocardiographie conventionnelle, la profondeur d'imagerie est réduite, limitant ainsi l'application du BST principalement à la cardiologie pédiatrique. En outre, l'*i*VFM vise à reconstruire le flux sanguin vectoriel à partir des images Doppler couleur. Plusieurs méthodes [5–8] ont été développées pour atteindre cet objectif, chacune avec ses propres hypothèses et techniques de décomposition du flux, allant

de la séparation du flux en composants de base et de vortex, à l'utilisation de l'équation de continuité 2D pour assurer la conservation de la masse dans le plan. Parmi celles-ci, la méthode de Vixège et al. [8], appelée *i*VFM-v3, résout un problème d'optimisation sous contrainte de l'équation de conservation de la masse et de la condition de glissement à la paroi (myocarde) avec les méthodes de multiplicateurs de Lagrange, réduisant le nombre de paramètres déterminés de manière supervisée à un seul. Nous nous sommes basés sur l'*i*VFM-v3 par la suite pour développer nos propres méthodes en utilisant les réseaux de neurones fondés sur la physique. Comme toutes les méthodes d'*i*VFM, l'*i*VFM-v3 nécessite des méthodes de prétraitement chronophages, manuelles ou semi-automatiques, telles que la segmentation du LV et le déaliasing des images Doppler couleur.

## Architecture des réseaux de neurones

Étant donné l'abondance de travaux dans le domaine de l'intelligence artificielle, nous nous concentrons sur certaines architectures de réseaux de neurones courantes (MLPs, CNNs, Segment Anything Model, approches de deep unfolding, et réseaux de neurones informés/guidés par la physique). Plusieurs architectures de réseaux ont été utilisées dans cette thèse pour atteindre l'objectif final de reconstruire le flux sanguin vectoriel intracardiaque, ce qui nécessite une segmentation préalable du LV et le déaliasing du Doppler couleur. nnU-Net [9], expliqué plus tard, a été largement étudié dans Chapter 3 pour segmenter le LV dans les séquences échocardiographiques en mode B avec une cohérence temporelle. Pour la tâche de déaliasing, la performance de nnU-Net, des méthodes basées sur les transformers et le deep unfolding ont été comparées dans Chapter 4. Enfin, des réseaux de neurones informés par la physique ont été développés dans Chapter 5 pour résoudre le même problème d'optimisation d'*i*VFM-v3 pour la reconstruction du flux vectoriel intracardiaque. De plus, nnU-Net, initialement conçu pour la segmentation, a été adapté et entraîné pour réaliser *i*VFM de manière supervisée.

Un perceptron multicouche (MLP) est un réseau de neurones pleinement connecté composé de plusieurs couches : une couche d'entrée, une ou plusieurs couches cachées, et une couche de sortie. Chaque neurone reçoit des informations des neurones connectés et les traite avant de les transmettre à d'autres neurones. Les poids, mis à jour pendant l'apprentissage, déterminent l'importance de chaque information. Les informations pondérées sont ensuite passées à travers des fonctions d'activation non linéaires. Les MLPs sont souvent utilisés dans les réseaux de neurones informés par la physique pour réaliser des tâches complexes.

Les réseaux de neurones convolutifs (CNNs) sont particulièrement bien adaptés à l'analyse d'images car ils apprennent automatiquement des hiérarchies spatiales de caractéristiques. Les CNNs utilisent des couches de convolution qui appliquent des filtres à l'image d'entrée,

capturant ainsi des caractéristiques de bas niveau dans les couches peu profondes et des caractéristiques de haut niveau dans les couches plus profondes. L'utilisation de poids partagés et de connexions locales réduit considérablement le nombre de paramètres par rapport aux réseaux entièrement connectés. Cela rend les CNNs plus efficaces sur le plan computationnel et améliore leur capacité à généraliser à partir de données d'entraînement limitées. En basant sur U-Net, le framework nnU-Net [9] détermine automatiquement les stratégies d'apprentissage et architectures de U-Net optimales en fonction de base de données d'entraînement.

Le Segment Anything Model (SAM) [10], proposé en 2023, s'inspire des grands modèles de langage en traitement automatique de langage naturel, qui montrent de fortes capacités de généralisation zéro-shot et few-shot. Bien que non implémenté ou testé dans cette thèse, SAM peut annoter des images à grande échelle pour améliorer les modèles de segmentation. SAM peut segmenter tout dans une image naturelle ou répondre à des invites comme des clics interactifs et des boîtes englobantes. En cas d'invites ambiguës, il génère plusieurs masques valides avec des scores de confiance associés. SAM comprend un encodeur d'image, un encodeur d'invite flexible, et un décodeur de masque rapide.

Le deep unfolding est une approche qui relie les algorithmes itératifs traditionnels aux techniques modernes d'apprentissage profond. Introduit en 2010, le deep unfolding transforme ces processus itératifs en réseaux de neurones entraînables [11]. Le deep unfolding mappe chaque itération d'un algorithme traditionnel à une couche dans un réseau neuronal. Ce processus affine itérativement la solution, similaire aux méthodes traditionnelles, mais les paramètres sont appris à partir des données. Dans notre recherche, nous avons utilisé un algorithme de deep unfolding basé sur le primal-dual appris pour le déaliasing.

Les réseaux de neurones informés par la physique (PINNs) [12] et les réseaux de neurones guidés par la physique (PGNNs) [13, 14] intègrent des connaissances physiques dans les réseaux de neurones. Les PINNs utilisent des MLPs pour résoudre des équations différentielles partielles décrivant divers phénomènes physiques. En intégrant des connaissances physiques dans l'optimisation, les PINNs assurent que les solutions respectent les lois physiques. Les PGNNs, en revanche, utilisent des données contraintes par la physique pour entraîner les réseaux de neurones. Cette approche repose sur des simulations physiques ou des expériences réelles pour générer des données d'entraînement conformes aux lois physiques.

# Contribution 1 : Segmentation du ventricule gauche en échocardiographie avec la cohérence temporelle

## Sommaire

Comme mentionné précédemment, l'*i*VFM est actuellement la seule méthode applicable en clinique pour reconstruire le flux sanguin vectoriel intracardiaque. Une étape cruciale de ce processus est la définition des conditions aux limites, nécessitant la segmentation de la cavité du LV tout au long du cycle cardiaque. Une meilleure délimitation de la région d'intérêt peut conduire à des conditions aux limites plus précises et, par conséquent, à des reconstructions du flux vectoriel plus exactes. La plupart des méthodes actuelles reposent sur une segmentation manuelle, souvent longue et inconsistente. L'objectif de notre contribution est de développer un outil entièrement automatisé pour la segmentation cohérente dans le temps (2D+t) des séquences mode B cardiaque en vue apicale trois-chambres (A3C). Pour ce faire, nous avons dû annoter notre propre base de données A3C, car il n'y a pas de base de données publique A3C disponible et les modèles entraînés sur des données en vues apicales deux (A2C) et quatre-chambres (A4C) ne fonctionnaient pas sur des données A3C en raison de la différence de forme du ventricule gauche.

Nous avons mené une étude de faisabilité de notre pipeline d'annotation de données sur les données A2C et A4C, où les données publiques sont disponibles. Pour ce faire, nous avons utilisé CASTOR [15] et une base de données échocardiographique public, CAMUS [16]. Avec l'architecture et les stratégies d'apprentissage optimales déterminées d'après notre étude, nous avons effectué la segmentation image par image en 2D sur une base de données privée non annotées, appelée CARDINAL. Les erreurs temporelles ont été corrigées automatiquement à l'aide de CASTOR et vérifiées par un expert humain. Les annotations post-traitées ont servi de référence pour entraîner notre algorithme de segmentation 2D+t, basé sur nnU-Net 3D [9]. Ce modèle prend une séquence d'images en mode B consécutives en entrée et considère la dimension temporelle comme la troisième dimension spatiale. Notre méthode a produit une segmentation précise et cohérente dans le temps, atteignant des erreurs inférieures à la variabilité intra-observateur et des résultats comparables à la plupart des méthodes spécialisées d'état de l'art lorsqu'elle a été testée sur la base de données CAMUS. Ce pipeline d'annotation a été adapté pour annoter nos données A3C interne non labellisées en utilisant l'apprentissage actif. Le nnU-Net 3D appris sur ces données a montré une performance suffisante pour extraire avec précision les conditions aux limites requises pour *i*VFM.

## Commentaire

Le travail présenté dans ce chapitre a été publié dans les actes de la conférence Functional Imaging and Modeling of the Heart (FIMH) en 2023. L'algorithme de segmentation 2D impliqué dans le pipeline d'annotation des données a été présenté lors du IEEE International Ultrasonics Symposium (IUS) en 2022. Les références complètes des deux articles sont fournies ci-dessous :

H. J. Ling, D. Garcia, et O. Bernard, "Reaching intra-observer variability in 2-D echocardiographic image segmentation with a simple U-Net architecture," dans IEEE International Ultrasonics Symposium (IUS), 2022 [17].

H. J. Ling, N. Painchaud, P.-Y. Courand, P.-M. Jodoin, D. Garcia, et O. Bernard, "Extraction of Volumetric Indices from Echocardiography: Which Deep Learning Solution for Clinical Use?," dans Functional Imaging and Modeling of the Heart (FIMH), 2023, pp. 245–254 [18].

## Contribution des articles

- Proposition d'un pipeline d'annotation pour annoter une base de données appelée CARDINAL, comprenant environ 200 séquences cardiaques A4C et A2C à cycle complet ;
- Proposition d'un nnU-Net 3D, entraîné à partir des données annotées de façon semi-automatique, permettant de produire des segmentations cohérentes dans le temps afin d'extraire de manière fiable les indices cliniques de la fonction cardiaque ;
- Adaptation du pipeline en intégrant l'apprentissage actif pour annoter nos données A3C ;
- Proposition d'un autre nnU-Net 3D, entraîné sur nos données A3C limitées en quantité, et utilisation de l'apprentissage par transfert depuis CARDINAL pour améliorer les performances de segmentation et extraire correctement les conditions aux limites.

## Conclusion

Ce chapitre a détaillé notre approche pour générer des pseudo-références pour le grand nombre de données non annotées CARDINAL et pour obtenir une segmentation cohérente dans le temps des séquences échocardiographiques A2C et A4C à l'aide de méthodes de segmentation 2D+t. Plus précisément, le nnU-Net 3D, qui considère la dimension temporelle

comme la troisième dimension spatiale, a pu produire des segmentations cohérentes dans le temps tout en étant cliniquement précises. Cette approche élimine le besoin d'algorithmes de post-traitement tels que CASTOR pour corriger les artefacts temporels, offrant deux avantages significatifs : i) étant donné que l'entrée de CASTOR consiste uniquement en des segmentations prédites, les erreurs temporelles et anatomiques peuvent être difficiles à corriger si les prédictions sont trop dégénérées ; ii) l'entraînement de CASTOR nécessite un réglage minutieux des hyperparamètres pour éviter la sur-régularisation. De plus, le nnU-Net 3D a montré une capacité remarquable à maintenir une qualité de segmentation élevée à travers différents ensembles de données.

En nous appuyant sur ces résultats, nous avons établi une base de données A3C privée pour relever le défi de la segmentation des vues A3C, qui ne sont pas couramment disponibles dans les bases de données publiques. En utilisant une stratégie d'apprentissage actif, nous avons considérablement réduit la charge de travail liée à l'annotation manuelle, permettant la création de notre base de données A3C robuste et entièrement annotée. L'apprentissage par transfert depuis CARDINAL a également été adopté pour améliorer les performances de notre modèle. Cela a permis de former un nnU-Net 3D avec des données A3C limitées, qui présentent une forme de ventricule gauche relativement différente de celle des données A2C et A4C, afin d'extraire les conditions aux limites nécessaires pour *i*VFM. La prochaine étape est de corriger les artefacts de repliement de phase (aliasing) sur les images Doppler couleur A3C pour obtenir des champs de vitesse Doppler corrects, permettant ainsi la reconstruction du flux sanguin vectoriel.

# Contribution 2 : Correction de l'aliasing en échocardiographie Doppler couleur avec apprentissage profond

## Sommaire

Le chapitre précédent a abordé la première étape essentielle de l'*i*VFM: segmenter automatiquement la paroi endocardique du LV pour extraire les conditions aux limites. La prochaine étape avant le processus d'*i*VFM est de corriger l'aliasing. En échographie Doppler couleur, les artefacts d'aliasing surviennent lorsque les vitesses sanguines dépassent la vitesse de Nyquist. Corriger cet aliasing est crucial pour les méthodes quantitatives basées sur le Doppler couleur, en particulier pour notre cas de l'*i*VFM, qui utilise les vitesses Doppler comme terme d'attache aux données.

Les méthodes traditionnelles de correction de l'aliasing sont peu fiables lorsqu'elles sont appliquées aux données Doppler couleur A3C, nécessitant l'exploration de solutions d'apprentissage supervisé par DL. Contrairement à la segmentation des structures cardiaques, qui bénéficie de jeux de données publics avec annotations, il n'existe pas de jeu de données Doppler couleur public avec à la fois des données aliasées et des labels sans aliasing. Pour y remédier, nous avons constitué notre propre jeu de données Doppler couleur A3C pour entraîner et évaluer les méthodes basées sur DL. Au lieu de prédire directement la vitesse sans aliasing, nous avons simplifié la tâche d'apprentissage en segmentant les pixels aliasés puis en les corrigeant par la suite. Plusieurs techniques de DL ont été explorées à cet effet, y compris les CNNs, les transformers et une approche de deep unfolding.

Malgré un nombre de paramètres à apprendre nettement inférieur, notre approche de deep unfolding basée sur le primal-dual a obtenu des performances similaires à celles des autres méthodes de DL. Nos résultats suggèrent que les méthodes basées sur DL peuvent efficacement éliminer les artefacts d'aliasing dans les images échocardiographiques Doppler couleur, surpassant DeAN, une technique semi-automatique de l'état de l'art.

## Commentaire

Le travail décrit dans ce chapitre a été publié dans les IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. Sa référence est la suivante :

H. J. Ling, O. Bernard, N. Ducros, et D. Garcia, "Phase Unwrapping of Color Doppler Echocardiography using Deep Learning," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 70, no. 8, pp. 810–820, août 2023 [19].

## Contribution de l'article

- Proposition d'un réseau primal-dual basé sur le deep unfolding, et comparaison avec les méthodes de segmentation DL de l'état de l'art et DeAN (voir Fig. 1 pour le pipeline général) ;
- Création d'un jeu de données privé d'échocardiographie Doppler couleur A3C (45 patients, 1338 images aliasées et 2379 images non aliasées) pour entraîner les réseaux de neurones et analyser leurs performances ;
- Investigation de la valeur ajoutée de l'inclusion de la puissance Doppler comme information d'entrée pour améliorer la déaliasing ;
- Introduction d'une stratégie d'augmentation des données générant des aliasing synthétiques, ce qui a résolu le problème du déséquilibre des classes et amélioré les performances de déaliasing sur les cas difficiles.

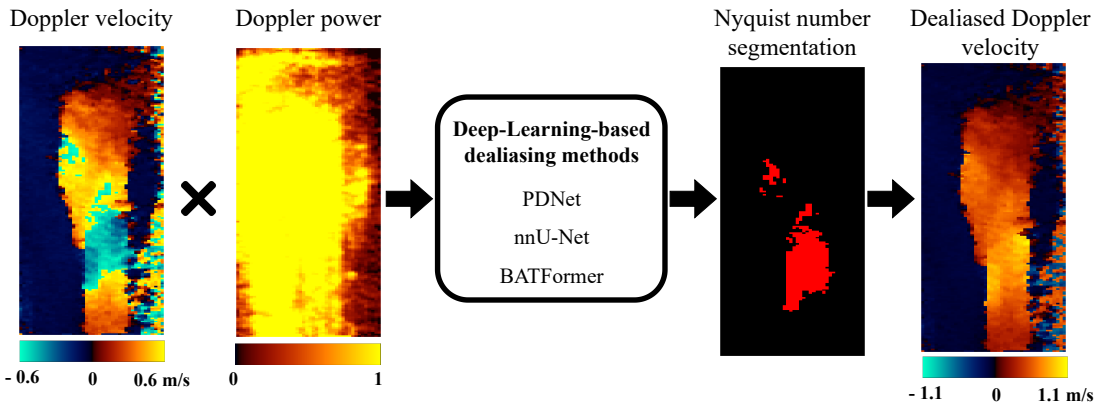



Fig. 1 Pipeline de déaliasing en échocardiographie Doppler couleur avec apprentissage profond. PDNet : approche du deep unfolding basée sur primal-dual ; BATFormer : méthode basée sur transformer

## Conclusion

Dans ce chapitre, nous avons introduit trois méthodes de DL pour le déaliasing des images d'échocardiographie Doppler couleur, chacune exploitant des approches distinctes pour résoudre le défi des artefacts d'aliasing. Nos méthodologies comprenaient un réseau d'optimisation primal-dual déroulé (PDNet), un réseau de segmentation à l'état de l'art (nnU-Net), et un réseau de segmentation basé sur un transformer (BATFormer).

Nous avons formulé le problème de déaliasing comme un problème inverse non linéaire et avons adapté OriPDNet pour le résoudre efficacement, aboutissant au développement de

PDNet. De plus, le déaliasing peut être vu comme un problème de segmentation multi-classe pour segmenter les pixels aliasés. Nous avons comparé les performances de PDNet avec nnU-Net et BATFormer, deux réseaux de segmentation puissants dans les tâches de segmentation d'images médicales.

Nos résultats expérimentaux ont démontré que les techniques de DL peuvent déaliaser efficacement les images d'échocardiographie Doppler couleur. Nos méthodes DL proposées ont surpassé la méthode non-DL DeAN, avec nnU-Net obtenant les meilleures performances, suivi de PDNet. De plus, l'incorporation des informations de puissance et l'augmentation artificielle de l'aliasing ont encore amélioré les résultats. L'application des techniques de DL à l'échocardiographie Doppler couleur est une approche prometteuse qui pourrait renforcer l'utilité clinique de cette modalité d'imagerie largement utilisée. Avec ces outils de segmentation et de déaliasing basés sur DL à notre disposition, nous sommes maintenant en mesure de réaliser *i*VFM en utilisant des réseaux de neurones aidés par la physique sur des données d'échocardiographie Doppler couleur A3C. Cela constitue la contribution finale de cette thèse, visant à développer un outil clinique entièrement automatique pour pouvoir extraire de nouveaux biomarqueurs basés sur le flux.

# Contribution 3 : Cartographie du flux vectoriel intraventriculaire avec réseaux de neurones guidés par la physique

## Sommaire

Les deux chapitres précédents ont décrit nos contributions à l'automatisation complète des deux étapes de prétraitement manuelles ou semi-automatiques pour l'*i*VFM. Tant pour la segmentation du ventricule gauche que pour le déaliasing du Doppler couleur, nnU-Net se distingue comme l'algorithme le plus performant. Ces deux outils basés sur l'apprentissage profond améliorent significativement la chaîne de traitement de *i*VFM-v3 [8], rendant l'ensemble du processus de reconstruction du flux sanguin vectoriel entièrement automatique.

Étant donné les performances des réseaux de neurones dans les tâches précédentes, nous avons proposé d'utiliser des approches "data-driven" pour réaliser l'*i*VFM. Dans ce chapitre, nous présentons notre contribution finale, en introduisant des alternatives novatrices au schéma d'optimisation traditionnel de l'*i*VFM en utilisant les PINNs et une approche supervisée basée sur nnU-Net guidée par la physique. Dans les PINNs, l'optimisation data-driven est utilisée pour résoudre le même problème d'optimisation contraint que dans *i*VFM-v3, tandis que dans l'apprentissage supervisé guidé par la physique, une base de données d'entraînement contrainte par la physique établissant une correspondance entre le champ de vitesse Doppler et le flux vectoriel intracardiaque est utilisé pour entraîner notre modèle de régression DL avec un terme de régularisation physique supplémentaire.

Lors de l'évaluation sur des images Doppler couleur simulées dérivées d'un modèle de dynamique des fluides computationnelle spécifique au patient et des acquisitions Doppler in vivo, les deux approches démontrent des performances de reconstruction comparables à l'algorithme original *i*VFM. L'efficacité des PINNs est renforcée grâce à une optimisation en deux étapes et à des poids pré-optimisés. D'autre part, la méthode nnU-Net excelle en termes de généralisation et de capacités en temps réel. Notamment, nnU-Net montre une robustesse supérieure sur les données Doppler éparses et tronquées tout en maintenant une indépendance vis-à-vis des conditions aux limites explicites.

## Commentaire

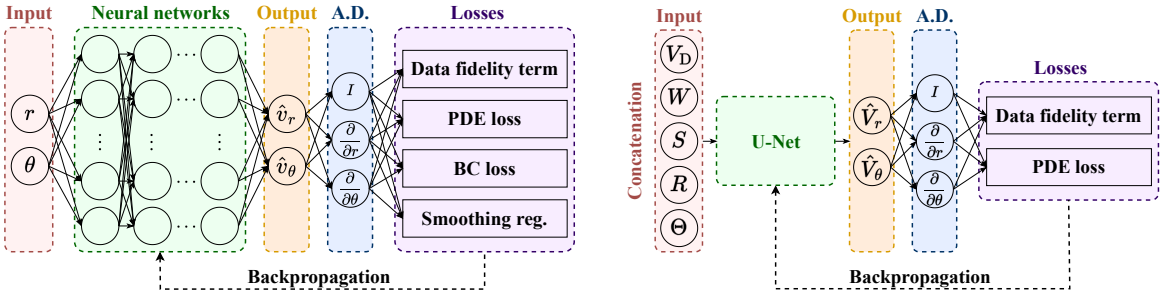
Le travail décrit dans ce chapitre a été publié dans IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control et est actuellement en accès anticipé. Sa référence est donnée ci-dessous :

H. J. Ling, S. Bru, J. Puig, F. Vixège, S. Mendez, F. Nicoud, P.-Y. Courand, O. Bernard,

and D. Garcia, “Physics-Guided Neural Networks for Intraventricular Vector Flow Mapping,” IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2024, doi: 10.1109/TUFFC.2024.3411718 [20].

## Contribution de l’article

- Développement d’une approche supervisée guidée par la physique basée sur nnU-Net, montrant une grande robustesse sur des données éparses et tronquées avec un temps d’inférence quasi en temps réel (voir Fig. 2b pour l’architecture) ;
- Mise en œuvre de deux variantes de PINNs basées sur la méthode de pénalité pour effectuer la cartographie du flux vectoriel, atteignant des performances comparables à l’algorithme original *i*VFM (voir Fig. 2a pour l’architecture) ;
- Utilisation de l’optimisation en deux étapes et des poids pré-optimisés à partir d’une image Doppler sélectionnée, ce qui a amélioré les performances des PINNs et réduit leur temps d’optimisation jusqu’à 3,5 fois.




(a) PINNs (b) nnU-Net guidé par la physique

Fig. 2 Architectures de PINNs et nnU-Net guidé par la physique. A.D. signifie la dérivation automatique. Dans (a), l’entrée 2D de PINNs est de taille $(B \times 2)$, où $B$ est la taille de batch. $r$ et $\theta$ représentent les coordonnées radiales and angulaires. Dans (b), nnU-Net prend en entrée un tenseur 4D de taille $(B \times 5 \times 192 \times 40)$, qui est la concaténation de la vitesse Doppler inversée $V_{\mathrm{D}}$, la matrice de poids $W$, la segmentation du ventricule gauche $S$, la matrice de coordonnées radiales $R$ et la matrice de coordonnées angulaires $\Theta$.

## Conclusion

Dans ce chapitre de contribution finale, nous avons proposé de nouvelles approches basées sur les réseaux de neurones pour la cartographie du flux vectoriel intraventriculaire, en utilisant l’optimisation basée sur les gradients à travers les PINNs et l’apprentissage supervisé guidé

par la physique (nnU-Net). Ces méthodes offrent des stratégies contrastées pour aborder le problème inverse mal posé de la reconstruction du flux vectoriel.

Pour l'approche d'optimisation utilisant les PINNs, nous avons décomposé le problème d'optimisation sous contraintes en une série de fonctions de perte sans contrainte. Nous avons examiné deux stratégies de pondération différentes, ReLoBRaLo et Lagrangien augmenté, pour équilibrer chaque fonction de perte, les deux produisant des résultats de reconstruction similaires. Étant donné que les méthodes basées sur les PINNs convergent lentement, nous avons utilisé des poids pré-optimisés et une stratégie d'optimisation en deux étapes pour améliorer leur efficacité.

Le nnU-Net guidé par la physique a été entraîné en utilisant un jeu de données contraint par la physique. Malgré la taille limitée de nos données d'entraînement, l'inclusion d'un terme de régularisation physique, l'équation de conservation de la masse, a significativement amélioré sa performance de reconstruction. nnU-Net a également montré une capacité quasi en temps réel, une robustesse avec des données Doppler éparses et une indépendance par rapport aux conditions aux limites explicites. Ces caractéristiques positionnent nnU-Net comme une solution prometteuse pour les applications cliniques en temps réel.

Dans les parties suivantes, nous discutons des limitations des études menées dans cette thèse et proposons des perspectives sur les améliorations potentielles, notamment pour *i*VFM utilisant les PINNs, ainsi que sur les orientations futures de la recherche.

# Conclusion

Dans la pratique clinique, l'analyse des images médicales est essentielle pour le diagnostic des pathologies. En particulier, l'échocardiographie est largement utilisée pour les pathologies cardiovasculaires en raison de sa portabilité et de son imagerie en temps réel. La segmentation des structures cardiaques a été considérablement améliorée par DL, augmentant ainsi la précision des indices cliniques, tels que la fraction d'éjection pour évaluer la fonction systolique.

Cependant, l'évaluation de la fonction diastolique, souvent basée sur les ratios E/A et E/e', peut donner des résultats diagnostiques discordants et manquer de sensibilité pour détecter les changements subtils du flux sanguin aux premiers stades des maladies cardiaques. L'analyse du flux sanguin intracardiaque, notamment les vortex, peut aider à détecter précocement ces maladies.

La méthode *i*VFM est une approche de post-traitement qui peut être appliquée directement aux acquisitions Doppler couleur cliniques sans équipements spécialisés. Toutefois, les techniques *i*VFM nécessitent souvent des étapes de prétraitement semi-automatiques ou manuelles, telles que la segmentation du LV et le déaliasing des images Doppler couleur, ce qui limite leur viabilité clinique. Automatiser ces étapes est crucial pour rendre *i*VFM un outil fiable pour explorer de nouveaux biomarqueurs basés sur le flux, potentiellement indicateurs précoces de maladies cardiovasculaires.

Cette thèse a d'abord développé un outil de segmentation basé sur nnU-Net 3D, produisant une segmentation cohérente dans le temps du LV en utilisant des convolutions 3D et le transfert d'apprentissage. La deuxième étape a consisté à automatiser le déaliasing des images Doppler couleur, en utilisant une approche de deep unfolding, une approche basée sur les transformers et nnU-Net, ce dernier montrant une meilleure performance.

Ensuite, des approches basées sur les PINNs et le nnU-Net guidé par la physique ont été explorées pour améliorer *i*VFM-v3. Les PINNs ont été utilisés pour résoudre des problèmes d'optimisation sous contraintes avec des résultats comparables à *i*VFM-v3, tandis que le nnU-Net guidé par la physique, avec une régularisation physique et une stratégie intelligente d'augmentation des données, a montré des avantages en termes de précision et de rapidité d'inférence.

Pour conclure, cette thèse représente une avancée dans le développement d'un outil clinique entièrement automatisé et rapide utilisant l'IA pour la détection précoce des maladies cardiovasculaires. Bien que les méthodes d'IA aient montré de bonnes performances, des améliorations sont encore possibles, notamment pour la segmentation du LV en A3C et la cartographie du flux vectoriel. Les perspectives pour de futurs travaux seront abordées dans la partie suivante.

# Perspectives

## Modèles de fondation pour l'annotation de données échocardiographiques à grande échelle

Notre pipeline d'annotation de données, nécessitant peu d'intervention humaine, a démontré son efficacité pour créer des annotations de référence pour des ensembles de données de petite à moyenne taille. Pour l'avenir, il est crucial de développer des modèles de fondation adaptés à la segmentation échocardiographique, tels que ceux basés sur SAM. Le développement de SAMUS [21] et MemSAM [22] représente une avancée majeure, bien que MemSAM n'atteigne pas encore la variabilité intra-observateur requise pour les métriques cliniques. Les cardiologues pourraient utiliser ces modèles semi-automatiques et faciles à prendre en main pour créer un consensus d'annotation. Cela peut réduire la variabilité inter-observateur et produire des labels de haute qualité et cohérents. En publiant davantage de jeux de données publics avec des annotations respectant le consensus, nous pourrions affiner ces modèles ou en former de nouveaux, comme nnU-Net, améliorant ainsi leur généralisation inter-vendeurs et inter-centres. Cela permettrait de développer des modèles robustes pour toutes les vues échocardiographiques, améliorant la précision des segmentations et le diagnostic des pathologies.

## Généralisation du déaliasing du Doppler couleur

Dans notre chapitre sur le déaliasing, nous avons formé des modèles d'apprentissage profond pour corriger les aliasings simples. Cependant, certaines maladies valvulaires, comme la régurgitation mitrale, peuvent provoquer des aliasings multiples, créant des jets fortement aliasés près des valves mitrales. Ces cas posent des défis, même pour le déaliasing manuel, et leur inclusion dans les ensembles d'entraînement est essentielle pour améliorer la généralisation des modèles. Une solution potentielle est l'utilisation de simulations, en utilisant une séquence multi-PRF pour obtenir des labels sans aliasing [23], tout en acquérant les données aliasées avec des séquences Doppler standard.

## Robustesse de la cartographie deu flux intraventriculaire

Notre étude a utilisé une contrainte physique simple, l'équation de conservation de la masse, en raison des limitations de fréquence d'images des acquisitions Doppler couleur conventionnelles. Pour utiliser des équations de dynamique des fluides, comme celles de Navier-Stokes, il est nécessaire d'augmenter la fréquence d'images avec des séquences

ultrarapides, telles que les ondes planes. Bien que ces séquences puissent produire des données plus bruitées, elles pourraient être lissées à l'aide des équations de Navier-Stokes, qui incluent davantage de termes de régularisation. Une autre approche consiste à combiner des images Doppler de plusieurs cycles cardiaques par interpolation de Fourier. Une fréquence d'images plus élevée permettrait l'intégration des équations de dynamique des fluides, et avec un terme d'attache aux données insensible à $2\pi$, nous pourrions éliminer l'étape de déaliasing [24].

## Validation clinique de la cartographie du flux intraventriculaire

La validation clinique de la cartographie du flux intraventriculaire est essentielle pour démontrer son utilité et sa pertinence clinique. Une caractérisation approfondie de la population est nécessaire pour évaluer la faisabilité de la détection des maladies cardiovasculaires à partir de biomarqueurs de flux. Les méthodes basées sur les réseaux de neurones développées peuvent être combinées pour créer un outil entièrement automatisé pour un usage clinique. Une étude préliminaire a montré que les modèles basés sur les CNNs 3D, entraînés sur des données limitées, peuvent distinguer les cas hypertensifs des cas sains à partir des séquences Doppler couleur avec une F1-score de 0.6. Une approche multimodale, combinant Doppler couleur et champ de vitesse vectorielle reconstruit, pourrait améliorer la classification des pathologies.

# Part II

# Background

# Introduction

The cardiovascular system consists of a complex network of organs and vessels. At the core of this system lies the heart, pumping blood and nutrients to every cell and tissue within the body. Understanding the cyclic operation of the heart is central to comprehending its function, which is characterized by two fundamental phases: systole and diastole. While systole denotes the contraction phase, expelling blood into the systemic and pulmonary circulations, diastole marks the period of relaxation and ventricular filling, crucial for maintaining adequate cardiac output and perfusion.

This thesis focuses primarily on the left ventricle, which pumps blood to the rest of the body. Quantifying the efficiency of both systolic and diastolic functions of the left ventricle is paramount for clinicians to accurately assess cardiac health. Measures such as ejection fraction and myocardial global longitudinal strain offer quantitative assessments of systolic function, reflecting the heart's ability to pump blood efficiently during contraction. Diastolic function, on the other hand, is evaluated through parameters like those involving velocities of the mitral flow and mitral annulus, which provide valuable information about the heart's ability to relax and fill adequately between beats. However, these parameters may lead to varying diagnostic conclusions, prompting the exploration of alternative biomarkers such as intracardiac flow patterns, vorticity, or pressure gradients for potentially more reliable diagnostics in clinical practice.

In cardiac diagnostics, echocardiography emerges as the preferred imaging modality in clinics due to its low cost, real-time capabilities, portability, and non-invasive nature. Color Doppler imaging mode, in particular, is used as a qualitative tool to visualize blood flow within the heart and vasculature. A major limitation of color Doppler imaging is that it only gives scalar information, which is the projection of actual blood flow velocities along the ultrasound scanlines.

Recently, a technique has been proposed to map the scalar field given by color Doppler to a vector velocity field, facilitating the quantification of intracardiac blood flow. However, this intraventricular vector flow mapping technique requires several semi-automatic, time-consuming pre-processing steps, such as segmenting the left ventricle for extracting the

boundary conditions and removing the aliasing artifacts on color Doppler images, and also lacks flexibility. Herein lies the motivation for the present study—to harness the power of deep learning to fully automate the pre-processing steps and use physics-based neural networks to improve the mapping and characterization of intraventricular vector flow by color Doppler echocardiography.

The subsequent chapters of this dissertation are structured to present key aspects of this study. Chapter 1 provides a comprehensive review of the cardiovascular system, and ultrasound imaging, explaining the mechanisms that underlie cardiac function and the principle of clinical ultrasound acquisition. Chapter 2 presents a state-of-the-art review of vector flow imaging techniques, detailing their theoretical foundations, applications, and inherent limitations. In the same chapter, we introduce deep learning methods for cardiac image analysis, such as convolutional neural networks, transformers, the Segment Anything Model, deep unfolding, and physics-based neural networks. Chapter 3 focuses on the segmentation of the left ventricle, while Chapter 4 addresses the challenge of dealiasing color Doppler images. In Chapter 5, we present the development and application of physics-informed and physics-guided neural networks for intraventricular flow mapping, showcasing the benefits of incorporating physical knowledge into neural networks. Finally, Part IV summarizes the main findings of this research and suggests future directions, such as integration of fluid dynamics equations in intraventricular vector flow mapping and its clinical validation, underscoring the potential contribution of this work in enhancing cardiovascular diagnostics for better and more personalized patient care.

# List of Publications

## Journal papers

1. **H. J. Ling**, O. Bernard, N. Ducros, and D. Garcia, "Phase Unwrapping of Color Doppler Echocardiography using Deep Learning," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 70, no. 8, pp. 810–820, Aug. 2023.

2. **H. J. Ling**, S. Bru, J. Puig, F. Vixège, S. Mendez, F. Nicoud, P.-Y. Courand, O. Bernard, and D. Garcia, "Physics-Guided Neural Networks for Intraventricular Vector Flow Mapping," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2024, doi: 10.1109/TUFFC.2024.3411718. (Early Access)

3. J. Puig, D. Friboulet, **H. J. Ling**, F. Varray, M. Mougharbel, J. Porée, J. Provost, D. Garcia, and F. Millioz, "Boosting Cardiac Color Doppler Frame Rates with Deep Learning," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2024, doi: 10.1109/TUFFC.2024.3424549. (Early Access)

## Conference papers

1. **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 (FIMH), Lyon, France, 2023, pp. 245–254. (Oral)

## Conference abstracts

1. **H. J. Ling**, D. Garcia, and O. Bernard, "Reaching intra-observer variability in 2-D echocardiographic image segmentation with a simple U-Net architecture," presented at IEEE International Ultrasonics Symposium (IUS), Venice, Italy, 2022. (Poster)

2. **H. J. Ling**, O. Bernard, and D. Garcia, "Dealiasing of color Doppler echocardiography using deep learning," presented at IEEE International Ultrasonics Symposium (IUS), Venice, Italy, 2022. (Oral)

3. **H. J. Ling**, O. Bernard, and D. Garcia, "Intraventricular Vector Flow Imaging using Physics-Informed Deep Learning," presented at IEEE International Ultrasonics Symposium (IUS), Montreal, Canada, 2023. (Oral)

4. **H. J. Ling**, O. Bernard, and D. Garcia, "Intraventricular Vector Flow Imaging using Physics-Informed Deep Learning," presented at Artimino Conference on Medical Ultrasound Technology, Artimino, Italy, 2023. (Oral)

5. **H. J. Ling**, O. Bernard, and D. Garcia, "Écoulement intraventriculaire vectoriel par apprentissage profond informé par la physique," presented at presented at Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale (IABM), Grenoble, France, 2024. (Poster)

# Chapter 1

# Clinical Context

Located in the middle compartment of the chest, slightly to the left, the human heart is the hardest working muscle in the body. It beats around 100 000 times and pumps up to 7500 liters of blood daily. Given that the heart is under constant workload, maintaining good cardiac health is essential to ensure an individual's quality of life. Unfortunately, according to the World Health Organization, cardiovascular diseases are the leading cause of death worldwide, accounting for one-third of all global deaths in 2019 [1].

To better detect and diagnose cardiovascular diseases, several examinations have been integrated into clinical routines for evaluating the efficiency of heart filling and ejection. These examinations are mostly performed using echocardiography, a non-invasive technique available in most clinics and hospitals. However, existing clinical metrics may not capture the subtle changes in intracardiac blood flow dynamics that typically occur in the early stages of cardiovascular diseases. Therefore, this thesis aims to estimate vector blood flows within the left ventricle cavity, from which we hope to derive new flow-based biomarkers for potential early disease detection.

In this chapter, we first introduce the fundamental concepts of the cardiovascular system (Section 1.1), including the heart and the cardiac cycle. Following this, Section 1.2 briefly describes the principles of ultrasound imaging (Section 1.2.1) and outlines the common imaging modes of transthoracic echocardiography, along with the associated clinical metrics to be measured (Section 1.2.2). Finally, Section 1.3 discusses the potential limitations of existing clinical metrics and the exploration of new biomarkers based on intracardiac blood flow, as well as their current clinical limitations.

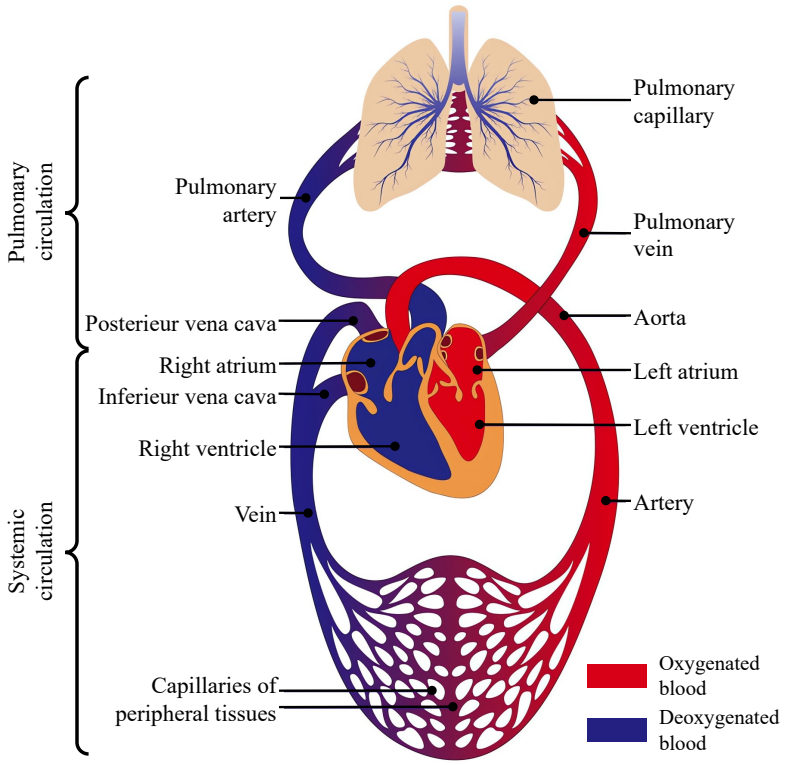



Fig. 1.1 Cardiovascular system. Adapted from: Quizlet.

## 1.1 Cardiovascular system

The cardiovascular system, including the heart and blood vessels, delivers nutrients and blood to our body. The heart's left and right parts play distinct roles in the blood circulatory system. The right heart takes deoxygenated blood to the lungs to be re-oxygenated and returned to the left heart. This process is called pulmonary circulation. The oxygenated blood is then pumped out of the left heart to the rest of our body via the arteries, and the deoxygenated blood is carried back to the right heart via the veins. This circuit loop is named systemic circulation. Fig. 1.1 shows a brief overview of the human cardiovascular system. The detailed blood circulation inside the heart is discussed in Section 1.1.1.

### 1.1.1 Heart anatomy

The human heart consists of four chambers: the left ventricle (LV), left atrium (LA), right ventricle (RV), and right atrium (RA), as shown in Fig. 1.2. The RA receives deoxygenated blood from the body via the superior and inferior vena cavae and pumps it into the RV. From there, the blood is pumped to the lungs for oxygenation. Oxygenated blood returns to the heart via the pulmonary veins, entering the LA, which then contracts to push the blood into

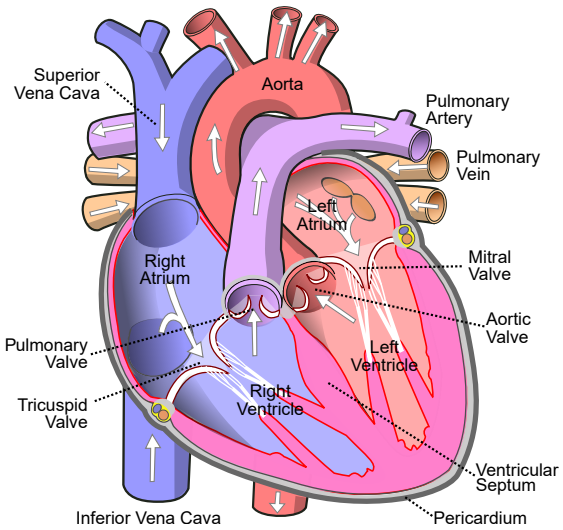



Fig. 1.2 Heart anatomy. Adapted from: Wikimedia Commons.

the LV. The LV, being the strongest chamber, pumps oxygenated blood out to the rest of the body through the aorta, the largest artery.

The heart's efficient functioning also relies on four crucial valves that ensure the unidirectional flow of blood through the heart chambers. The atrioventricular valves, namely the tricuspid valve on the right side and the mitral valve on the left side, separate the atria from the ventricles. They open during ventricular relaxation (diastole) to allow blood to flow from the atria into the ventricles and close during ventricular contraction (systole) to prevent backflow into the atria. Similarly, the semilunar valves, including the pulmonary valve and the aortic valve, guard the exits of the pulmonary trunk and the aorta, respectively. These valves open during ventricular contraction to allow blood to be ejected into the pulmonary artery and aorta, and close during ventricular relaxation to prevent blood from flowing back into the ventricles.

The heart is composed primarily of cardiac muscle tissue, known as myocardium, which contracts rhythmically to propel blood. A fibrous membrane called the septum separates the left and right sides of the heart, preventing the mixing of oxygen-rich and oxygen-poor blood.

The heartbeat is controlled by the cardiac conduction system, a network of specialized cells that generate and transmit electrical impulses. These impulses regulate the timing and coordination of cardiac contractions, ensuring the heart's rhythmic pumping action.

### 1.1.2 Cardiac cycle

The human heart beats around 60 to 80 times per minute at rest. The cardiac cycle refers to the sequence of events that occur during one complete heartbeat, including both systole and diastole phases of the heart. The heart's electrical activity through repeated cardiac cycles can be captured by electrocardiography by placing electrodes on the chest surface and limbs. The recorded activity is known as electrocardiogram (ECG). On an ECG, the P wave represents depolarization of the atria; the QRS complex represents depolarization of the ventricles; the T wave represents repolarization of the ventricles. Fig 1.3 shows a Wiggers diagram, illustrating different cardiac events with electrographic trace lines.

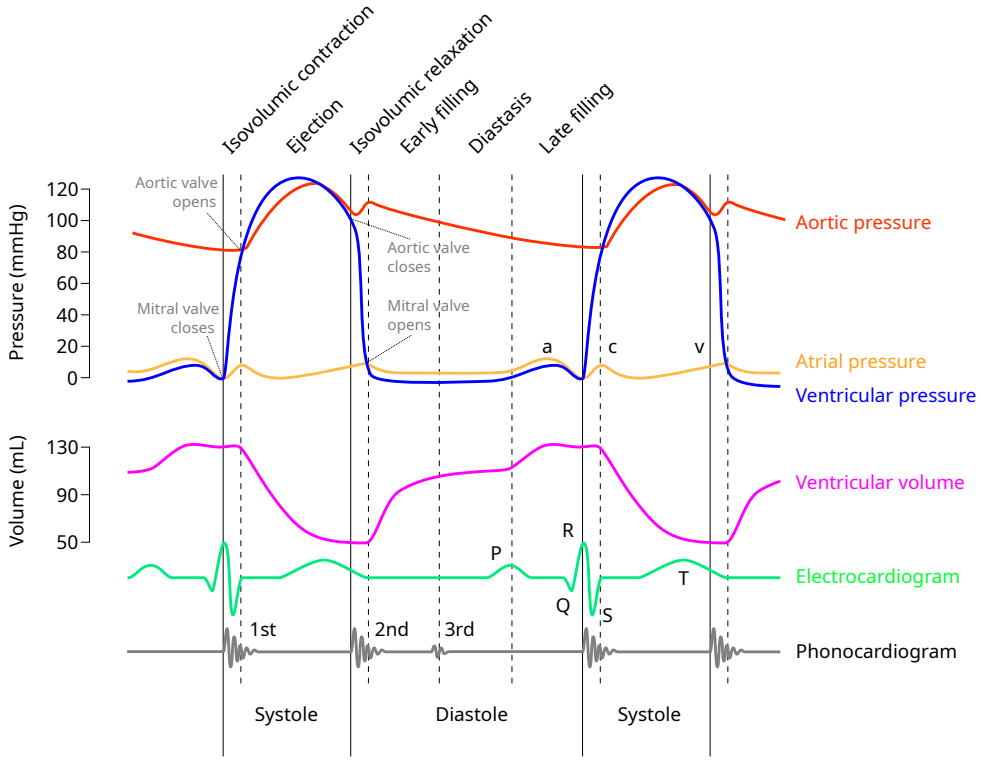



Fig. 1.3 Wiggers diagram. Adapted from: Wikimedia Commons.

The cardiac cycle can be divided into several key phases: ejection, early filling, diastasis, and late filling:

1. Ejection phase (QRS complex): The cardiac cycle begins with ventricular contraction, known as systole. During this phase, the ventricles contract forcefully, generating pressure that exceeds the pressure in the arteries. As a result, the semilunar valves open, allowing blood to be ejected from the ventricles into the pulmonary artery and

aorta, respectively. This phase represents the period of highest blood flow velocity and is crucial for maintaining adequate blood circulation throughout the body.

2. Early filling phase: Following ventricular contraction, the ventricles enter a period of relaxation, known as early diastole. During this phase, the pressure in the ventricles drops below that of the atria, causing the atrioventricular valves to open. Blood rapidly flows from the atria into the ventricles, filling them to approximately 70-80% of their capacity.

3. Diastasis phase: Following rapid ventricular filling, here comes the middle stage of the diastole known as diastasis. During diastasis, the passive filling of the heart's ventricles slows down as the pressure in the atria and ventricles reaches equilibrium.

4. Late filling phase (P wave): As the atria begin to contract (atrial systole), additional blood is pushed into the ventricles, completing ventricular filling. This late filling phase contributes the remaining 20-30% of ventricular volume. It's important to note that during late diastole, the ventricles are maximally filled with blood, preparing them for the upcoming ejection phase.

Together, these phases of the cardiac cycle ensure efficient pumping of blood throughout the body, providing oxygen and nutrients to tissues while removing metabolic waste products. The coordinated interplay between atrial and ventricular contractions, as well as relaxation, enables the heart to maintain its essential function as a continuous and effective pump.

In this thesis, we focus on two phases: early filling and diastasis. The intracardiac large vortex starts forming at the end of early filling and becomes more visible during diastasis. Details on the formation of the large vortex are given later in Section 1.3.

## 1.2 Ultrasound imaging: echocardiography

Echocardiography, an ultrasound-based imaging modality, is widely utilized in clinical practice for heart examination due to its real-time capabilities, bedside portability, and affordability. While echocardiography offers immediate imaging results, cardiac magnetic resonance imaging (CMRI) provides superior image quality. However, CMRI suffers from longer acquisition times, typically around thirty minutes.

The following subsections explain the principles of ultrasound imaging, discuss the different views of cardiac imaging, and elaborate on different imaging modes.

### 1.2.1 Principles of ultrasound imaging

Ultrasound imaging relies on the transmission and reception of high-frequency sound waves to generate real-time images of internal body structures. The acquisition theory of ultrasound imaging, involving the principles of acoustic wave propagation, transducer technology, wave transmission and reception, and image formation, is explained below.

#### 1.2.1.1 Acoustic wave properties

Ultrasound imaging uses acoustic waves with frequencies exceeding the upper limit of human hearing (typically above 20 kHz). In medical ultrasound, frequencies commonly used range from 2 to 18 MHz, with higher frequencies providing better spatial resolution but reduced tissue penetration. In cardiac imaging, the acquisition frequencies are usually between 2 and 5 MHz.

#### 1.2.1.2 Ultrasound probes and application

Ultrasound probes, also known as transducers, are composed of several essential components, including piezoelectric elements, backing material, an acoustic matching layer, and an acoustic lens, as illustrated in Fig. 1.4.

The piezoelectric elements, serve as both transmitters and receivers of ultrasound waves. Typically made of ceramics or polymers, these elements convert electrical energy into mechanical vibrations during transmission (transmit mode) and vice versa during reception (receive mode). When an electric voltage is applied to the piezoelectric material, it undergoes mechanical deformation, generating ultrasonic waves. Conversely, when ultrasound waves strike the piezoelectric material, it generates electrical signals proportional to the received acoustic energy. Fig. 1.5 provides a schematic representation of a probe with piezoelectric elements. The spacing between these elements is known as the “kerf”, while the distance between the center of two adjacent elements is referred to as the “pitch”.

The backing material is positioned behind the piezoelectric elements to prevent their excessive oscillation. This helps generate ultrasound waves with shorter pulse lengths, resulting in higher resolution in imaging and accurately determining the distance to objects in images.

To optimize the transmission of ultrasound waves to the object being imaged, an acoustic matching layer is inserted. This layer minimizes the acoustic impedance difference between the probe and the imaged object.

Additionally, the acoustic lens is used to focus ultrasound waves in the slice direction, preventing them from spreading and enhancing resolution. By focusing the ultrasound waves,

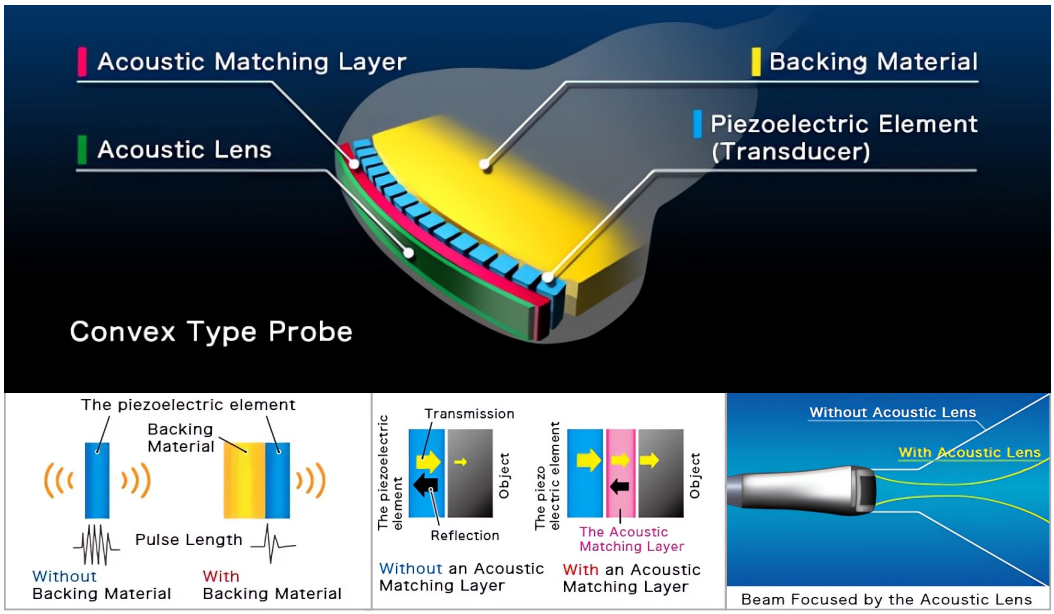



Fig. 1.4 Schematic representation of different materials inside a convex array probe. Reproduced from: NDK website.

the acoustic lens improves the sharpness and clarity of the resulting images, making them more clinically valuable for diagnostic purposes.

Different types of ultrasound probes are available, each designed for specific clinical applications. Fig. 1.6 illustrates the three most commonly used ultrasound probes in clinical settings, each with distinct characteristics:

1. Linear Array Probes (5 to 15 MHz): Linear array probes consist of approximately 64 to 256 aligned elements and emit narrow, focused beams suitable for imaging superficial structures with high resolution. Their maximal imaging depth is around 8 cm, and they are commonly used for musculoskeletal imaging, vascular studies, and superficial soft tissue examinations.

2. Convex Array Probes (2 to 5 MHz): Convex array probes are equipped with around 64 to 192 elements arranged along a curved surface. They produce wider beams suitable for imaging larger anatomical areas with greater depth penetration from 15 to 30 cm. Convex array probes are extensively used in abdominal imaging, obstetrics, gynecology, and prostate imaging.

3. Phased Array Probes (1 to 5 MHz): Phased array probes, often referred to as “cardiac probes”, are smaller compared to linear probes, with a maximal imaging depth ranging from 12 to 16 cm. They typically contain approximately 16 to 64 elements and utilize

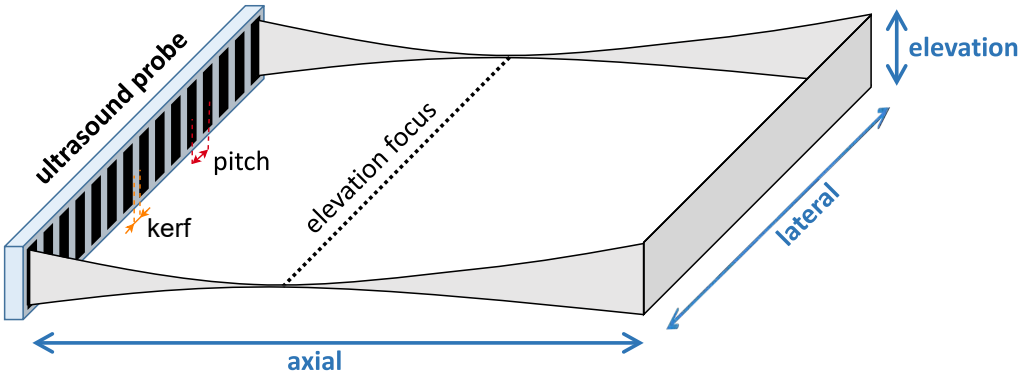



Fig. 1.5 Schematic representation of a linear array probe with the pitch and kerf parameters along with the axial, lateral, and elevation axes. Adapted from: Biomecardio.

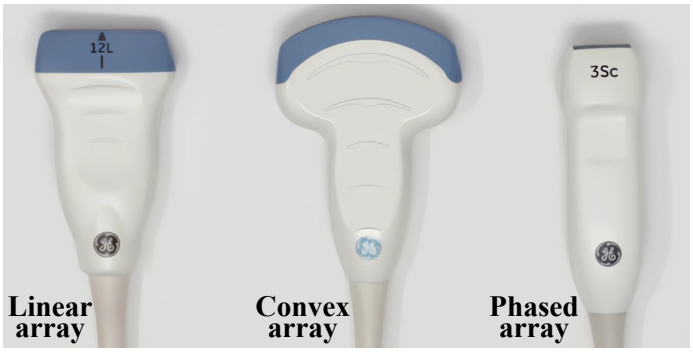



Fig. 1.6 Three most commonly used ultrasound probes. Adapted from: Pocus 101.

multiple elements to electronically steer and focus the ultrasound beam. Their compact design facilitates maneuverability in tight spaces such as between the ribs. Phased array probes find widespread use in cardiac imaging, vascular imaging, and guiding interventional procedures.

#### 1.2.1.3 Wave Transmission and Reception

Due to impedance mismatch, ultrasound waves emitted by transducers transmit poorly through the air. Therefore, during clinical ultrasound acquisitions, ultrasound gel is applied to eliminate air between the ultrasound probe and the contact point on the skin.

The ultrasound probe emits short pulses of ultrasound waves into the body tissues. Transmission delay can be individually applied to each element in the ultrasound probe to create various sequences, including focused, plane, diverging, or multi-line transmit waves. Conventional clinical ultrasound probes use focused waves for higher-quality images at the cost of having a lower frame rate than the other three more research-oriented ultrafast

sequences (plane, diverging, and multi-line transmit). Fig. 1.7 shows the application of transmission delay to create an ultrasound wave focusing at a specific focal point. Fig. 1.8 illustrates the acoustic pressure fields of the three ultrafast sequences.

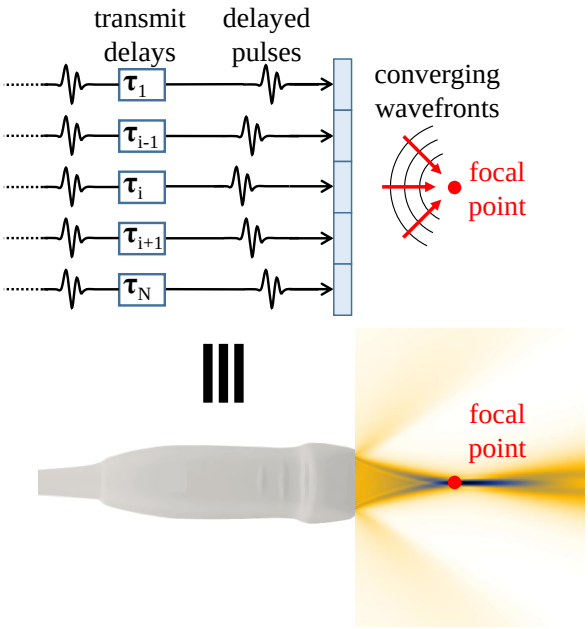



Fig. 1.7 Generation of a focused ultrasound wave (top) with the corresponding acoustic pressure field (bottom). Adapted from: Biomecardio.

The emitted waves propagate through tissues until they encounter interfaces between media of different material properties, such as density and acoustic impedance. At these interfaces, different phenomena occur depending on tissue dimensions and boundary smoothness. Specular reflection (see Fig. 1.9a) happens when a wave strikes a tissue with a smooth boundary larger than the wavelength of the incident wave. Conversely, diffuse reflection (Fig. 1.9b) takes place when a wave encounters a large tissue with a rough boundary, reflecting the wave in multiple directions. Additionally, scattering (Fig. 1.9c) occurs when ultrasound waves interact with tissues smaller than their wavelength.

Most human tissue structures are significantly smaller than the ultrasound wavelength and can be viewed as scatterers. Since the sizes of these tissues, especially red blood cells, are typically less than one-tenth of the ultrasound wavelength, Rayleigh scattering occurs. This causes the incident waves to scatter equally in all directions with reduced amplitude, resulting in multiple echoes from numerous scatterers. These scattered echoes interact, causing constructive and destructive interference, leading to “speckles” in the resultant image. These speckles appear as a granular noise texture commonly seen on ultrasound Brightness-mode (B-mode) images.

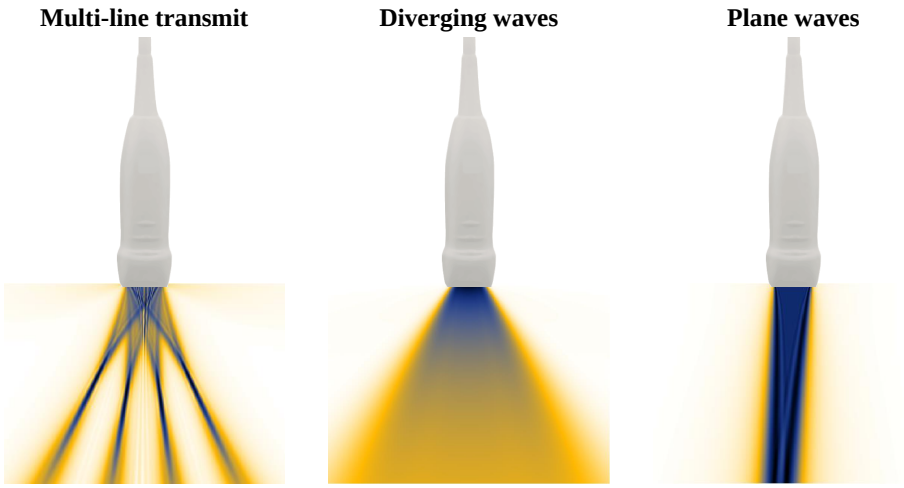
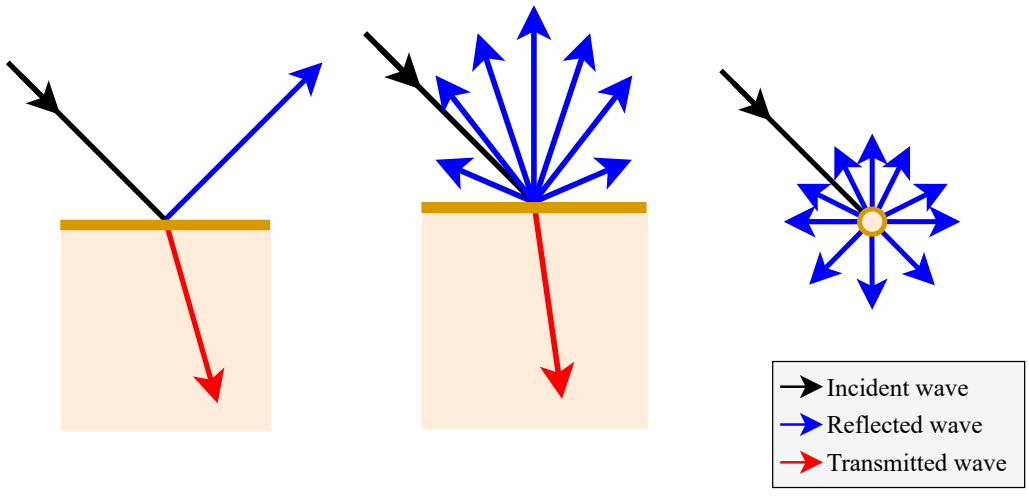



Fig. 1.8 Ultrafast ultrasound sequences. From left to right: multi-line transmit, diverging, and plane wave sequences. Adapted from: Biomecardio.




(a) Specular reflection (b) Diffuse reflection (c) Scattering

Fig. 1.9 Different phenomena occurring at tissue interfaces during ultrasound imaging.

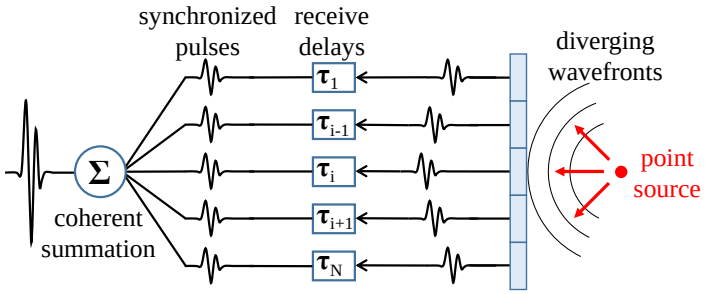



Fig. 1.10 Example of delay-and-sum during the reception of echoes produced by a point source scatterer. Reproduced from: Biomecardio.

To correctly combine the echoes received by each element, coming from the same point source, a beamforming process is necessary. Delay-and-sum (DAS) is one of the most used beamforming techniques to beamform the received signals. DAS consists of applying reception delays before summing to ensure signals from the same point source are in phase. The reception delay for each element can be obtained by taking into account the two-way travel times of the wavefront, from the transducer to a scatterer and from the scatterer to each element. Fig. 1.10 illustrates an example of DAS during the reception of echoes from a point source.

#### 1.2.1.4 Image Formation

As most conventional clinical ultrasound systems utilize focused waves, this section only discusses the image formation of such sequences. To image a given area, the focused beam is swept along the imaging sector by adjusting the transmission delays. Although the beam is focused on a focal point, the ultrasound waves insonify a large zone in the direction of the focal point, as shown in Fig. 1.7. This allows for image formation along the entire zone or scanline. The image along each scanline is constructed successively and combined before displaying the complete image on the ultrasound scanner. The detailed image formation along a scanline is explained below.

Once the radiofrequency (RF) signals are beamformed, they undergo further processing to generate an image. This processing includes time gain compensation (TGC) to overcome ultrasound attenuation, analog-to-digital conversion, and in-phase and quadrature (I/Q) demodulation. The magnitude and phase of the resulting I/Q signals are used for B-mode and color Doppler visualization, respectively. The magnitude is: 1) log-compressed to improve image visualization by reducing the dynamic range of the signals, and 2) filtered by applying edge enhancement and speckle reduction techniques.

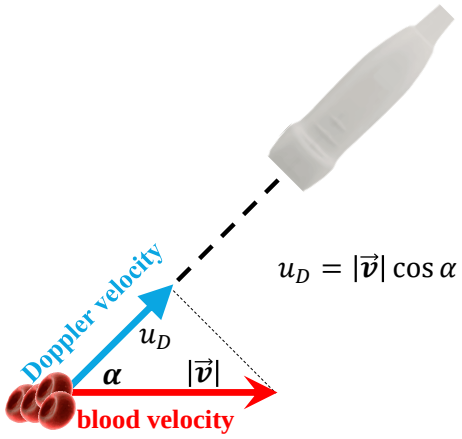



Fig. 1.11 Example of the projection of blood velocity along the ultrasound axis. Adapted from: Biomecardio.

Forming a color Doppler image requires successive firings, typically 8, of focused beams along a scanline, also known as a packet size, to compute the phase shift. The successive I/Q signals are initially clutter-filtered to retain only the blood signals before computing the lag-1 autocorrelation to estimate the Doppler velocity, $u_{\mathrm{D}}$. Doppler velocity represents the projection of the actual blood velocity along the propagation axis of the ultrasound wave (ultrasound axis). An example of this projection is illustrated in Fig. 1.11. In color Doppler imaging, the Doppler frequency shift or the difference between the transmitting ($f_{\mathrm{TX}}$) and receiving frequency ($f_{\mathrm{RX}}$), $\Delta f$, can be expressed as:

$$\Delta f = f_{\mathrm{TX}} - f_{\mathrm{RX}} = \frac{2 f_0}{c} |\vec{v}| \cos\alpha, \qquad [\mathrm{Hz}] \quad (1.1)$$

where $f_0$ is the center frequency, $c$ represents the speed of sound, and $\alpha$ denotes the angle between the target's moving direction and the ultrasound axis. Since $\alpha$ is often unknown, the velocity estimation by color Doppler is limited to Doppler velocity:

$$u_{\mathrm{D}} = |\vec{v}| \cos\alpha = \frac{c}{2} \frac{\Delta f}{f_0} = \frac{c}{2} \frac{\Delta \varphi}{2\pi f_0} \mathrm{PRF}, \qquad [\mathrm{m\,s^{-1}}] \quad (1.2)$$

with $\Delta\varphi$ being the phase shift obtained via autocorrelation, and PRF representing the pulse repetition frequency (see Fig. 1.12 for illustrative example), i.e., the number of ultrasound firing pulses generated per second. Here, the phase shift is used instead of the frequency shift because of the insufficient sampling for accurate spectral analysis.

Equation (1.2) can be rewritten as:

$$u_{\mathrm{D}} = \frac{c \times \mathrm{PRF}}{4 f_0} \frac{\Delta\varphi}{\pi} = V_{\mathrm{N}} \frac{\Delta\varphi}{\pi}, \qquad [\mathrm{m\,s^{-1}}] \quad (1.3)$$

where $V_{\mathrm{N}}$ represents the Nyquist velocity, which is the maximum Doppler velocity that can be reliably measured. This is because $\Delta\varphi$ falls within the range of $[-\pi, \pi]$, implying that $u_{\mathrm{D}}$ lies within the interval $[-V_{\mathrm{N}}, V_{\mathrm{N}}]$. Velocities with a magnitude exceeding $V_{\mathrm{N}}$ result in wrapping around in the opposite direction, causing aliasing artifacts. These artifacts can be mitigated by increasing the PRF, at the sacrifice of reducing the maximum imaging depth.

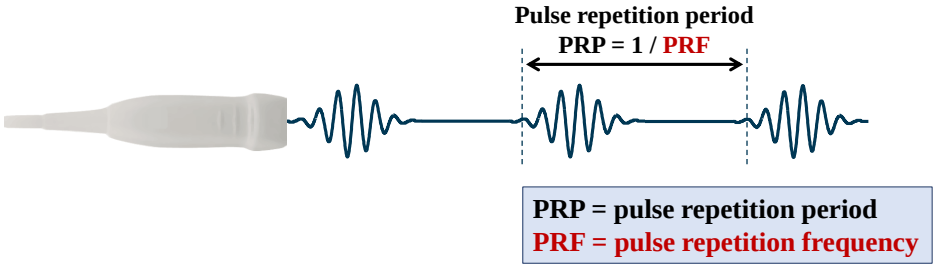



Fig. 1.12 Example of ultrasound pulses emitted at a given pulse repetition frequency. Adapted from: Biomecardio.

## 1.2.2 Transthoracic echocardiography

Transthoracic echocardiography, the standard form of echocardiography, involves placing the ultrasound probe on the chest wall (thorax). Cardiologists acquire various views of the heart depending on diagnostic requirements. These views include apical, parasternal, and subcostal views. Apical views, such as the two- (A2C), three- (A3C), and four-chamber (A4C) views, are obtained from the apex of the heart. Parasternal long-axis (PLAX) views are captured alongside the sternum, while subcostal views are taken from beneath the edge of the last rib. Fig. 1.13 illustrates examples of echocardiographic images acquired from different views.

A complete transthoracic echocardiography involves a series of acquisitions to obtain clinical metrics for assessing the functioning of different cardiac structures. This section primarily focuses on apical acquisitions and the LV, which are the main subjects of interest in this thesis, particularly the A3C view.

### 1.2.2.1 Apical B-mode

Fig. 1.14 shows examples of A2C, A3C, and A4C B-mode images, respectively. It is important to note that in apical B-mode images, the apex is located at the top. A typical

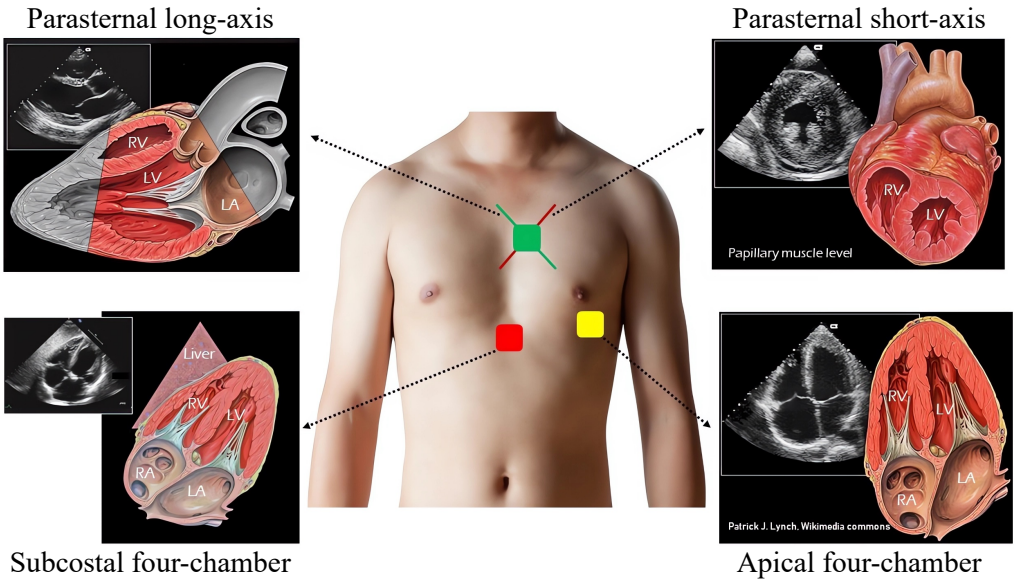



Fig. 1.13 Acquisition of various views of the heart by echocardiography. Adapted from: Renal Fellow Network.

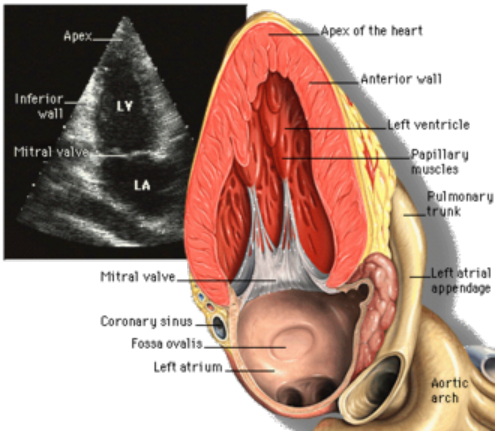
echocardiographic B-mode acquisition of 100 scanlines utilizing focused waves with an imaging depth of 15 cm generates approximately 60 frames per second (fps).

B-mode images provide valuable information for detecting cardiac diseases such as cardiomyopathy. Cardiomyopathy is a disease of the heart muscles, usually reflected by dilated, stretched, or thickened myocardium. Fig. 1.15 compares a healthy heart with one affected by amyloid cardiomyopathy, displaying a thickened myocardium and a dilated LA in the pathological case.

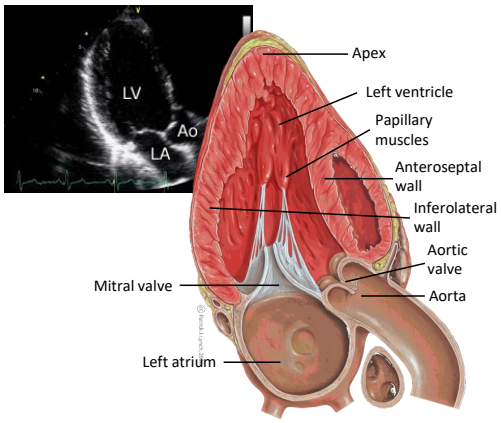
Quantitative analysis on B-mode images involves measuring parameters such as left ventricle ejection fraction (LVEF), a crucial metric for assessing systolic function. LVEF represents the percentage of blood pumped out of the LV with each heartbeat, calculated using the end-diastolic volume (EDV) and end-systolic volume (ESV) of the LV as shown in Equation (1.4). Cardiologists segment the LV endocardial wall on A2C and A4C acquisitions at end-diastole (ED) and end-systole (ES) instants (see Fig. 1.16), respectively, to derive these volumes using Simpson's biplane rule [25]. LVEF for a healthy adult heart ranges from 53 to 73% [26].

$$\text{LVEF} = \frac{\text{EDV} - \text{ESV}}{\text{EDV}} \times 100 \qquad [\%] \quad (1.4)$$

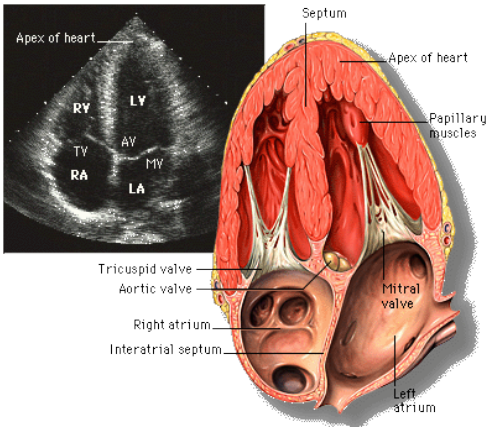
Global Longitudinal Strain (GLS) has emerged as a significant metric for evaluating myocardial function in transthoracic echocardiography. Utilizing speckle tracking techniques, GLS quantifies the percentage change in myocardial deformation along the longitudinal axis




(a) Apical two-chamber view




(b) Apical three-chamber view

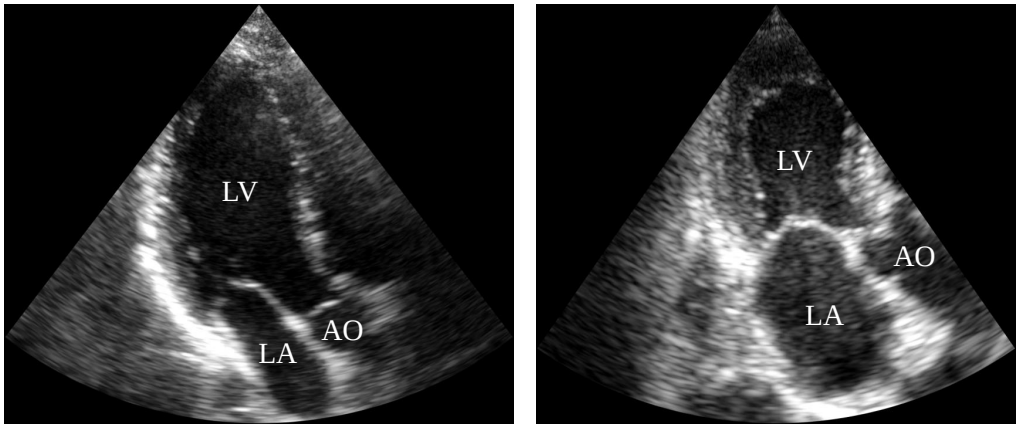



(c) Apical four-chamber view

Fig. 1.14 Apical two-, three-, and four-chamber B-mode images. (a) and (c) are reproduced from Wikimedia Commons; (b) is adapted from Wikimedia Commons.

of the LV throughout the cardiac cycle. Mathematically, GLS is defined as the difference between the myocardial length at end-diastole ($ML_d$) and end-systole ($ML_s$), divided by $ML_d$, as shown in (1.5). Since $ML_s$ is smaller than $ML_d$, peak GLS is typically a negative value. A peak GLS of approximately -20% is considered normal in healthy individuals [26]. A lower absolute value of peak GLS indicates a greater likelihood of cardiac abnormality. Fig. 1.17 showcases GLS plots for a normal subject (peak GLS = -23%) and a patient with reduced systolic function (peak GLS = -14%).

$$\mathrm{GLS} = \frac{\mathrm{ML_s} - \mathrm{ML_d}}{\mathrm{ML_d}} \times 100 \quad [\%] \quad (1.5)$$




(a) Healthy subject (b) Amyloid cardiomyopathy

Fig. 1.15 A healthy subject versus a patient with amyloid cardiomyopathy acquired in A3C view.

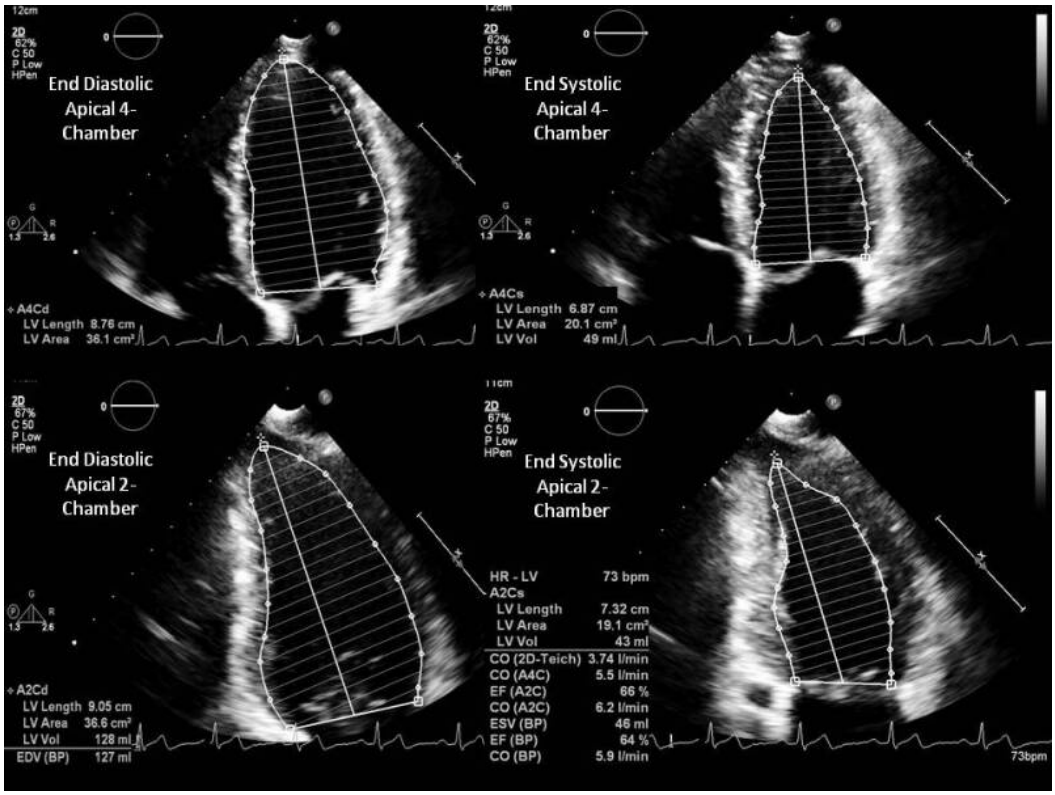



Fig. 1.16 Segmentation of LV endocardial wall on end-diastolic and end-systolic A2C and A4C view B-mode images. Reproduced from: NIH-NLM-NCBI.

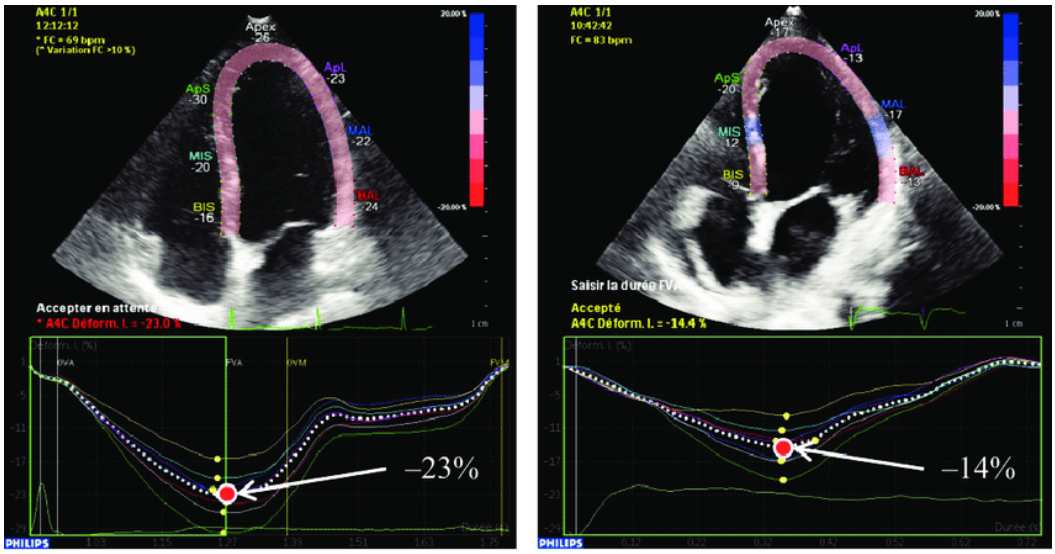



Fig. 1.17 GLS plots and peak GLS of a healthy subject (left) versus a patient with reduced systolic function (right). Reproduced from [2, Figure 26.6].

#### 1.2.2.2 Color Doppler

Color Doppler imaging (CDI) provides valuable information about blood flow dynamics. In these images, red and blue hues represent blood moving towards and away from the ultrasound probe, respectively, as illustrated in Fig. 1.18. The intensity of these colors correlates with the blood flow velocity, with brighter shades indicating higher Doppler velocities. The superimposition of color Doppler images onto B-mode images, known as duplex images, facilitates the simultaneous visualization of anatomical structures and blood flow patterns. In conventional echocardiographic color Doppler imaging, the frame rate is typically around 15 fps, which is significantly lower than B-mode acquisition, primarily due to the necessity for multiple firings per scanline.

CDI is widely used in clinical settings for qualitative analysis of blood flow patterns. However, these images may contain aliasing artifacts that can be corrected to enhance flow visualization, as demonstrated in Fig. 1.19. This can be achieved by adjusting imaging parameters or through posterior image processing, resulting in clearer and more accurate representations of blood flow dynamics for clinicians.

During specific clinical examinations, such as the detection of mitral regurgitation, cardiologists intentionally reduce the Nyquist velocity to increase aliasing. Mitral regurgitation occurs when blood flows back from the LV into the LA due to improper closure of the mitral valves. This condition is visualized as a highly aliased jet leaking from the mitral valve area, as seen in Fig. 1.20. The severity of mitral regurgitation can be characterized with the help of quantitative measures, such as the Proximal Isovelocity Surface Area (PISA) method.

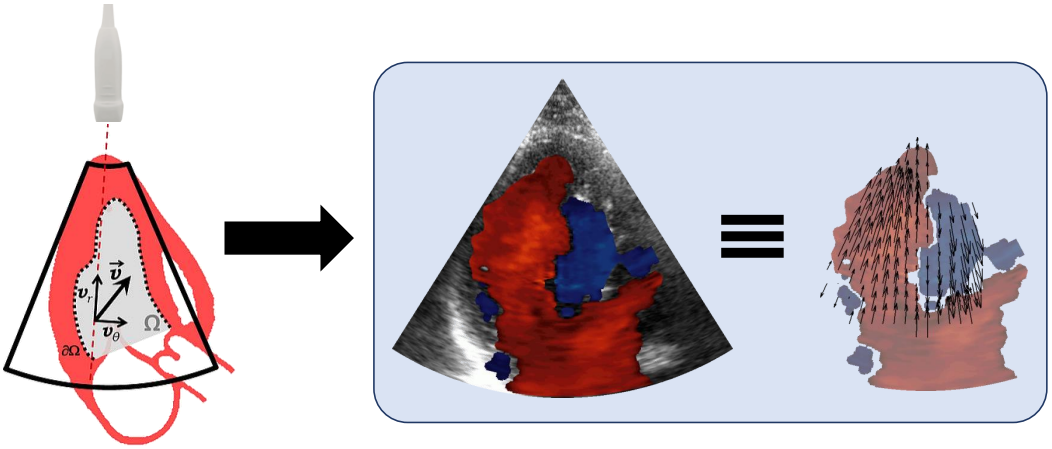


Fig. 1.18 Example of an A3C duplex image, showing the scalar velocity field along the ultrasound scanlines. Red indicates blood movement toward the ultrasound probe, while blue indicates blood moving away from the ultrasound probe. Adapted from: Biomecardio.

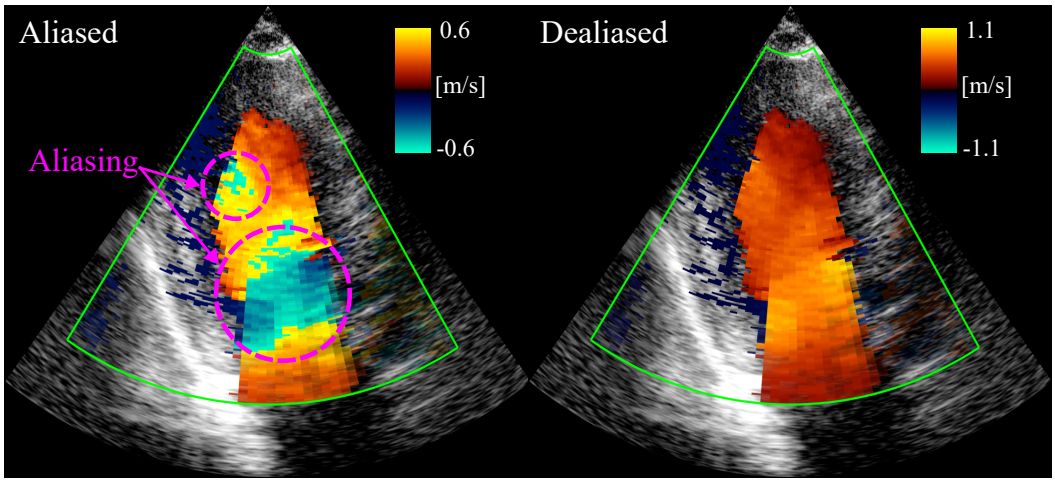



Fig. 1.19 Example of aliasing correction in an A3C duplex image acquired during diastole. Aliased zones are marked with dotted purple circles. Adapted from [19, Fig. 4].

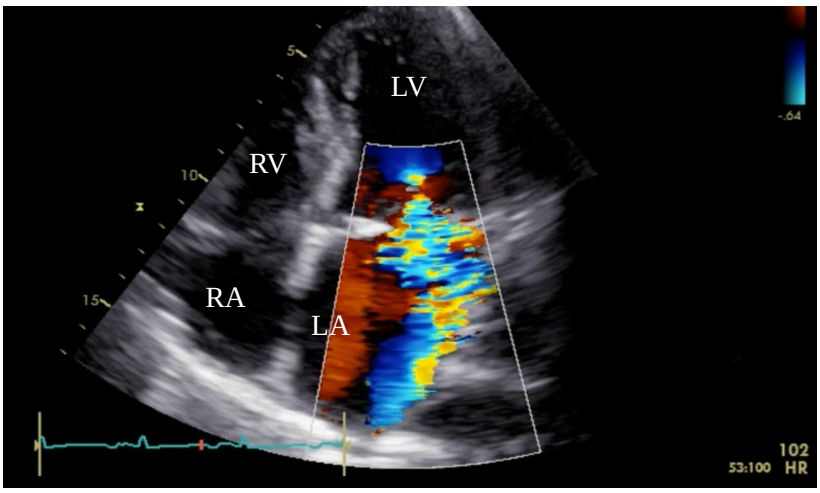



Fig. 1.20 Visualization of mitral regurgitation using color Doppler in A4C view. The Nyquist velocity is reduced to 0.4 m/s. Adapted from: eMedical Academy.

Additionally, color Doppler imaging is useful in detecting intracardiac shunt, which are abnormal pathways for blood flow in the heart. Clinicians can diagnose conditions such as atrial septal defects or ventricular septal defects by visualizing abnormal flow patterns, such as a jet of blood crossing a septal defect. Fig 1.21 illustrates an example of intracardiac shunt between the left and right ventricles, visualized using parasternal long-axis color Doppler acquisition.

#### 1.2.2.3 Pulsed-wave Doppler

Unlike CDI, which estimates Doppler velocity over a wide area, pulsed-wave (PW) Doppler focuses on a specific sample volume (SV) along a single scanline. This narrower focus boosts the temporal resolution of PW Doppler to more than 100 frames per second (fps), enabling a more detailed analysis of the velocity profile at the selected SV.

PW Doppler is particularly useful for evaluating the transmitral flow by measuring the E/A ratio and the deceleration time, both of which are crucial for assessing the diastolic function of the heart. The E/A ratio compares the early (E-wave) and late (A-wave) diastolic filling velocities of the LV, while the deceleration time measures the time required for equalizing the pressure difference between the LV and LA during early filling (refer to Fig. 1.22). These measurements are typically performed in the apical four-chamber (A4C) view by specifying the measuring point or the SV at the mitral valve. Fig. 1.23 shows the spectral curves of normal left ventricular filling. In this image, the SV is indicated by the two lines perpendicular to the chosen scanline. Although age- and sex-dependent, a normal E/A ratio ranges between 0.75 and 1.5, with a deceleration time of 160–260 ms [27].

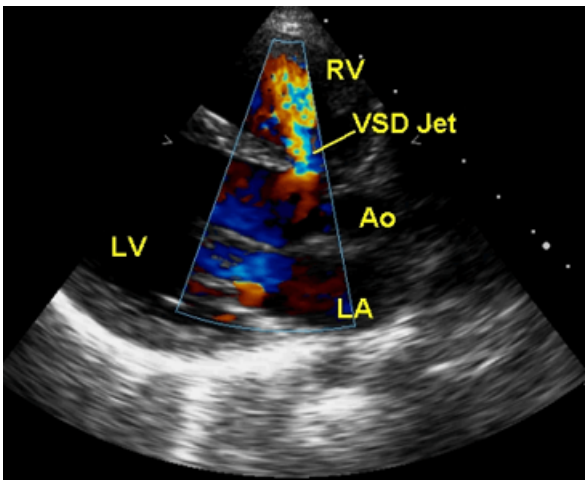



Fig. 1.21 Visualization of intracardiac shunt between the left and right ventricles using color Doppler in a parasternal long-axis view. Reproduced from: Biomecardio.

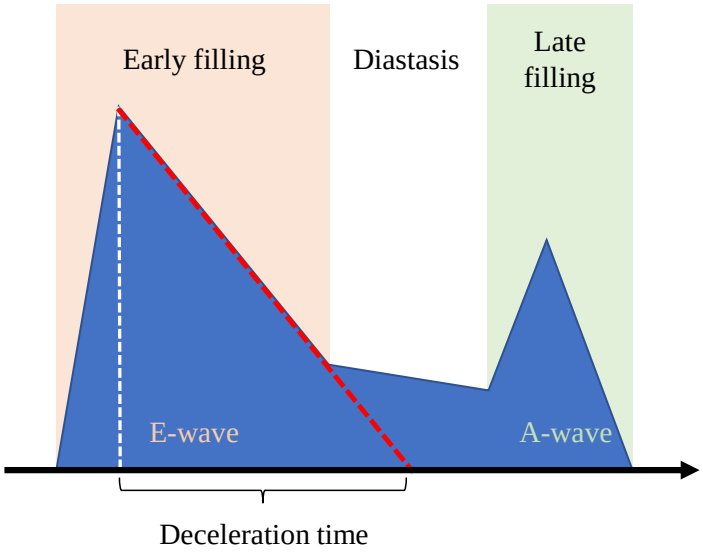



Fig. 1.22 Measurement of the deceleration time on the velocity curve given by PW Doppler.

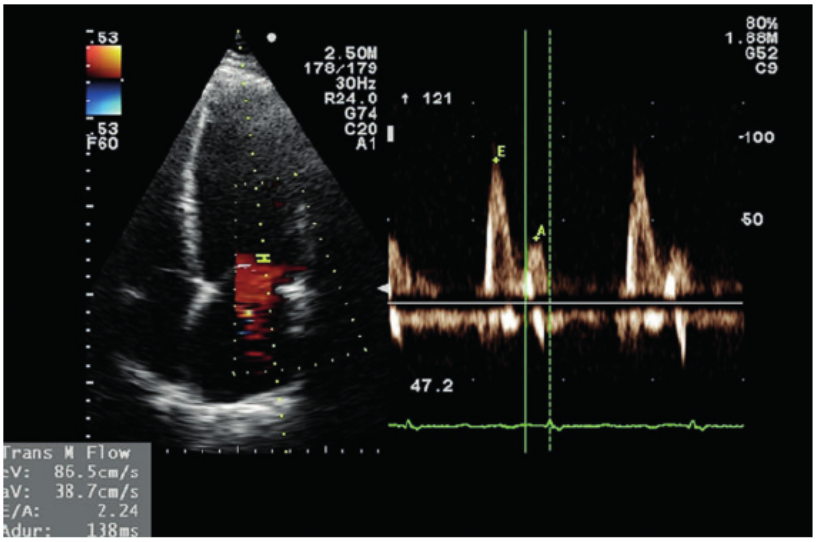



Fig. 1.23 PW Doppler measurement of flow velocities through the mitral valve annulus. Reproduced from [28, FIGURE 2].

#### 1.2.2.4 Pulsed Tissue Doppler

Pulsed tissue Doppler imaging (TDI) uses techniques similar to PW Doppler to estimate the velocity of moving tissues, particularly myocardial tissue. By placing the SV for velocity measurements at the base of the mitral annulus, specifically in the septal and lateral points of the A4C view (as shown in Fig. 1.24), TDI provides information on the myocardial motion.

Three fundamental measures obtained from TDI are the peak annular velocities during systole (S'), early filling (e'), and late filling (a'). By combining TDI with PW Doppler, the E/e' ratio can be derived. The E/e' ratio correlates with atrial pressure and serves as a metric for assessing diastolic function. A normal left ventricular filling pressure typically results in an E/e' ratio of less than 8 [29].

## 1.3 Alternative flow-based biomarkers

Cardiac diseases frequently alter heart geometry and function, which directly influence intracardiac blood flow patterns. Conventional echocardiography assesses cardiac structures and hemodynamics using parameters like ejection fraction, peak velocities, and pressure gradients. However, these methods may not capture subtle changes in blood flow dynamics early in the disease process. Moreover, the existing metrics used for evaluating diastolic function may result in inconsistent diagnoses, prompting the need for more reliable biomarkers in cardiac function assessment.

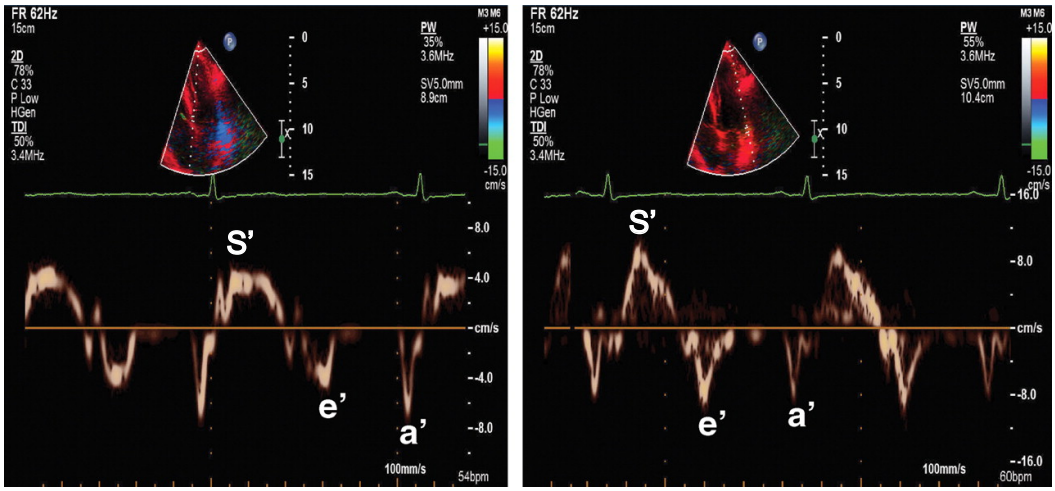



Fig. 1.24 Septal (left) and lateral (right) tissue Doppler spectral curves from a patient with an anteroseptal myocardial infarction. Reproduced from [29, Figure 10].

Recent research [30, 31] suggests that flow-based biomarkers may be an early predictor of cardiovascular diseases. These biomarkers provide a more comprehensive understanding of cardiac physiology and pathology by analyzing intracardiac blood flow patterns.

One such biomarker involves the assessment of vortex formation and flow turbulence within the LV, particularly during diastole [31, 32]. Fig. 1.25 illustrates the normal vector blood flow and vortex formation in the LV during diastole. As blood enters the LV and interacts with the mitral valve, a vortex ring is created, initiating a rotational motion from the center outwards. Meanwhile, blood continues to fill the LV through the mitral valve, pushing the vortex toward the apex while maintaining the rotational motion of the flow. During this trajectory, the vortex ring hits the lateral wall, causing part of the vortex to be crushed against the wall while allowing the other part to expand until it fills the LV, forming a large vortex. This global vortex, rotating in the natural direction of flow, conserves kinetic energy and redirects blood to the left ventricular outflow tract [30, 33]. Disruptions in normal flow patterns, often seen in conditions such as valvular disorders and cardiomyopathies, can lead to the formation of abnormal vortices and turbulence. Quantifying vortex characteristics, including location, size, and vorticity, can provide valuable diagnostic and prognostic insights [33].

Additionally, the analysis of intracardiac flow energetic properties has emerged as a promising approach for detecting early signs of cardiac dysfunction [35, 36]. Changes in flow energy parameters, such as kinetic energy, viscous dissipation, and flow efficiency, may reflect alterations in myocardial contractility and valve function.

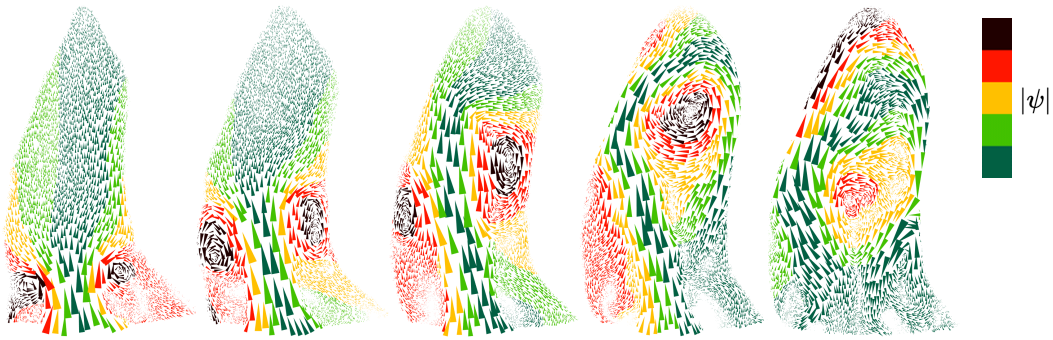



Fig. 1.25 Velocity fields and stream functions (their absolute values, $|\psi|$) during diastole given by a patient-specific computational fluid dynamics (CFD) model [34]. Vortex formation: from the two vortices alongside the mitral leaflets during early filling to the large global vortex during diastasis.

While these flow-based biomarkers hold great promise for improving early disease detection and prognosis, their implementation in clinical practice is challenging. Obtaining the intracardiac vector velocity field required for accurate analysis often relies on advanced imaging techniques that are not readily available in routine clinical settings. Although color Doppler imaging can provide partial information on intracardiac blood flow, reconstructing the complete 2D velocity field often involves time-consuming manual or semi-automatic preprocessing steps [7, 8].

This thesis aims to address the aforementioned gap by fully automating the intracardiac vector blood flow reconstruction process using color Doppler echocardiography. This could contribute toward the development of the first robust clinical tool for advancing the accessibility of flow-based biomarker analysis.

# Chapter 2

# State-of-the-art Methods

In the previous chapter, we presented the clinical context of this thesis and discussed the need to explore new flow-based biomarkers for better accessing the diastolic function due to the limitation of the existing clinical metrics. To achieve this, it is essential to analyze intracardiac vector blood flow within the LV. Therefore, we begin this chapter (Section 2.1) with an overview of various families of techniques for intraventricular vector flow imaging, covering their principles and limitations.

The ultimate goal of this thesis is to fully automate the entire intraventricular vector flow mapping pipeline (*i*VFM), as detailed in Section 2.1.3, using neural networks. This includes both the preprocessing and flow reconstruction steps, from LV segmentation and color Doppler aliasing correction to intracardiac vector blood flow reconstruction. Hence, we provide an overview of popular neural network architectures in Section 2.2, with an emphasis on their applications in echocardiography analysis, such as segmentation.

## 2.1 Intraventricular vector flow imaging

As outlined in Section 1.3, traditional ultrasound imaging techniques used in clinical practice cannot reconstruct the vector velocity field within the LV. Various studies have been proposed to address this limitation, including echocardiographic particle image velocimetry, blood speckle tracking, and *i*VFM.

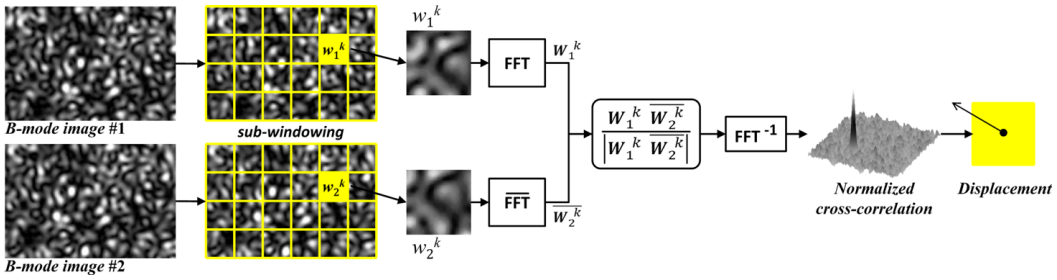



Fig. 2.1 Speckle tracking using block-matching algorithm with normalized cross-correlation implemented in the Fourier domain. Reproduced from [2, Figure 26.4].

## 2.1.1 Echocardiographic particle image velocimetry

### 2.1.1.1 Principles

Echocardiographic particle image velocimetry, also known as echo-PIV, is derived from optical PIV for measuring 2D or 3D velocity fields within the cardiovascular system. In optical PIV, two laser pulses sequentially illuminate a flow containing light-scattering microparticles, which are assumed to follow the flow dynamics. Depending on the fluid nature, appropriate particles are chosen to best match the fluid properties to ensure accurate flow tracking. High-speed digital cameras then capture the resulting scattered light in two successive images. A displacement field of the particles can be computed using block-matching algorithms. These algorithms involve dividing two successive frames into subwindows and computing the displacement vectors for each subwindow using cross-correlation techniques. Fluid velocities are then derived by taking into account the delay between two laser pulses. The optical PIV setup must be transparent due to the use of laser lights.

Echo-PIV extends this approach to ultrasound imaging through contrast-enhanced ultrasound (CE-US). CE-US uses gas-filled microbubbles as contrast agents, which are injected intravenously into human blood. These microbubbles, typically less than 6 µm in size, can freely travel through pulmonary and systemic capillaries. Unlike optical PIV, which produces images of particles, echo-PIV generates images of the interference of the scattered ultrasound waves (speckles). Consequently, speckle tracking is performed on the resulting B-mode images instead of particle tracking. Figure 2.1 illustrates speckle tracking using a block-matching algorithm with cross-correlation.

Several studies have demonstrated the feasibility of applying echo-PIV to track blood motion within the LV and assess the clinical significance of estimated velocity fields [37–39]. Fig. 2.2 shows the 2D intraventricular velocity field estimation in a healthy subject using echo-PIV in the A3C view, highlighting the presence of vortices within the LV.

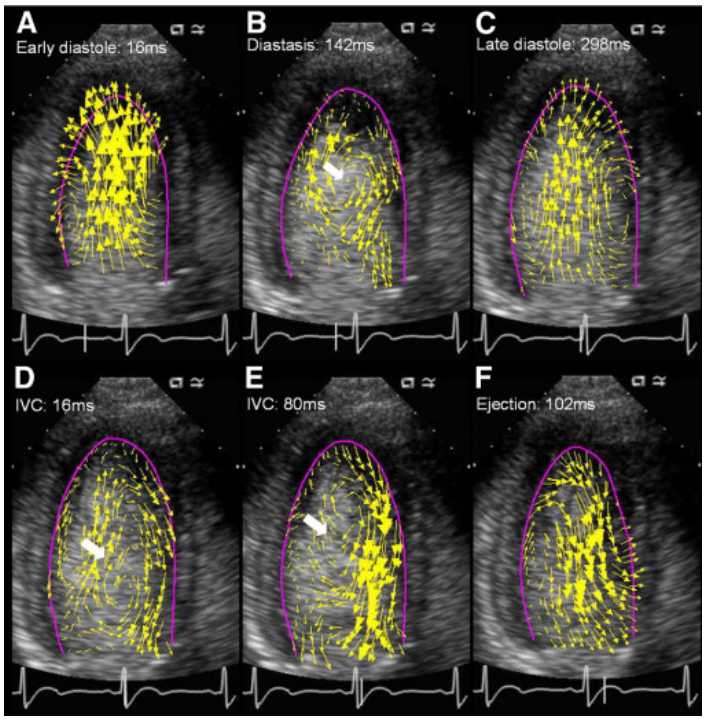



Fig. 2.2 2D intraventricular velocity field estimation in a healthy subject using echo-PIV in the A3C view. Vortices are indicated with white arrows. Reproduced from [37, Figure 5].

#### 2.1.1.2 Limitations

A major limitation of echo-PIV is the requirement for microbubble injection, which can be time-consuming and necessitates additional staff compared with conventional echocardiography. Consequently, echo-PIV is not used as a routine clinical method.

### 2.1.2 Blood speckle tracking

#### 2.1.2.1 Principles

Blood speckle tracking (BST) is a technique that directly tracks the speckles of blood using ultrafast ultrasound sequences, eliminating the need for a contrast agent [3, 4]. Typically, the blood signal is invisible on B-mode images due to the stronger echo signals produced by tissues. Therefore, a wall filter is applied to isolate the blood signals by attenuating the tissue signals. BST can be combined with CDI to enhance computational efficiency and reduce variance. This combination utilizes the radial velocity obtained from CDI and the angular velocity estimated through BST. Additionally, missing information near the wall caused by the wall filter can be restored by incorporating wall motion information via flow regularization. However, flow regularization is not mandatory for accurate flow reconstruction. Fig. 2.3

illustrates examples of intraventricular vector flow within the LV in a healthy newborn child using BST in the A3C view without regularization.

#### 2.1.2.2 Limitations

As previously mentioned, BST relies on ultrafast ultrasound sequences. However, due to the higher acquisition frequency compared to conventional echocardiography, the imaging depth is reduced (10–12 cm). This limitation restricts the application of BST primarily to pediatric cardiology, precluding its use in routine clinical practice.

### 2.1.3 Intraventricular vector flow mapping

#### 2.1.3.1 Principles

Unlike echo-PIV or BST, which track the movement of microbubbles or blood speckles, the *i*VFM method aims to reconstruct vector blood flow from color Doppler images. As explained in Section 1.2.1.4, Doppler velocity is the projection of the actual blood velocity along the ultrasound axis. In a polar coordinate system, the sign-inverted Doppler velocity (positive velocities for movement away from the probe), $v_{\mathrm{D}} = -u_{\mathrm{D}}$, provides the noisy radial component, $v_r$:

$$v_{\mathrm{D}} = v_r + \epsilon, \qquad [\mathrm{m\,s^{-1}}] \quad (2.1)$$

where $\epsilon$ represents the acquisition noise. To construct the 2D vector field, all that remains is to estimate the angular or polar velocity component, $v_\theta$.

The first *i*VFM method was proposed in 2006 by Ohtsuki et al. [40]. In this method, intracardiac vector blood flow is characterized by two velocity vector components: the velocity in the beam direction ($u$) and the velocity perpendicular to the beam direction ($v$).

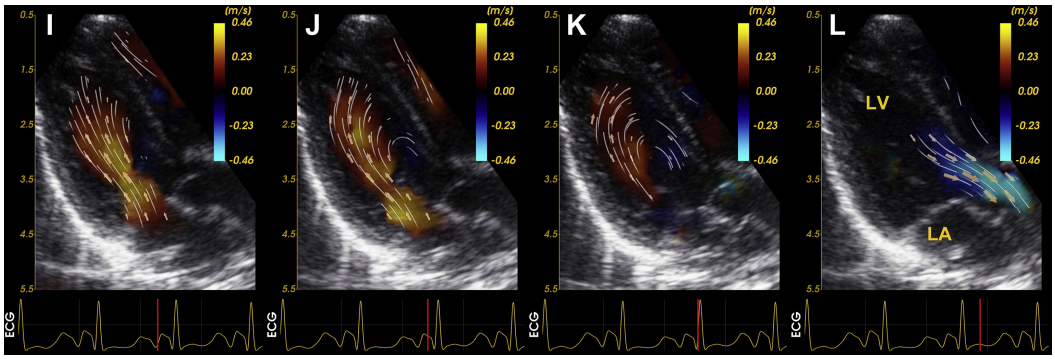



Fig. 2.3 2D intraventricular velocity field estimation in a healthy newborn child using BST in the A3C view without regularization. Reproduced from [3, Supplemental Figure 2].

Each of these components is further decomposed into a vortex flow component ($u_s$ and $v_s$) and a base flow component ($u_b$ and $v_b$). The vortex component represents the swirling or circulating part of the flow and is derived using the stream function, $S(x, y)$ [40, Eq. (5)]. Specifically, $u_s = \frac{\partial S}{\partial y}$ and $v_s = -\frac{\partial S}{\partial x}$. This implies that the stream function is calculated based on the integration of the velocity component in the beam direction along the perpendicular axis. On the other hand, the base flow component represents the main directional flow and is derived from the flow function [40, Eq. (6)]. The flow function is calculated by integrating the beam direction velocity component ($u$) along the perpendicular axis, enabling the computation of the ratio $k$, which represents the ratio of the positive flux of the vortex to the positive portion of the total flux passing through the integration boundary. The base flow velocity component ($u_b$) is then determined by $u_b = u(1 - k)$. Subsequently, the base flow streamline can be computed by normalizing the flow function by the total flux at each position. The $v_b$ component, perpendicular to $u_b$, can be obtained by $v_b = u_b \tan\theta$, where $\theta$ is the angle between the beam direction and a tangent to the base flow streamline. Finally, the total velocity components are represented as the sum of the vortex and base flow components: $u = u_s + u_b$ and $v = v_s + v_b$.

This method was subsequently improved by Uejima et al. in 2010 [41]. In this version, the vortex flow is assumed to be symmetric along iso-radial lines (white circular arcs in Fig. 2.4 (a1) and (a2)), meaning that the positive fraction of the vortex is equal to the negative fraction, as illustrated in Fig. 2.4 (b). This hypothesis considers the vortex flow as pure rotational motion. However, this strong assumption is not typically verified under physiological conditions. Studies using echo-PIV have demonstrated that the ring-shaped vortex, which is symmetrical at the start of diastole, deforms and elongates to fill the ventricular chamber, thereby losing its symmetrical property [37, 42, 43].

In 2012, Pedrizzetti et al. [44] proposed a technique similar to Ohtsuki's, but with a different decomposition approach. Instead of separating the flow into base and vortex components, the method decomposes the flow into Doppler and irrotational components. The irrotational flow is calculated so that its curl ($\omega = \nabla \times \vec{V}$) is zero, while the final flow, which is the sum of Doppler and irrotational flows, has zero divergence to ensure mass conservation in the plane. This approach ensures that the radial component of the rotational flow is zero, so the radial component of the reconstructed flow matches that of the radial velocity component. However, this reconstruction assumption is not based on any physical principle and tends to underestimate the norm of the curl. Consequently, it also reduces the value of the vorticity used to estimate the vortex, as it is derived from the curl value.

Unlike the flow decomposition techniques discussed above, Garcia et al. [5] proposed reconstructing the intraventricular vector blood flow based on the 2D continuity equation,

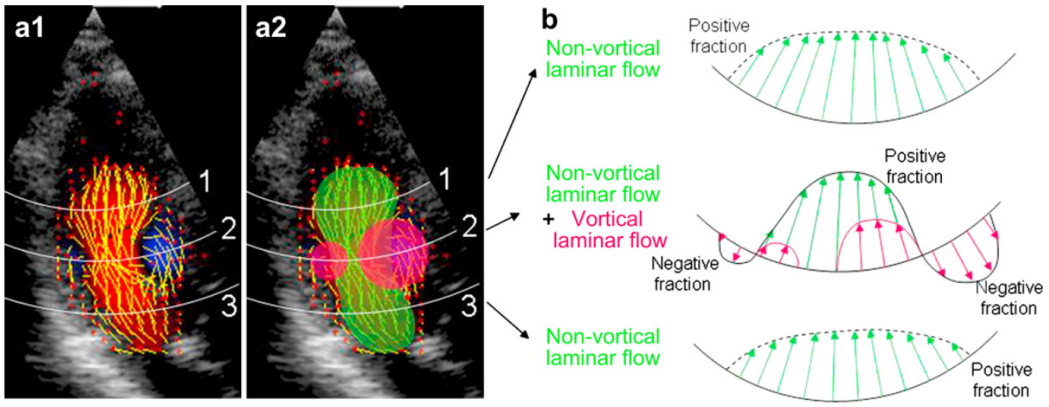



Fig. 2.4 (a1): Example of vector flow mapping produced by Uejima's method [41]. (a2): Decomposition of intracardiac flow into a base flow (green) and two vortex flows (pink). (b): Color Doppler velocity decomposition into basic flow and vortex flow so that the vortex flow component is bilaterally asymmetrical on each arc shown in a1 and a2. The velocity field is estimated at the start of diastole. Reproduced from [41, Fig. 3].

also known as the mass conservation equation. This approach is hereinafter referred to as *i*VFM-v1. The underlying hypothesis is that blood is an incompressible fluid and out-of-plane velocity components are negligible. The continuity equation in polar coordinates relates the angular ($V_\theta$) and radial ($V_r$, given by color Doppler data) velocities through a first-order partial differential equation:

$$\frac{\partial V_\theta(r,\theta)}{\partial \theta} = -r\frac{\partial V_r(r,\theta)}{\partial r} - V_r(r,\theta). \qquad [\mathrm{m\,s^{-1}}] \quad (2.2)$$

This equation can be solved if a boundary condition is provided. In Garcia's method, the time-varying boundary condition is obtained using a speckle-tracking algorithm that is commercially available in most clinical echocardiographs. It is defined as the fluid angular velocity at the myocardium level by imposing free-slip boundary conditions (normal component equals zero) because the resolution of the color Doppler is insufficient to capture the thin boundary layers. Given the boundary condition, two angular velocities are estimated by integrating from the left (inferolateral wall) and right (anteroseptal wall) along each iso-radial line. A weight function, which decreases from 1 to 0 from one wall to the other, is then used to combine these angular velocity estimations to reduce computational error. Additionally, to overcome the limitation of the temporal resolution in color Doppler imaging, Doppler velocities over several successive cardiac cycles are merged and smoothed.

One limitation of this algorithm lies in the line-by-line reconstruction of the velocity field along each iso-radial line within the LV. Another limitation is that this method does not consider the neighboring Doppler velocities in the radial direction during the reconstruction,

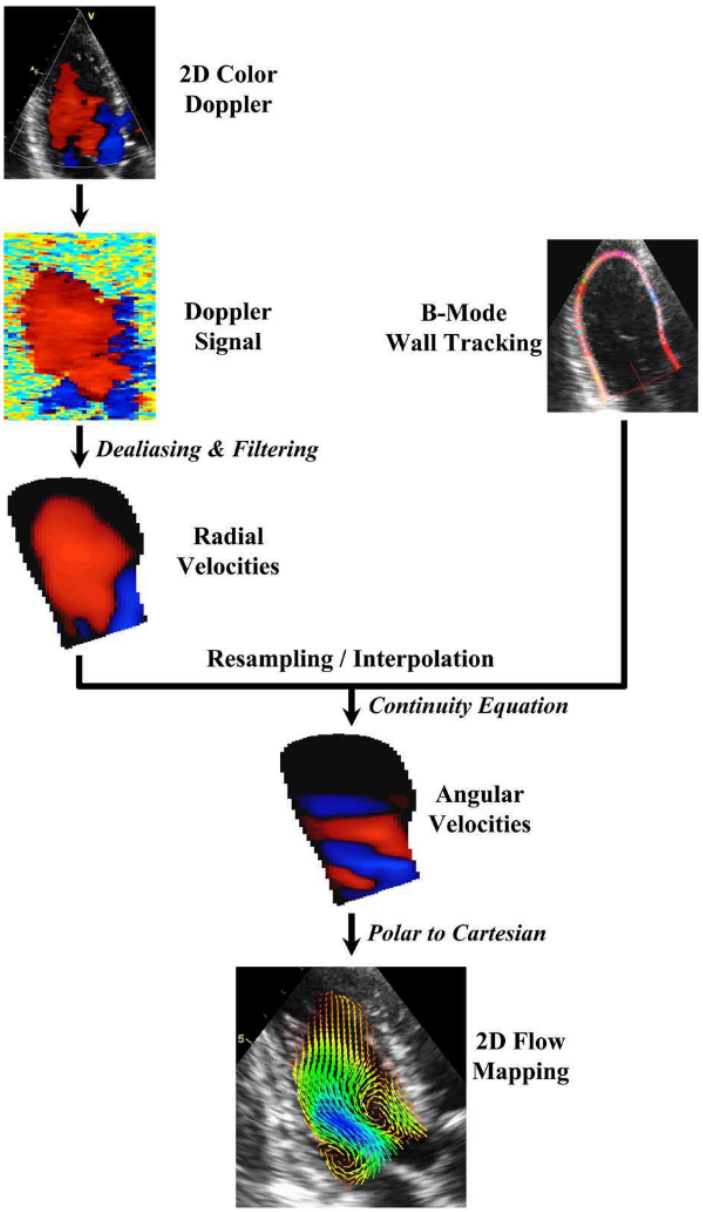



Fig. 2.5 Flowchart for reconstructing the angular velocity from color Doppler in Garcia's *i*VFM-v1 method. Reproduced from [5, Fig. 2].

which may lead to discontinuities in the reconstructed vector field. Fig. 2.5 summarizes the intraventricular flow mapping method introduced by Garcia et al.

This algorithm has also been integrated into Fujifilm Healthcare (previously Hitachi) [45] and Esaote [46] clinical echocardiography scanners. In 2020, Zhuang et al. [47] proposed improving the wall tracking algorithm used by Garcia et al. through deep learning (DL), aiming to obtain more accurate boundary angular velocities.

Meyers et al. [6] developed another vector flow mapping method based on the vorticity equation and the stream function. This approach enforces a divergence-free velocity field through the Laplace operator. However, the no-slip boundary condition is used rather than the free-slip boundary condition, implying that fluid near the boundary follows the wall velocity. This assumption is inappropriate as the boundary layer is not capturable by CDI, as explained previously. Furthermore, the raw Doppler velocity is directly used to impose the radial velocity, which may introduce noise into the reconstructed velocity field. Smoothing might be required to reduce the noise in the estimated velocity field.

In 2017, Assi et al. [7] proposed another vector flow mapping technique (*i*VFM-v2) by utilizing the 2D continuity equation, free-slip boundary conditions, and a spatial smoothing regularization. The vector field, $\hat{\boldsymbol{v}} = (\hat{v}_r, \hat{v}_\theta)$, is reconstructed by minimizing the following cost function:

$$\mathcal{L}_{\mu_1,\mu_2,\mu_3}(\boldsymbol{v}) = \underbrace{\int_\Omega (v_r - v_\mathrm{D})^2}_{\substack{\text{closely match}\\ \text{the Doppler data}}} + \mu_1 \underbrace{\int_\Omega \operatorname{div}(\boldsymbol{v})^2}_{\substack{\text{divergence-free}\\ \text{constraint}}} + \mu_2 \underbrace{\int_{\partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n}_\mathrm{W})^2}_{\substack{\text{free-slip}\\ \text{boundary conditions}}} + \mu_3 \underbrace{\int_\Omega S(\boldsymbol{v})^2}_{\substack{\text{smoothing}\\ \text{regularization}}}, \tag{2.3}$$

where $\Omega$ is the domain of interest (LV cavity), $\boldsymbol{n}_\mathrm{W} = (n_{\mathrm{W}r}, n_{\mathrm{W}\theta})$ represents a unit vector perpendicular to the endocardial wall, $S(\cdot)$ signifies the spatial smoothing function, and $\mu_1, \mu_2, \mu_3 \in \mathbb{R}_{>0}$ are the penalty coefficients. To solve $\arg\min_{(v_r, v_\theta)} \mathcal{L}_{\mu_1,\mu_2,\mu_3}(\boldsymbol{v})$, the problem is rewritten in matrix form, and the finite difference technique is used to compute the derivatives. This forms a linear problem, $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{b}$, which is then solved using the least squares approach. In this matrix equation, $\boldsymbol{A}$ is a symmetric hollow matrix, $\boldsymbol{v}$ is a column vector containing the reconstructed vector field, and $\boldsymbol{b}$ is another column vector consisting of Doppler velocities. The free-slip boundary conditions seek to impose a flow parallel to the bounding surface near the endocardial wall. In other words, the blood is free to move along the wall but does not necessarily follow the myocardium movement as the LV expands or contracts. Fig. 2.6 shows the *i*VFM-v2 pipeline, from endocardial wall delineation to vector flow reconstruction. A major limitation of this technique is the use of the L-curve [48] to obtain the optimal penalty coefficients, meaning several calculations must be done beforehand to find these coefficients.

Later in 2021, Vixège et al. [8] introduced *i*VFM-v3 by imposing physical constraints as compared to *i*VFM-v2. This is achieved by solving the following optimization problem governed by two equality constraints, $C_1$ (2D mass conservation equation) and $C_2$ (free-slip boundary conditions):

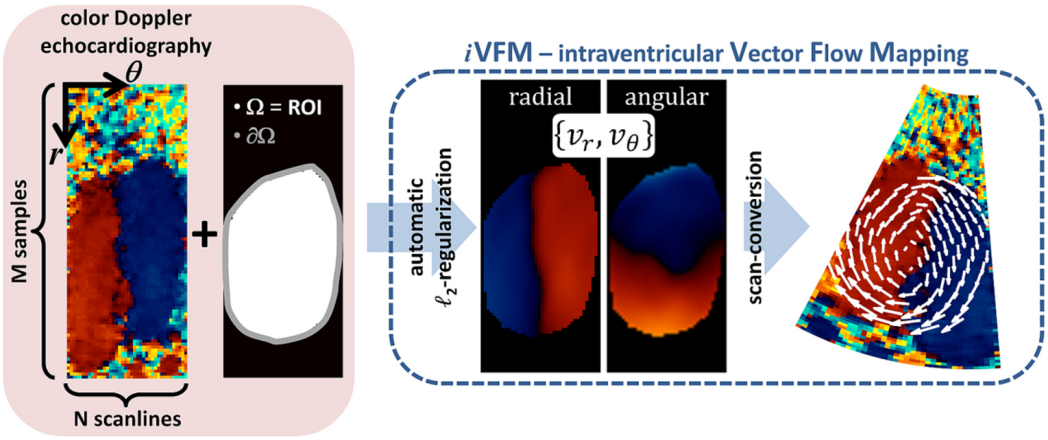



Fig. 2.6 *i*VFM-v2 method proposed by Assi et al.. Left: delineation of the endocardial wall; right: reconstructed intracardiac vector blood flow. Reproduced from [7, Figure 2].

$$\hat{\boldsymbol{v}} = (\hat{v}_r, \hat{v}_\theta) = \arg\min_{(v_r, v_\theta)} \underbrace{\int_\Omega \omega \, \|v_r - v_\mathrm{D}\| \, d\Omega}_{\text{closely match the Doppler data}}$$

subject to: (2.4)

$$\begin{cases} C_1 = r\mathrm{div}(\hat{\boldsymbol{v}}) = r\dfrac{\partial \hat{v}_r}{\partial r} + \hat{v}_r + \dfrac{\partial \hat{v}_\theta}{\partial \theta} = 0 & \text{on } \Omega \\ C_2 = (\hat{\boldsymbol{v}} - \boldsymbol{v}_\mathrm{W}) \cdot \boldsymbol{n}_\mathrm{W} = (\hat{v}_r - v_{\mathrm{W}_r}) n_{\mathrm{W}_r} + (\hat{v}_\theta - v_{\mathrm{W}_\theta}) n_{\mathrm{W}_\theta} = 0 & \text{on } \partial\Omega, \end{cases}$$

where, $(\hat{v}_r, \hat{v}_\theta)$ represent the estimated radial and angular blood velocity components. The term $\omega$ indicates the weights of the data fidelity term, equalling normalized Doppler power values in the range of $[0, 1]$, as they reflect the reliability of the Doppler velocity. The vector $\boldsymbol{v}_\mathrm{W} = (v_{\mathrm{W}_r}, v_{\mathrm{W}_\theta})$ is a velocity vector of the endocardial wall.

Instead of relying on the L-curve to obtain the optimal penalty coefficients, Vixège et al. opted for the Lagrange multipliers method. This reduces the number of supervisedly determined parameters to just one, i.e., the smoothing regularization weight. The Lagrangian function of the constrained minimization problem (2.4) is given by:

$$\mathcal{L}(\boldsymbol{v}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}) = \omega \, \|v_r - v_\mathrm{D}\|^2 + \langle \boldsymbol{\lambda_1}, \boldsymbol{C_1} \rangle + \langle \boldsymbol{\lambda_2}, \boldsymbol{C_2} \rangle. \tag{2.5}$$

Here, $\lambda_1$ and $\lambda_2$ are real Lagrange multipliers to be determined, associated with the two constraints: the mass conservation, $C_1$, and the free-slip boundary condition, $C_2$. The notation $\langle \cdot, \cdot \rangle$ refers to the inner product of two vectors. Subsequently, $\mathcal{L}(\boldsymbol{v}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2})$ is expressed in a linear least squares form. Solving $\nabla_{v,\lambda_1,\lambda_2} \mathcal{L}(\boldsymbol{v}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}) = 0$ results in a linear system, $\boldsymbol{A'x} = \boldsymbol{b'}$, that contains the solution to the constrained minimization problem. Given that

Doppler velocities and endocardial wall velocities can be relatively noisy, a smoothing regularization, $S(\cdot)$, similar to the one used by Assi et al. is added to the linear system. The final system is then solved using a regularized least squares approximation:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{ \left\| \boldsymbol{A}'\boldsymbol{x} - \boldsymbol{b}' \right\|^2 + \alpha \left\| S(\boldsymbol{x}) \right\|^2 \right\}, \tag{2.6}$$

where the regularization weight $\alpha$ is determined with the L-curve.

Even though 2D *i*VFM methods provide a clear visualization, they are limited to flow reconstruction on a selected plan. Several works have been proposed to reconstruct the 3D vector blood flow in a full intraventricular volume, either by triplane [49] or full-volume color Doppler imaging [50, 51]. Since this manuscript only focuses on 2D *i*VFM, the methodological details of these 3D+t *i*VFM techniques will not be discussed.

#### 2.1.3.2 Limitations

Most 2D *i*VFM techniques require manual or semi-automatic preprocessing steps, which often involve the segmentation of the LV endocardial wall (for the boundary conditions) and the correction of the aliasing artifacts on the color Doppler images (for the correct data fidelity term). These time-consuming steps can be fully automated using artificial intelligence, including the vector flow reconstruction itself.

## 2.2 Artificial intelligence for echocardiography analysis

Artificial neural networks (NNs) consist of interconnected neurons that mimic their biological counterparts in the human brain. NNs have been proven to be universal function approximators [52], capable of learning complex function mappings from one domain to another.

Although NNs were introduced as early as the mid-1900s, they only started gaining popularity in the early 2000s, especially after their achievement in the ImageNet computer vision classification challenge in 2012 [53]. Starting from the early multi-layer perceptrons (MLPs) to the convolutional neural networks (CNNs) that revolutionized image analysis using DL, the recent introduction of transformers and diffusion models further pushes the limits of natural language processing and generative artificial intelligence (AI) domains, giving rise to large language models (LLMs) such as GPT-4o by OpenAI, Gemini by Google, and Mistral by Mistral AI.

Given the overwhelming amount of work that has been proposed in the AI domain, the following subsections focus on some common NN architectures (MLPs, CNNs, transformers, Segment Anything Model, deep unfolding approaches, and physics-informed/-guided neural

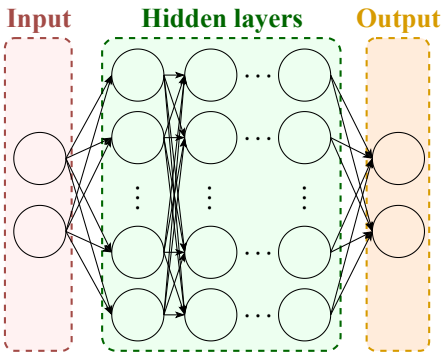



Fig. 2.7 Example of an MLP architecture that consists of an input layer, several hidden layers, and an output layer.

networks) and their application in echocardiography analysis, specifically B-mode and color Doppler. Several network architectures were used in this thesis to achieve the ultimate goal of reconstructing intracardiac vector blood flow, which requires prior LV segmentation and color Doppler dealiasing. nnU-Net [9], explained further in Section 2.2.2, was extensively studied in Chapter 3 for segmenting the LV in echocardiographic B-mode sequences with temporal consistency. For the dealiasing task, the performance of nnU-Net, transformer-, and deep unfolding-based methods were compared in Chapter 4. Finally, physics-informed neural networks were developed in Chapter 5 to solve (2.4) for intracardiac vector flow reconstruction. Additionally, nnU-Net initially made for segmentation, was adapted and trained to perform *i*VFM in a supervised manner.

### 2.2.1 Multi-layer perceptrons

An MLP is a fully connected NN typically consisting of several layers: an input layer, one or more intermediate (hidden) layers, and an output layer. Each neuron receives information from connected neurons and processes it before passing the processed information to other connected neurons. The importance of each piece of information from connected neurons is determined by a parameter called weight, which is updated during the learning process. The weighted information is summed, sometimes with the addition of learnable bias parameters, and then passed through non-linear functions known as activation functions. Some commonly used activation functions are sigmoid, Rectified Linear Units (ReLU) [54], and Gaussian Error Linear Units (GELU) [55]. Given their specific property as universal function approximators, MLPs are often used in physics-informed neural networks, as explained in Section 2.2.6, which we used to perform *i*VFM in Chapter 5.

Fig. 2.7 illustrates an example of an MLP architecture. The number of trainable parameters of an MLP with $n$ hidden layers and learnable biases can be calculated using the following formula:

$$\#\text{ trainable param.} = i \times h_1 + \sum_{k=1}^{n-1} \big(h_k \times h_{k+1}\big) + h_n \times o + \sum_{k=1}^{n} h_k + o, \tag{2.7}$$

where $i$ is the number of neurons in the input layer, $h_k$ is the number of neurons in the $k^{th}$ hidden layer, and $o$ represents the number of neurons in the output layer.

### 2.2.2 Convolutional neural networks

Although MLPs can learn features from input data, they quickly become impractical for high-dimensional inputs like high-resolution images due to the exponential growth in the number of parameters. This is where CNNs come into play, as CNNs can efficiently process grid-like data structures such as images.

CNNs are particularly well-suited for image analysis because they are designed to automatically and adaptively learn spatial hierarchies of features across multiple layers using the concept of the receptive field. The receptive field of a neuron in a convolutional layer refers to the region of the input image that it "sees" or responds to. Unlike MLPs, CNNs use convolutional layers that apply filters to the input image. Shallow layers in a CNN have small receptive fields, allowing them to capture low-level features such as edges, textures, and simple patterns. As the network goes deeper, the receptive fields expand, enabling neurons to include information over larger areas of the image. This allows deeper layers to capture higher-level features such as shapes, objects, and even semantic content.

A key advantage of CNNs is their use of shared weights and local connections, which significantly reduces the number of parameters compared to fully connected networks. This not only makes CNNs more computationally efficient but also improves their ability to generalize from limited training data. The use of pooling layers to aggregate local information further reduces the dimensionality of the data, focusing on the most relevant features and improving computational efficiency.

When paired with a fully connected layer at the output, CNNs can be used to perform classification. In 2012, AlexNet [53] marked a milestone in computer vision (CV) by winning the ImageNet classification challenge, proving that deeper architectures lead to better performance, especially given a large dataset. To address the vanishing gradient problem in deep networks, residual connections [56] were proposed. These connections directly link the

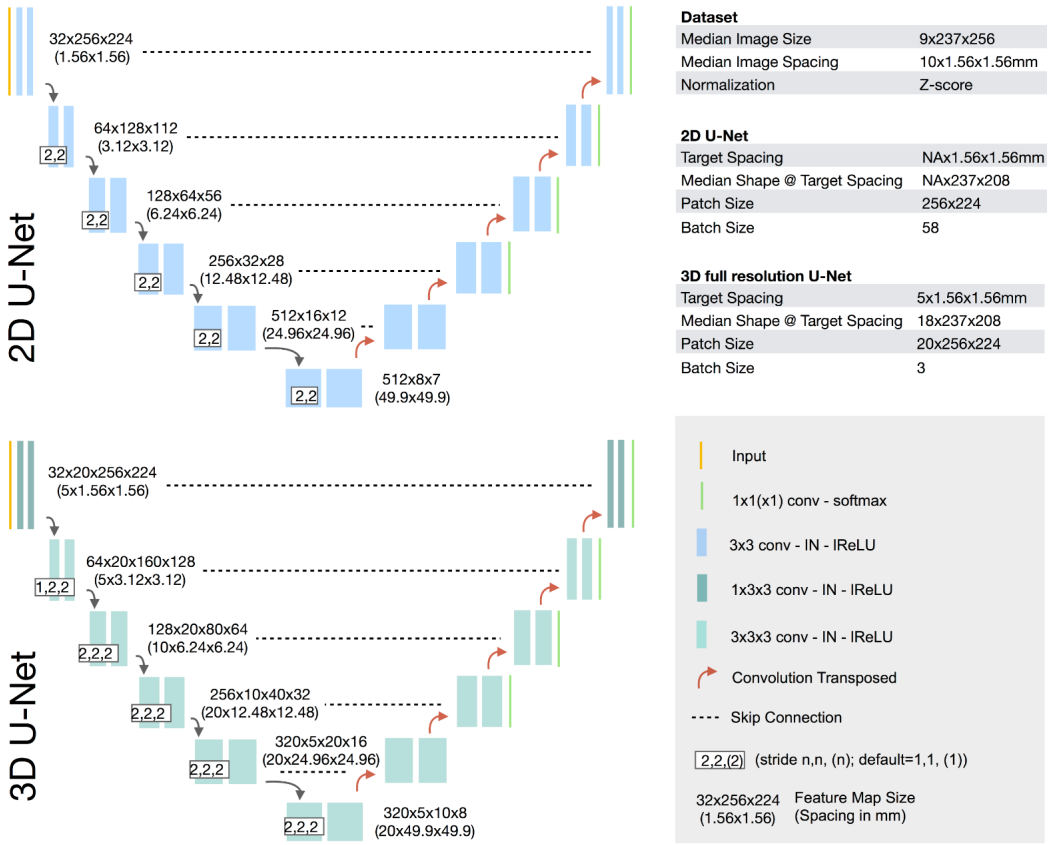



Fig. 2.8 Example of 2D and 3D U-Net architectures generated by nnU-Net for ACDC dataset. IN indicated instance normalization while lReLU signifies Leaky ReLU activation function. Reproduced from [9, Figure C.1].

input of a subnetwork with its output, enabling the training of substantially deeper models and further improving performance in tasks such as image classification and object detection.

For segmentation tasks, the introduction of U-Net [57] popularized the encoder-decoder architecture with additional symmetric skip connections between the encoder and decoder. This design allows for precise localization by combining low-level spatial information from the encoding path with high-level semantic information from the decoding path. U-Net's versatility and accuracy have been further improved by nnU-Net [9]. Through the extraction of dataset fingerprints, such as image modality, spacing, and size, nnU-Net automatically determines the optimal U-Net architectures and training strategies (pre- and postprocessing, data augmentations, model ensembling, etc.), making it the state-of-the-art framework that has won over 20 public biomedical segmentation challenges. nnU-Net uses a patch-wise approach for both training (one patch per image or volume) and inference (sliding window approach). The patch size is chosen based on the median image shape of the training dataset, resampled to the median voxel spacing. Fig. 2.8 shows an example of the 2D and 3D U-Net

architectures generated by nnU-Net for the Automated Cardiac Diagnosis Challenge (ACDC) dataset [58].

### 2.2.3 Transformers

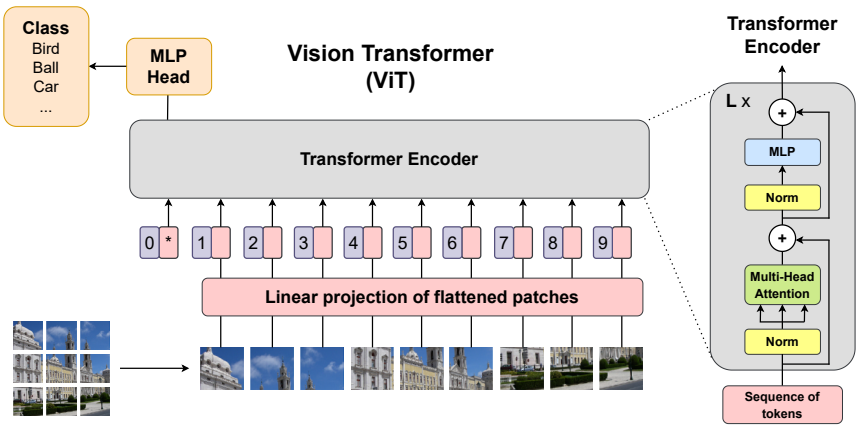
If U-Net is the architecture that pioneered segmentation in the medical domain, transformers have surely revolutionized natural language processing (NLP). Introduced by Vaswani et al. in 2017 [59], transformers convert input data into a sequence of tokens and rely entirely on attention mechanisms to process input data. This allows the modeling of complex dependencies without the need for recurrent layers.

Dividing the input data into smaller units, equivalent to embeddings in a $D$-dimensional space, is achieved through the tokenization process. Tokenization involves a simple linear projection and a position embedding step that encodes the relative or absolute position of the tokens in the sequence. Unlike CNNs that consider the spatial connectivity of the input, transformers learn the relations and interactions between tokens.

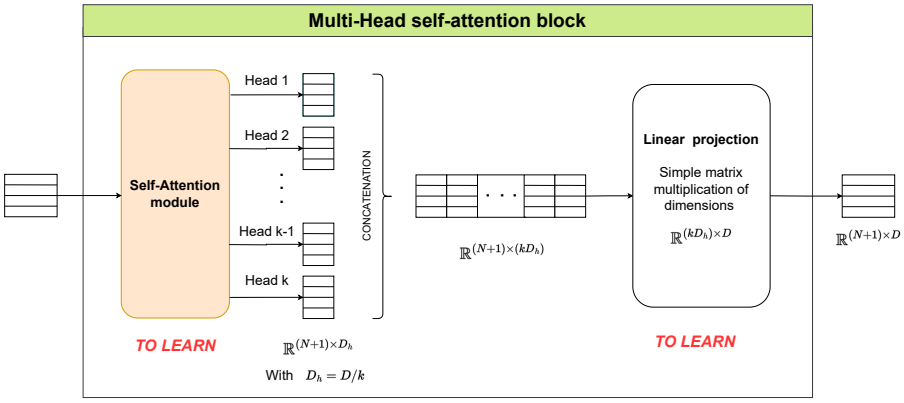
The core component of a transformer is the self-attention mechanism, also known as scaled dot-product attention. Self-attention is the process by which transformers compute the relevance of each input token to every other token in the sequence. This is achieved through three main steps: creating queries ($Q$), keys ($K$), and values ($V$) from the input embeddings. The attention scores are then computed by taking the dot product of $Q$ with $K^{\mathsf{T}}$ and scaling the result by the square root of the dimension of the key vectors. These attention scores are then normalized using the *softmax* function to obtain the attention weights matrix, $A$, representing the importance of each token relative to others. The final output of the self-attention mechanism is the dot product between $A$ and $V$.

Instead of relying on a single self-attention module, multi-head self-attention is proposed to jointly attend to information from different representation subspaces at different positions. This is achieved by using multiple self-attention modules running in parallel to compute multiple sets of queries, keys, and values, and concatenating their outputs. A transformer encoder is typically composed of a stack of identical layers, with each layer consisting of a multi-head self-attention mechanism, a simple MLP, and a normalization layer, as shown in Fig. 2.9a.

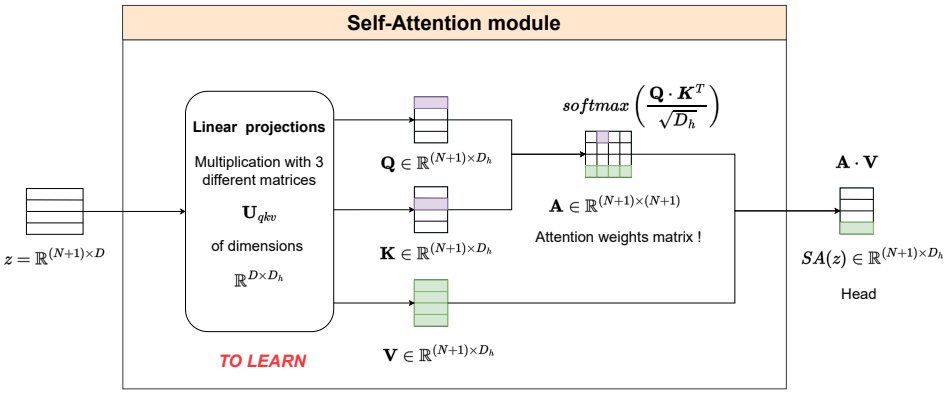
While initially designed for NLP, transformers have been adapted for other domains, including computer vision (Vision Transformer or ViT [60]), time series analysis [61], and reinforcement learning [62]. For example, ViT applies the transformer architecture to image patches, treating them as sequences of tokens similar to words in a sentence. This approach has demonstrated competitive performance with traditional CNNs on image classification tasks and has paved the way for transformers' application in other image-based tasks. Fig. 2.9




(a) ViT architecture overview




(b) Multi-head self-attention mechanism




(c) Self-attention module

Fig. 2.9 (a): ViT architecture overview. The input image is divided into patches and linearly projected to create a sequence of tokens. A learnable class token (token 0 in the schema) is appended to other tokens before being fed to the transformer encoder. At the output of the transformer encoder, only the class token is kept and passed to the classification head to learn the classification. (b): multi-head self-attention mechanism; (c): self-attention module. Adapted from MYRIAD Team Website.

provides a schematic overview of ViT architecture along with the multi-head and self-attention mechanisms described previously.

In contrast to the fixed filter weights in CNNs after the training, the attention weights matrix is dynamically computed based on the given input, making transformers more expressive than CNNs. Furthermore, transformers typically have a larger receptive field compared to CNNs, allowing them to better focus on the global information in the input data. By leveraging the attention weights matrix, transformers are proven to have better explainability than CNNs [63].

Despite the mentioned advantages, transformers suffer from some drawbacks. This includes the limited length of the sequence of tokens, as the computation complexity and memory increase quadratically with the length. Another known limitation is the data hungriness since a huge amount of data is needed to train the transformers due to the lack of assumptions about the interactions in the data.

### 2.2.4 Segment Anything Model

In NLP, LLMs pre-trained on huge datasets exhibit strong zero-shot and few-shot generalization capabilities [64]. Zero-shot learning allows a trained model to generalize to new and unseen categories without requiring any annotated examples during training. Few-shot learning, on the other hand, involves the model adapting to new tasks with only a few annotated examples. These models are known as foundation models [65]. Drawing inspiration from LLMs, the Segment Anything Model (SAM) [10] was proposed in 2023 as a foundation model for segmentation tasks. Although not implemented or tested in this thesis, SAM-based models can be useful in annotating large-scale images to improve the performance of segmentation models, as discussed further in the "Foundation models for large-scale echocardiographic data annotation" section in "Perspectives for Future Work" of Part IV.

SAM is a promptable segmentation tool that can segment everything in a natural image or respond to prompts such as interactive clicks and bounding boxes. In cases of ambiguous prompts, it can generate multiple valid masks with associated confidence scores. Along with SAM, the authors released the massive SA-1B dataset, containing 11M images and 1.1B masks automatically segmented by SAM. Initially, SAM was trained on publicly available segmentation datasets and used to produce masks for a subset of the SA-1B dataset, which were then corrected by professional annotators. After accumulating enough annotated data, SAM was retrained using these newly annotated masks. This retrained SAM was then used to annotate additional images and objects. Through this iterative retraining process, SAM became powerful enough to fully and automatically annotate all 11M images in the SA-1B dataset, producing 1.1B high-quality masks.

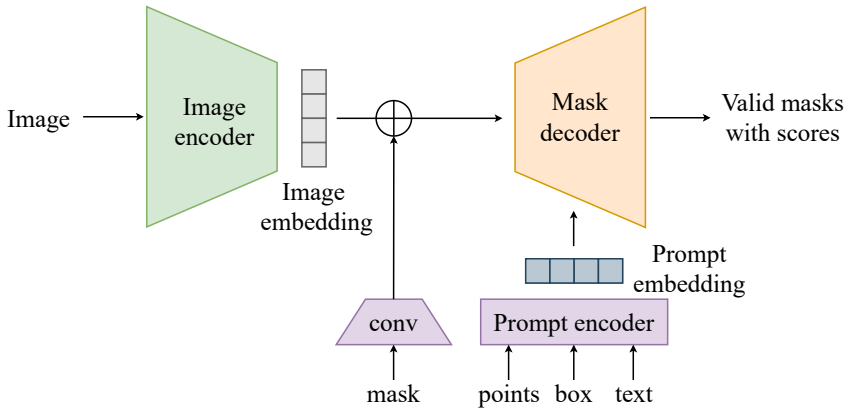



Fig. 2.10 SAM overview. A ViT-based image encoder outputs an image embedding that can be efficiently queried by different input prompts to generate object masks at nearly real-time speed. SAM can produce multiple valid masks and associated confidence scores for ambiguous prompts corresponding to more than one object. See [10, Figure 4] for the original schema.

SAM's architecture comprises three components, as illustrated in Fig. 2.10: an image encoder, a flexible prompt encoder, and a fast mask decoder. The image encoder is a masked-autoencoder [66] pre-trained ViT adjusted to process high-resolution inputs. Point and box-like prompts are represented by positional encodings summed with learned embeddings for each prompt type, while text prompts are embedded using the off-the-shelf text encoder from CLIP [67]. Mask prompts are embedded using convolutions and summed element-wise with the image embedding. The lightweight mask decoder is based on a modified transformer decoder block followed by a dynamic mask prediction head.

Shortly after the release of SAM, several works have extended SAM to medical applications, such as MedSAM [68], SAM-Med2D [69], and SAM-Med3D [70]. These models utilize SAM's pre-trained weights for fine-tuning on 2D or 3D medical datasets, demonstrating its versatility and potential for specialized segmentation tasks.

### 2.2.5 Deep unfolding

Deep unfolding, also known as deep unrolling, is an approach that bridges the gap between traditional iterative algorithms and modern DL techniques. Traditional algorithms for solving inverse problems often rely on iterative methods, which can be computationally expensive and slow to converge. Deep unfolding, introduced in 2010 [11], proposes to transform these iterative processes into trainable NNs. In our research, we utilized a learned primal-dual-based deep unfolding algorithm to address the inverse problem of dealiasing, which is elaborated in detail in Chapter 4.

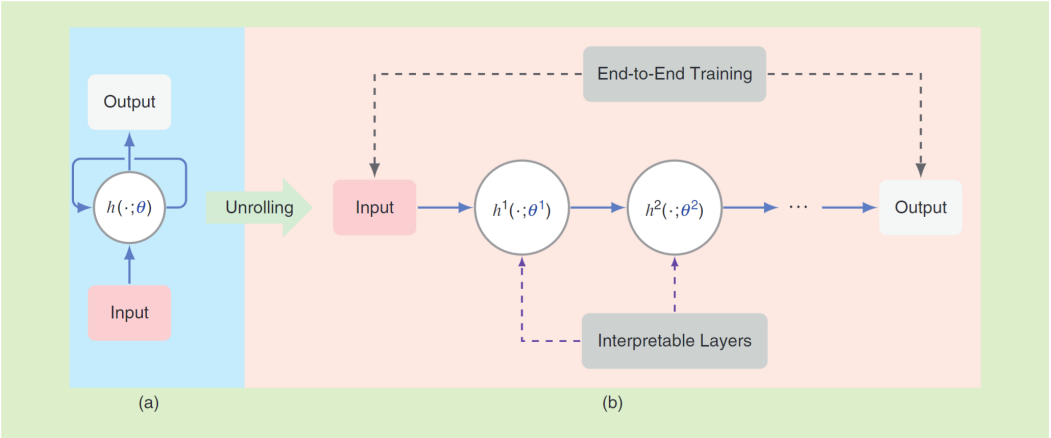



Fig. 2.11 A high-level overview of deep-unfolding algorithm. (a): Traditional iterative algorithm. (b): Unrolled algorithm into a stack of algorithm's iterations, $h$. The algorithm parameters $\theta$ are transferred to learnable network parameters $\theta_1, \theta_2, \ldots$ from training data sets through end-to-end supervised training. Reproduced from [71, Figure 1].

The main idea of deep unfolding is to map each iteration of a traditional algorithm to a layer in a neural network. This process begins with an initial guess and iteratively refines the solution, similar to traditional methods. However, in deep unfolding, the parameters of each iteration (or layer) are learned from data rather than being hand-crafted. Importantly, only certain components of the traditional algorithm are replaced by neural network elements, allowing the model to retain the beneficial properties of the original method while gaining the adaptability and learning capability of NNs. This not only accelerates convergence but also improves the quality of the solution by leveraging NNs. Fig. 2.11 shows an overview of the deep unrolling of an iterative algorithm.

Consider an iterative algorithm such as the iterative shrinkage-thresholding algorithm (ISTA) for solving sparse coding problems. In deep unfolding, ISTA can be unrolled into a neural network where each layer corresponds to an iteration of ISTA. The network is then trained end-to-end using backpropagation to optimize the parameters, including the shrinkage thresholds and the step sizes, which are traditionally fixed.

In applications such as image denoising [72], restoration [73], and super-resolution [74], deep unfolding can effectively integrate denoisers within the iterative framework. Apart from learning the parameters, each iteration is further enhanced by embedding a learnable denoiser, typically an NN, into the process. The denoiser refines the intermediate image estimates by reducing noise or artifacts at each step, thereby improving the quality of the final output. Unrolling such an iterative process into a neural network and training it end-to-end usually leads to better performance in image enhancement tasks as the model learns optimal denoising parameters and strategies from data.

In deep unfolding, the domain knowledge can be incorporated into the learning process. The structure of the iterative algorithm imposes a useful inductive bias, guiding the network towards physically plausible solutions and improving generalization, especially in scenarios with limited training data. Moreover, deep unfolding offers interpretability, making it a less black-box-like approach than conventional DL models. Since each layer of the network corresponds to a specific iteration of the traditional algorithm, the unfolding process can be traced back to its algorithmic origin, providing insights into the network's decision-making process.

Despite its advantages, deep unfolding also faces challenges [71]. One notable issue is the need for expert knowledge to design unrolled networks to avoid highly nonlinear or nonsmooth operators. Additionally, the architecture and the training of a deep unfolding network are often customized to a specific application, which might not be well-suited for other applications.

### 2.2.6 Physics-informed and -guided neural networks

Unlike deep unfolding that includes NNs in traditional iterative methods, physics-informed neural networks (PINNs) [12] and physics-guided neural networks (PGNNs) [13, 14] seek to incorporate physical prior knowledge into NNs. Our use of both PINNs and PGNNs to reconstruct intracardiac vector blood flow is detailed in Chapter 5.

PINNs represent a data-driven optimization framework primarily used to solve partial differential equations (PDEs) that describe various physical phenomena, including fluid dynamics [75], heat transfer [76], and solid mechanics [77]. Essentially, PINNs function as NN-based solvers, typically utilizing simple MLPs. By incorporating known physics into the optimization process, PINNs ensure that NN solutions adhere to the underlying physical laws. This integration is achieved by embedding the PDEs, initial, and boundary conditions directly into the loss function, alongside traditional data-driven loss terms. The partial derivatives are computed using automatic differentiation [78], which is also the technique used to compute the gradients during backpropagation. In PINNs, the learned function mapping by NNs is decomposed into a sequence of elementary arithmetic operations and elementary functions. By applying the chain rule repeatedly to these operators, partial derivatives of arbitrary order can be computed automatically and accurately to working precision. This technique is more precise than traditional numerical methods like finite difference, which can introduce round-off errors due to the discretization process [78].

However, if the initial or boundary conditions change, the optimization process must be repeated to obtain solutions that comply with the new conditions. This limitation can be addressed by using PGNNs. PGNNs refer to the supervised training of NNs using

physics-constrained datasets and are typically employed in scenarios where collecting labeled data that adhere to physical laws is feasible. This approach leverages physical simulations or real-world experiments to generate training data that already incorporate physical constraints. Once trained, a PGNN can directly perform inference on unseen data. Nonetheless, like all supervised learning methods, PGNN approaches rely on large and realistic training datasets to generalize effectively to real-world data [14].

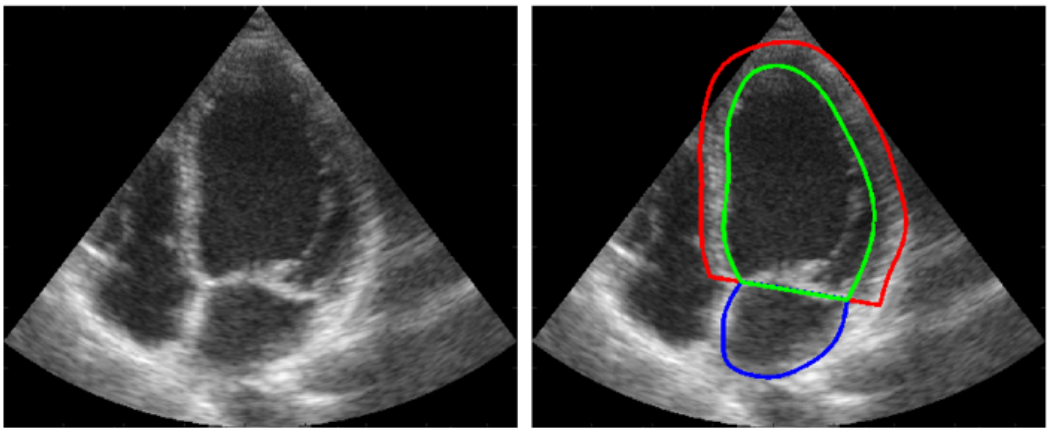
### 2.2.7 Application of artificial intelligence in echocardiography

Here, we describe the application of AI in echocardiography. The automation of the preprocessing steps of *i*VFM involves LV segmentation and color Doppler dealiasing. In this context, we provide a non-exhaustive selection of important and interesting papers focusing on echocardiographic image segmentation (cardiac structure segmentation on B-mode images) and image enhancement (artifact correction on both B-mode and color Doppler images). Although not covered in this work, we also include a section on AI-based pathology diagnosis using B-mode and color Doppler data, which could be insightful for validating *i*VFM methods in the future.

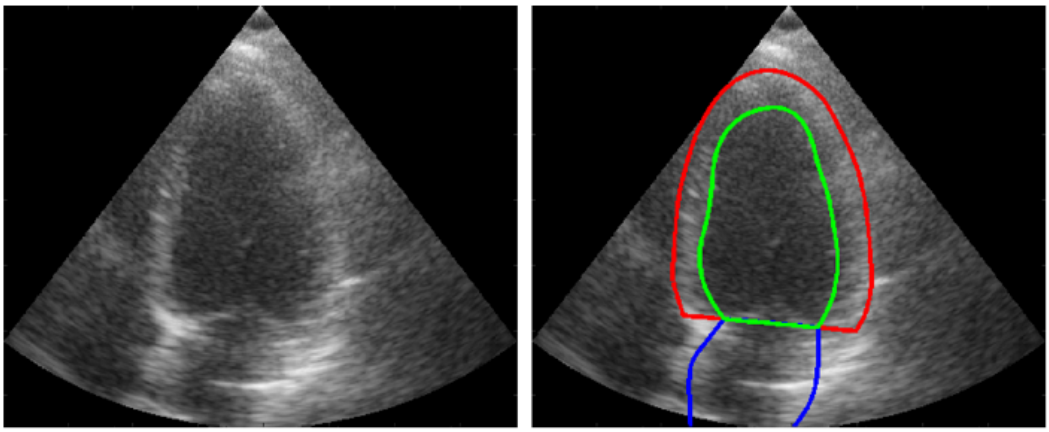
#### 2.2.7.1 Segmentation

Training DL models for echocardiography segmentation was a non-trivial task due to the lack of properly annotated datasets and the highly operator-dependent image quality. In 2019, a turning point was marked with the release of the first large public echocardiographic dataset, namely CAMUS [16]. The CAMUS dataset contains 500 patients, with expert annotation of the LV, LA, and myocardium at the ED and ES instants of the cardiac cycle in both A2C and A4C views. Besides introducing this public dataset, Leclerc et al. achieved competitive results compared to intra-observer variability using different fine-tuned U-Net variants. Examples of the B-mode images and the corresponding expert annotations are provided in Fig. 2.12. In a follow-up work, Leclerc et al. proposed a two-step segmentation method [79], first detecting the region of interest (ROI) and then performing segmentation within the ROIs, leading to improved segmentation results.

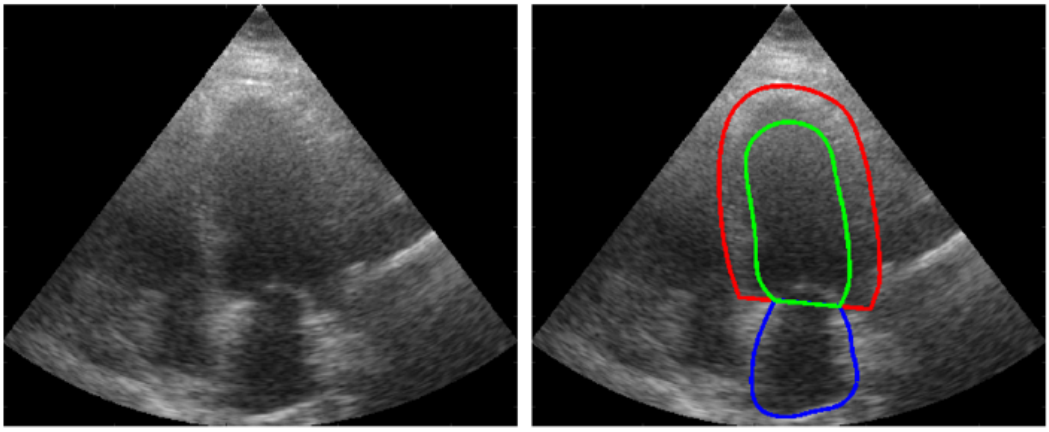
In 2019, another large-scale echocardiographic dataset, EchoNet-Dynamic [80], was published. This dataset contains 10 030 A4C B-mode sequences, with manually segmented left ventricles on the ED and ES instants. The objective of the EchoNet-Dynamic was to segment the LV on A4C sequences and predict the EF. In 2020, Ouyang et al. [81] proposed to use Deeplabv3 to perform segmentation and another CNN-based spatiotemporal encoder to predict the EF of each cineloop. A recent study by Lin et al. [21] introduced SAMUS, which



(a) Good image quality



(b) Medium image quality



(c) Poor image quality

Fig. 2.12 Examples of CAMUS B-mode images (left) and expert annotations (right) of the LV (in red), LA (in green), myocardium (in blue). (a), (b), and (c) represent different image quality, ranging from good to poor. Reproduced from [16, Fig. 1].

fine-tuned SAM (as detailed in Section 2.2.4) using a combination of several public ultrasound datasets, including CAMUS. SAMUS reduces the input size and performs natural to medical image domain adaptation using a position and a feature adapter. An additional CNN branch is used to inject local image features into the image encoder via cross-branch attention. Only point prompts are utilized, with the points being placed randomly on different foreground positions. Their ablation study showed that 5 point prompts yield the best segmentation results. Despite achieving worse performance than specialized models trained exclusively on echocardiographic datasets, SAMUS shows a strong generalization ability.

Despite the great performance produced by most DL models, the predicted segmentations are often subject to anatomical errors. This can be resolved by including prior information during the training of segmentation methods to constrain the network prediction, or through post-processing of the network output. For the former, Oktay et al. [82] used an anatomically constrained neural network to ensure that the segmentation output closely respects the non-linear embedding of the underlying anatomy, derived from an autoencoder network. For the latter, Painchaud et al. [83] opted for a constrained variational autoencoder (VAE) [84] to learn a representation of valid cardiac shapes through a smooth and constrained latent space, which can be used to correct anatomically incorrect segmentations produced by any model.

While most works seek to perform 2D segmentation on echocardiographic images, it is important to note that echocardiographic acquisition provides a cineloop of images. Few studies have leveraged the temporal information of these cineloops. Wei et al. introduced CLAS, a segmentation network based on 3D U-Net that aims to produce temporally consistent segmentations using only 2D ED and ES annotations. This is done by predicting the frame-to-frame deformation to correctly propagate the annotations from the ED frame to the ES frame. These pseudo annotations are used to guide the network to segment frames between ED and ES in a completely self-supervised manner. In a follow-up work, they proposed MCLAS [85], a multitask learning model by adding an acquisition view classification and EF regression branches. This approach further improves the temporal consistency of the predicted segmentations from ED to ES, producing smaller errors in both geometric and clinical metrics than the intra-observer variability.

Another technique proposed by Smistad et al. [86] and Hu et al. [87] uses convolutional long short-term memory (LSTM) modules [88], relying on annotations over the entire cardiac cycle rather than only ED and ES. During training and inference, the cineloop is passed to the model, where the LSTM modules keep track of the information from previously segmented frames, reducing temporal errors. A temporal loss is also included in the training loss function to constrain the maximum displacement between consecutive segmentations. Painchaud et al. [15] proposed CASTOR, a post-processing algorithm based on the Attribute-Regularized

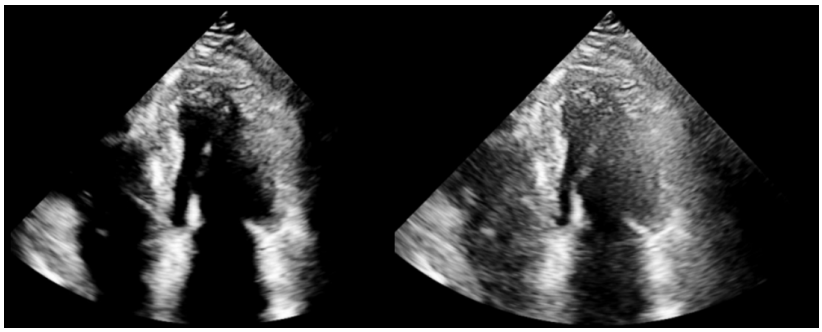


Fig. 2.13 Example of hazy (right) and dehazed (left) cardiac B-mode image by diffusion models. Reproduced from [91, Fig. 15].

VAE [89]. The main idea is to learn to correlate certain dimensions in the latent space to physiological information, such as the LV area, myocardial area, etc. Segmentations are encoded frame-by-frame into the latent space, where smoothing the correlated dimensions corrects temporal inconsistencies. Finally, Deng et al. [22] extended SAMUS [21] described earlier for echocardiographic cineloop segmentation by adding a space-time memory network. Based on Gated Recurrent Units (GRUs) [90], the memory network creates the memory prompt that contains both spatial and temporal information from the previous frames to prompt the segmentation of the current frame. Only the segmentation of the first frame requires a single-point prompt on the foreground regions, while the segmentation of the subsequent frames uses memory prompts. Moreover, a memory reinforcement module is included to combine the output probability map and the image embedding to improve the memory quality, making the model focus its attention on the left ventricular area and become more robust to speckle noise and artifacts.

#### 2.2.7.2 Image enhancement

Image enhancement in the context of echocardiography involves improving the visual quality of ultrasound images, which can be crucial for accurate diagnosis and analysis. Echocardiographic images often suffer from noise, low contrast, and various artifacts due to the inherent limitations of ultrasound imaging techniques and the variability in patient anatomy and scanning conditions. These challenges can be addressed either by improving the front-end acquisition or by post-processing after the image formation. Only AI-based post-processing techniques are discussed here, as access to front-end acquisition modules from clinical echocardiography devices is limited.

Stevens et al. [91] leveraged score-based diffusion models [92, 93] to effectively remove haze and noise artifacts in cardiac B-mode images due to clutter. This is achieved by a

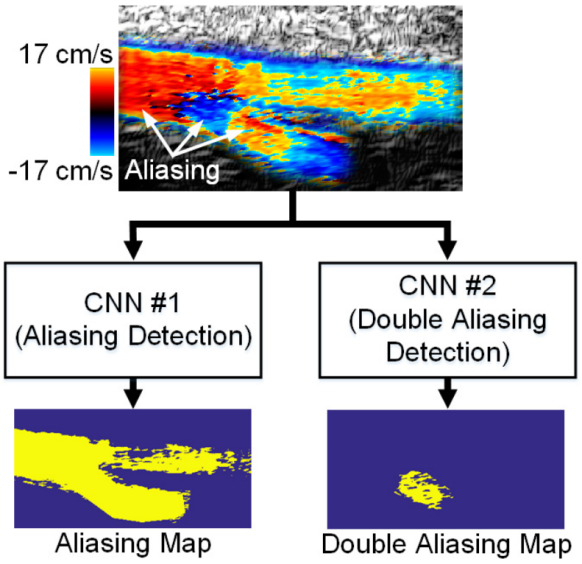



Fig. 2.14 femoral bifurcation CDI using two U-Nets. Reproduced from [94, Fig. 1].

joint posterior sampling framework that combines two separate diffusion models to model the distribution of both haze-free ultrasound and haze in an unsupervised manner. Their experiments showed that training using RF signals as input is more beneficial than using image input data. Fig. 2.13 shows an example of an A4C image before and after dehazing.

Like B-mode imaging, color Doppler imaging is also prone to artifacts, with aliasing being the main one. Nahas et al. [94] proposed a U-Net-based dealiasing method to resolve double aliasing artifacts in femoral bifurcation CDI. Their method involves training two U-Nets: the first one is trained to detect single aliasing, and the second one is trained to identify double-aliased regions. An example of the segmentation of single- and double-aliased pixels is illustrated in Fig. 2.14. Although this approach has not yet been applied to echocardiography, it appears to be easily transferable to cardiac imaging.

#### 2.2.7.3 Pathology diagnosis

This thesis is part of a long journey to automatically extract quantitative flow-based biomarkers from echocardiography, with the goal of enabling the early detection of cardiovascular diseases like hypertension. To validate *i*VFM approaches clinically, we conducted a preliminary study using DL methods trained on a small dataset to classify healthy and hypertensive patients directly from color Doppler images. More details of this study are given in the "Clinical validation of intraventricular vector flow mapping" section in "Perspectives for Future Work" of Part IV. Although the results from this limited sample size are not exceptional, they are encouraging and suggest that performance may be improved by incorporating more training

data and additional information such as vector blood flow fields reconstructed by *i*VFM and vorticity.

To further inspire future research, we present a summary of existing DL-based methods for pathology detection using cardiac B-mode and color Doppler images below. These methods can generally be categorized into two approaches: post-processing the segmentation results from DL models to compute clinical metrics, or directly classifying pathologies from the input images. The latter approach, while potentially faster, often lacks the clinical interpretability provided by the former method.

Ghorbani et al. [95] utilized CNNs to identify local cardiac structures (such as the presence of intracardiac devices like pacemakers, severe left atrial dilation, and left ventricular hypertrophy) and estimate cardiac function (ESV, EDV, and EF of the LV) using B-mode images. Despite the good performance of their model, the results are not yet interpretable, even with the use of attention maps, which highlight the regions the models focus on for making decisions.

Zhang et al. [96] proposed two frameworks for processing B-mode images, with the first focusing on segmenting cardiac structures and the second dedicated to classifying three cardiac diseases: hypertrophic cardiomyopathy, cardiac amyloid, and pulmonary arterial hypertension. Their automated computation of clinical metrics based on the segmentations highly correlates with those provided by commercial software. Instead of training a single model to detect all three diseases, they trained separate models for each disease, achieving a high area ($> 0.85$) under the receiver operating characteristic curve (AUROC).

Unlike the previous methods that use only B-mode images, Yang et al. [97] introduced a three-step framework to grade the severity of mitral stenosis, mitral regurgitation, aortic stenosis, and aortic regurgitation using a combination of input images (B-mode, color Doppler, and PW Doppler acquisitions in different views). This proposed framework mimics the normal clinical workflow for diagnosis through multitask learning. To detect regurgitation, the color Doppler images are passed to an encoder for feature extraction. The extracted features are then used to classify the presence and type of regurgitation, identify the frame showing the regurgitation, and segment the regurgitation area. Clinical metrics are derived from these predictions to grade the regurgitation severity. For stenosis detection, two separate encoders extract features from A4C and PLAX cineloops. These features are concatenated before being used to detect the presence of stenosis. If stenosis is found, the associated A4C PW Doppler acquisition is used to quantify the lesion severity through triangular wave segmentation, contour extraction, and key point detection.

## 2.3 Conclusion

In this chapter, we presented the state-of-the-art (SOTA) methods for intraventricular vector flow imaging and their limitations, with a focus on the *i*VFM approach, which is directly applicable to clinical color Doppler sequences produced by conventional echocardiography machines. Among all *i*VFM techniques, *i*VFM-v3 stood out by imposing physical constraints, specifically the mass conservation equation and free-slip boundary conditions, during the vector flow reconstruction. Consequently, we extended *i*VFM-v3 in Chapter 5 by leveraging physics-based neural networks, specifically PINNs and PGNNs.

Like most *i*VFM techniques, *i*VFM-v3 requires time-consuming manual or semi-automatic preprocessing steps, such as LV segmentation on B-mode images and dealiasing of color Doppler images. Therefore, we also discussed the fundamental principles of some common NN architectures and their application in echocardiography analysis. The promising results of the nnU-Net framework make it a solid candidate, which we adapted for tackling various problems in the following chapters: temporally consistent LV segmentation (Chapter 3), dealiasing of color Doppler echocardiography (Chapter 4), and vector flow mapping (Chapter 5). The recently proposed MemSAM shows great generalization and temporally consistent segmentation performance with the incorporation of a space-time memory network. However, this method was proposed only after the development of our own segmentation model.

# Part III

# Contributions

# Chapter 3

# Temporally Consistent Echocardiographic Left Ventricular Segmentation

As mentioned in the previous chapter, *i*VFM is the only method currently applicable in clinical settings for reconstructing intracardiac blood flow. A crucial step in this process is defining the boundary condition, which requires segmenting the LV cavity throughout the cardiac cycle. Better delineation of the ROI can lead to more accurate boundary conditions and, consequently, more precise vector flow reconstructions.

Most methods presented in Section 2.1.3 rely on manual segmentation, which can be time-consuming and inconsistent. Therefore, the goal of our first contribution is to develop a fully automated tool for temporally consistent (2D+t) segmentation of cardiac B-mode cineloops. It is important to note that *i*VFM is usually performed on A3C acquisitions, where we can observe blood entering and exiting the LV. Unlike A2C and A4C data, for which public datasets exist, there is no public A3C dataset available. Moreover, models trained only on A2C and A4C data did not work on A3C data due to the LV shape difference in this view. Therefore, it was necessary to annotate our own A3C dataset. Before that, we conducted a feasibility study of our data annotation pipeline on A2C and A4C data, where public datasets are available, and investigated the potential benefits brought by 2D+t segmentation compared with pure 2D segmentation.

To establish an in-house A2C and A4C 2D+t large dataset, we leveraged CASTOR [15] and existing public echocardiographic datasets, such as CAMUS [16]. To maintain a simple network architecture, we investigated the extra components and training strategies needed to improve the performance of a simple 2D U-Net architecture to lie within intra-observer variability. With the optimal architecture and training strategies, we performed 2D frame-by-frame segmentation on unannotated in-house echocardiographic sequences, namely CARDINAL. The temporal errors were then corrected automatically using CASTOR

and verified by a human expert. Any remaining segmentation errors or inconsistencies were corrected manually. The post-processed annotations served as the gold standard for training our 2D+t segmentation algorithm.

Our 2D+t model, based on 3D nnU-Net, takes a sequence of consecutive B-mode images as input and considers the temporal dimension as the third spatial dimension. This approach is justified by the relatively smooth LV movement over the cardiac cycle.

We evaluated the segmentation performance of our 2D+t segmentation algorithm using both geometric and clinical metrics. Our proposed method produced precise yet temporally consistent segmentation, achieving lower errors than intra-observer variability when evaluated on a hold-out CARDINAL test set. Regarding generalizability, the results of our proposed method were comparable to intra-observer variability and most SOTA methods when tested on the CAMUS dataset.

Finally, we adapted this annotation pipeline for annotating our in-house unlabeled A3C dataset, as CASTOR, originally made for post-processing A2C/A4C segmentations, could not be applied to A3C data with a different LV shape. Despite having slightly worse evaluation scores than models implemented on CARDINAL, the performance of the 3D nnU-Net trained with limited A3C samples and with transfer learning from the CARDINAL dataset was sufficient to accurately extract boundary conditions required for *i*VFM.

## Remarks

The work presented in this chapter was published in the proceedings of the 2023 Functional Imaging and Modeling of the Heart (FIMH) conference. The 2D segmentation algorithm involved in the data annotation pipeline was presented at the 2022 IEEE International Ultrasonics Symposium (IUS) conference. Full references for both papers are provided below:

H. J. Ling, D. Garcia, and O. Bernard, “Reaching intra-observer variability in 2-D echocardiographic image segmentation with a simple U-Net architecture,” in IEEE International Ultrasonics Symposium (IUS), 2022 [17]

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 (FIMH), 2023, pp. 245–254 [18].

## 3.1 Introduction

Echocardiographic imaging has undergone significant advances in recent years, particularly due to the integration of DL methodologies. A primary focus has been the automated extraction of clinical indices through the segmentation of cardiac structures, which has led to substantial breakthroughs in the field. Key to these advancements has been the availability of open-access annotated datasets, such as CETUS (45 patients, 3D images annotated at ED and ES) [98], CAMUS (500 patients, 2D images annotated at ED and ES in A2C and A4C views) [16], EchoNet-Dynamic (10 036 patients, 2D sub-sampled images annotated at ED and ES in A4C view) [81], HMC-QU (109 patients, 2D sequences annotated in A4C view) [99], and TED (98 patients from the CAMUS dataset, 2D sequences annotated in A4C view) [15].

These datasets have facilitated effective and fair comparisons of various methods, including both generic image segmentation models [16, 17] and those specifically tailored for echocardiographic images [100, 85, 101]. The performance of current SOTA methods on the CAMUS dataset has confirmed the superiority of DL-based approaches, which have matched inter- and intra-observer variability in most geometric (Dice score, Hausdorff distance—HD, mean absolute distance—MAD) and clinical metrics (EF, EDV, and ESV).

While these results are highly promising and represent a significant step towards the automation of echocardiographic image analysis, they are not yet sufficient to warrant full confidence in automated methods for clinical use. Two critical challenges remain underexplored in the field: i) ensuring the frame-by-frame temporal consistency of the predictions, and ii) generalizing the methods across different datasets. To address these challenges, we proposed the following contributions:

1. We investigated the performance of two generic architectures based on common temporal data processing techniques on 2D echocardiography sequences and compared them to current SOTA methods in the field.

2. We introduced a new private dataset called CARDINAL (240 patients, 2D sequences annotated in A4C and A2C views) and assess the impact of training our methods exclusively on CARDINAL while testing on the CAMUS dataset.

## 3.2 Reaching intra-observer variability in 2D echocardiographic image segmentation

As introduced in Section 2.2.2, nnU-Net is a self-configuring segmentation framework built on a U-Net architecture, enforcing best practices such as pre- and post-processing, data

augmentations, and a patch-wise approach tailored to the type of input data [9]. nnU-Net is capable of adapting to various biomedical datasets without manual tuning. Hence, we conducted an extensive study to determine whether a simple U-Net architecture could achieve segmentation performance comparable to intra-observer variability, guiding us to the optimal 2D U-Net architecture for the data annotation pipeline.

### 3.2.1 Proposed methods

We proposed four models trained on the CAMUS dataset, progressively incorporating the core elements of 2D nnU-Net (53M parameters) into a simple 2D U-Net (Model #1): deep supervision (Model #2), data augmentation in training (Model #3), test-time augmentation (TTA), and a patch-wise approach (Model #4).

The data augmentations included rotations, scaling, Gaussian noise, etc. TTA involved performing successive inferences on augmented input data with various flips along different axes, followed by probability averaging to obtain the final segmentation. In the patch-wise approach, a patch size of $1024 \times 640$ pixels (median image shape) was used during training, while sliding windows with Gaussian weighting aggregated the predicted patches during inference. For Models #1 and #2, the input images were resized to $256 \times 256$ pixels. The number of starting filters was 16 for Models #1 to 4 and 32 for 2D nnU-Net, with the maximum number of filters capped at 480. Models #1 to 3 had five downsampling layers, while Model #4 had six, and 2D nnU-Net had eight downsampling layers.

To enhance the generalization ability of the models, we used a specific optimization scheme: a small batch size of 2 and a reduced number of iterations of 250 per epoch. The geometric metrics used were HD and ASSD. For clinical evaluation, correlations (Corr) and mean absolute errors (MAE) of the EDV, ESV, and EF were computed. Models #3 and #4 were evaluated both with and without TTA.

### 3.2.2 Results and discussions

Table 3.1 presents the 10-fold cross-validation results of Models #1 to 4 and nnU-Net versus inter-observer variability. In terms of segmentation performance (2nd column of Table 3.1), starting with Model #3 + TTA, the models significantly outperformed CLAS and intra-observer variability (p-value $< 0.05$). Consistent with clinical scores, the estimated volumes were highly correlated with the ground truths (Corr $> 0.97$), although they were less accurate than CLAS for estimating the ejection fraction due to the lack of temporal consistency. Our study indicates that data augmentation during training and TTA, combined with a well-matched optimization scheme, are crucial to achieving intra-observer variability. Consequently, a

Table 3.1 10-fold cross-validation results of Model #1 to 4 and nnU-Net versus inter-observer variability on CAMUS dataset.

| Methods (# param.) | $LV_{endo}$ & $LV_{epi}$ | | | EDV | | ESV | | EF | |
|---|---|---|---|---|---|---|---|---|---|
| | HD [mm] (↓) ($\bar{x} \pm \sigma$) | ASSD [mm](↓) ($\bar{x} \pm \sigma$) | Outliers [%](↓) ($\bar{x} \pm \sigma$) | Corr (↑) | MAE [mL](↓) ($\bar{x} \pm \sigma$) | Corr (↑) | MAE [mL](↓) ($\bar{x} \pm \sigma$) | Corr (↑) | MAE [%](↓) ($\bar{x} \pm \sigma$) |
| Intra-observer | 4.7 ± 2.0 | 1.5 ± 0.7 | - | 0.98 | 6.5 ± 4.4 | 0.98 | 4.5 ± 3.9 | 0.90 | 4.7 ± 4.1 |
| CLAS [100] | 4.8 | 1.5 | - | 0.96 | - | 0.98 | – | 0.93 | - |
| Model # 1 (2M) | 5.4 ± 3.2 (n.s.) | 1.6 ± 0.9 (n.s.) | 4.3 | 0.96 | 8.0 ± 7.3 | 0.96 | 6.6 ± 6.0 | 0.84 | 5.1 ± 4.3 |
| Model # 2 (2M) | 5.3 ± 3.1 (n.s.) | 1.6 ± 0.9 (n.s.) | 4.3 | 0.97 | 8.0 ± 7.4 | 0.97 | 6.3 ± 5.6 | 0.85 | 5.0 ± 4.7 |
| Model # 3 (7M) | 4.8 ± 2.5 (n.s.) | 1.4 ± 0.7 (**) | 2.7 | 0.97 | 7.2 ± 5.9 | 0.97 | 5.7 ± 5.9 | 0.85 | 4.9 ± 4.2 |
| Model # 3 + TTA (7M) | 4.5 ± 2.1 (**) | 1.4 ± 0.7 (****) | 1.8 | 0.97 | 6.8 ± 6.1 | 0.97 | 5.6 ± 4.8 | 0.86 | 4.6 ± 4.0 |
| Model # 4 (30M) | 4.5 ± 1.9 (*) | 1.4 ± 0.7 (**) | 1.5 | 0.97 | 6.7 ± 5.9 | 0.97 | 5.5 ± 5.1 | 0.84 | 4.9 ± 4.3 |
| Model # 4 + TTA (30M) | 4.4 ± 1.9 (***) | 1.4 ± 0.7 (****) | 1.5 | 0.97 | 6.6 ± 5.7 | 0.97 | 5.5 ± 4.8 | 0.84 | 4.7 ± 4.4 |
| 2D nnU-Net (53M) | 4.3 ± 1.9 (****) | 1.3 ± 0.7 (****) | 1.5 | 0.98 | 6.5 ± 5.6 | 0.98 | 5.3 ± 4.6 | 0.88 | 4.4 ± 3.6 |

# param. stands for the number of parameters.
The statistical test was the left-tailed two-sample t-test between each model and intra-observer variability.
(*): p-value < 0.05; (**): p-value < 0.01; (***): p-value < 0.001; (****): p-value < 0.0001.

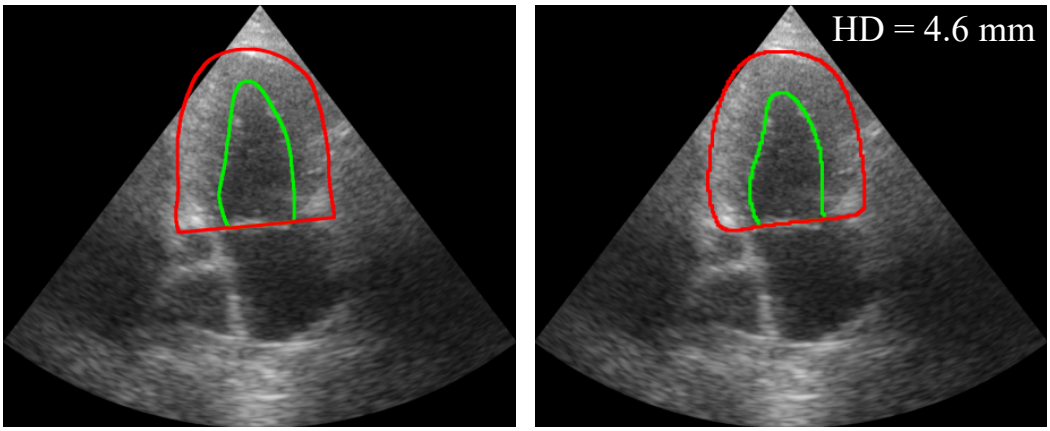
simple U-Net architecture (7M parameters) can yield high-quality segmentation and accurate volume estimation. Fig. 3.1 shows an example of the segmentation done by Model #3 + TTA versus ground truth annotation.

With 53M parameters, 2D nnU-Net demonstrated the best segmentation results, falling within the intra-observer variability for all metrics except for the EF estimation. Hence, 2D nnU-Net was chosen as the optimal 2D segmentation architecture to generate gold standard annotations for the CARDINAL dataset.

## 3.3 Benchmarked methods for echocardiographic sequence segmentation

The CAMUS dataset served as the primary evaluation platform for effectively comparing the performance of segmentation methods[1], with the platform operational until January 2023. Therefore, we relied on this dataset to select the methods retained in this study.

[1]https://www.creatis.insa-lyon.fr/Challenge/camus/results.html




(a) Ground truth annotation (b) Segmentation by Model #3 + TTA

Fig. 3.1 Example of segmentation contour predicted by Model #3 + TTA versus ground truth annotation on a CAMUS data, with a mean HD of 4.6 mm for LV endocardial (green) and epicardial wall (red).

### 3.3.1 2D DL methods

As presented in Section 3.2.2, 2D nnU-Net is the best-performing segmentation model on the CAMUS dataset, sitting at the top of the CAMUS challenge leaderboard. As a result, it was used as the baseline for comparison. It is important to note that we did not perform any model ensembling with nnU-Net.

Another 2D method used for comparison is GUDU, developed by Sfakianakis et al. [101]. GUDU is based on three key aspects: i) tailored data augmentations specific to ultrasound acquisition, including variations in contrast between the myocardial tissue and the LV cavity, ii) random rotations from the origin of the sectorial shape to mimic different probe positioning, and iii) perspective transformations to simulate probe twisting. Inspired by ensemble methods, the authors trained five U-Nets with different architectures and averaged their outputs during inference to compute the final prediction. They also proposed a novel loss function that considers the relative positions of cardiac structures with respect to each other. The authors demonstrated the effectiveness and complementarity of each contribution in an ablation study.

### 3.3.2 2D+t DL methods

By their very nature, echocardiographic sequences exhibit regular properties along the time axis. Therefore, it seems logical to consider 2D ultrasound sequences as complete volumes containing coherent 3D shapes and to extract 3D features using 3D convolutional layers to promote temporal consistency. Thus, we proposed to train a 3D nnU-Net to segment the

complete cardiac sequences in a single run. We hypothesized that this model will inherently learn temporal consistency while maintaining a high level of segmentation accuracy.

We compared the 3D nnU-Net to U-Net LSTM [86, 87], a different approach where LSTM blocks are added to the encoder of U-Net, treating the input as sequential data. Instead of processing a single frame, U-Net LSTM takes a series of frames as input and stores the extracted features over time to produce the final segmentation of the entire sequence. Such a strategy can reduce segmentation shifts from one frame to another.

## 3.4 Experimental setup

### 3.4.1 CARDINAL dataset

#### 3.4.1.1 Acquisition protocol

The proposed dataset consists of clinical examinations of 240 patients, acquired at the University Hospital of Lyon (Croix-Rousse Lyon Sud, France) under the regulation of the local ethics committee of the hospital. The complete dataset was acquired with GE ultrasound scanners. For each patient, 2D A4C and A2C view sequences were exported from the EchoPAC analysis software. Each exported sequence corresponds to a set of B-mode images expressed in polar coordinates. The same interpolation procedure as used for the CAMUS dataset was applied to express all sequences in Cartesian coordinates with a single grid resolution of 0.31 $mm^2$. Each sequence in the CARDINAL dataset corresponds to a complete cardiac cycle defined as the interval between peaks of maximal LV cavity surface area. This resulted in 480 full cardiac cycle sequences. Given the high number of frames and sequences, no reference annotation was available.

#### 3.4.1.2 Data annotation pipeline

To tackle the total number of frames to be annotated, we proposed the following data annotation pipeline. First, we trained two 2D nnU-Net on the CAMUS dataset with 1800 2D echocardiographic images with expert annotation, to learn the best possible models for segmenting A2C and A4C sequences. These models were then used to predict 2D frame-by-frame segmentations on the unannotated CARDINAL data. Afterward, we used CASTOR to correct the temporal and anatomical artifacts present in the predictions. The resulting segmentations were validated by an expert, rather than annotating the whole dataset from scratch. Once validated, these segmentations served as the gold-standard annotations to train the 3D nnU-Net and U-Net LSTM.

Table 3.2 Implementation details of the three methods for echocardiographic sequence segmentation.

| Configurations | 3D nnU-Net | 2D nnU-Net | U-Net LSTM |
|---|---|---|---|
| Patch size (pixels) | $320 \times 256 \times 24$ | $640 \times 512$ | $256 \times 256 \times 24$ |
| Batch size | 2 | 2 | 2 |
| Nb. feature maps | $32 \downarrow 480 \uparrow 32$ | $32 \downarrow 480 \uparrow 32$ | $32 \downarrow 480 \uparrow 32$ |
| Lowest resolution[a] | $10 \times 8 \times 6$ | $5 \times 4$ | $8 \times 8 \times 24$ |
| Downsampling scheme | Stride pooling | | |
| Upsampling scheme | Transposed convolution | | |
| Normalization scheme | Instance normalization | | |
| Optimization scheme[b] | SGD + 0.01 + polynomial decay | | |
| Loss function | Cross entropy + Dice | | |
| Number of parameters | 41.3 M | 30.4 M | 49 M |
| Training duration (hours)[c] | 22.8 | 8 | 69.5 |

[a] Size of the lowest resolution of feature maps in pixels.
[b] Optimizer + initial learning rate + learning rate scheduler used.
[c] Number of hours required to train each model for 1000 epochs.
The configurations shared between models are only shown once in their respective rows.

### 3.4.2 Implemented DL methods

For a fair comparison, we implemented the 2D nnU-Net, U-Net LSTM, and 3D nnU-Net described in Section 3.3 using the same Python library called ASCENT[2]. These models shared the following training hyperparameters: batch size of 2, SGD optimizer with a learning rate of 0.01 coupled with a polynomial decay scheduler, and 1000 training epochs. The 2D and 3D nnU-Net used a patch-wise approach to avoid resizing the images, thus preserving the native image resolution. To train the U-Net LSTM, the input images were resized to $256 \times 256$ pixels and 24 consecutive frames were randomly selected and fed to the model to produce the corresponding segmentations. For inference, the sliding window approach with a Gaussian importance map along with TTA was used. The prediction was given by the average of the *softmax* probabilities of all windows. More implementation details for each model can be found in Table 3.2.

[2]https://github.com/creatis-myriad/ASCENT

## 3.5 Results and discussions

We evaluated the methods described in Section 3.3 using three types of measures to get a complete picture of their performance in terms of segmentation accuracy (Table 3.3), extraction of clinical indices (Table 3.4) and temporal consistency (Table 3.5). In each of these tables, we group the methods according to the datasets on which they were trained and tested (CARDINAL is abbreviated as CL, and CAMUS is abbreviated as CS) to make it easier to observe the change in performance when generalizing to a new dataset.

### 3.5.1 Geometric and clinical accuracy

Table 3.3 shows the segmentation accuracy computed from the CARDINAL and CAMUS datasets for the five algorithms described in Section 3.3. The values in bold correspond to the best scores for each metric for a given training/test dataset setup. From the results on the CARDINAL dataset (CL/CL case), we can see that the 3D nnU-Net has the best segmentation scores for all metrics, for both ED and ES. It is also interesting to note that the two temporal consistency methods (3D nnU-Net and U-Net LSTM) produce better results than the 2D nnU-Net method. This can be explained by the fact that the reference segmentation has regular properties along the temporal axis due to the annotation process. Methods that integrate the temporal dimension into their architecture are therefore more likely to produce segmentation results that are closer to the manual references.

It is worth mentioning that methods trained and tested on the same dataset (Sections CL/CL and CS/CS in Table 3.3) get overall better results up to 1.7× for the HD and MAD metrics. One reason for such an improvement is the larger amount of annotated training images for CARDINAL (18 793 images from 190 training/validation patients, reference frames for the full cardiac cycle in A2C and A4C views) than for CAMUS (1800 images from 450 training/validation patients, reference frames at ED and ES in A2C and A4C views).

Table 3.4 reports the clinical metrics for the 5 methods. As in Table 3.3, the methods enforcing temporal consistency get the best results on CARDINAL, especially for the EF for which temporal consistency is essential (mean correlation score of 0.917). Furthermore, the best models trained on CARDINAL or CAMUS produce similar results for volume estimation (average correlation of 0.978), revealing a limit reached by these approaches, certainly due to the resolution of the imaging systems.

Table 3.3 LV endocardial segmentation accuracy of the benchmarked methods, on different subsets of frames.

| Methods | Train/test | Dice (↑) | | | HD [mm](↓) | | | MAD [mm](↓) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | All | ED | ES | All | ED | ES | All | ED | ES |
| 3D nnU-Net | CL/CL | **.969** | **.968** | **.960** | **2.3** | **2.7** | **2.5** | **0.7** | **0.8** | **0.7** |
| 2D nnU-Net | | .957 | .961 | .942 | 2.9 | 3.1 | 3.1 | 0.9 | 1.0 | 1.1 |
| U-Net LSTM | | .964 | .964 | .956 | 2.5 | 2.8 | 2.6 | 0.8 | 0.9 | 0.8 |
| 3D nnU-Net | CL/CS | - | **.939** | **.926** | - | 5.2 | **4.6** | - | **1.6** | **1.5** |
| 2D nnU-Net | | - | .934 | .921 | - | **4.9** | **4.6** | - | 1.8 | 1.6 |
| U-Net LSTM | | - | .925 | .903 | - | 6.0 | 5.8 | - | 2.1 | 2.1 |
| 2D nnU-Net | CS/CS | - | **.952** | **.935** | - | **4.3** | **4.2** | - | **1.3** | **1.3** |
| CLAS | | - | .947 | .929 | - | 4.6 | 4.6 | - | 1.4 | 1.4 |
| GUDU | | - | .946 | .929 | - | 4.7 | 4.7 | - | 1.4 | 1.4 |

CL:CARDINAL; CS:CAMUS.
Columns All, ED, and ES indicate results averaged over all frames, only ED frames, and only ES frames, respectively. Since CAMUS only provides annotation for ED/ES frames, results over all frames are not available when testing on it.

Table 3.4 Clinical metrics of the benchmarked methods.

| Methods | Train/test | EF | | EDV | | ESV | |
|---|---|---|---|---|---|---|---|
| | | Corr (↑) | MAE [%](↓) | Corr (↑) | MAE [mL](↓) | Corr (↑) | MAE [mL](↓) |
| 3D nnU-Net | CL/CL | .913 | 2.9 | **.978** | **3.3** | **.974** | **2.7** |
| 2D nnU-Net | | .850 | 3.8 | .967 | 4.4 | .957 | 3.2 |
| U-Net LSTM | | **.922** | **2.7** | .973 | 3.4 | .969 | 2.8 |
| 3D nnU-Net | CL/CS | **.869** | **5.3** | **.974** | **9.6** | **.976** | **4.9** |
| 2D nnU-Net | | .810 | 7.0 | .970 | 12.8 | .959 | 6.2 |
| U-Net LSTM | | .822 | 11.1 | .879 | 15.9 | .903 | 8.2 |
| 2D nnU-Net | CS/CS | .857 | 4.7 | **.977** | **5.9** | **.987** | **4.0** |
| CLAS | | **.926** | **4.0** | .958 | 7.7 | .979 | 4.4 |
| GUDU | | .897 | **4.0** | **.977** | 6.7 | .981 | 4.6 |

Corr: Correlation between the EF/EDV/ESV derived from the predicted/reference segmentation.
MAE: Mean Absolute Error between the EF/EDV/ESV derived from predicted/reference segmentation.
EDV and ESV were computed from both the predicted and reference masks using Simpson's biplane method.

Table 3.5 Temporal consistency of the benchmarked methods, as defined in [15].

| Methods | Train/test | Nb of seq. w/ err.[a] | % of frames w/ err.[b] | Err. to thresh. ratio[c] |
|---|---|---|---|---|
| 3D nnU-Net | CL/CL | **4** | **4** | **.045** |
| 2D nnU-Net | | 100 | 30 | .210 |
| U-Net LSTM | | 98 | 13 | .110 |
| 3D nnU-Net | CL/CS | **28** | **12** | **.095** |
| 2D nnU-Net | | 85 | 21 | .162 |
| U-Net LSTM | | 83 | 16 | .114 |

[a] Number of sequences (out of the 100 testing sequences) where at least one frame is temporally inconsistent.
[b] Percentage of frames that are inconsistent in the sequences with at least one temporally inconsistent frame.
[c] Average ratio between the measure used to identify temporal inconsistencies and the threshold for temporal inconsistencies. A lower value indicates "smoother" temporal segmentations.

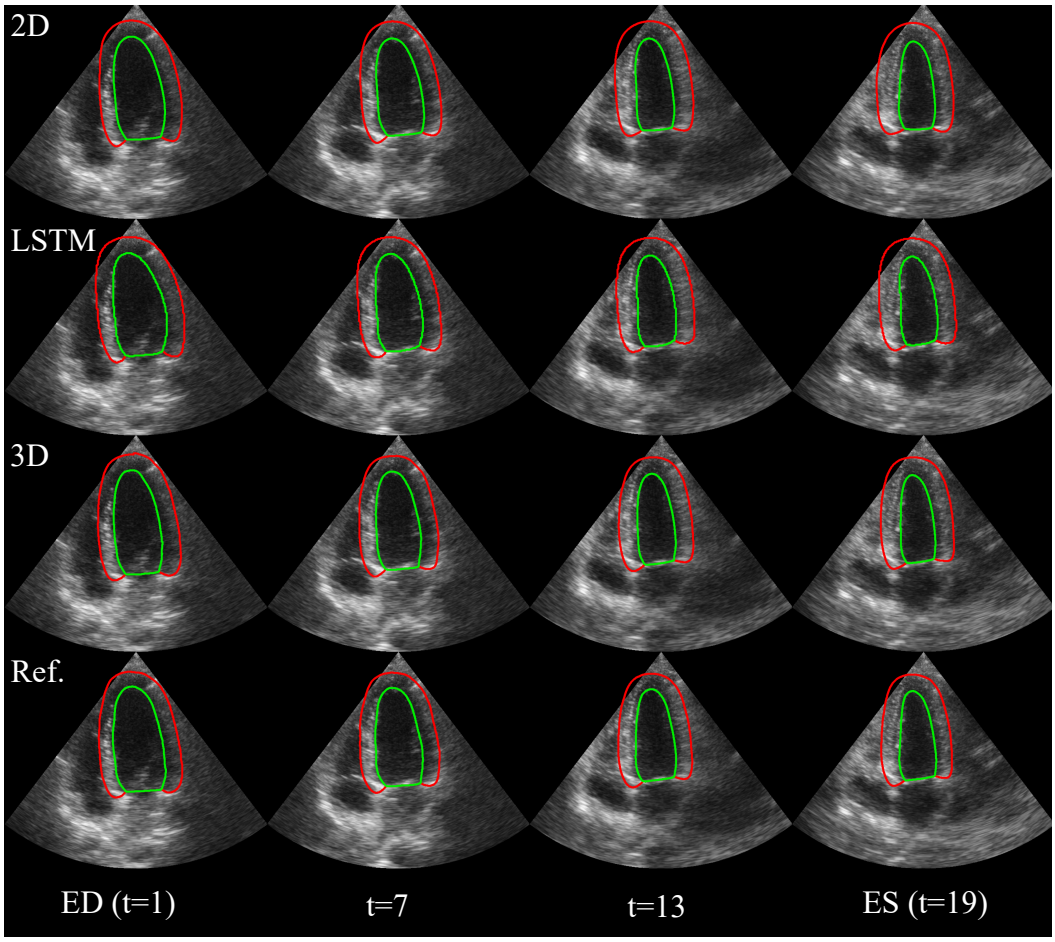
### 3.5.2 Integration of temporal consistency

Table 3.5 allows a better investigation of the temporal performance of the methods by providing additional information on the number/percentage of frames considered temporally inconsistent w.r.t. their neighboring frames. As expected, the methods incorporating temporal persistence produced fewer temporal errors. Looking at the number of sequences with at least one temporally inconsistent frame, the 3D nnU-Net outperforms U-Net LSTM, with only 4 inconsistent sequences over 100 compared to 98 sequences for U-Net LSTM. This result illustrates the greater ability of features computed from 3D convolutional layers to extract relevant spatio-temporal information. The few remaining temporal errors for the 3D nnU-Net are more an indication that the metrics we used are (overly) strict on temporal smoothness.

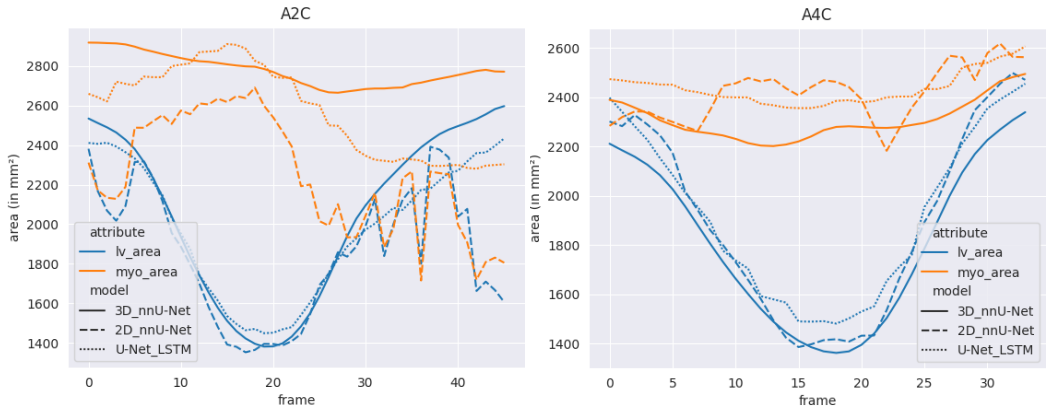
Indeed, the 3D nnU-Net temporal "inconsistencies" appear invisible to the expert eye. As a qualitative evaluation, Fig. 3.2 illustrates in detail the temporal consistency of our methods on one patient from the CARDINAL test set.

### 3.5.3 Generalization across datasets

The ability to generalize across datasets is crucially important to gauge the capacity of a method to properly analyze data affected by a distributional shift. To this end, the models trained on CARDINAL were also evaluated on the CAMUS test set without any fine-tuning. The results are reported in the "CL/CS" sections of Tables 3.3 to 3.5. Among the methods evaluated, 3D nnU-Net is the undisputed best. It even produces competitive geometric and clinical scores compared with SOTA methods trained directly on CAMUS. Thanks to the integration of temporal consistency, the 3D nnU-Net trained on CARDINAL also




(a) Segmentation from ED to ES




(b) LV and myocardial surfaces over the cardiac cycle

Fig. 3.2 Visualization of the temporal consistency of the segmentations on one patient from the CARDINAL test set. (a): Frames sampled between ED and ES, with segmentation masks from our methods + reference; (b): Curves of the LV and myocardium surfaces w.r.t. frame in the sequence.

Table 3.6 LV endocardial segmentation accuracy of 2D and 3D nnU-Net on A3C test set of 8 sequences (173 frames).

| Methods | Dice (↑) ($\bar{x} \pm \sigma$) | HD [mm] (↓) ($\bar{x} \pm \sigma$) | MAD [mm] (↓) ($\bar{x} \pm \sigma$) |
|---|---|---|---|
| 3D nnU-Net | 0.907 ± 0.025 | 6.2 ± 1.4 | 2.3 ± 0.8 |
| 3D nnU-Net* | 0.909 ± 0.029 | 5.6 ± 1.0 | 2.2 ± 0.8 |
| 2D nnU-Net | **0.916** ± 0.036 | **5.4** ± 1.9 | **2.0** ± 0.9 |

* Transfer learning from 3D nnU-Net trained on the CARDINAL dataset.

produces one of the best correlation scores for the EF calculated on the CAMUS dataset, even when compared to SOTA methods trained directly on CAMUS. Given these results, and considering that the annotation process between the two databases was not identical and was validated by different experts (which inevitably introduces a bias during the learning phase), the generalization capacity of the 3D nnU-Net model seems remarkable.

## 3.6 Transferability to apical three-chamber data

*i*VFM is typically performed on A3C data, where the blood inflow and outflow in the left ventricle can be observed. However, most public echocardiographic datasets only contain A2C and/or A4C data. Therefore, training a DL segmentation model on A3C data required creating our in-house A3C dataset consisting of 80 unannotated A3C B-mode sequences, totaling 1782 frames.

CASTOR, tailored for A2C/A4C views, could not be applied directly due to the distinct LV shape in the A3C view. To expedite the manual data annotation process, we employed active learning. Initially, we manually segmented a subset of the A3C data. Using this subset, we trained a 2D nnU-Net model, as a 2D model allowed us to utilize the maximum number of frames for training. This trained model was then used to segment the remaining unannotated data. We performed a qualitative analysis of the model's predictions, selecting segmentations that required minimal manual correction. These selected segmentations were verified and corrected manually to create gold-standard annotations. We retrained the model using the newly verified data and repeated this process of prediction, verification, and retraining multiple times until all remaining data were segmented accurately. This iterative active learning approach efficiently accelerated the annotation process. Finally, we trained a 3D nnU-Net on the fully annotated A3C dataset. Given that our A3C dataset of 80 sequences

was relatively small to efficiently train the 3D nnU-Net without overfitting, we leveraged transfer learning by utilizing weights pre-trained on the CARDINAL dataset for initialization.

Table 3.6 shows the segmentation performance of 3D nnU-Net with and without transfer learning versus a 2D nnU-Net trained from scratch. With significantly more training samples, the 2D nnU-Net outperformed the 3D nnU-Net on all geometric scores. Nevertheless, with transfer learning, the 3D nnU-Net's performance was not far behind the 2D nnU-Net.

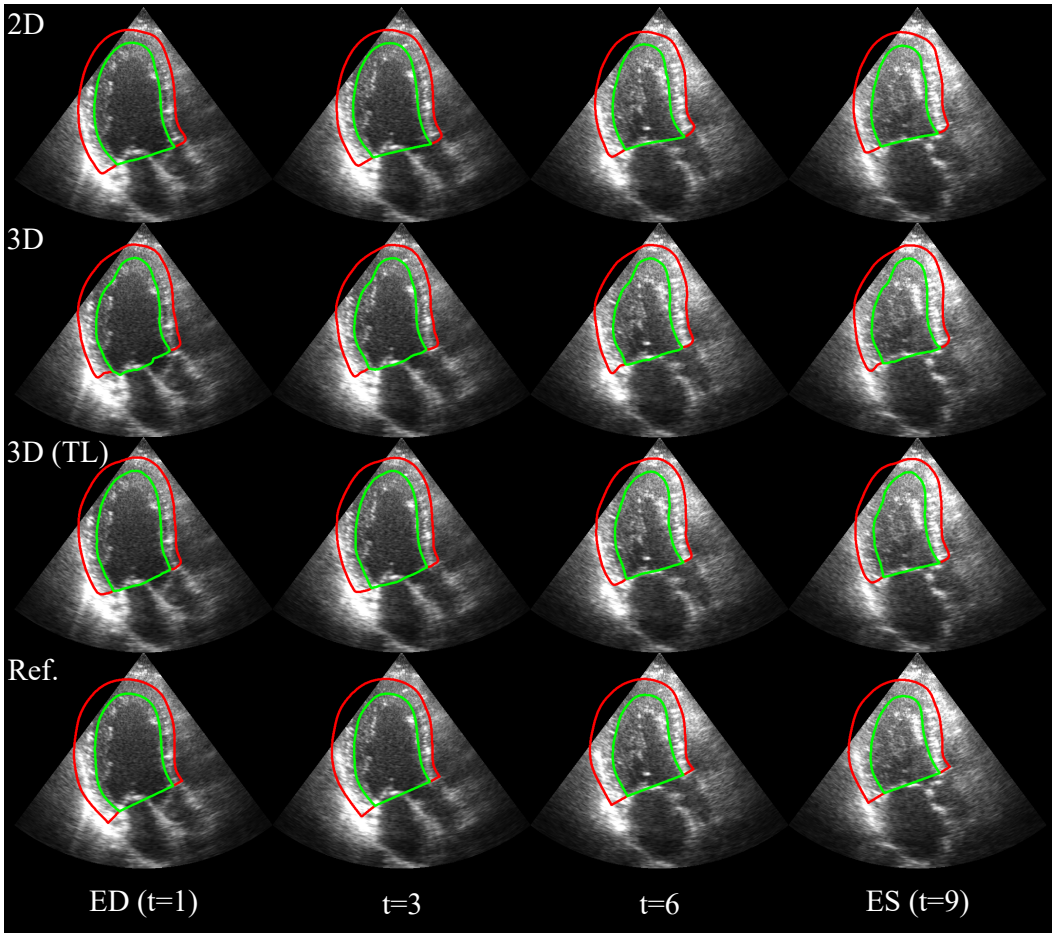
Qualitative segmentation results on one patient from the A3C test set by each method and the reference, along with plots of LV and myocardial surface areas over the cardiac cycle, are illustrated in Fig. 3.3. The 3D nnU-Net trained from scratch showed some segmentation artifacts, especially on the first frame, which were not present in the segmentation mask produced by the 3D nnU-Net with transfer learning. Although the 2D nnU-Net had the best geometric scores, its temporal consistency was inferior to both 3D approaches, as reflected by the less smooth surface area curves.

Despite having worse performance than the other approaches implemented on the CARDINAL dataset due to the significantly smaller dataset size, the predicted segmentations of the 3D nnU-Net model with transfer learning were precise and accurate enough for extracting the boundary conditions.

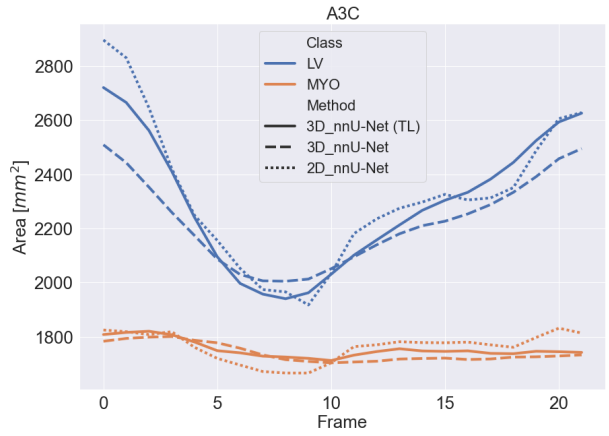
## 3.7 Conclusion

This chapter has detailed our approach to generating pseudo or gold-standard labels for the large unannotated CARDINAL dataset and achieving temporally consistent segmentation of A2C and A4C echocardiographic sequences through 2D+t segmentation methods. Specifically, the 3D nnU-Net, which considers the temporal dimension as the third spatial dimension, was able to produce temporally coherent yet clinically accurate segmentations. This approach eliminates the need for post-processing algorithms like CASTOR to correct temporal artifacts, offering two significant advantages: i) Given that CASTOR's input consists solely of predicted segmentations, temporal and anatomical errors can be difficult to correct if the predictions are too degenerate; ii) Training CASTOR requires careful hyperparameter tuning to prevent over-regularization. Furthermore, the 3D nnU-Net demonstrated a remarkable ability to maintain high cross-dataset segmentation quality.

Building on these findings, we developed an in-house A3C dataset to address the challenge of segmenting A3C views, which are not commonly found in public datasets. Using an active learning strategy, we significantly reduced the manual annotation workload, enabling the creation of a robust and fully annotated A3C dataset. Transfer learning from the CARDINAL dataset was also adopted to improve our model's performance. This allowed for the training




(a) Segmentation from ED to ES




(b) LV and myocardial surfaces over the cardiac cycle

Fig. 3.3 Visualization of the temporal consistency of the segmentations on one patient from the A3C test set. (a): Frames sampled between ED and ES, with segmentation masks from 2D nnU-Net, 3D nnU-Net, and 3D nnU-Net with transfer learning (TL) + reference; (b): Curves of the LV and myocardium surfaces w.r.t. frame in the sequence.

of the 3D nnU-Net on limited A3C data, which have a relatively different LV shape compared to A2C and A4C data, to extract the boundary conditions needed for *i*VFM. The next step is to correct the aliasing artifacts on the A3C color Doppler images to obtain clean and accurate Doppler velocity maps for reconstructing the vector blood flow.

# Chapter 4

# Phase Unwrapping of Color Doppler Echocardiography using Deep Learning

The previous chapter addressed the first essential step of *i*VFM: automatically segmenting the LV endocardial wall to extract the boundary conditions. The remaining step before the *i*VFM process is correcting phase wrapping or aliasing artifacts. In CDI, aliasing artifacts occur when blood speeds exceed the Nyquist velocity. Correcting this aliasing is crucial for quantitative methods based on color Doppler, especially in our case of *i*VFM, which uses Doppler velocities as the data fidelity term.

Traditional phase unwrapping or dealiasing methods are unreliable when applied to A3C color Doppler data, necessitating the exploration of supervised DL solutions. Unlike cardiac structure segmentation, which benefits from public datasets with annotations, there is no public color Doppler dataset with both aliased data and alias-free labels.

To address this, we established our in-house A3C color Doppler dataset to train and evaluate DL-based methods. Instead of directly predicting the dealiased velocity, we simplified the learning task by segmenting the aliased pixels and correcting them afterward. We explored several DL techniques for this purpose, including CNNs, transformers, and a deep unfolding approach.

Despite having significantly fewer learnable parameters, our primal-dual-based deep unfolding approach achieved similar performance to other DL methods. Our results suggest that DL-based methods can effectively remove aliasing artifacts in color Doppler echocardiographic images, outperforming a SOTA semi-automatic technique.

## Remarks

The work described in this chapter was published in the IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. Its reference is as follows:



## 4.1 Introduction

Color Doppler ultrasound is a widely accepted clinical imaging modality for non-invasive, real-time analysis of cardiovascular blood flow. However, the occurrence of aliasing artifacts, resulting from insufficient slow-time sampling, can affect the qualitative and quantitative analysis of color Doppler data. This issue arises when the PRF is unable to capture high axial velocities effectively. Consequently, the Doppler velocity is wrapped to the opposite side of the Doppler spectrum when its absolute value exceeds the Nyquist velocity. Experienced clinicians can easily identify zones of aliasing in most color Doppler images, where the color-coded velocities shift from red to blue or vice versa. Aliasing can be removed in Doppler echocardiography by designing multi-PRF sequences, as described by Posada et al. [23]. However, this approach requires control of the ultrasound machine and is primarily suitable for high-frame-rate echocardiography. When clinical scanners are used, aliasing must be corrected by post-processing the color Doppler fields. While a number of unwrapping algorithms have been proposed for dealiasing data maps in atmospheric science, geodesy, and optical interferometry [102–104], this problem has received less attention in color Doppler imaging.

Inspired by traditional radar approaches, Muth et al. [105] developed a segmentation-based method for color Doppler dealiasing using statistical region merging, called DeAN. This unsupervised method uses a scalar hyperparameter to control the segmentation process. An optimal parameter was determined from a supervised analysis of 50 color Doppler datasets. However, it turns out that the DeAN method fails in difficult cases, as shown in [105, Fig. 11], and that supervised corrections are still necessary in some situations. With the goal of developing imaging tools that quantify blood flow from color Doppler, we propose a DL approach to correct the aliased areas of echocardiographic color Doppler maps. DL has been proposed for color Doppler dealiasing in vascular flow imaging by Nahas et al. [106]. Their

approach aimed to solve the double aliasing problem using two U-Nets. The first U-Net detected the presence of single aliasing, while the second U-Net was trained to identify and segment double-aliased pixels. They evaluated the performance of their model by training it with different types of ultrasound information. They found that the model trained with a combination of Doppler frequency, power, and bandwidth performed the best for dealiasing in the femoral bifurcation.

In this chapter, we focused on Doppler echocardiography. In contrast to vascular flow imaging, cardiac color flow imaging can be subject to substantial clutter signals originating from the myocardium, which tend to spread the aliased patterns. With the goal of proposing a robust DL method that correctly handles aliasing in most situations, we developed and compared several architectures. Our main contributions are:

1. We designed a primal-dual network based on the idea of deep unfolding, and compared it with SOTA DL segmentation methods and DeAN.

2. We used a private color Doppler echocardiographic dataset acquired in apical three-chamber view (45 patients, 1338 aliased and 2379 non-aliased frames) to train the NNs and analyze their performance.

3. We investigated the value of adding Doppler power as input information to improve dealiasing.

4. We introduced a data augmentation strategy that generates synthetic aliasing, which solved the class imbalance problem and improved dealiasing performance on difficult color Doppler images.

## 4.2 Methods

Aliasing artifacts occur when axial blood speeds (velocity magnitudes) exceed the Nyquist velocity $V_{\mathrm{N}}$. The acquired Doppler velocity $u_{\mathrm{D}}$ can be written as a function of the unwrapped or alias-free Doppler velocity $V_{\mathrm{u}}$ as follows:

$$u_{\mathrm{D}} = V_{\mathrm{u}} - 2 \times n_{\mathrm{N}} V_{\mathrm{N}}, \qquad [\mathrm{m\,s^{-1}}] \quad (4.1)$$

where $n_{\mathrm{N}}$ is an integer called the Nyquist number, representing the number of times the signal wraps around the Nyquist limit. The Nyquist number reads (see [23] for the demonstration)

$$n_{\mathrm{N}} = \mathrm{floor}\left(\frac{V_{\mathrm{u}} + V_{\mathrm{N}}}{2V_{\mathrm{N}}}\right). \quad (4.2)$$

Except for highly turbulent flows that may occur in transvalvular or transseptal jets, there is no multiple aliasing in the adult LV scanned in the apical long-axis view, i.e., the integer $n_N$ belongs to $\{-1, 0, 1\}$. In adult echocardiography with a 3 MHz phased array, Nyquist velocities typically range from 0.55 to 0.7 m/s. As a result, single (i.e., $n_N = -1$ or 1) or no (i.e., $n_N = 0$) wrapping occurs as long as the actual blood speed is less than 1.65–2.1 m/s (see (4.4)). This means that multiple aliasing is uncommon in the LV in most situations without valvular disease or cardiac shunt. Equations (4.1) and (4.2) can be rewritten to express $u_D$ as a wrapped version of $V_u$:

$$u_D = K(V_u) = (V_u + V_N) \bmod (2V_N) - V_N, \tag{4.3}$$

where mod is the modulo operation. In particular, for $n_N \in \{-1, 0, 1\}$, the wrapping function $K$ becomes

$$u_D = K(V_u) = \begin{cases} V_u - 2V_N & \text{if } V_N < V_u < 3\,V_N \\ V_u & \text{if } -V_N \leq V_u \leq V_N \\ V_u + 2V_N & \text{if } -3\,V_N < V_u < -V_N \end{cases}. \tag{4.4}$$

This representation implies that the dealiasing problem can be approached in two different ways: i) inverting the wrapping function (4.3) to recover $V_u$ from the Doppler velocities $u_D$ by changing absolute jumps greater than $V_N$ to their $2{\times}V_N$ complement; ii) a multi-class segmentation approach that assigns a Nyquist number $n_N$ (4.2) to each pixel of the input image then computes the actual unwrapped velocities using (4.1). We investigated three DL models for dealiasing color Doppler. The first method, derived from the deep unfolding/unrolling framework, solved the inverse problem defined in (4.3) to estimate the actual velocities. We faced a nonlinear inverse problem on non-trivial data, whose solution may contain phase jumps at the blood/myocardium interfaces. Unrolled methods are well suited for solving inverse problems, and primal-dual optimization is useful for nonlinear problems. For these reasons, we tested a learned primal-dual algorithm inspired by Adler et al. [107], as described in the following subsection. The other two methods were SOTA networks adapted for determining Nyquist numbers in color Doppler images. Fig. 4.1 illustrates the pipeline used for all three methods, whose input data were the Doppler velocity multiplied by the Doppler power before scan-conversion.

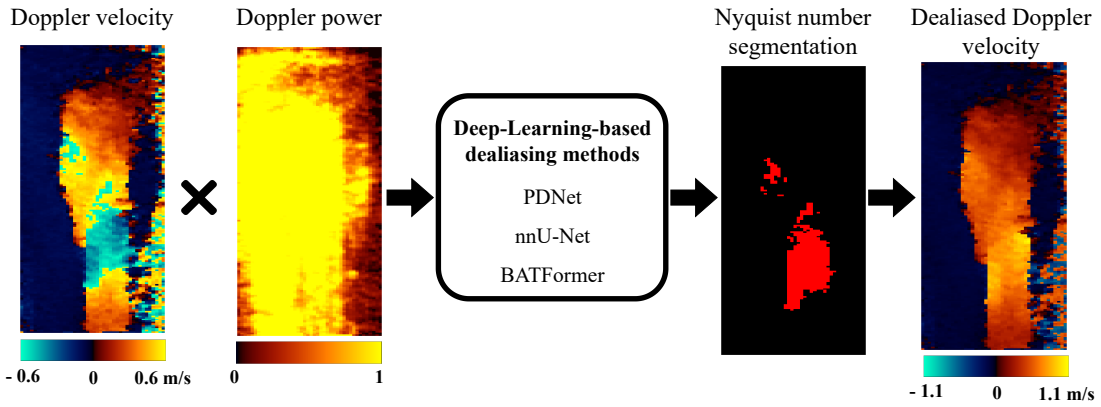



Fig. 4.1 Pipeline of the DL-based methods for color Doppler dealiasing.

### 4.2.1 OriPDNet: A primal-dual-based deep unfolding network to solve inverse problems

To solve our nonlinear inverse problem (4.3), we used OriPDNet (refer to Fig. 2 in [107] for the network architecture), a deep unfolding network based on a primal-dual optimization scheme [107]. Given a general inverse problem aiming to obtain the solution $f$ from the measurement $g$ with the forward operator $K$:

$$g = K(f), \tag{4.5}$$

the outline of OriPDNet to solve this problem is presented in Algorithm 1.

**Algorithm 1:** OriPDNet (Original primal-dual network)

Initialize $\boldsymbol{f_0}, \boldsymbol{h_0} = [\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}] \in \mathbb{R}^{M \times N \times 5}$
**for** $i = 1, \ldots, I$ **do**
  $\boldsymbol{h_i} \leftarrow \Gamma_{\theta_i^d}(\boldsymbol{h_{i-1}}, K(\boldsymbol{f_{i-1}^{(2)}}), \boldsymbol{g})$
  $\boldsymbol{f_i} \leftarrow \Lambda_{\theta_i^p}(\boldsymbol{f_{i-1}}, [\partial K(\boldsymbol{f_{i-1}^{(1)}})]^* (\boldsymbol{h_i^{(1)}}))$
**end**
**return** $\boldsymbol{f_I^{(1)}}$

OriPDNet involves several variables and operators, including the forward operator $K$, the adjoint of its Fréchet derivative $[\partial K]^*$, the input measured data $g$, the primal and dual variables $f_i$ and $h_i$, and the learned primal and dual proximal operators $\Lambda_{\theta_i^p}$ and $\Gamma_{\theta_i^d}$. Convolutional layers are used to learn these proximal operators. The hyperparameter $I$, which determines the number of iterations, requires careful tuning for each specific problem. The primal and dual variables, $f_i$ and $h_i$, are initialized then iteratively updated using the learned primal and

dual proximal operators, $\Lambda_{\theta_i^p}$ and $\Gamma_{\theta_i^d}$. The solution to the inverse problem (4.5) is obtained by extracting the first element of the primal variables, $f_I^{(1)}$. In [107], the authors recommended setting the dimension of the primal and dual spaces to five as the best compromise between memory usage and reconstruction quality, i.e., $f_i = [f_i^{(1)}, f_i^{(2)}, f_i^{(3)}, f_i^{(4)}, f_i^{(5)}] \in \mathbb{R}^{M\times N\times 5}$ and $h_i = [h_i^{(1)}, h_i^{(2)}, h_i^{(3)}, h_i^{(4)}, h_i^{(5)}] \in \mathbb{R}^{M\times N\times 5}$, where $(M \times N)$ is the size of the input data. We conducted preliminary testing and validated the use of five spaces, in accordance with their suggestion.

### 4.2.2 PDNet: A deep unfolding network for color Doppler dealiasing

To deal with our specific inverse problem (4.3) for color Doppler dealiasing, we adapted OriPDNet. The modified version was named as PDNet and summarized in Algorithm 2, with the main changes highlighted in blue.

**Algorithm 2:** PDNet (Proposed primal-dual network)

Initialize $\boldsymbol{V_0}$, $\boldsymbol{h_0} = \mathbf{[0, 0, 0, 0, 0]} \in \mathbb{R}^{M\times N\times 5}$
**for** $i = 1, \ldots, I$ **do**
  $\boldsymbol{h_i} \leftarrow \Gamma_{\theta^d}(\boldsymbol{h_{i-1}}, K(\boldsymbol{V_{i-1}^{(2)}}), \boldsymbol{u_D})$
  $\boldsymbol{V_i} \leftarrow \Lambda_{\theta^p}(\boldsymbol{V_{i-1}}, \boldsymbol{h_i^{(1)}})$
**end**
**return** $C_\theta(\boldsymbol{V_I^{(1)}})$, with $\boldsymbol{V_u} = \boldsymbol{V_I^{(1)}}$ and $\boldsymbol{n_N} = C_\theta(\boldsymbol{V_I^{(1)}})$

Specifically, we defined the forward operator $K$ as a wrapping function given by (4.3). Despite the discontinuity of this function at each $V = V_N \pm 2kV_N$ (with $k \in \mathbb{Z}^*$), its derivative was an identity function, i.e., $\partial K(V) = \mathrm{id}$. Thus, its adjoint $[\partial K(V)]^*$ was also an identity function. Unlike the original approach (see Algorithm 1), we used the same feature maps for each iteration of the main loop, which significantly reduced the number of parameters to learn (30 000 instead of 30 000 $\times$ $I$, with $I$ being the number of iterations) while maintaining the same accuracy. We made this change to avoid training instabilities that we observed while experimenting with OriPDNet. We also added a convolutional layer $C_\theta$ at the end of the network to output the Nyquist number from the estimated velocities $V_I^{(1)}$. The main reason for this was to avoid non-integer Nyquist numbers due to the regressed velocities. For a fair comparison between PDNet and OriPDNet, the same convolution layer $C_\theta$ was also applied to the output of OriPDNet.

Table 4.1 Main configurations of the three methods evaluated in this study.

| Methods | Number of feature maps[a] | Lowest resolution | Down. scheme[b] | Up. scheme[c] | Normalization scheme | Batch size | Optimization scheme[d] | Loss function | # param.[e] |
|---|---|---|---|---|---|---|---|---|---|
| PDNet | 32 → 32 → 5 | 192×40 | - | - | - | 4 | ADAM[108] + 0.001 + cosine annealing | Cross entropy + Dice | 0.03M |
| nnU-Net | 32 ↓ 480 ↑ 32 | 12×5 | Stride pooling | Transposed conv. | Instance norm. | 4 | SGD + 0.01 + polynomial decay | Cross entropy + Dice | 7M |
| BATFormer | 16 ↓ 256 ↑ 16 | 16×16 | Max pooling | $2 \times 2$ repeats | Batch norm. | 4 | ADAM + 0.001 | Cross entropy + Dice + smooth L1 | 1.2M |

[a] Size of the lowest resolution of feature maps in pixels.
[b] Downsampling scheme.
[c] Upsampling scheme.
[d] Optimizer + initial learning rate (+ learning rate scheduler used).
[e] Number of trainable parameters.

### 4.2.3 Segmentation networks for color Doppler dealiasing

nnU-Net is currently one of the best-performing approaches for medical image segmentation [9]. In this study, we addressed the dealiasing of color Doppler as a three-class segmentation problem with nnU-Net, where each class corresponded to a Nyquist number $n_{\mathrm{N}} \in \{-1, 0, 1\}$. Our network included four stages in the encoder/decoder parts and had an input size of $192 \times 40$ pixels, which was the median image size of our dataset. Table 4.1 provides more details about the architecture and the training scheme (see [9, Fig. C.1] for an illustration of the nnU-Net architecture).

Recently, transformer-based approaches have been shown to outperform the nnU-Net model in some medical challenges [109]. These models attempt to solve the segmentation problem in a different way, using attention mechanisms with receptive fields that cover the entire image. Among the best-performing models, we chose to train the BATFormer architecture [110] for the color Doppler dealiasing task, using a segmentation-based technique. This model employs a multiscale approach based on a U-Net architecture, with transformer blocks added to the decoder. This strategy results in an efficient and lightweight architecture (1.2M parameters) that is suitable for learning from small to medium-sized datasets. The main configurations of BATFormer are listed in Table 4.1, and an illustration of its architecture can be found in [110, Fig. 2].

### 4.2.4 Input data strategy

Color Doppler echocardiography produces two types of information: i) Doppler velocity, which can be corrupted by aliasing in regions of high blood speed, and ii) Doppler power,

which provides insight into the regions where velocity measurements are reliable. Using both Doppler power and velocity as input to DL models allowed them to learn how to limit the dealiasing process in regions of interest and identify ambiguous areas. Therefore, we performed an ablation study to evaluate the potential improvement provided by Doppler power. This study was conducted using the nnU-Net architecture, known for its stability in training and optimal configurations. Specifically, three nnU-Net models were trained with three combinations of input data: 1) nnU-Net #1 trained with Doppler velocity only, 2) nnU-Net #2 trained with the concatenation of Doppler velocity with Doppler power, and 3) nnU-Net #3 trained with the multiplication of Doppler velocity by Doppler power. After determining the best candidate for the input data, we trained the three models, PDNet, nnU-Net, and BATFormer, using this input combination to compute the Nyquist numbers $n_{\mathrm{N}}$ (4.2) from which the unwrapped Doppler velocities $V_{\mathrm{u}}$ (4.1) were derived. Our goal was to increment or decrement the Doppler velocities by $2n_{\mathrm{N}}V_{\mathrm{N}}$, not to modify them by smoothing, for example.

### 4.2.5 Artificial aliasing augmentation strategy

Color Doppler images may exhibit aliasing only in localized regions or frames, resulting in datasets that are often imbalanced, with most pixels belonging to the background class (i.e., without aliasing). To address this issue, we used standard data augmentation techniques such as rotation, flipping, etc., during training. We also proposed an additional data augmentation technique, which we called artificial aliasing augmentation, to improve the generalizability of our algorithms. This technique involved identifying regions with high Doppler velocity and power on alias-free Doppler images, and applying a wrapping function defined by (4.3) with a lower Nyquist velocity to create artificial aliasing artifacts, followed by normalization. The ground-truth references of these artificially aliased frames were created on the fly by comparing the Doppler velocities before and after this augmentation. By creating realistic artificially aliased frames, as shown in Fig. 4.2, this strategy enabled us to balance the classes in training batches. To evaluate the potential benefits of artificial aliasing augmentation, we conducted an additional ablation study where we tested the three DL models with and without this technique during training.

### 4.2.6 DeAN: State-of-the-art non DL-based dealiasing method for CDI

DeAN, mentioned in Section 4.1, is currently one of the most powerful non DL-based methods for color Doppler dealiasing. It is a semi-supervised method with a hyperparameter, $Q$. To unwrap aliased pixels, DeAN first segments color Doppler images using a region-merging scheme based on the Hoeffding's probability inequality. Then, DeAN compares

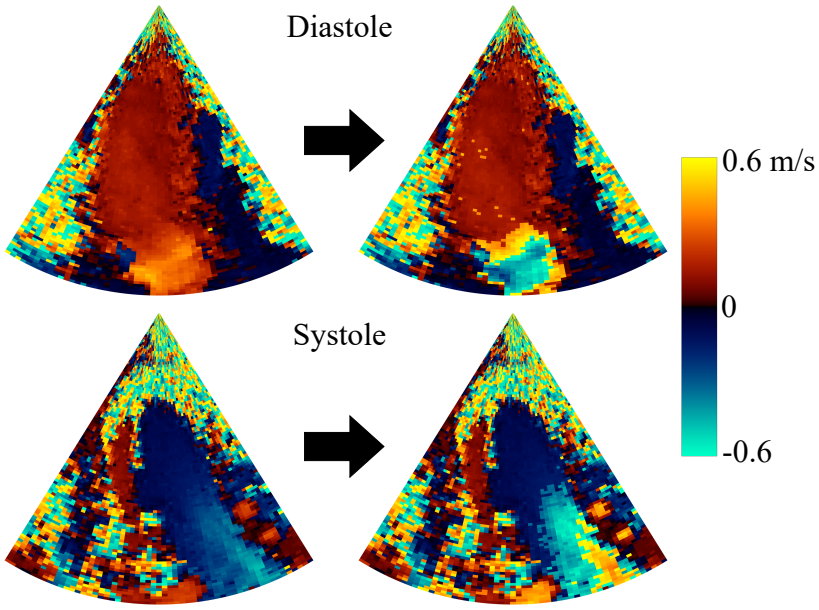



Fig. 4.2 Generation of synthetic images with artificial aliasing artifacts (right column) from non-aliased images (left column).

each segmented region with its nearest neighbors and performs dealiasing if necessary. This step is based on the assumption that the largest segment is not aliased and is repeated until all the segments have been analyzed. The main drawback of this method is the need to manually search for the optimal $Q$ hyperparameter for each frame to obtain the best dealiasing results. We compared the dealiasing performance of DeAN with both the default $Q = 10$ and with the manually optimized $Q$ hyperparameter, against the three DL methods.

### 4.2.7 Evaluation metrics

All three DL models were designed to output the Nyquist numbers $n_{\mathrm{N}}$; the dealiased velocity maps, $V_{\mathrm{u}}$, were recovered using (4.1). To evaluate the accuracy of the dealiased velocity maps and the Nyquist numbers outputted by each method, we computed four evaluation metrics.

We compared the dealiased Doppler velocity maps $V_{\mathrm{u}}$ with the ground-truth alias-free Doppler velocity maps $V_{\mathrm{ref}}$ by computing the cosine similarity index:

$$\mathrm{Cosim}(V_{\mathrm{u}}, V_{\mathrm{ref}}) = \frac{V_{\mathrm{u}} \cdot V_{\mathrm{ref}}}{\|V_{\mathrm{u}}\| \cdot \|V_{\mathrm{ref}}\|} \tag{4.6}$$

Cosine similarity is a commonly used similarity measure for comparing text data or images. We used this similarity index in our previous work on color Doppler dealiasing [105]. In addition, we computed three classification metrics to verify whether each pixel was classified

correctly on color Doppler images. The first classification metric was the balanced accuracy score, which is more suitable for unbalanced datasets than the classical accuracy score. It was calculated using the following formula:

$$\text{Accuracy} = \frac{1}{2} \times \left( \frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right) \tag{4.7}$$

where TP, FN, TN and FP refer to true positives, false negatives, true negatives, and false positives, respectively. Besides, the classical recall ($\frac{\text{TP}}{\text{TP+FN}}$) and precision ($\frac{\text{TP}}{\text{TP+FP}}$) metrics were also computed to evaluate the overall performance of the methods.

To ensure the reliability and relevance of the results, we conducted a 9-fold cross-validation to compute the scores presented in each table in Section 4.3.2. For each fold, we split the dataset into training, validation, and test sets using a ratio of 36/4/5 patients. This resulted in an average of 2974, 330, and 413 color Doppler frames for the training, validation, and test sets, respectively.

## 4.3 Experiment setup and results

### 4.3.1 Dataset and training strategies

#### 4.3.1.1 Color Doppler dataset

To evaluate the performance of our methods, we used a color Doppler echocardiographic dataset of 45 patients that were acquired using a Vivid 7 ultrasound system (GE Healthcare, USA) and a GE 5S cardiac sector probe (bandwidth = 2–5 MHz). Doppler velocity and power data prior to scan conversion were extracted into HDF formats using EchoPAC software (GE Healthcare). The EchoPAC software returned unitless power data in the range of 1-100. The power data, $P$, were compressed by taking the logarithm and then scaled to [0,1]: $\log(P)/2 \leftarrow P$. The cardiologist used the default settings (center frequency, pulse length, PRF, packet size, clutter filter, etc.). These proprietary parameters are masked and could not be extracted. In most cases, the Nyquist velocity was about 0.6 m/s. Assuming a center frequency of 3 MHz, the pulse repetition frequency was approximately 4500 Hz. These anonymized data came from a previous published study [111]. The patients were examined in the echocardiography laboratory under standard medical conditions. As a result, some patients had significant heart disease, while others had no visible pathology. The random selection of patients and their anonymization prevent us from knowing their demographic and pathological status. The sequences were acquired in the apical three-chamber view and included both Doppler velocity and power information. Each sequence covered at

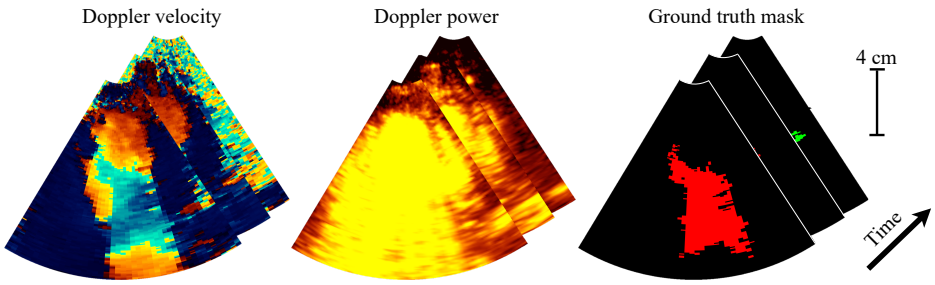



Fig. 4.3 Generation of the ground-truth masks from the Doppler velocities. The red, black, and green segmentation masks correspond to Nyquist numbers $n_{\mathrm{N}} = 1$, 0, and $-1$, respectively.

least one complete cardiac cycle, resulting in a total of 1338 aliased frames and 2379 non-aliased frames. Since color Doppler has a relatively low frame rate of 10 to 15 frames per second in clinical echocardiography, we considered each frame to be independent. To avoid interpolation artifacts, the data were collected and processed in polar coordinates (i.e., before scan-conversion), but for better visualization, the results were presented in Cartesian coordinates.

The training, validation, and test data sets, i.e., the pairs of original and alias-free Doppler velocity maps, were generated by an experienced analyzer. For this task, the non-scan-converted clinical Doppler maps were oversegmented and labeled using a statistical growing region method (see [105, Fig. 1.B]). The analyst manually identified the aliased regions, specifically focusing on those related to intraventricular blood flow, which were then corrected by applying $\pm 2V_{\mathrm{N}}$. The noisy regions associated with low Doppler power were left unchanged. Examples of color Doppler frames along with their reference segmentations are illustrated in Fig. 4.3.

#### 4.3.1.2 Training strategies

To train the DL methods described in Section 4.2, we performed supervised learning using the ground-truth segmentations from our in-house dataset. Besides applying the data augmentation strategies mentioned in Section 4.2.5, we further addressed the class imbalance of our dataset by ensuring that each batch contained at least one aliased image, whether real or synthetic. The BATFormer model was designed using the official implementation proposed in its GitHub repository[3]. This model took color Doppler images resized to 256 × 256 pixels as input and was trained for 400 epochs. On the other hand, nnU-Net and PDNet were implemented using the ASCENT[4] framework. For these two approaches, we used a patch-wise approach to preserve the resolution of the input data. The models were trained

[3]https://github.com/xianlin7/BATFormer
[4]https://github.com/creatis-myriad/ASCENT

Table 4.2 Dealiasing by nnU-Net trained with different combinations of input data using 9-fold cross-validation.

| Methods | Full dataset (Number of frames = 3717) | | | | Difficult fold* (Number of frames = 413) | | | |
|---|---|---|---|---|---|---|---|---|
| | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) |
| nnU-Net #1 *(vel. only)*[a] | **0.99** ±0.01 | **0.95** ±0.03 | 0.90 ±0.07 | **0.92** ±0.06 | 0.98 ±0.01 | 0.93 ±0.06 | 0.86 ±0.13 | **0.95** ±0.01 |
| nnU-Net #2 *(vel. ⊕ pow.)*[b] | **0.99** ±0.01 | **0.95** ±0.03 | **0.91** ±0.07 | 0.91 ±0.05 | 0.98 ±0.01 | **0.94** ±0.07 | 0.88 ±0.13 | 0.92 ±0.05 |
| nnU-Net #3 *(vel. × pow.)*[c] | **0.99** ±0.01 | **0.95** ±0.06 | 0.89 ±0.13 | 0.90 ±0.11 | **0.99** ±0.01 | **0.94** ±0.07 | **0.89** ±0.14 | 0.92 ±0.02 |

* A challenging testing fold containing color Doppler frames with aliased and non-aliased regions of similar hue.
[a] nnU-Net trained only with Doppler velocity as input data.
[b] nnU-Net trained with the concatenation of Doppler velocity with Doppler power as input data.
[c] nnU-Net trained with the multiplication of the Doppler velocity by Doppler power as input data.

for 1000 epochs to prevent any potential under/overfitting. More details on the BATFormer, nnU-Net, and PDNet architectures are provided in Table 4.1.

## 4.3.2 Experimental results

### 4.3.2.1 Doppler power information was useful in dealiasing difficult case

Table 4.2 reports the results of the ablation study aimed at identifying the optimal combination of input data. The results indicate that the three nnU-Net models performed similarly across all metrics, implying that incorporating Doppler power information in the input data did not substantially improve the models' performance. However, upon evaluation on a challenging test set (right part of Table 4.2), the model that was trained with the multiplication of Doppler velocity and Doppler power (nnU-Net #3) demonstrated better performance for all metrics except precision. The last two columns of Fig. 4.4 show two samples taken from the difficult fold, where the aliased and non-aliased regions had similar hues. This made the correction of the aliased velocities difficult. Thus, although not critical, using the Doppler velocity-Doppler power product as input data is recommended as it can enhance the models' generalization ability, especially for challenging data. For subsequent experiments and results, we trained all DL methods with this input combination.

### 4.3.2.2 PDNet outperformed its original counterpart

We conducted a study to determine the optimal number of iterations for updating the primal and dual variables in both OriPDNet and PDNet, given the sensitivity of this type of method

Table 4.3 9-fold cross-validation dealiasing results of OriPDNet and PDNet.

| # iter.[a] | OriPDNet[*]: Different feature maps per iteration | | | | PDNet[†]: Same feature maps for each iteration | | | |
|---|---|---|---|---|---|---|---|---|
| | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) |
| 1 | 0.95 ±0.03 | 0.86 ±0.06 | 0.71 ±0.13 | 0.51 ±0.09 | 0.95 ±0.03 | 0.87 ±0.06 | 0.74 ±0.12 | 0.52 ±0.10 |
| 10 | **0.98** ±0.02 | **0.92** ±0.06 | **0.84** ±0.13 | 0.78 ±0.11 | **0.98** ±0.02 | 0.92 ±0.06 | 0.84 ±0.12 | 0.80 ±0.10 |
| 20 | **0.98** ±0.02 | **0.92** ±0.06 | **0.84** ±0.13 | **0.80** ±0.11 | **0.98** ±0.01 | **0.94** ±0.06 | **0.87** ±0.13 | **0.83** ±0.11 |
| 30 | - | - | - | - | 0.97 ±0.02 | 0.91 ±0.07 | 0.82 ±0.13 | 0.77 ±0.14 |

[*] Original primal-dual approach with different feature maps per iteration.
[†] Proposed primal-dual approach using the same feature maps for each iteration.
[a] Number of iterations for updating the primal and dual variables.

to this parameter. The results are shown in Table 4.3. From this table, we can see that OriPDNet reached a plateau after 10 iterations, and beyond 20 iterations, it became unstable during training and failed to produce results. On the other hand, PDNet exhibited greater training stability and reached a performance plateau after 20 iterations. Moreover, PDNet achieved better optimal results than OriPDNet for all metrics except for the cosine similarity index, where both methods performed equally well. These results suggest that using the same feature maps in the primal-dual approach is more suitable for the dealiasing task and support the use of our deep unfolding network. Additionally, it is worth mentioning that PDNet had only 30 000 parameters, making it the lightest of the three DL models tested, as detailed in Table 4.1. The significant reduction in parameters of PDNet as compared with other models was due to the inclusion of the forward operator as prior information and the use of the same feature maps per iteration.

#### 4.3.2.3 Artificial aliasing augmentation improved the performance of segmentation-based networks

The results presented in Table 4.4 show that the use of artificial aliasing augmentation during training had varying effects on the performance of the different DL models. For nnU-Net, there was a slight improvement in all metrics except precision. In contrast, BATFormer showed significant improvement in accuracy, recall, and precision metrics, with values increasing from 0.88, 0.76, and 0.85 to 0.91, 0.81, and 0.91, respectively. However, for PDNet, the use of artificial aliasing augmentation resulted in degraded performance, with

Table 4.4 9-fold cross-validation dealiasing results of the three implemented DL solutions trained with and without the proposed artificial aliasing augmentation strategy.

| Methods | Without artificial aliasing augmentation | | | | With artificial aliasing augmentation | | | |
|---|---|---|---|---|---|---|---|---|
| | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) |
| PDNet | **0.98** ±0.02 | **0.94** ±0.06 | **0.87** ±0.12 | 0.83 ±0.10 | **0.98** ±0.01 | 0.88 ±0.08 | 0.77 ±0.16 | **0.84** ±0.09 |
| nnU-Net | **0.99** ±0.01 | 0.95 ±0.03 | 0.89 ±0.09 | **0.90** ±0.06 | **0.99** ±0.01 | **0.96** ±0.03 | **0.91** ±0.06 | 0.89 ±0.09 |
| BATFormer | **0.98** ±0.02 | 0.88 ±0.08 | 0.76 ±0.17 | 0.85 ±0.11 | **0.98** ±0.02 | **0.91** ±0.07 | **0.81** ±0.14 | **0.91** ±0.07 |

accuracy and recall decreasing from 0.94 and 0.87 to 0.88 and 0.77, respectively. These results highlight the challenge that the primal-dual-based regression methods face when generalizing to different types of aliasing. Based on the findings shown in Tables 4.2 to 4.4, we determined that the multiplication of Doppler velocity and power should be used as input for all methods, while artificial aliasing augmentation should be applied during training only for the segmentation-based techniques, i.e., nnU-Net and BATFormer.

#### 4.3.2.4 nnU-Net gave the best dealiasing results

Table 4.5 presents the final results of our study, where we compare the three DL methods with their optimal configurations against the DeAN algorithm. We observe that all three DL methods outperformed the DeAN algorithm, even the version with the manually chosen optimal $Q$ hyperparameter. This outcome confirms the potential of DL methods for color Doppler dealiasing. Among the DL methods, nnU-Net achieved the highest scores overall, with a cosine similarity close to 1, an accuracy of 0.96, a recall of 0.91, and a precision of 0.89. Therefore, we conclude that nnU-Net is the best DL approach currently available for dealiasing tasks in echocardiography. Additionally, it is interesting to note that BATFormer showed a clear improvement when we increased the amount of synthetic data, indicating that this type of approach requires a larger dataset to improve its performance for the dealiasing task. Finally, it is worth noting that PDNet achieved promising results with 233 times fewer parameters compared to nnU-Net, highlighting the potential of incorporating analytical context into the DL framework to regularize the solution space.

We also provide a visual inspection of the performance of the various methods on aliased images with different degrees of difficulty in Fig. 4.4. We can see that the DL methods performed similarly well on the easy and moderate cases (first two columns), but nnU-Net produced the closest results to the reference on the more challenging case (third column).

Table 4.5 9-fold cross-validation final dealiasing results of DL methods with their best configurations against non-DL DeAN method.

| Methods | Cosim (↑) ($\bar{x} \pm \sigma$) | Accuracy (↑) ($\bar{x} \pm \sigma$) | Recall (↑) ($\bar{x} \pm \sigma$) | Precision (↑) ($\bar{x} \pm \sigma$) |
|---|---|---|---|---|
| DeAN (Q=10) | 0.95 ± 0.03 | 0.81 ± 0.08 | 0.62 ± 0.16 | 0.53 ± 0.20 |
| DeAN (Optimized Q) | 0.98 ± 0.01 | 0.91 ± 0.04 | 0.83 ± 0.08 | 0.80 ± 0.13 |
| PDNet | 0.98 ± 0.02 | 0.94 ± 0.06 | 0.87 ± 0.12 | 0.83 ± 0.10 |
| nnU-Net | **0.99** ± 0.01 | **0.96** ± 0.06 | **0.91** ± 0.06 | 0.89 ± 0.06 |
| BATFormer | 0.98 ± 0.02 | 0.91 ± 0.07 | 0.81 ± 0.14 | **0.91** ± 0.07 |

This finding is consistent with the quantitative results presented in Table 4.5. The last column in Fig. 4.4 shows an example where no method was able to handle aliasing correctly. This example is similar to the one in the first column, but with a more pronounced level of aliasing. In this particular case, the DeAN method gave the best results. This suggests that it would be advisable to supplement the training set with challenging configurations.

## 4.4 Discussions

Color Doppler imaging takes high-pass-filtered I/Q data of the same region of interest acquired along the slow-time axis and differentiates them pairwise using a lag-1 autocorrelator. The resulting maps show blood displacements between two consecutive slow-time samples. By its very nature, color Doppler imaging is an interferometric technique that enables the measurement of displacements with a precision that can reach fractions of the center wavelength. Similarly, synthetic aperture radar interferometry (InSAR), a remote sensing technique used to map the Earth's surface deformations, generates interferograms that display ground-surface displacements. Like color Doppler, most interferometric imaging techniques in fields such as medical imaging, remote sensing, and optical metrology (e.g., phase-contrast MRI, InSAR, holographic interferometry) are subject to aliasing, i.e., jumps that occur whenever the phase shift equals $\pm\pi$. Our study aimed to address the issue of phase jumps. Among the traditional methods for phase unwrapping, one can mention: i) graph cuts [112, 113], which involve representing the wrapped phase data as a graph and determining the minimum cut that separates the known and unknown phase values; ii) least-squares approaches, which minimize the differences between partial derivatives of the wrapped phase and those of the unwrapped solution [114, 102]. Specifically for color Doppler echocardiography, Muth et al. developed DeAN, a dealiasing algorithm based on statistical region merging [105], which was used in this study for comparative purposes.

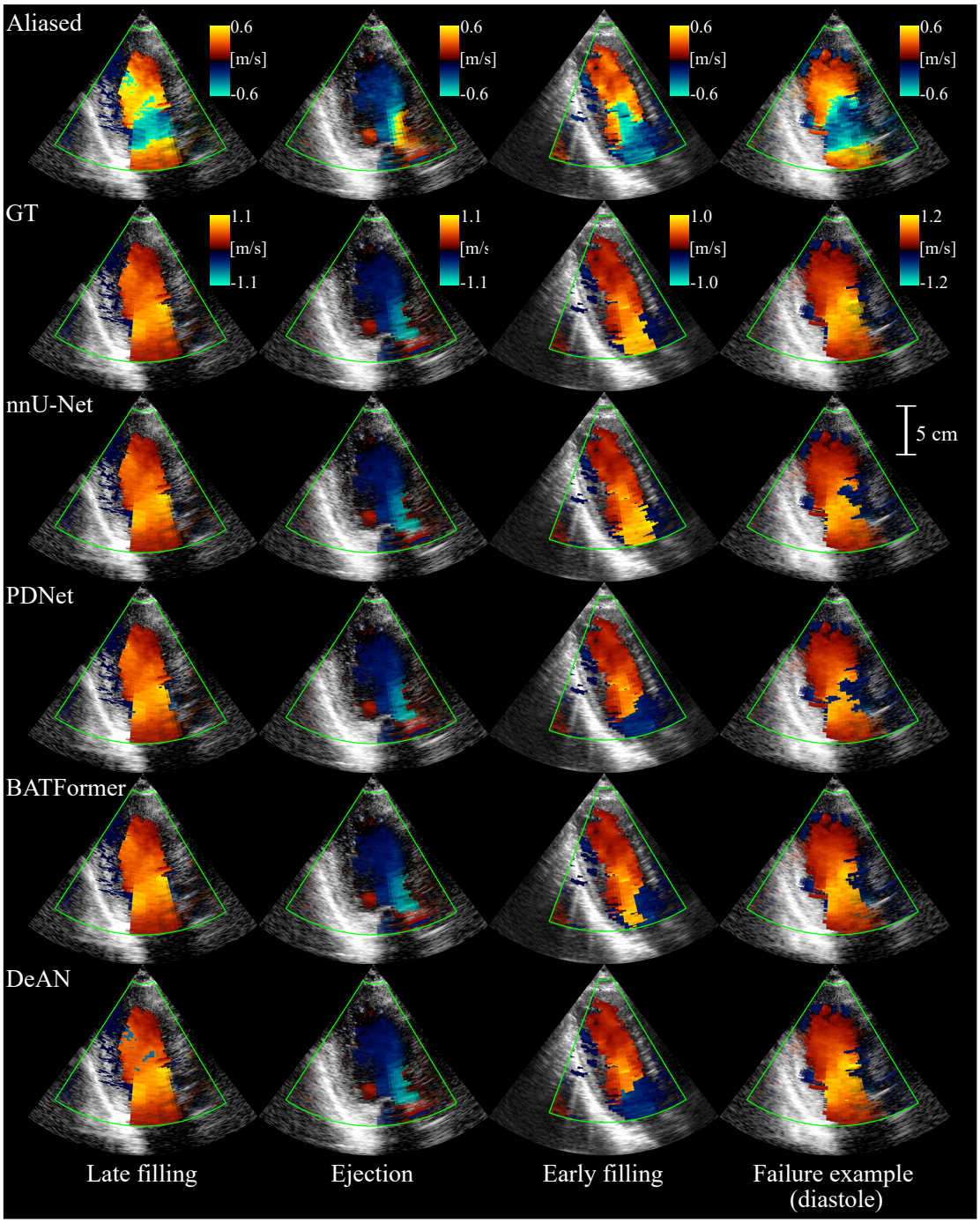



Fig. 4.4 Color Doppler images acquired during: from left to right column, late diastolic filling, systolic ejection, early diastolic filling, and diastole (a failed case). They were dealiased by DL-based methods and by DeAN with optimized $Q$ hyperparameters. First row: aliased raw color Doppler. Second row: alias-free ground truth (GT).

Recently, DL techniques have been used to improve traditional methods in phase unwrapping [115, 116, 106, 117]. Our goal was to obtain alias-free color Doppler echocardiography by applying DL to dealias clinical Doppler velocity fields. DL-based approaches have been introduced for 2D phase unwrapping in InSAR [117]. Unlike echocardiographic images, InSAR interferograms are subject to multiple wraps, making the networks proposed in our study not suited since we focused only on single aliasing. On the other hand, InSAR images are not subject to significant clutter, whereas clutter in Doppler echocardiography generates substantial noise near moving tissues, making 2D phase unwrapping challenging. As a result, non-DL approaches such as graph cuts or least-squares methods, which work well for InSAR interferograms, are not effective for echocardiographic Doppler fields. Although the DeAN technique largely solved the problem, it still fails in some situations, as shown in our study. Therefore, we turned to DL and conducted an in-depth analysis and comparison of three architectures, including PDNet, which utilizes an unfolding framework. In addition, we illustrated the potential benefits of incorporating Doppler power information, since low power generally indicates poor blood Doppler signal. To better balance the aliased and non-aliased input data during training, we resorted to data augmentation by generating synthetic aliasing.

### 4.4.1 Comparison of DL methods

Our study found that the three DL methods we tested (PDNet, nnU-Net, and BATFormer) outperformed the non-DL DeAN method for color Doppler dealiasing. Notably, we observed that nnU-Net had the best performance, suggesting that the 2D U-Net architecture used in nnU-Net may be particularly well-suited for this task due to its ability to effectively capture spatial features. For example, in a challenging case where the aliased and non-aliased regions had similar hues (Fig. 4.4, third column), nnU-Net was able to unwrap correctly while other DL methods failed or were less successful. Similar structures corrected by an expert were part of the training dataset, which implies that nnU-Net probably learned the flow patterns and leveraged this knowledge to achieve successful outcomes. PDNet also performed well, requiring $> 200$ times fewer parameters than nnU-Net, highlighting the potential for simpler DL models to achieve competitive results in color Doppler processing. Further exploration of this type of unfolding approach, including more complex modeling of the forward operator, is needed.

Although the third input strategy (velocity-power multiplication) contained less information than the second (velocity-power concatenation), it performed slightly better in the difficult fold (last row of Table 4.2). The multiplication strategy largely suppressed velocity discontinuities in noisy regions, making the training task easier. In contrast, the concatenation

of Doppler power and velocity allowed the model to learn the best strategy for combining these two inputs, which could be beneficial in larger datasets.

While we did not observe significant performance gains with BATFormer, the addition of synthetic data improved the outcomes, indicating that BATFormer also has the potential for color Doppler dealiasing, especially when more data is available. Our results demonstrate that DL methods can significantly improve upon traditional methods for color Doppler processing. Additionally, they underscore the importance of further investigating the performance of different DL architectures for this task and finding ways to effectively exploit the strengths of each architecture.

### 4.4.2 Limitations and perspectives

Color Doppler aliasing in the LV mainly occurs in the mitral jet during early and late filling, as well as during ejection into the ventricular outflow tract. As depicted in the figures, the aliasing in our study was single. However, in certain valvular diseases, such as mitral stenosis or aortic regurgitation, multiple aliasing can occur in the intraventricular cavity due to the high fluctuating velocities of the turbulent jet. The nature of multiple aliasing in this context differs from that observed in InSAR, requiring specific studies to assess the feasibility of removing aliasing in areas with significant local flow perturbations. Although this remains to be verified, it is likely that a similar strategy could also work with disturbed flows, provided that we have access to Doppler data with their alias-free references. Such ground truths could be obtained by supervised correction, as in this study, and by simulations [118].

Since we used a clinical ultrasound system with a color Doppler rate of 10 to 15 frames per second, our study did not exploit temporal information. In the context of high-frame-rate echocardiography [119], NNs with enforced temporal consistency [15] or 3D U-Net [18] could potentially improve dealiasing performance by leveraging temporal information. This approach would be especially relevant as high-frame-rate color Doppler is subject to more noise related to clutter signals.

### 4.4.3 Applications in quantitative color Doppler

Once corrected, color Doppler images can be used to visualize and quantify intracardiac blood flow. As mentioned in Section 4.1, *i*VFM is an approach to obtain comprehensive flow information, from which hemodynamic parameters can be estimated. Using a color M-mode, it is also possible to estimate the pressure difference between the apex and the mitral base, which reflects the cardiac filling [120, 121]. However, prior dealiasing is required for this method [121]. To this end, the approaches outlined in this study could be used with color

M-mode images. In a more ambitious perspective, it would be conceivable to develop NNs that can directly infer velocity vector fields or relative pressure fields from color Doppler images, once properly trained. In this case, the dealiasing process would be intrinsically integrated into the network. The main difficulty would lie in obtaining paired input data that provides the reference values. Simulations combining flows and acoustics could provide a relevant avenue for this purpose [122, 118].

## 4.5 Conclusion

In this chapter, we introduced three DL methods for the dealiasing of color Doppler echocardiography images, each leveraging distinct approaches to address the challenge of aliasing artifacts. Our methodologies included an unrolled primal-dual optimization network (PDNet), a SOTA segmentation network (nnU-Net), and a transformer-based segmentation network (BATFormer).

We formulated the dealiasing problem as a nonlinear inverse problem and adapted OriPDNet to solve it efficiently, resulting in the development of PDNet. Additionally, dealiasing can be seen as a multi-class segmentation problem for segmenting the aliasing pixels. We compared the performance of PDNet with nnU-Net and BATFormer, two powerful segmentation networks in medical image segmentation tasks.

Our experimental results demonstrated that DL techniques can achieve alias-free color Doppler echocardiography. Our proposed DL methods outperformed the non-DL DeAN method, with nnU-Net achieving the best performance, followed by PDNet. In addition, the incorporation of power information and artificial aliasing augmentation further improved the results. The application of DL techniques to color Doppler echocardiography is a promising approach that could enhance the clinical utility of this widely used imaging modality.

With these DL-based LV segmentation and dealiasing tools at our disposal, we are now positioned to perform *i*VFM using NNs aided by physics on A3C color Doppler echocardiographic data. This marks the final contribution of this thesis, pushing toward the development of a fully automatic clinical tool for extracting new flow-based biomarkers.

# Chapter 5

# Physics-Guided Neural Networks for Intraventricular Vector Flow Mapping

Chapters 3 and 4 described our contributions to fully automating the two manual or semi-automatic prerequisite steps for *i*VFM. In both LV segmentation and color Doppler dealiasing, nnU-Net stands out as the best-performing algorithm. These two DL-based tools significantly enhance the *i*VFM-v3 [8] pipeline, making the entire vector blood flow reconstruction process fully automatic.

Given the performance of NNs in previous tasks, we proposed to use data-driven approaches to perform *i*VFM. In this chapter, we present our final contribution, introducing novel alternatives to the traditional *i*VFM optimization scheme by utilizing PINNs and a physics-guided nnU-Net-based supervised approach. In PINNs, data-driven optimization is used to solve the same constrained optimization problem as in *i*VFM-v3, while in physics-guided supervised learning, a training dataset mapping from the color Doppler input to the vector blood flow output is used to train our DL regression model with an additional physical regularization term.

When evaluated on simulated color Doppler images derived from a patient-specific computational fluid dynamics model and in vivo Doppler acquisitions, both approaches demonstrate comparable reconstruction performance to the original *i*VFM algorithm. The efficiency of PINNs is boosted through dual-stage optimization and pre-optimized weights. On the other hand, the nnU-Net method excels in generalizability and real-time capabilities. Notably, nnU-Net shows superior robustness on sparse and truncated Doppler data while maintaining independence from explicit boundary conditions.

## Remarks

The work described in this chapter was published in the IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control and is currently in early access. Its reference is given below:



## 5.1 Introduction

Among the methods aiming to perform intracardiac flow imaging by color Doppler, *i*VFM [7, 8, 123, 6, 4] stands out as a post-processing approach applicable to clinical color Doppler acquisitions. The *i*VFM algorithm relies on a constrained least squares optimization scheme (see Section 5.2.1 for more details).

Recently, PINNs [12] have emerged as a novel approach for data-driven optimization by integrating neural networks and the laws of physics during the optimization process. The physical laws, often described by partial differential equations (PDEs), are incorporated into the loss function to enforce the correctness of the solutions. Automatic differentiation [78] has proven to be efficient in computing partial derivatives in PINNs. In cases involving strong non-linear PDEs in the spatiotemporal domain, extensions to PINNs, such as conservative PINNs (cPINNs) and extended PINNs (XPINNs), have been proposed [75].

PINNs have found applications predominantly in fluid mechanics [75]. In the medical field, Arzani et al. [124] utilized PINNs to recover blood flow from sparse data in 2D stenosis and aneurysm models. Kissas et al. [125] applied PINNs to predict arterial blood pressure from 4D flow MRI data. In the ultrasound domain, PINNs have been primarily used for modeling wave propagation [126], shear wave elastography [127], and regularizing velocity field given by ultrafast vector flow imaging [128].

While PINNs have demonstrated effectiveness on sparse and incomplete data, their application remains unexplored in scenarios where one or more velocity components are missing, as is the case in *i*VFM. CDI provides only scalar information—the Doppler velocity, representing the noisy radial velocity—from which we aim to derive both the radial and angular velocity components of intraventricular blood flow.

PINNs often require re-optimization for new cases with different initial or boundary conditions, which can be time-consuming. A potential solution is physics-guided supervised

learning [14], which produces output that adheres to the laws of physics by using a physics-constrained training dataset, with optional physical regularization terms. Once trained, inference can be performed seamlessly on unseen data, provided their distribution closely resembles that of the training dataset.

In this chapter, we investigated the feasibility of using physics-based NNs for vector flow mapping, exploring both a physics-guided supervised approach implemented through the nnU-Net framework [9] and two variants of PINNs. Our contributions included:

1. Training a physics-guided supervised approach based on nnU-Net, which showed high robustness on sparse and truncated data with nearly real-time inference speed;
2. Implementing two PINNs variants based on the penalty method to perform vector flow mapping, achieving performance comparable to the original *i*VFM algorithm;
3. Utilizing dual-stage optimization and pre-optimized weights from a selected Doppler frame, which enhanced PINNs' performance and reduced the optimization time of PINNs by up to 3.5 times.

## 5.2 Related Work

### 5.2.1 Intraventricular vector flow mapping (*i*VFM)

As introduced in [8], a vector blood flow map within the left ventricle can be obtained from clinical color Doppler echocardiography by solving a minimization problem. This optimization task is governed by two equality constraints: $C_1$, representing the mass conservation equation, and $C_2$, representing the free-slip boundary conditions. The first constraint ensures the 2D free-divergence of the optimized velocity field, while the second constraint enforces that the normal component of the blood velocity is zero relative to the endocardial surface. Additionally, a smoothing regularization, further detailed in Section 5.3.1, is incorporated to impose spatial smoothness of the velocity field. Equation (5.1) expresses the mathematical formulation of this problem.

In (5.1), $(\hat{v}_r, \hat{v}_\theta)$ denote the estimated radial and angular blood velocity components. Here, $\Omega$ stands for the domain of interest, i.e., the left ventricle cavity, with its endocardial boundary denoted by $\partial\Omega$. The term $\omega$ indicates the weights of the data fidelity term. Weights equal to normalized Doppler power values in the range of $[0, 1]$ were used with *in vivo* Doppler data, as they reflect the reliability of the Doppler velocity. For simulated data, we used $\omega$ equal to one. $v_\mathrm{D}$ refers to the sign-inverted Doppler velocity (positive velocities for movement away from the probe) to ensure the sign compatibility between $v_\mathrm{D}$ and the $v_r$. The

vectors $\boldsymbol{n}_{\mathrm{W}} = (n_{\mathrm{W}_r}, n_{\mathrm{W}_\theta})$ and $\boldsymbol{v}_{\mathrm{W}} = (v_{\mathrm{W}_r}, v_{\mathrm{W}_\theta})$ represent a unit vector perpendicular to the endocardial wall and a velocity vector of the endocardial wall, respectively.

The *i*VFM method [8] linearizes the constrained problem (5.1) and solves it using Lagrange multipliers and a least squares optimization scheme, which reduces the number of supervisedly determined parameters to just one, namely, the smoothing regularization weight.

$$\hat{\boldsymbol{v}} = (\hat{v}_r, \hat{v}_\theta) = \arg\min_{(v_r, v_\theta)} \underbrace{\int_\Omega \omega \, \|v_r - v_{\mathrm{D}}\| \, d\Omega}_{\text{closely match the Doppler data}}$$

subject to: (5.1)

$$\begin{cases} C_1 = r\mathrm{div}(\hat{\boldsymbol{v}}) = r\dfrac{\partial \hat{v}_r}{\partial r} + \hat{v}_r + \dfrac{\partial \hat{v}_\theta}{\partial \theta} = 0 & \text{on } \Omega \\ C_2 = (\hat{\boldsymbol{v}} - \boldsymbol{v}_{\mathrm{W}}) \cdot \boldsymbol{n}_{\mathrm{W}} = (\hat{v}_r - v_{\mathrm{W}_r}) n_{\mathrm{W}_r} + (\hat{v}_\theta - v_{\mathrm{W}_\theta}) n_{\mathrm{W}_\theta} = 0 & \text{on } \partial\Omega \end{cases}$$

### 5.2.2 Physics-informed neural networks (PINNs)

Unlike the conventional least squares Lagrangian optimization, such as in *i*VFM, PINNs use an iterative scheme with MLPs to iteratively refine and reach the optimal solution. PINNs offer the advantage of being flexible, especially in the optimization scheme, regardless of the linearity of the problem [12]. When additional complex physical constraints need to be incorporated, PINNs require minimal architectural modifications, typically requiring only the adaptation of the loss function. However, this process can be challenging when applied to standard approaches involving non-linear constraints.

When addressing a constrained optimization problem using PINNs, the problem is reformulated into a series of loss functions, which often involve conflicting objectives. To manage multi-objective optimization in PINNs, a linear scalarization of the losses is commonly used:

$$\mathcal{L}_\mu(\theta_{\mathrm{NN}}) = \sum_{j=1}^{N} \mu_j \mathcal{L}_j(\theta_{\mathrm{NN}}), \quad \mu_j \in \mathbb{R}_{>0}, \tag{5.2}$$

with $\mu_j$ representing the penalty coefficients or the loss weights, $\mathcal{L}_{1,\cdots,N}$ being the multiple losses derived from the original constrained optimization problem, and $\theta_{\mathrm{NN}}$ representing the network parameters. All the losses involved in PINNs' optimization are functions of $\theta_{\mathrm{NN}}$. However, for better readability of the equations, $\theta_{\mathrm{NN}}$ is omitted in the subsequent loss expressions.

Among the methods with linear scalarization, two notable approaches are the soft constraints and penalty methods. The soft constraints approach uses fixed penalty coefficients throughout the optimization. However, determining optimal coefficients can be challenging,

especially as the number of objectives ($N$) increases, making this approach generally less favored in PINNs optimization.

In contrast to the soft constraints approach, the penalty method involves varying coefficients. Various techniques have been proposed to adapt these coefficients during optimization. Examples include GradNorm [129], SoftAdapt [130] or ReLoBRaLo [131]. GradNorm and SoftAdapt dynamically adjust the loss weights based on the relative training rates of different losses. ReLoBRaLo can be seen as a combination of the former two techniques, which incorporates a moving average for loss weights and a random look-back mechanism. The random look-back mechanism is controlled by a variable that determines whether the loss statistics of the previous steps or those of the first step are used to compute the coefficients.

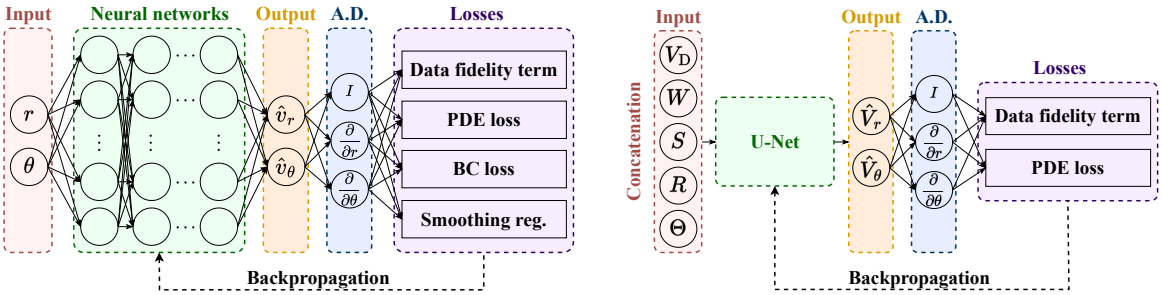
Recently, an alternative method called Augmented Lagrangian (AL) has been proposed for solving constrained optimization problems using PINNs [132, 133]. Similar to the penalty method, the AL approach involves penalty terms, but it also introduces a term designed to mimic a Lagrange multiplier.

In situations where optimizing initial or boundary conditions is difficult, some studies have proposed imposing those conditions as hard constraints by utilizing a distance function and an analytical approximation of the conditions [134, 135]. Although it is feasible to impose hard constraints on the output of neural networks, it is often challenging and more appropriate for problems with several initial or boundary conditions.

## 5.3 Methods

In this study, we addressed the constrained optimization problem of *i*VFM (5.1) through neural networks aided by physics: a physics-guided supervised approach based on nnU-Net and PINNs using the penalty method. Specifically, we studied two variants of PINNs: 1) PINNs with the ReLoBRaLo weight-adapting strategy (RB-PINNs); 2) Augmented Lagrangian PINNs (AL-PINNs). Schematic representations of the general architectures of PINNs and nnU-Net for intraventricular vector flow reconstruction are shown in Fig. 5.1a and Fig. 5.1b, respectively.

The following subsections introduce the loss functions to be optimized in PINNs (Section 5.3.1), provide implementation details for PINNs (Sections 5.3.2 to 5.3.5), discuss the physics-guided nnU-Net approach (Section 5.3.6), and present the evaluation metrics (Section 5.3.7).




(a) PINNs (b) Physics-guided nnU-Net

Fig. 5.1 Architectures of PINNs and physics-guided nnU-Net. A.D. refers to automatic differentiation. In (a), the 2D input of PINNs has a shape of $(B \times 2)$, where $B$ is the batch size. $r$ and $\theta$ denote radial and angular coordinates. In (b), nnU-Net takes a 4D input of shape $(B \times 5 \times 192 \times 40)$, which is the concatenation of sign-inverted dealiased Doppler velocity $V_{\mathrm{D}}$, weight matrix $W$, left ventricular segmentation $S$, radial coordinate array $R$, and angular coordinate array $\Theta$.

### 5.3.1 PINNs' loss functions

In line with the previous *i*VFM method, we decomposed the mathematical formulation in (5.1) into several objectives to be optimized ((5.3) to (5.6)): 1) $\mathcal{L}_1$: data fidelity term; 2) $\mathcal{L}_2$: mass conservation residual loss (PDE loss); 3) $\mathcal{L}_3$: boundary condition residual loss (BC loss); 4) $\mathcal{L}_4$: smoothing regularization.

For the PDE loss, namely $\mathcal{L}_2$, the partial derivatives were computed using automatic differentiation . On the contrary, for $\mathcal{L}_4$, the partial derivatives were obtained using finite difference methods with 2D convolution kernels, as spatial smoothness could not be computed with automatic differentiation. This was done by setting the weights of the $3 \times 3$ convolution kernels to the central finite difference coefficients with second-order accuracy.

The norm $\|\cdot\|$ used for computing the $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ was the Smooth L1 loss or Huber loss with $\beta = 1.0$ and *sum* reduction over all the samples unless otherwise stated. The Smooth L1 loss uses a squared term if the absolute error falls below $\beta$ and an absolute term otherwise, making it less sensitive to outliers than the mean squared error.

$$
\begin{cases}
\mathcal{L}_1 = \omega \left\| \hat{v}_r - v_{\mathrm{D}} \right\| & \text{on } \Omega \quad (5.3) \\
\mathcal{L}_2 = \left\| r \dfrac{\partial \hat{v}_r}{\partial r} + \hat{v}_r + \dfrac{\partial \hat{v}_\theta}{\partial \theta} \right\| & \text{on } \Omega \quad (5.4) \\
\mathcal{L}_3 = \left\| (\hat{v}_r - v_{\mathrm{W}_r}) n_{\mathrm{W}_r} + (\hat{v}_\theta - v_{\mathrm{W}_\theta}) n_{\mathrm{W}_\theta} \right\| & \text{on } \partial\Omega \quad (5.5) \\
\mathcal{L}_4 = \displaystyle\sum_{k \in \{r,\theta\}} \left\{ \left( r^2 \frac{\partial^2 v_k}{\partial r^2} \right)^2 + 2 \left( r \frac{\partial^2 v_k}{\partial r \partial \theta} \right)^2 + \left( \frac{\partial^2 v_k}{\partial \theta^2} \right)^2 \right\} & \text{on } \Omega \quad (5.6)
\end{cases}
$$

### 5.3.2 RB-PINNs

#### 5.3.2.1 Global Loss

The global loss function to be optimized in RB-PINNs was defined as:

$$\mathcal{L}_{\mu,\theta_{\text{NN}}} = \underbrace{\mu_1\mathcal{L}_1}_{\substack{\text{data fidelity}\\ \text{term}}} + \underbrace{\mu_2\mathcal{L}_2 + \mu_3\mathcal{L}_3}_{\text{PDE \& BC losses}} + \underbrace{\mu_4\mathcal{L}_4}_{\substack{\text{smoothing}\\ \text{reg.}}}, \tag{5.7}$$

where $\mu_1, \mu_2, \mu_3 \in \mathbb{R}_{>0}$ are adaptive penalty coefficients, and $\mu_4$ is the smoothing regularization weight. We heuristically set $\mu_4$ to $10^{-7.5}$.

**Algorithm 3:** ReLoBRaLo update strategy

Initialize $\mu_j(1) = 1$, $j \in \{1, 2, 3\}$.
**for** $i = 1, \ldots, I$ **do**
  Forward pass and compute losses
  $\mathcal{L}_j(i) \leftarrow \mathcal{L}_j$
  **if** $i >= 2$ **then**
    $\hat{\mu}_j^{(i,i-1)} \leftarrow n_{\text{loss}} \times \text{Softmax}\left(\frac{\mathcal{L}_j(i)}{\mathcal{T}\mathcal{L}_j(i-1)+\epsilon}\right)$
    $\hat{\mu}_j^{(i,1)} \leftarrow n_{\text{loss}} \times \text{Softmax}\left(\frac{\mathcal{L}_j(i)}{\mathcal{T}\mathcal{L}_j(1)+\epsilon}\right)$
    $\mu_j(i) \leftarrow \alpha\left(\rho\mu_j(i-1) + (1-\rho)\hat{\mu}_j^{(i,1)}\right)$
      $+(1-\alpha)\hat{\mu}_j^{(i,i-1)}$
  **end**
  Compute final loss using (5.7) and do backpropagation to update network parameters: $\boldsymbol{\theta}_{\text{NN}} \leftarrow \boldsymbol{\theta}_{\text{NN}} - \eta_{\theta_{\text{NN}}}\nabla_{\theta_{\text{NN}}}\mathcal{L}_{\mu,\theta_{\text{NN}}}(i)$
**end**

#### 5.3.2.2 Update strategy for loss weights

Algorithm 3 details the ReLoBRaLo update strategy introduced in [131] for determining the loss weights, i.e., $\mu_1$, $\mu_2$, and $\mu_3$. In this algorithm, $n_{\text{loss}}$ represents the total number of losses for which the loss weights are updated; in our case, $n_{\text{loss}} = 3$. $\hat{\mu}_j^{(i,i')}$ computes the scaling based on the relative improvement of $\mathcal{L}_j$ between the iterations $i'$ and $i$. $\mu_j(i)$ is defined as the weight for $\mathcal{L}_j$ at the $i^{th}$ iteration, obtained through an exponential decay. The algorithm's hyperparameters consist of $\alpha$ for the exponential decay rate, $\rho$ for a Bernoulli random variable with an expected value close to 1, $\mathcal{T}$ for temperature, and $I$ for the total number of iterations. We heuristically set $\alpha = 0.999$, $\mathbb{E}(\rho) = 0.999$, and $\mathcal{T} = 1.0$, as this

combination yielded the best results for our problem. $\theta_{\text{NN}}$ denotes the learnable network parameters, $\eta_{\theta_{\text{NN}}}$ is the learning rate used for updating network parameters, and $\nabla_{\theta_{\text{NN}}}$ is the gradient of the final loss with respect to $\theta_{\text{NN}}$.

### 5.3.3 AL-PINNs

#### 5.3.3.1 Global loss

For AL-PINNs, we defined its global loss as:

$$\mathcal{L}_{\lambda,\mu,\theta_{\text{NN}}} = \underbrace{\mathcal{L}_1}_{\substack{\text{data fidelity}\\ \text{term}}} + \underbrace{\left\langle \boldsymbol{\lambda_1}, \boldsymbol{C_1} \right\rangle + \left\langle \boldsymbol{\lambda_2}, \boldsymbol{C_2} \right\rangle}_{\substack{\text{PDE \& BC losses}\\ \text{(Lagrange multipliers)}}} + \underbrace{0.5 \times \mu \times (\mathcal{L}_2 + \mathcal{L}_3)}_{\text{PDE \& BC losses (penalty)}} + \underbrace{\mu_4 \mathcal{L}_4}_{\substack{\text{smoothing}\\ \text{reg.}}}. \tag{5.8}$$

In this equation, $\lambda_1$ and $\lambda_2$ are learnable real Lagrange multipliers related to the two constraints: the mass conservation, $C_1$, and the free-slip boundary condition, $C_2$. The notation $\left\langle \cdot, \cdot \right\rangle$ refers to the inner product of two vectors. The learnable penalty coefficient for the two physical constraints is denoted by $\mu \in \mathbb{R}_{>0}$. Similar to RB-PINNs, $\mu_4$ was set heuristically to $10^{-7.5}$.

**Algorithm 4:** Augmented Lagrangian update strategy

Initialize $\boldsymbol{\lambda_j} = \vec{0}, \mu = 2, j \in \{1, 2\}$.
**for** $i = 1, \ldots, I$ **do**
  Forward pass and compute losses
  $\mathcal{L}_{\lambda,\mu,\theta_{\text{NN}}}(i) \leftarrow \mathcal{L}_{\lambda,\mu,\theta_{\text{NN}}}$
  Compute final loss using (5.8) and do backpropagation to simultaneously update network parameters, learnable $\boldsymbol{\lambda_j}$ as well as $\mu$:

$$\begin{cases} \boldsymbol{\theta}_{\text{NN}} \leftarrow \boldsymbol{\theta}_{\text{NN}} - \eta_{\theta_{\text{NN}}} \nabla_{\theta_{\text{NN}}} \mathcal{L}_{\lambda,\mu,\theta_{\text{NN}}}(i) \\ \boldsymbol{\lambda_j} \leftarrow \boldsymbol{\lambda_j} + \eta_{\lambda} \nabla_{\lambda_j} \mathcal{L}_{\lambda,\mu,\theta_{\text{NN}}}(i) \\ \mu \leftarrow \mu + \eta_{\mu} \nabla_{\mu} \mathcal{L}_{\lambda,\mu,\theta_{\text{NN}}}(i) \end{cases}$$

**end**

#### 5.3.3.2 Update strategy for loss weights

We applied the gradient ascent method [133] to update $\lambda_1$, $\lambda_2$, and $\mu$, as decribed in Algorithm 4. In the original approach [133], $\mu$ remains constant throughout the optimization process, but we proposed to update this penalty coefficient to better adhere to the original AL method [136]. In our experiments, the gradient ascent method showed higher optimization stability and was less prone to gradient explosion compared to the original AL update rule proposed in [136]. In Algorithm 4, $\theta_{\text{NN}}$, $\eta_{\theta_{\text{NN}}}$, and $\nabla_{\theta_{\text{NN}}}$ represents the learnable network parameters,

the learning rate for updating network parameters, and the gradient of the final loss with respect to $\theta_{\mathrm{NN}}$, respectively. $\nabla_{\lambda_j}\cdot$ and $\nabla_{\mu}\cdot$ denote the gradient of the final loss with respect to $\lambda_j$ and $\mu$; $I$ indicates the total number of iterations; $\eta_\lambda$ and $\eta_\mu$ are the learning rates for the learnable Lagrange multipliers $\lambda_j$ and $\mu$. The selection of appropriate learning rates is critical in preventing gradient overflow when dealing with physical losses that involve unbounded Lagrange multipliers. Their values are discussed in Section 5.4.2.1.

### 5.3.4 Dual-stage optimization

To improve the convergence of our PINNs, we introduced a dual-stage optimization strategy: 1) an optimization stage using the AdamW [137] optimizer for the first 90% of the iterations to converge to a rough solution; 2) a fine-tuning stage using the L-BFGS [138] optimizer, which is not sensitive to learning rates, for the remaining iterations to obtain an optimal final solution. This approach aimed to reduce the optimization time of PINNs. An ablation study was performed using RB-PINNs to assess the potential improvement of this strategy.

### 5.3.5 PINNs' architecture, weight initialization, and sampling strategy

#### 5.3.5.1 Network Architecture

For both PINNs implemented in this paper, we utilized an MLP with six hidden layers, each containing 60 neurons, with the *tanh* activation function. This architecture resulted in approximately 18.6k trainable parameters.

#### 5.3.5.2 Weight initialization

We first applied the dual-stage optimization to a Doppler frame selected at the end of the early filling phase using RB-PINNs. The resulting weights were then saved as pre-optimized weights and used as initialization for all subsequent PINNs models before optimization on new Doppler data. This initialization technique aimed to accelerate the optimization process of our PINNs and enhance their performance. A second ablation study was carried out to justify this choice.

#### 5.3.5.3 Sampling strategy

Leveraging the regularly spaced polar grid of color Doppler imaging and its relatively small size, we utilized all sample points within the left ventricle on the grid for both data and collocation points. These data points were used to compute the data fidelity term, while the

PDE loss and smoothness term were evaluated from the collocation points. For the boundary condition residual loss, all extracted points on the boundary were considered.

### 5.3.6 Physics-guided nnU-Net

We trained a physics-guided nnU-Net (refer to Fig. 5.1b for its architecture) with configurations similar to those described in [19, Table I] on both simulations and in vivo data. We adapted the loss function to L1 loss for supervised regression. Additionally, to enforce mass conservation in the predicted velocity field, we incorporated $\mathcal{L}_2$ in the loss function as a physical regularization term with a weight of $\gamma = 10^{-3}$. Both supervised and regularization terms were masked with the left ventricular binary segmentation, restricting loss computation to the region of interest. The final loss was expressed as:

$$\mathcal{L}_{\gamma,\theta_{\mathrm{NN}}} = \underbrace{\left\|\hat{V}_r - V_{r_{\mathrm{ref}}}\right\|_1 + \left\|\hat{V}_\theta - V_{\theta_{\mathrm{ref}}}\right\|_1}_{\text{data fidelity term}} + \gamma \underbrace{\left\| r\frac{\partial \hat{V}_r}{\partial r} + \hat{V}_r + \frac{\partial \hat{V}_\theta}{\partial \theta}\right\|_1}_{\text{PDE loss}}, \tag{5.9}$$

where $(V_{r_{\mathrm{ref}}}, V_{\theta_{\mathrm{ref}}})$ represent the reference velocity field given by simulations or predicted velocity field by *i*VFM [8], considered as the gold standard for in vivo data.

Unlike PINNs that directly take coordinates $(r, \theta)$ as input, nnU-Net requires image data. Our nnU-Net's input was a concatenation of: 1) dealiased color Doppler image before scan-conversion for in vivo data or alias-free image for simulated data; 2) a weight matrix with normalized Doppler powers in the range of $[0, 1]$ for in vivo data or containing ones for simulated data; 3) binary segmentation of the left ventricle cavity; 4) radial coordinate array; 5) angular coordinate array. More details about the training dataset are given in Section 5.4.1.2.

During training, we applied data augmentations, including random rotation ($[-15, 15]^\circ$), random zoom ($[0.7, 1.4]$), and random scanline masking. With the latter, a block of $n$ consecutive scanlines was randomly masked out with a step size of $m$. In our experiments, $m = 10$ and $n$ was a random integer between 0 and 9. This strategy simulated sparse Doppler data, enhancing the model's robustness and generalizability.

### 5.3.7 Evaluation metrics

We assessed the performance of RB-PINNs, AL-PINNs, and nnU-Net using simulated Doppler images from a patient-specific computational fluid dynamics (CFD) model (see Section 5.4.1.1). Evaluation metrics, including squared correlation ($r^2$) and normalized

root-mean-square error (nRMSE), were computed by comparing predicted and ground truth velocity fields within the left ventricle.

#### 5.3.7.1 Squared correlation

We defined the squared correlation as follows:

$$r_{v_k}^2 = \mathrm{Corr}(\hat{v}_k, v_{k_{\mathrm{CFD}}})^2, \quad k \in \{r, \theta\}, \tag{5.10}$$

where Corr is the Pearson correlation coefficient.

#### 5.3.7.2 nRMSE

For both the radial and angular components, we computed the root-mean-square errors normalized by the maximum velocity defined by:

$$\mathrm{nRMSE} = \frac{1}{\max \|v_{\mathrm{CFD}}\|_2} \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left\| \hat{v}_k - v_{\mathrm{CFD}_k} \right\|_2^2}, \tag{5.11}$$

where $n$ stands for the number of velocity samples in the left ventricular cavity. For the nRMSE metrics shown in Tables 5.1, 5.2 and 5.4, we considered both velocity components and reported them as $(\tilde{x} \pm \sigma_{\mathrm{rob}})$. Here, $\tilde{x}$ signifies the median, while $\sigma_{\mathrm{rob}} = 1.4826 \times \mathrm{MAD}$ represents the robust standard deviation (std.), with MAD denoting the mean absolute deviation.

## 5.4 Experimental setup and results

### 5.4.1 Dataset

#### 5.4.1.1 Patient-specific computational fluid dynamics (CFD) heart model

To validate our approaches, we utilized a new patient-specific physiological CFD model of cardiac flow developed by the IMAG laboratory. This model features a more realistic mitral valve compared to the previous version [139, 34]. We followed the same method described in [8, Sec. 2.3] to generate 100 simulated Doppler images evenly distributed over a cardiac cycle, with a signal-to-noise ratio (SNR) equal to 50 dB. Each image comprised 80 scanlines with 200 samples per scanline. As Doppler power information was not available, we set the weight for the data fidelity term in both PINNs and *i*VFM to one, i.e., $\omega = W = 1$.

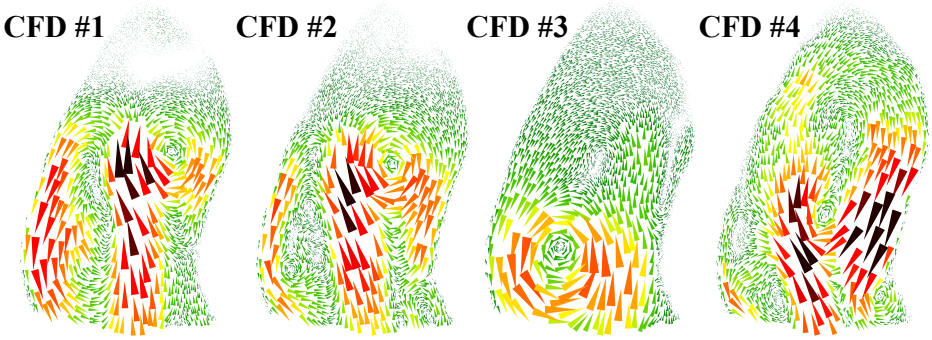



Fig. 5.2 Simulated color Doppler image during early filling derived from patient-specific CFD heart models with four variants of mitral valves. CFD #1-3 represent cases following mitral valve replacement with a bioprosthetic valve, while CFD #4 is a normal case.

#### 5.4.1.2 CFD and in vivo apical three-chamber (A3C) training dataset for nnU-Net

Based on the CFD model described in Section 5.4.1.1, we introduced variations to the mitral valve geometry, by modifying opening angle and orientation, to simulate three distinct cases following mitral valve replacement with a bioprosthetic valve (CFD #1-3) [140]. CFD #1 and #2 corresponded to two different inflow jet orientations, with a slight increase in mitral valve cross-sectional area in CFD #2. CFD #3 mimicked blood flow with a wide-opened mitral valve, resulting in a weak jet with limited penetration. CFD #1-3 were included as training/validation data for the nnU-Net, while the unaltered model representing a normal case (CFD #4) was used for testing. Fig. 5.2 provides an example of a simulated color Doppler image during early filling for each CFD model. Although these four models shared the underlying cardiac geometry, modifications to the mitral valve resulted in sufficiently diverse intraventricular flows to reduce the training bias.

Due to the limited availability of simulated Doppler data for training a supervised model, we chose to include in vivo A3C duplex (B-mode + color Doppler) data in our training dataset. This decision was further elaborated in Section 5.5.1. These data aligned with the dataset of prior studies [19, 111], acquired using a Vivid 7 ultrasound system (GE Healthcare, USA) with a GE 5S cardiac sector probe (bandwidth = 2–5 MHz). Further details about this dataset can be found in [19, Sec. III-A.1].

We processed the in vivo A3C data to ensure high-quality training data. We initially filtered out low-quality data, resulting in a compilation of 92 Doppler echocardiographic cineloops from 37 patients, totaling 2,668 frames. Subsequently, we performed preprocessing using ASCENT [18, 19]. This process involved segmenting the left ventricle cavity on B-mode images to define the region of interest and boundary conditions, which varied across frames, and correcting aliased pixels on the corresponding color Doppler images. Finally, we

applied the *i*VFM method to reconstruct the 2D vector field in the left ventricle, serving as the gold standard for training the nnU-Net. For training purposes, the physics-constrained dataset was split subject-wise into 74/10/8 clinical cineloops plus 2/1/1 CFD simulations, resulting in 2037/434/197 in vivo and 200/100/100 simulated images for training/validation/testing.

## 5.4.2 Training Strategies

All methods were implemented in the same PyTorch-based framework to ensure consistent training and optimization. The training configurations for each approach were as follows:

### 5.4.2.1 RB-PINNs and AL-PINNs

For both RB-PINNs and AL-PINNs, we used a dual-stage optimization strategy, involving two stages with a total of $I = 2500$ iterations. In the first stage, we applied AdamW optimization for $0.9 \times I$ iterations, updating all learnable parameters with a learning rate of $\eta_{\theta_{\text{NN}}} = \eta_\lambda = \eta_\mu = 10^{-5}$. Then, in the fine-tuning stage, which comprised the remaining 10% of the iterations, we utilized L-BFGS optimization. In this stage, only the network parameters were updated, while the learnable loss weights from the first stage were retained. The L-BFGS optimizer was configured with a maximum of ten iterations per optimization step, and the strong Wolfe line search conditions.

This strategy ensured a balance between accuracy and optimization duration, enhancing the stability and efficiency of PINNs' optimization process. The advantage of this strategy was further demonstrated in Section 5.4.3.1.

### 5.4.2.2 Physics-Guided nnU-Net

Our physics-guided nnU-Net underwent 1000 epochs of training with the following configurations: a patch size of $(192 \times 40)$ pixels, a batch size of 4, and SGD optimizer with an initial learning rate of 0.01, paired with a linear decay scheduler.

## 5.4.3 Experimental Results

### 5.4.3.1 Pre-Optimized weights and dual-stage optimization enhanced PINNs' performance

The ablation study presented in Table 5.1 highlights that the combination of pre-optimized weights and dual-stage optimization in RB-PINNs yielded the best performance within a fixed optimization time. The dual-stage optimization strategy significantly reduced the

Table 5.1 Ablation study on 100 simulated Doppler images using RB-PINNs. Optimization time is consistent (100 seconds per frame) across all configuration combinations. Default settings are highlighted in purple.

| Pre-optimized weights | Dual-stage optimization | $r^2(\uparrow)$ | | nRMSE [%]($\downarrow$) ($\tilde{x} \pm \sigma_{\text{rob}}$) |
|---|---|---|---|---|
| | | $v_r$ | $v_\theta$ | |
| ✘ | ✘ | 0.88 | 0.23 | 4.3 ± 2.2 |
| ✘ | ✔ | 0.97 | 0.57 | 2.8 ± 1.2 |
| ✔ | ✘ | 0.96 | 0.58 | 2.4 ± 1.0 |
| ✔ | ✔ | **0.99** | **0.66** | **2.2** ± 1.0 |

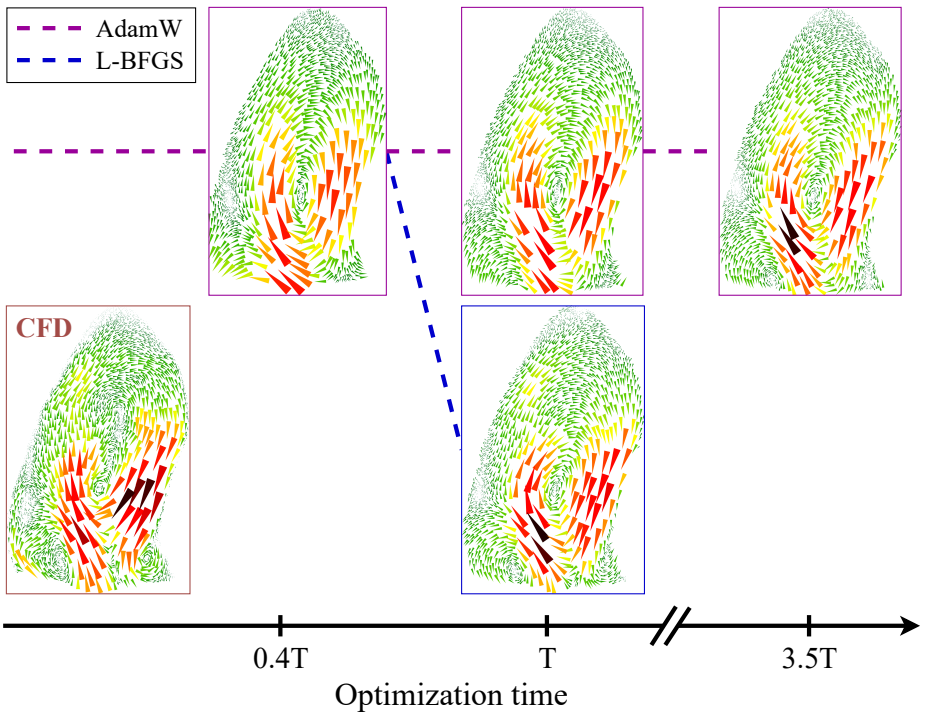



Fig. 5.3 Dual-stage (AdamW + L-BFGS) versus single-stage (AdamW only) optimization using RB-PINNs initialized with pre-optimized weights. T refers to the total amount of time required for dual-stage optimization. In this example, 3.5 × more time is needed for single-stage optimization (top right) to converge to a similar solution given by dual-stage optimization (bottom right).

optimization time of PINNs methods. Figure 5.3 provides a visual representation of a case where optimization was conducted using pre-optimized weights with and without dual-stage optimization. In this example, single-stage optimization with AdamW required 3.5× more optimization time to achieve a visually similar solution compared to dual-stage optimization. Subsequent experiments with PINNs followed this optimization strategy—dual-stage optimization with pre-optimized weight initialization.

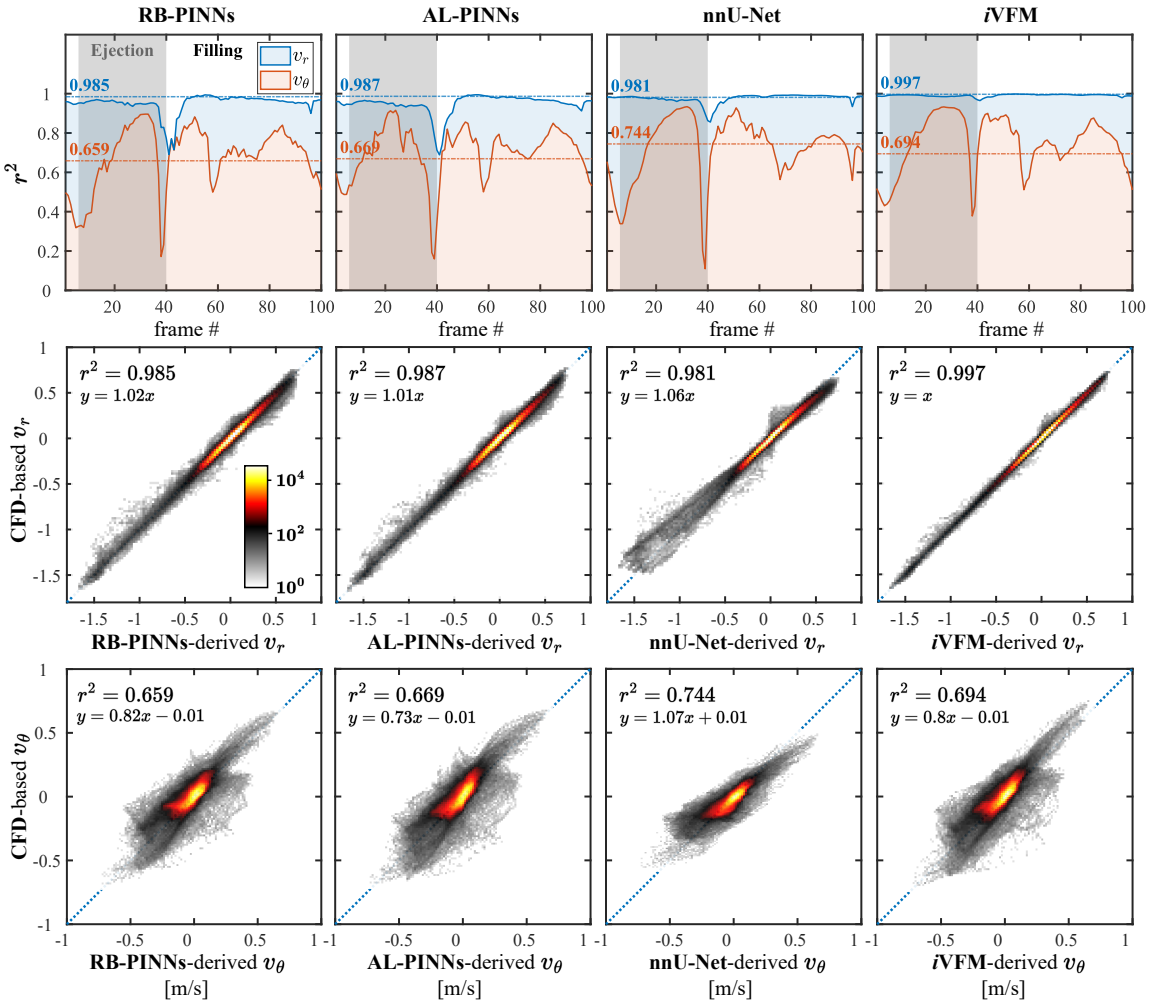



Fig. 5.4 Top row: time-varying squared correlation between CFD-based velocities and reconstructed velocities by each method; mid and bottom rows: CFD-based velocities versus estimated velocities derived from various methods. For mid and bottom rows, velocity data from 100 simulated color Doppler images were pooled. The binned scatter plots show the number of velocity occurrences.

#### 5.4.3.2 NN-based approaches aligned with the original *i*VFM

All methods achieved high correlation in the radial velocity estimation on the 100 simulated color Doppler images derived from CFD #4, with $r^2_{v_r} > 0.98$ (see Fig. 5.4). For angular velocity correlation, both PINNs, RB-PINNs and AL-PINNs performed similarly to *i*VFM, $r^2_{v_\theta} = 0.659$ and 0.669 versus 0.694, while nnU-Net surpassed *i*VFM ($r^2_{v_\theta} = 0.744$). This suggests the effectiveness of a supervised approach in learning intraventricular blood flow patterns. However, nnU-Net tended to be less precise when estimating highly negative radial velocities on simulated Doppler data, potentially due to the limited CFD training samples. Interestingly, all NN methods exhibited more errors for the radial component than *i*VFM, highlighting the high robustness and precision of the physics-constrained *i*VFM approach (see Fig. 5.5). The nRMSE of *i*VFM ranged between 0.2%–1.6% and 1.4%-21.3% for the radial and angular velocities, respectively. Among all methods, AL-PINNs had the highest nRMSE for angular velocities (2.2% to 23.3%). Despite having the highest nRMSE for radial velocities (3.8% to 6.7%), nnU-Net produced the least errors in angular velocity estimation (2.3% to 13.3%). A cineloop showing the reconstructed field by PINNs and nnU-Net versus CFD can be found in the Supplementary Material.

The final optimized values of the penalty coefficients were (median ± robust std. [min, max]): $\mu_1 = 0.90 \pm 0.11$ [0.32, 1.00], $\mu_2 = 0.90 \pm 0.12$ [0.32, 1.02], $\mu_3 = 1.19 \pm 0.22$ [0.99, 2.36] for RB-PINNs, and $\mu = 2.02 \pm 0.01$ [2.01, 2.07] for AL-PINNs.

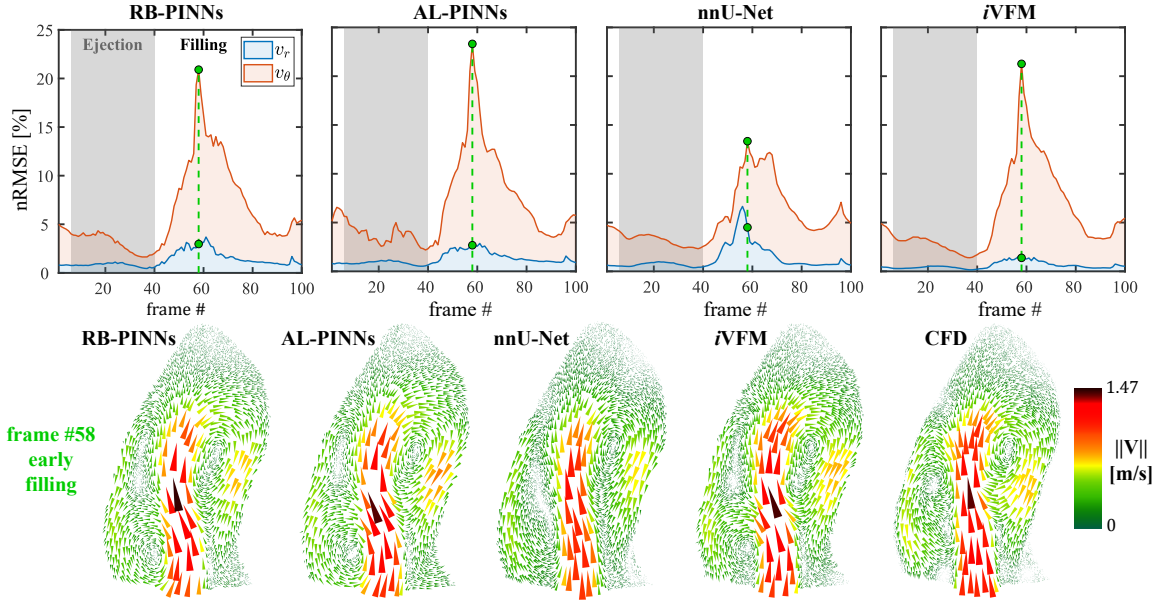



Fig. 5.5 Normalized root-mean-square errors (nRMSE) between CFD-based and estimated velocity vectors by different techniques.

Table 5.2 Metrics computed on 100 full and sparse simulated Doppler images

| Methods | Full data | | | Sparse data* | | |
|---|---|---|---|---|---|---|
| | $r^2(\uparrow)$ | | nRMSE [%]($\downarrow$) ($\tilde{x} \pm \sigma_{\text{rob}}$) | $r^2(\uparrow)$ | | nRMSE [%]($\downarrow$) ($\tilde{x} \pm \sigma_{\text{rob}}$) |
| | $v_r$ | $v_\theta$ | | $v_r$ | $v_\theta$ | |
| *i*VFM | **1.00** | 0.69 | **1.7** ± 1.0 | 0.85 | 0.58 | 3.2 ± 2.3 |
| RB-PINNs | 0.99 | 0.66 | 2.2 ± 1.0 | 0.86 | 0.56 | 3.4 ± 1.6 |
| AL-PINNs | 0.99 | 0.67 | 2.5 ± 1.0 | 0.80 | 0.58 | 3.5 ± 1.4 |
| nnU-Net‡ | 0.99 | 0.60 | 2.3 ± 0.9 | 0.67 | 0.12 | 6.0 ± 3.0 |
| nnU-Net† | 0.99 | **0.74** | 2.1 ± 0.9 | 0.64 | 0.49 | 6.0 ± 3.3 |
| nnU-Net | 0.98 | **0.74** | 2.1 ± 0.9 | **0.88** | **0.71** | **2.4** ± 1.0 |

* indicates data masked every 9 out of 10 scanlines from the center to the borders.
‡ means training without both physical regularization term (PDE loss) and random scanline masking augmentation.
† signifies training with physical regularization term (PDE loss), but without random scanline masking augmentation.

Table 5.3 Comparison of training, optimization, and inference times for NN-based methods and *i*VFM.

| Methods | Device | No. trainable parameters | Training time | Optimization/ inference time per frame |
|---|---|---|---|---|
| *i*VFM | CPU | - | - | 0.2 s |
| RB-PINNs | GPU | 18.6 k | - | 100 s |
| AL-PINNs | GPU | 18.6 k | - | 100 s |
| nnU-Net | GPU | 7 M | 12 h | **0.05 s** |

#### 5.4.3.3 nnU-Net demonstrated better generalizability and robustness on sparse Doppler data

Table 5.2 presents metrics for each method on both full and sparse simulated Doppler images. In the evaluation on full data, *i*VFM achieved the highest correlation for radial velocities and the lowest nRMSE, while nnU-Net, trained with the physical regularization term (PDE loss), excelled in the correlation of angular velocities. This implies that incorporating physical regularization helps constrain the nnU-Net's output to better adhere to the laws of physics.

As expected, the performance of all methods was significantly impacted when evaluated on sparse data, where nine out of ten scanlines were masked. Remarkably, nnU-Net, which was trained with both the physical regularization term and the random scanline

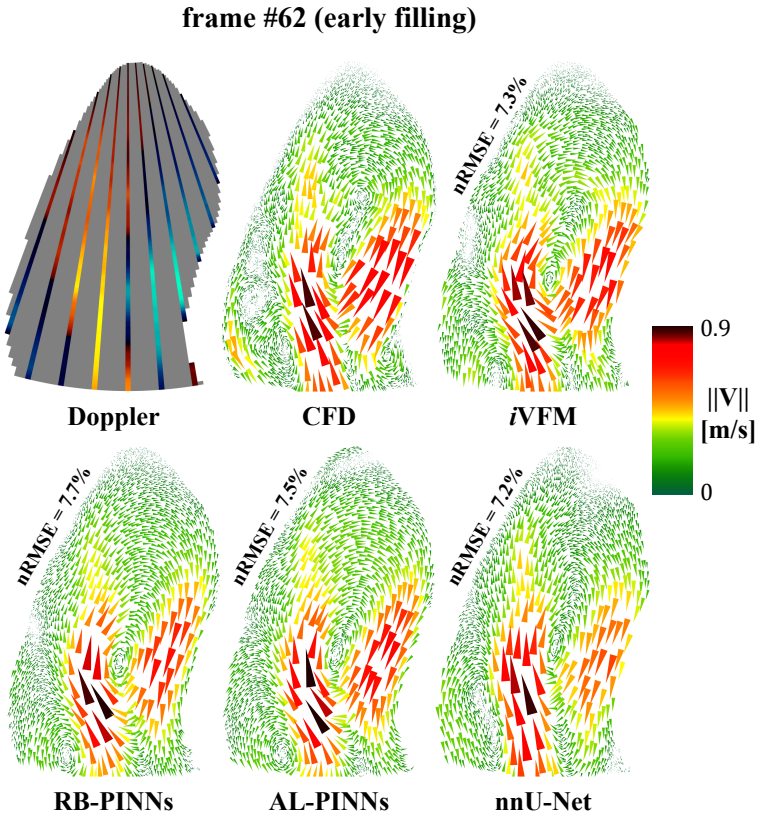



Fig. 5.6 Reconstruction of intraventricular vector blood flow from a CFD-simulated sparse Doppler with only 10 scanlines using NN-based approaches and *i*VFM.

masking augmentation, demonstrated the least decline in performance. It maintained a high correlation for both radial and angular velocities while achieving the least nRMSE. This finding underscores the benefit of this augmentation in supervised learning for enhanced generalization. Despite having lower correlations than nnU-Net, *i*VFM remained robust, producing a lower nRMSE than RB-PINNs and AL-PINNs. Fig. 5.6 shows the vector blood flow reconstruction performed by each method from a CFD-simulated sparse Doppler with only 10 scanlines during early filling. In this example with more scanlines, all methods performed similarly, with nnU-Net having a slight lead in the nRMSE. Another example of vector flow reconstruction from the same frame but with further reduced scanlines down to six is illustrated in Fig. 5.7. In this extreme scenario, RB-PINNs, AL-PINNs, and *i*VFM generated similar flow patterns with comparable nRMSE values. Although not perfect, the flow derived from nnU-Net had the lowest nRMSE when compared to the CFD-based velocities.

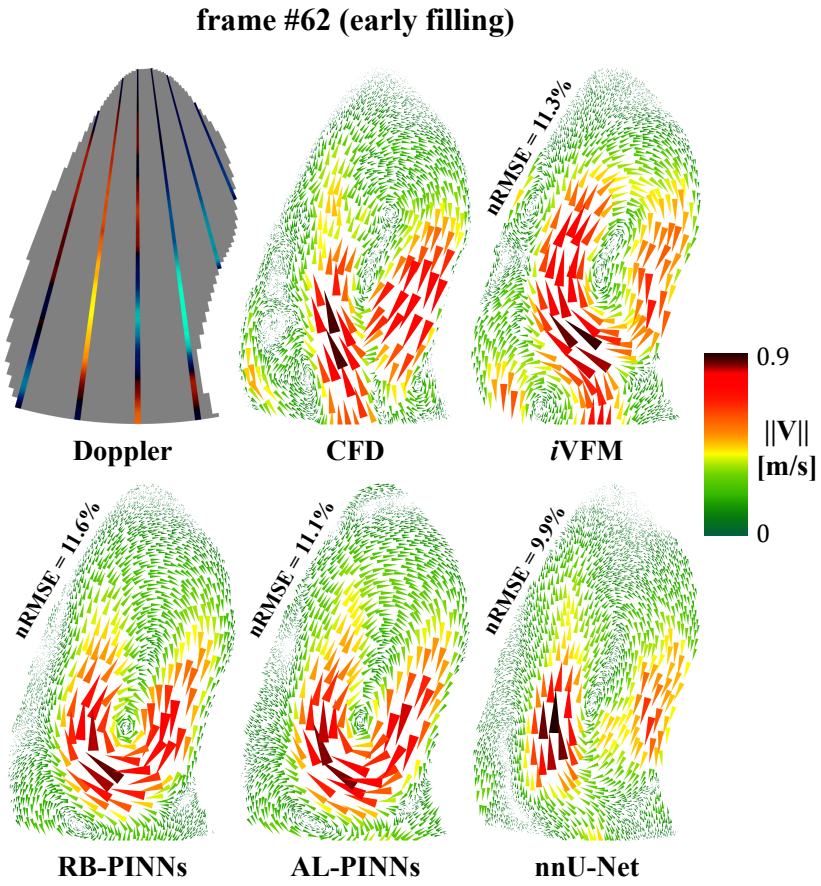



Fig. 5.7 Reconstruction of intraventricular vector blood flow from a CFD-simulated sparse Doppler with only 6 scanlines using NN-based approaches and *i*VFM.

#### 5.4.3.4 nnU-Net exhibited superior reconstruction speed

Table 5.3 provides a comparison of the training, optimization, and inference times for NN-based approaches against *i*VFM. These metrics were computed using a 16 GB V100 GPU for NN methods and an Intel i5-11500H CPU for *i*VFM. As the only supervised approach in the comparison, nnU-Net required 12 hours of training but achieved the fastest per-frame inference time, taking only 0.05 second. *i*VFM ranked second, with a reconstruction time of 0.2 second per frame. Notably, both PINNs necessitated longer optimization times, around 100 seconds, which is a recognized drawback of this approach.

#### 5.4.3.5 Clinical application of vector blood flow mapping

Fig. 5.8 showcases the intraventricular vector blood flow mapping by various methods on an in vivo case at different cardiac phases, including ejection, early filling, diastasis, and late filling. The reconstructed flow patterns by all methods appear relatively similar, with

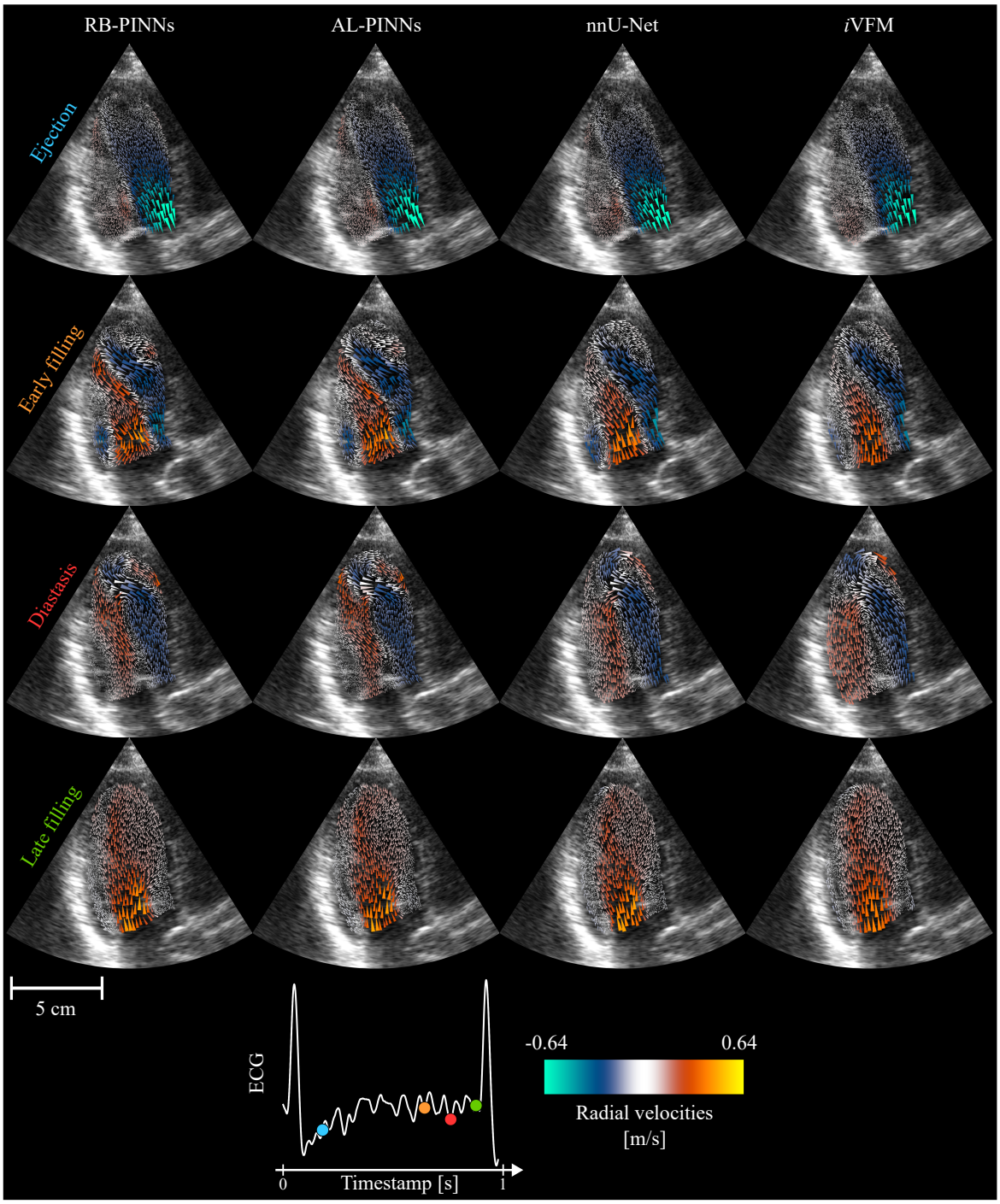



Fig. 5.8 Reconstruction of intraventricular vector blood flow in a patient using NN-based approaches and *i*VFM. The color of the arrows represents the estimated radial velocity fields.

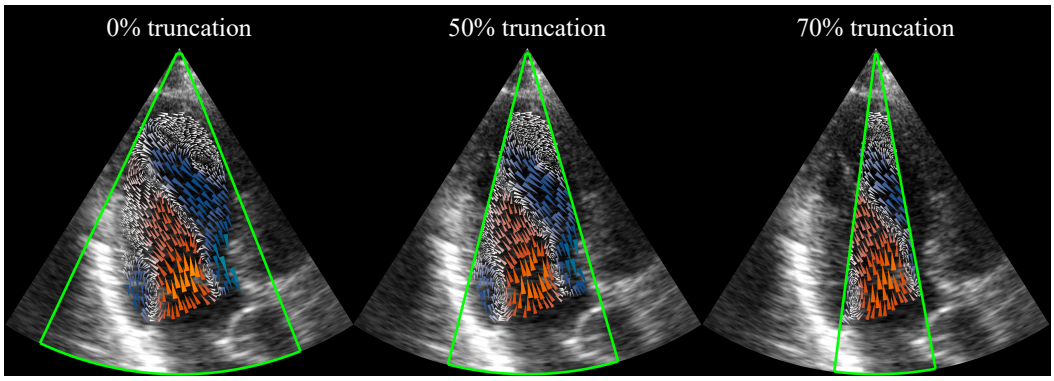



Fig. 5.9 Intraventricular vector blood flow reconstruction from Doppler data with varying percentages of scanline truncation using physics-guided nnU-Net. The color of the arrows represents the estimated radial velocity fields.

*i*VFM generating the smoothest flow patterns. Although the prominent vortex was less visible during early filling, it became more pronounced at the center of the left ventricle cavity during diastasis. Another example of vector blood flow reconstruction by all four methods on in vivo data is given in the Supplementary Material ▶.

#### 5.4.3.6 Robustness of nnU-Net on truncated clinical Doppler data

Unlike other methods that required explicit boundary conditions, nnU-Net learned these conditions implicitly during training. This offered an advantage as it reduced the need for specific knowledge about the flow at the endocardium. Table 5.4 illustrates nnU-Net's behavior under Doppler scanline truncation, achieved by progressively cutting scanlines from both sides towards the center. The metrics were computed within the common region remaining after truncation. The results show stable performance up to a 50% reduction. However, beyond this threshold, a more pronounced decrease in performance was observed. Fig. 5.9 visually demonstrates nnU-Net's ability to consistently produce accurate intraventricular vector blood flow reconstructions, even with a significant 70% truncation.

## 5.5 Discussions

Our study introduces alternative approaches to the physics-constrained *i*VFM algorithm [8], leveraging the power of neural networks: physics-informed neural networks (RB-PINNs and AL-PINNs) and a physics-guided supervised technique (nnU-Net). These methods offer distinct strategies for the inherent constrained optimization problem in *i*VFM.

Table 5.4 nnU-Net's metrics computed on 8 in vivo test cineloops of 197 frames with different percentages of scanline truncation

| Percentage of truncation [%] | $r^2(\uparrow)$ | | nRMSE [%]($\downarrow$) ($\tilde{x} \pm \sigma_{\text{rob}}$) |
|---|---|---|---|
| | $v_r$ | $v_\theta$ | |
| 20 | 0.97 | 0.96 | $2.4 \pm 1.2$ |
| 40 | 0.99 | 0.92 | $3.7 \pm 1.5$ |
| 50 | 0.99 | 0.87 | $4.3 \pm 1.6$ |
| 60 | 0.97 | 0.75 | $6.5 \pm 2.7$ |
| 70 | 0.94 | 0.59 | $8.4 \pm 2.9$ |

Note: Comparison made with nnU-Net's estimated velocity fields on full scanline data.

In the PINNs framework, we addressed the same optimization problem as in *i*VFM, but solved it differently using gradient descent and neural networks. By incorporating governing equations, such as mass conservation and boundary conditions, PINNs inherently enforce physical laws during optimization, potentially leading to optimized intraventricular vector velocity fields.

On the other hand, the supervised approach (nnU-Net) was trained on patient-specific CFD-derived simulations and *i*VFM-estimated velocity field on in vivo Doppler data. The network learned the underlying flow patterns while adhering to physical principles through the use of physics-constrained labels and the physical regularization term in the loss function. This approach demonstrated robustness to data limitations, such as missing scanlines.

### 5.5.1 PINNs versus physics-guided nnU-Net

The application of PINNs deviates from the conventional optimization methods by leveraging neural networks to find the optimal solution, which can be advantageous in complex physical problems. Although PINNs may not necessarily outperform analytical or numerical methods in computational efficiency, approximation accuracy, or convergence guarantees [141], they offer a unique advantage in terms of flexibility. This flexibility allows their architecture to remain relatively consistent across various physical optimization problems by adapting the loss functions to be optimized.

In our case of intraventricular blood flow reconstruction, we successfully improved the computational efficiency of PINNs while maintaining accuracy comparable to *i*VFM. This was achieved by implementing a dual-stage optimization with the use of pre-optimized weights. For future exploration, both PINNs architectures could benefit from imposing hard boundary conditions rather than optimizing them in the form of soft constraints. Additionally, better

strategies for automatically determining or learning the optimal smoothing regularization weight, rather than relying on heuristic search methods, will be investigated to further improve the robustness of PINNs.

Unlike PINNs, nnU-Net operates within a supervised learning framework, heavily relying on labeled training data. In this study, CFD-derived simulations played a crucial role in providing ground truth velocity fields for training. This explains the high squared correlation achieved by nnU-Net, as the simulated training samples shared the exact heart geometry despite variations in flow patterns due to different mitral valve conditions.

However, when trained exclusively on simulated data, nnU-Net struggled to correctly estimate vector blood flow in in vivo color Doppler data due to a distribution shift between the simulated and real data. This arose from limitations in the current physiological spectrum of the simulations. To bridge this gap, we included *i*VFM-estimated velocities in our training data, allowing nnU-Net to learn from solutions representative of clinical Doppler data generated by the established *i*VFM method. As shown in Fig. 5.8, nnU-Net effectively learned the underlying flow pattern and physical properties, i.e., the free-divergence and boundary conditions, from the training samples generated by *i*VFM.

Moreover, as illustrated in Fig. 5.9, nnU-Net can precisely reconstruct intraventricular flow on truncated Doppler data, where PINNs and *i*VFM cannot be directly applied as such due to incomplete and unknown boundary conditions. This potentially makes nnU-Net the preferred candidate for clinical applications, especially considering that most clinical Doppler acquisitions do not capture the entire left ventricular cavity due to limitations in probe placement or patient anatomy. With the added advantage of the shortest inference time, nnU-Net has real-time capabilities suitable for clinical settings. Future work will focus on generating more patient-specific CFD models and creating more realistic simulated Doppler data [118] to avoid the bias associated with using *i*VFM estimates as a reference to train our model.

### 5.5.2 Limitations of color Doppler and vector flow mapping

Conventional color Doppler echocardiography is subject to various limitations, posing challenges for accurate vector flow mapping. These limitations include clutter signals arising from myocardial tissue and valve leaflets, aliasing artifacts caused by Doppler velocity overshooting beyond Nyquist velocity, and low spatial and temporal resolutions. While the impact of clutter signal filtering on flow reconstruction is acknowledged, its specific effects were not investigated in this study. Our input data were already clutter-filtered (*in vivo* non scan-converted data from a GE scanner) or clutter-free (simulated data). Some researchers address clutter filtering in color flow imaging using DL techniques [142].

The clinical color Doppler data used for training and testing in this study contained only single aliasing, corrected by a DL-based unwrapping algorithm from our previous work [19]. It is important to note that this algorithm may face limitations in scenarios involving multiple aliasing, particularly in valvular disease. A potential solution could involve training a supervised DL model with multi-aliased data and their alias-free labels, such as in interferometric imaging [143] or color Doppler imaging of the femoral bifurcation [106].

The temporal resolution of clinical color Doppler, typically ranging from 10 to 15 frames per second, significantly restricts the use of temporal information. Consequently, the application of physical constraints is limited to those that are not time-dependent, such as mass conservation for an incompressible fluid. Around 25 harmonics are necessary to accurately record pressure time derivatives within the left ventricle with a 5% margin of error [144]. Hence, a color Doppler frame rate of 25 should be ideally sought to characterize intracardiac blood flow. Although we did not address such a strategy in this study, one approach would be to use Doppler information from two or three successive cardiac cycles. To overcome these limitations, advances in increasing the frame rate of color Doppler imaging, such as diverging scan sequences [23, 145] or DL [146], have been explored. Another promising avenue involves utilizing multi-line transmission [147], especially given the robust performance of our models on sparse Doppler data, notably nnU-Net. Higher frame rates offer the possibility of incorporating more complex physical constraints, including vorticity, Euler, or Navier-Stokes equations, potentially enhancing flow reconstruction accuracy. While this might pose challenges for the original *i*VFM due to the non-linear terms in the equations, it aligns seamlessly with PINNs, requiring minimal changes to the loss function.

### 5.5.3 Future directions

The successful application of PINNs in mapping intraventricular vector flow from color Doppler paves the way for future investigations. Upcoming research will prioritize the integration of high-frame-rate color Doppler with PINNs while incorporating the governing Navier-Stokes equations. This combination aims to leverage the temporal information to obtain a more accurate velocity field and the pressure gradient within the left ventricle.

Furthermore, with the development of our fully automated and robust tools, including left ventricular segmentation, dealiasing, and velocity field reconstruction using neural networks, we anticipate extracting potential biomarkers from intracardiac vector blood flow for enhanced clinical insights and diagnostic capabilities.

## 5.6 Conclusion

In this final contribution chapter, we proposed novel neural network-based approaches for intraventricular vector flow mapping, utilizing gradient-based optimization through PINNs (RB-PINNs and AL-PINNs) and physics-guided supervised learning (nnU-Net). These methods offer contrasting strategies to tackle the ill-posed inverse problem of vector flow reconstruction.

For the data-driven optimization approach using PINNs, we decomposed the constrained optimization problem into a series of unconstrained loss functions. We examined two different weighting strategies, ReLoBRaLo and AL, to balance each loss function, both yielding similar reconstruction results. Given that PINN-based methods were slow to converge, we employed pre-optimized weights and a dual-stage optimization strategy to enhance their efficiency.

In contrast, the physics-guided nnU-Net was trained using a physics-constrained dataset. Despite the limited size of our training dataset, the inclusion of a physical regularization term—specifically, the mass conservation equation—significantly enhanced its reconstruction performance. nnU-Net also demonstrated quasi-real-time capability, robustness with sparse Doppler data, and independence from explicit boundary conditions. These characteristics position nnU-Net as a promising solution for real-time clinical applications.

In the following part, “Conclusion and Perspectives”, we discuss the limitations of the studies conducted within this thesis and provide perspectives on potential improvements, particularly for *i*VFM using PINNs, as well as future research directions.

# Part IV

# Conclusion and Perspectives

# Conclusion

In clinical practice, medical image analysis plays a pivotal role in diagnosing various pathologies. Cardiovascular pathology diagnosis, in particular, relies heavily on echocardiography due to its affordability, portability, and real-time imaging capabilities. Over the past decade, the segmentation of cardiac structures such as the LV, myocardium, and LA on echocardiographic B-mode images using DL has been extensively researched. This has contributed to increasing the accuracy and robustness of estimating clinical indices like EF and GLS, which are crucial for evaluating systolic function.

Conversely, diastolic function is typically assessed using parameters such as the E/A and E/e' ratios. However, these parameters can sometimes give distinct diagnostic conclusions and might not detect subtle changes in blood flow dynamics that occur in the early stages of cardiac diseases. Analyzing intracardiac blood flow, particularly the characteristics of vortices, can aid in the early detection of cardiovascular diseases. This is because even slight changes in heart geometry due to diseases could have a direct and immediate impact on the flow patterns, notably during diastole.

Among the methods for deriving intracardiac vector blood flow, *i*VFM is uniquely positioned as the only postprocessing method directly applicable to clinical color Doppler acquisitions without necessitating specialized ultrasound machines like BST or microbubble injections as in echo-PIV. However, most *i*VFM techniques depend on time-consuming semi-automatic or manual preprocessing steps—notably, LV segmentation and the dealiasing of color Doppler images—to accurately reconstruct the vector velocity field. Automating these preprocessing procedures is crucial for making *i*VFM a viable and reliable tool in clinical settings for discovering and extracting new flow-based biomarkers, which may be an early predictor of cardiovascular diseases like hypertension.

To this end, this thesis started by developing a robust and precise DL-based segmentation tool, which produces temporally consistent LV segmentation. This was achieved by considering the temporal dimension of the echocardiographic cineloops as the third dimension since the sequences present smooth and regular properties along the temporal axis. The proof of concept of learning temporal consistency through 3D convolutions in 3D nnU-Net

was done on an in-house A2C and A4C CARDINAL dataset, which were annotated and corrected automatically with DL models. Due to the lack of public A3C datasets, the view in which *i*VFM is usually performed, we resorted to active learning for annotating a new private A3C dataset. Given the relatively small dataset size, transfer learning from the CARDINAL dataset improved our 3D nnU-Net performance in better segmenting the LV, serving as the boundary conditions for *i*VFM.

The second step consisted of removing the aliasing artifacts on color Doppler images automatically. Dealiasing can be done in two ways: either by solving an inverse problem or a multi-class segmentation problem to segment the aliased pixels. For the former, we utilized a primal-dual-based deep unfolding technique, namely PDNet. As for the latter, we opted for 2D nnU-Net and BATFormer. Again, nnU-Net came up on top for color Doppler dealiasing. PDNet, with 233 times fewer parameters, followed closely behind nnU-Net, showing that simpler DL models were capable of achieving comparable performance in color Doppler dealiasing by unrolling a well-defined inverse problem.

With these two automated tools being developed, we then explored PINNs and physics-guided nnU-Net to further enhance the *i*VFM-v3 approach [8]. We relied on PINNs to solve the same constrained optimization as in *i*VFM-v3 and obtained similar results. To improve the computational efficiency of our PINNs, we adopted a dual-stage optimization strategy and pre-optimized weights. Physics-guided nnU-Net, on the other hand, was trained in a supervised manner using a physics-constrained dataset and incorporated a physical regularization term, which helped the model better estimate the intracardiac vector blood flow. We introduced a novel random scanline masking strategy that enhanced the generalizability of nnU-Net on sparse Doppler data. Furthermore, our physics-guided nnU-Net possessed two unbeatable advantages over the PINN-based and *i*VFM-v3 approaches: i) independence from explicit boundary conditions, making it applicable to truncated color Doppler data with high accuracy; ii) the shortest inference time of only 0.05 s.

This thesis marks a step forward in developing a fully automatic, fast, and robust clinical tool using AI for extracting new flow-based biomarkers for the early detection of cardiovascular diseases. While our results showcased the fairly good performance of AI-based methods for each task described earlier, there is still room for improvement, especially for the A3C LV segmentation and vector flow mapping. In the next and final section, we elaborate on some potential perspectives for future work.

# Perspectives for Future Work

## Foundation models for large-scale echocardiographic data annotation

Our data annotation pipeline, with minimal human intervention as explained in Section 3.4.1.2, has been proven to be effective in generating gold standard annotations for small to medium-scale datasets. Building on the success of our 3D nnU-Net and other DL models mentioned in Section 2.2.7.1, which consistently achieved performance within the intra-observer variability, future research should focus on developing and refining foundational models specifically tailored for echocardiographic segmentation. These models, usually publicly available, could facilitate the annotation of larger datasets, particularly with SAM-based models using point and bounding box prompts.

The development of SAMUS and MemSAM, as described in Section 2.2.7.1, marks a significant milestone in training foundation models for ultrasound image segmentation. SAMUS fine-tuned the pre-trained SAM weights on public ultrasound segmentation datasets, while MemSAM further refined SAMUS weights for echocardiographic segmentation. Despite showing excellent generalizability with the introduction of memory prompts, MemSAM's performance did not reach the intra-observer variability, especially for clinical metrics. This indicates that the currently available public echocardiographic datasets may be limited in size, variability, and most importantly, segmentation label quality. Public dataset labels are often annotated by different experts, resulting in segmentation masks with high variability, which complicates training a robust segmentation model.

Cardiologists should utilize these SAM-based models as an easy-to-use semi-automatic tool to establish an annotation consensus. This approach would help reduce inter-observer variability in segmentation labels, creating data with high-quality and consistent labels. With the release of more public datasets with labels that strictly follow the annotation consensus, we could fine-tune the aforementioned foundational models or train specialized models like nnU-Net from scratch. This should enhance the models' cross-vendor and cross-center

generalizability, easing the challenges of segmenting images acquired with varying qualities and parameters.

Following this path, robust echocardiographic segmentation models for all views—not just the common A2C and A4C—could be developed and deployed in clinics. This would improve the accuracy and consistency of echocardiographic image segmentation, leading to better analysis and ultimately better patient outcomes.

## Generalizability of color Doppler dealiasing

In Chapter 4, we trained DL models for color Doppler dealiasing, but these models are limited to correcting single aliasing. Certain valvular diseases can lead to multiple aliasing in the left ventricle cavity. For example, mitral regurgitation can produce a highly aliased jet near the mitral valves due to the backflow of blood from the LV into the LA. Such cases may be difficult even for manual dealiasing, posing challenges in including such samples in the training dataset to extend the generalizability of the models to pathological cases.

One potential solution is simulation. The alias-free label can be obtained using a multi-PRF sequence [23], avoiding the compromise of reduced imaging depth by increasing the Nyquist velocity during acquisition. The aliased data can be acquired using standard color Doppler sequences.

## Robustness of intraventricular vector flow mapping

Our study in Chapter 5 used a simple physical constraint, the mass conservation equation, due to the frame rate limitations of conventional color Doppler acquisitions. To utilize fluid dynamics equations that better represent real-world blood flows, such as vorticity and Navier-Stokes (or Euler) equations, we need to increase the frame rate. This can be achieved through ultrafast sequences like plane waves, diverging waves, or multi-line transmissions. While plane and diverging scan sequences can boost the color Doppler frame rate, they might produce noisier Doppler data near the mitral valves due to clutters caused by the mitral leaflets. However, this may not be an issue for *i*VFM incorporated with Navier-Stokes or Euler equations, which include more regularization terms than mass conservation equations. Multi-line transmissions can also be considered, given the great reconstruction performance of our proposed methods (RB-PINNs, AL-PINNs, and physics-guided nnU-Net) on sparse Doppler data mimicking multi-line transmission acquisitions.

Another possible avenue is merging color Doppler frames from multiple cardiac cycles by performing Fourier interpolation, as proposed in a recent study [24]. This study has also

shown that prior dealiasing of color Doppler images is not necessary with the incorporation of Navier-Stokes equations and a $2\pi$-insensitive data fidelity term.

Incorporating these fluid dynamic equations requires minimal changes to the loss function to be optimized in our PINNs, allowing for the output of the pressure field up to a constant factor. Pressure gradients derivable from the pressure field can be biomarkers for various cardiac diseases [120], including aortic stenosis [148], dilated cardiomyopathy [149], diastolic dysfunction [150], and heart failure with preserved EF [151].

Training a physics-guided supervised model to output pressure fields or gradients is far more complicated than optimizing PINNs, as it relies on a training dataset with pressure field or gradient labels, which can only be obtained via simulations, since measuring pressure in vivo requires invasive cardiac catheterization. Our current CFD simulations can provide pressure field labels but are limited in terms of physiological spectrum to generate realistic color Doppler data similar to clinical acquisitions. Potential solutions include adding myocardial muscles to simulate myocardial movement and papillary muscles to have a more realistic intracardiac blood flow.

## Clinical validation of intraventricular vector flow mapping

Clinical validation of *i*VFM is essential to demonstrate the utility and clinical significance of reconstructing intracardiac vector blood flow. Therefore, a thorough population characterization is necessary to assess the feasibility of detecting cardiovascular diseases using flow-based biomarkers or directly from color Doppler data. The NN-based methods developed in this thesis can be combined to create a fully automated *i*VFM tool for clinical use. This enables automatic data processing to generate datasets with reconstructed vector blood flow and patient diagnosis.

A preliminary study at the end of this thesis showed that 3D CNN-based models could distinguish hypertensive and healthy cases directly from color Doppler sequences (2D+t input) with an F1-score of 0.6 by learning from only 22 samples. More accurate analyses and conclusions can be drawn when more data is available. A multi-modal approach is also interesting for pathology classification by combining color Doppler with the reconstructed vector velocity field and other parameters derived from the vector field, such as vorticity and stream function. Combined with attribution techniques and attention heatmap generation, we can aim to explain the models' predictions and decisions given the input features. Existing methods include Grad-CAM [152], which is based on backpropagation of the gradients of the deepest NN layers, and the Deep Taylor Decomposition-based technique [63], tailored for attention-based architectures like transformers.

Analyzing the attention heatmaps of neural networks can provide insights into the local features of the input data that DL models use for their classification decisions. Superimposing these heatmaps on the reconstructed vector velocity field or flow-based parameters like stream function can help identify potential biomarkers for classifying pathologies.

# Part V

# Appendices

# Appendix A

# Financial Support

This Ph.D. thesis is funded in part by:

- The MEGA Doctoral School (ED 162);
- The French National Research Agency (ANR) through the “4D-iVFM” [ANR-21-CE19-0034-01] and “ORCHID” [ANR-22-CE45-0029-01] projects;
- The LABEX PRIMES [ANR-11-LABX-0063] and LABEX CELYA [ANR-10-LABX-0060] of Université de Lyon within the “Investissements d’Avenir” [ANR-11-IDEX-0007] program.

Most of the work in this thesis has been performed using the High-Performance Computing (HPC) resources of IDRIS under the grants 2023-AD010313603 and 2023-AD010313603R1 made by GENCI.

# Appendix B

# IEEE Notice

With respect to IEEE copyrighted material which is used with permission in this thesis, IEEE does not endorse any products or services of INSA Lyon. Internal or personal use of this material is permitted. For those who are interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes, or to create new collective works for resale or redistribution, please go to https://www.ieee.org/publications/rights/rights-link.html to learn how to obtain a license from RightsLink.

# References


[1] World Health Organization, "Cardiovascular diseases (CVDs)," Jun. 2021, retrieved from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] D. Garcia, P. Lantelme, and E. Saloux, "Introduction to speckle tracking in cardiac ultrasound imaging," in *Handbook of Speckle Filtering and Tracking in Cardiovascular Ultrasound Imaging and Video*. IET Digital Library, Jan. 2018, pp. 571–598.

[3] S. A. Nyrnes, S. Fadnes, M. S. Wigen, L. Mertens, and L. Lovstakken, "Blood Speckle-Tracking Based on High–Frame Rate Ultrasound Imaging in Pediatric Cardiology," *Journal of the American Society of Echocardiography*, vol. 33, no. 4, pp. 493–503.e5, Apr. 2020.

[4] A. S. Daae, M. S. Wigen, S. Fadnes, L. Løvstakken, and A. Støylen, "Intraventricular Vector Flow Imaging with Blood Speckle Tracking in Adults: Feasibility, Normal Physiology and Mechanisms in Healthy Volunteers," *Ultrasound in Medicine & Biology*, vol. 47, no. 12, pp. 3501–3513, Dec. 2021.

[5] D. Garcia *et al.*, "Two-Dimensional Intraventricular Flow Mapping by Digital Processing Conventional Color-Doppler Echocardiography Images," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1701–1713, Oct. 2010.

[6] B. A. Meyers, C. J. Goergen, P. Segers, and P. P. Vlachos, "Colour-Doppler echocardiography flow field velocity reconstruction using a streamfunction–vorticity formulation," *Journal of The Royal Society Interface*, vol. 17, no. 173, p. 20200741, Dec. 2020.

[7] K. C. Assi *et al.*, "Intraventricular vector flow mapping-a Doppler-based regularized problem with automatic model selection," *Physics in Medicine & Biology*, vol. 62, no. 17, pp. 7131–7147, Aug. 2017.

[8] F. Vixège *et al.*, "Physics-constrained intraventricular vector flow mapping by color Doppler," *Physics in Medicine & Biology*, vol. 66, no. 24, p. 245019, Dec. 2021.

[9] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[10] A. Kirillov *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 4015–4026.

[11] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Jun. 2010, pp. 399–406.

[12] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, Feb. 2019.

[13] A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling," Sep. 2021, arXiv:1710.11431 [cs.LG].

[14] S. A. Faroughi *et al.*, "Physics-Guided, Physics-Informed, and Physics-Encoded Neural Networks and Operators in Scientific Computing: Fluid and Solid Mechanics," *Journal of Computing and Information Science in Engineering*, vol. 24, no. 040802, Jan. 2024.

[15] N. Painchaud, N. Duchateau, O. Bernard, and P.-M. Jodoin, "Echocardiography Segmentation With Enforced Temporal Consistency," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2867–2878, Oct. 2022.

[16] S. Leclerc *et al.*, "Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019, conference Name: IEEE Transactions on Medical Imaging.

[17] H. J. Ling, D. Garcia, and O. Bernard, "Reaching intra-observer variability in 2-D echocardiographic image segmentation with a simple U-Net architecture," in *IEEE International Ultrasonics Symposium (IUS)*, 2022.

[18] 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*, 2023, pp. 245–254.

[19] H. J. Ling, O. Bernard, N. Ducros, and D. Garcia, "Phase Unwrapping of Color Doppler Echocardiography using Deep Learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 8, pp. 810–820, Aug. 2023.

[20] H. J. Ling *et al.*, "Physics-Guided Neural Networks for Intraventricular Vector Flow Mapping," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024, doi: 10.1109/TUFFC.2024.3411718.

[21] X. Lin, Y. Xiang, L. Zhang, X. Yang, Z. Yan, and L. Yu, "SAMUS: Adapting Segment Anything Model for Clinically-Friendly and Generalizable Ultrasound Image Segmentation," Sep. 2023, arXiv:2309.06824 [cs].

[22] X. Deng, H. Wu, R. Zeng, and J. Qin, "MemSAM: Taming Segment Anything Model for Echocardiography Video Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 9622–9631.

[23] D. Posada *et al.*, "Staggered Multiple-PRF Ultrafast Color Doppler," *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1510–1521, Jun. 2016.

[24] B. Maidu *et al.*, "Super-resolution Left Ventricular Flow and Pressure Mapping by Navier-Stokes-Informed Neural Networks," Apr. 2024, bioRxiv 2024.04.12.589319.

[25] E. D. Folland, A. F. Parisi, P. F. Moynihan, D. R. Jones, C. L. Feldman, and D. E. Tow, "Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. A comparison of cineangiographic and radionuclide techniques," *Circulation*, vol. 60, no. 4, pp. 760–766, Oct. 1979.

[26] R. M. Lang *et al.*, "Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging," *Journal of the American Society of Echocardiography*, vol. 28, no. 1, pp. 1–39.e14, Jan. 2015.

[27] P. M. Mottram and T. H. Marwick, "Assessment of diastolic function: what the general cardiologist needs to know," *Heart*, vol. 91, no. 5, pp. 681–695, May 2005.

[28] T. Marinko, J. Dolenc, and C. Bilban-Jakopin, "Cardiotoxicity of concomitant radiotherapy and trastuzumab for early breast cancer," *Radiology and Oncology*, vol. 48, no. 2, pp. 105–112, Jun. 2014.

[29] S. F. Nagueh *et al.*, "Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography," *European Journal of Echocardiography*, vol. 10, no. 2, pp. 165–193, Mar. 2009.

[30] M. Gharib, E. Rambod, A. Kheradvar, D. J. Sahn, and J. O. Dabiri, "Optimal vortex formation as an index of cardiac health," *Proceedings of the National Academy of Sciences*, vol. 103, no. 16, pp. 6305–6308, Apr. 2006.

[31] G. Pedrizzetti, G. La Canna, O. Alfieri, and G. Tonti, "The vortex—an early predictor of cardiovascular outcome?" *Nature Reviews Cardiology*, vol. 11, no. 9, pp. 545–553, Sep. 2014.

[32] P. M. Arvidsson *et al.*, "Vortex ring behavior provides the epigenetic blueprint for the human heart," *Scientific Reports*, vol. 6, no. 1, p. 22021, Feb. 2016.

[33] M. S. M. Elbaz *et al.*, "Assessment of viscous energy loss and the association with three-dimensional vortex ring formation in left ventricular inflow: In vivo evaluation using four-dimensional flow MRI," *Magnetic Resonance in Medicine*, vol. 77, no. 2, pp. 794–805, Feb. 2017.

[34] C. Chnafa, S. Mendez, and F. Nicoud, "Image-Based Simulations Show Important Flow Fluctuations in a Normal Left Ventricle: What Could be the Implications?" *Annals of Biomedical Engineering*, vol. 44, no. 11, pp. 3346–3358, Nov. 2016.

[35] D. Mele *et al.*, "Intracardiac Flow Analysis: Techniques and Potential Clinical Applications," *Journal of the American Society of Echocardiography*, vol. 32, no. 3, pp. 319–332, Mar. 2019.

[36] J. Eriksson, P. Dyverfeldt, J. Engvall, A. F. Bolger, T. Ebbers, and C. J. Carlhäll, "Quantification of presystolic blood flow organization and energetics in the human left ventricle," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 300, no. 6, pp. H2135–H2141, Jun. 2011.

[37] G.-R. Hong *et al.*, "Characterization and Quantification of Vortex Flow in the Human Left Ventricle by Contrast Echocardiography Using Vector Particle Image Velocimetry," *JACC: Cardiovascular Imaging*, vol. 1, no. 6, pp. 705–717, Nov. 2008.

[38] H. Abe *et al.*, "Contrast echocardiography for assessing left ventricular vortex strength in heart failure: a prospective cohort study," *European Heart Journal - Cardiovascular Imaging*, vol. 14, no. 11, pp. 1049–1060, Nov. 2013.

[39] P. Chen *et al.*, "Blood flow patterns estimation in the left ventricle with low-rate 2D and 3D dynamic contrast-enhanced ultrasound," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105810, Jan. 2021.

[40] S. Ohtsuki and M. Tanaka, "The flow velocity distribution from the doppler information on a plane in three-dimensional flow," *Journal of Visualization*, vol. 9, no. 1, pp. 69–82, Mar. 2006.

[41] T. Uejima *et al.*, "A New Echocardiographic Method for Identifying Vortex Flow in the Left Ventricle: Numerical Validation," *Ultrasound in Medicine & Biology*, vol. 36, no. 5, pp. 772–788, May 2010.

[42] G. Pedrizzetti and F. Domenichini, "Nature Optimizes the Swirling Flow in the Human Left Ventricle," *Physical Review Letters*, vol. 95, no. 10, p. 108101, Sep. 2005.

[43] A. Kheradvar *et al.*, "Echocardiographic Particle Image Velocimetry: A Novel Technique for Quantification of Left Ventricular Blood Vorticity Pattern," *Journal of the American Society of Echocardiography*, vol. 23, no. 1, pp. 86–94, Jan. 2010.

[44] G. Pedrizzetti and G. Tonti, "Method for transforming a Doppler velocity dataset into a velocity vector field," EP Patent EP2 514 368A1, Oct., 2012.

[45] T. Tanaka *et al.*, "Intracardiac VFM technique using diagnostic ultrasound system," *Hitachi Review*, vol. 64, no. 8, pp. 488–492, 2015.

[46] Esaote, "HyperDoppler Intracardiac Flow Dynamics," Esaote SpA, White Paper, Nov. 2019.

[47] Z. Zhuang *et al.*, "Cardiac VFM visualization and analysis based on YOLO deep learning model and modified 2D continuity equation," *Computerized Medical Imaging and Graphics*, vol. 82, p. 101732, Jun. 2020.

[48] P. C. Hansen and D. P. O'Leary, "The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.

[49] F. Vixège *et al.*, "Full-volume three-component intraventricular vector flow mapping by triplane color Doppler," *Physics in Medicine and Biology*, vol. 67, no. 9, Apr. 2022.

[50] A. Gomez, K. Pushparajah, J. M. Simpson, D. Giese, T. Schaeffter, and G. Penney, "A sensitivity analysis on 3D velocity reconstruction from multiple registered echo Doppler views," *Medical Image Analysis*, vol. 17, no. 6, pp. 616–631, Aug. 2013.

[51] A. Gomez *et al.*, "4D Blood Flow Reconstruction Over the Entire Ventricle From Wall Motion and Blood Velocity Derived From Ultrasound Data," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2298–2308, Nov. 2015.

[52] P. C. Petersen, "Neural network theory," *University of Vienna*, vol. 535, 2020.

[53] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012.

[54] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," Feb. 2019, arXiv:1803.08375 [cs.NE].

[55] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)," Jun. 2023, arXiv:1606.08415 [cs.LG].

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234–241.

[58] O. Bernard *et al.*, "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.

[59] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[60] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations*, Oct. 2020.

[61] Q. Wen *et al.*, "Transformers in time series: a survey," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, Aug. 2023, pp. 6778–6786.

[62] L. Chen *et al.*, "Decision Transformer: Reinforcement Learning via Sequence Modeling," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 15 084–15 097.

[63] H. Chefer, S. Gur, and L. Wolf, "Transformer Interpretability Beyond Attention Visualization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 782–791.

[64] T. Brown *et al.*, "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1877–1901.

[65] R. Bommasani *et al.*, "On the Opportunities and Risks of Foundation Models," Jul. 2022, arXiv:2108.07258 [cs].

[66] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked Autoencoders Are Scalable Vision Learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 16 000–16 009.

[67] A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," in *Proceedings of the 38th International Conference on Machine Learning*, Jul. 2021, pp. 8748–8763.

[68] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, p. 654, Jan. 2024.

[69] J. Cheng *et al.*, "SAM-Med2D," Aug. 2023, arXiv:2308.16184 [cs].

[70] H. Wang *et al.*, "SAM-Med3D," Oct. 2023, arXiv:2310.15161 [cs].

[71] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, Mar. 2021.

[72] Z. Zhang, Y. Liu, J. Liu, F. Wen, and C. Zhu, "AMP-Net: Denoising-Based Deep Unfolding for Compressive Image Sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 1487–1500, 2021.

[73] C. Mou, Q. Wang, and J. Zhang, "Deep Generalized Unfolding Networks for Image Restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 17 399–17 410.

[74] K. Zhang, L. V. Gool, and R. Timofte, "Deep Unfolding Network for Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[75] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, "Physics-informed neural networks (PINNs) for fluid mechanics: a review," *Acta Mechanica Sinica*, vol. 37, no. 12, pp. 1727–1738, Dec. 2021.

[76] S. Cai, Z. Wang, S. Wang, P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks for Heat Transfer Problems," *Journal of Heat Transfer*, vol. 143, no. 060801, Apr. 2021.

[77] E. Haghighat, M. Raissi, A. Moure, H. Gomez, and R. Juanes, "A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics," *Computer Methods in Applied Mechanics and Engineering*, vol. 379, p. 113741, Jun. 2021.

[78] A. Baydin, B. Pearlmutter, A. Radul, and J. Siskind, "Automatic differentiation in machine learning: A survey," *Journal of Machine Learning Research*, vol. 18, pp. 1–43, Apr. 2018.

[79] S. Leclerc *et al.*, "LU-Net: A Multistage Attention Network to Improve the Robustness of Segmentation of Left Ventricular Structures in 2-D Echocardiography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2519–2530, Dec. 2020.

[80] D. Ouyang *et al.*, "EchoNet-Dynamic: a Large New Cardiac Motion Video Data Resource for Medical Machine Learning," 2019, available: https://echonet.github.io/dynamic/NeuroIPS_2019_ML4H%20Workshop_Paper.pdf.

[81] D. Ouyang *et al.*, "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, no. 7802, pp. 252–256, Apr. 2020.

[82] O. Oktay *et al.*, "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, Feb. 2018.

[83] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, "Cardiac Segmentation With Strong Anatomical Guarantees," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3703–3713, Nov. 2020.

[84] I. Higgins *et al.*, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework," in *International Conference on Learning Representations*, Jul. 2022.

[85] H. Wei, J. Ma, Y. Zhou, W. Xue, and D. Ni, "Co-learning of appearance and shape for precise ejection fraction estimation from echocardiographic sequences," *Medical Image Analysis*, vol. 84, p. 102686, Feb. 2023.

[86] E. Smistad, I. M. Salte, H. Dalen, and L. Lovstakken, "Real-time temporal coherent left ventricle segmentation using convolutional LSTMs," in *2021 IEEE International Ultrasonics Symposium (IUS)*. IEEE, Sep. 2021, pp. 1–4.

[87] 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.

[88] X. SHI, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. WOO, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[89] A. Pati and A. Lerch, "Attribute-based regularization of latent spaces for variational auto-encoders," *Neural Computing and Applications*, vol. 33, no. 9, pp. 4429–4444, May 2021.

[90] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," Dec. 2014, arXiv:1412.3555 [cs].

[91] T. S. Stevens, F. C. Meral, J. Yu, I. Z. Apostolakis, J.-L. Robert, and R. J. Van Sloun, "Dehazing Ultrasound using Diffusion Models," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024.

[92] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6840–6851.

[93] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations," Feb. 2021, arXiv:2011.13456 [cs.LG].

[94] H. Nahas, J. S. Au, T. Ishii, B. Y. S. Yiu, A. J. Y. Chee, and A. C. H. Yu, "A Deep Learning Approach to Resolve Aliasing Artifacts in Ultrasound Color Flow Imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2615–2628, Dec. 2020.

[95] A. Ghorbani *et al.*, "Deep learning interpretation of echocardiograms," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–10, Jan. 2020, publisher: Nature Publishing Group.

[96] J. Zhang *et al.*, "Fully Automated Echocardiogram Interpretation in Clinical Practice," *Circulation*, vol. 138, no. 16, pp. 1623–1635, Oct. 2018.

[97] F. Yang *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, Apr. 2022.

[98] O. Bernard *et al.*, "Challenge on endocardial three-dimensional ultrasound segmentation (CETUS)," *Proceedings MICCAI challenge on echocardiographic three-dimensional ultrasound segmentation (CETUS)*, pp. 1–8, 2014.

[99] A. Degerli *et al.*, "Early Detection of Myocardial Infarction in Low-Quality Echocardiography," *IEEE Access*, vol. 9, pp. 34 442–34 453, 2021.

[100] H. Wei *et al.*, "Temporal-Consistent Segmentation of Echocardiography with Co-learning from Appearance and Shape," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel *et al.*, Eds. Cham: Springer International Publishing, 2020, pp. 623–632.

[101] C. Sfakianakis, G. Simantiris, and G. Tziritas, "GUDU: Geometrically-constrained Ultrasound Data augmentation in U-Net for echocardiography semantic segmentation," *Biomedical Signal Processing and Control*, vol. 82, p. 104557, Apr. 2023.

[102] D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *Journal of the Optical Society of America A*, vol. 11, no. 1, p. 107, 1994.

[103] A. P. Shanker and H. Zebker, "Edgelist phase unwrapping algorithm for time series InSAR analysis," *Journal of the Optical Society of America A*, vol. 27, no. 3, pp. 605–612, 2010.

[104] J. Martinez-Carranza, K. Falaggis, and T. Kozacki, "Fast and accurate phase-unwrapping algorithm based on the transport of intensity equation," *Applied Optics*, vol. 56, no. 25, pp. 7079–7088, Sep. 2017.

[105] S. Muth, S. Dort, I. A. Sebag, M.-J. Blais, and D. Garcia, "Unsupervised dealiasing and denoising of color-doppler data," *Medical Image Analysis*, vol. 15, no. 4, pp. 577–588, 2011.

[106] H. Nahas, J. S. Au, T. Ishii, B. Y. S. Yiu, A. J. Y. Chee, and A. C. H. Yu, "A deep learning approach to resolve aliasing artifacts in ultrasound color flow imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2615–2628, 2020.

[107] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[108] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[109] M. Antonelli *et al.*, "The Medical Segmentation Decathlon," *Nature Communications*, vol. 13, no. 1, p. 4128, 2022.

[110] X. Lin, L. Yu, K.-T. Cheng, and Z. Yan, "BATFormer: Towards Boundary-Aware Lightweight Transformer for Efficient Medical Image Segmentation," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–13, 2023.

[111] F. Mehregan *et al.*, "Doppler vortography: a color Doppler approach for quantification of the intraventricular blood flow vortices," *Ultrasound in medicine & biology*, vol. 40, no. 1, pp. 210–221, 2014.

[112] J. Dong, F. Chen, D. Zhou, T. Liu, Z. Yu, and Y. Wang, "Phase unwrapping with graph cuts optimization and dual decomposition acceleration for 3d high-resolution mri data," *Magnetic resonance in medicine*, vol. 77, no. 3, pp. 1353–1358, 2017.

[113] J. M. Bioucas-Dias and G. Valadao, "Phase unwrapping via graph cuts," *IEEE Transactions on Image processing*, vol. 16, no. 3, pp. 698–709, 2007.

[114] B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *Journal of the Optical Society of America*, vol. 69, no. 3, pp. 393–399, 1979.

[115] G. Spoorthi, S. Gorthi, and R. K. S. S. Gorthi, "Phasenet: A deep convolutional neural network for two-dimensional phase unwrapping," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 54–58, 2018.

[116] K. Wang, Y. Li, Q. Kemao, J. Di, and J. Zhao, "One-step robust deep learning phase unwrapping," *Optics express*, vol. 27, no. 10, pp. 15 100–15 115, 2019.

[117] H. Zhou *et al.*, "The phu-net: A robust phase unwrapping method for mri based on deep learning," *Magnetic Resonance in Medicine*, vol. 86, no. 6, pp. 3321–3333, 2021.

[118] Y. Sun *et al.*, "A pipeline for the generation of synthetic cardiac color doppler," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 3, pp. 932–941, 2022.

[119] J. Faurie *et al.*, "Intracardiac vortex dynamics by high-frame-rate doppler vortography—in vivo comparison with vector flow mapping and 4-d flow mri," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, no. 2, pp. 424–432, 2016.

[120] K. Sørensen *et al.*, "Assessment of early diastolic intraventricular pressure difference in children by blood speckle-tracking echocardiography," *Journal of the American Society of Echocardiography*, 2023.

[121] A. Hodzic *et al.*, "Echocardiographic evidence of left ventricular untwisting-filling interplay," *Cardiovascular Ultrasound*, vol. 18, no. 1, pp. 1–12, 2020.

[122] S. Shahriari and D. Garcia, "Meshfree simulations of ultrasound vector flow imaging using smoothed particle hydrodynamics," *Physics in Medicine & Biology*, vol. 63, no. 20, p. 205011, 2018.

[123] R. Asami, T. Tanaka, K.-i. Kawabata, K. Hashiba, T. Okada, and T. Nishiyama, "Accuracy and limitations of vector flow mapping: left ventricular phantom validation using stereo particle image velocimetory," *Journal of Echocardiography*, vol. 15, no. 2, pp. 57–66, Jun. 2017.

[124] A. Arzani, J.-X. Wang, and R. M. D'Souza, "Uncovering near-wall blood flow from sparse data with physics-informed neural networks," *Physics of Fluids*, vol. 33, no. 7, p. 071905, Jul. 2021.

[125] G. Kissas, Y. Yang, E. Hwuang, W. R. Witschey, J. A. Detre, and P. Perdikaris, "Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 358, p. 112623, Jan. 2020.

[126] L. Wang, H. Wang, L. Liang, J. Li, Z. Zeng, and Y. Liu, "Physics-informed neural networks for transcranial ultrasound wave propagation," *Ultrasonics*, vol. 132, p. 107026, Jul. 2023.

[127] Z. Yin, G.-Y. Li, Z. Zhang, Y. Zheng, and Y. Cao, "SWENet: a physics-informed deep neural network (PINN) for shear wave elastography," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2023.

[128] H. Guan, J. Dong, and W.-N. Lee, "Towards Real-time Training of Physics-informed Neural Networks: Applications in Ultrafast Ultrasound Blood Flow Imaging," Sep. 2023, arXiv:2309.04755 [cs.CE].

[129] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks," in *International conference on machine learning (ICML)*, Jul. 2018, pp. 794–803.

[130] A. A. Heydari, C. A. Thompson, and A. Mehmood, "SoftAdapt: Techniques for Adaptive Loss Weighting of Neural Networks with Multi-Part Loss Functions," Dec. 2019, arXiv:1912.12355 [cs.LG].

[131] R. Bischof and M. Kraus, "Multi-Objective Loss Balancing for Physics-Informed Deep Learning," Nov. 2021, arXiv:2110.09813 [cs.LG].

[132] L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, and S. G. Johnson, "Physics-Informed Neural Networks with Hard Constraints for Inverse Design," *SIAM Journal on Scientific Computing*, vol. 43, no. 6, pp. B1105–B1132, Jan. 2021.

[133] H. Son, S. W. Cho, and H. J. Hwang, "Enhanced physics-informed neural networks with Augmented Lagrangian relaxation method (AL-PINNs)," *Neurocomputing*, vol. 548, p. 126424, Sep. 2023.

[134] J. Berg and K. Nyström, "A unified deep artificial neural network approach to partial differential equations in complex geometries," *Neurocomputing*, vol. 317, pp. 28–41, Nov. 2018.

[135] N. Sukumar and A. Srivastava, "Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 389, p. 114333, Feb. 2022.

[136] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, no. 5, pp. 303–320, Nov. 1969.

[137] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations (ICLR)*, 2019.

[138] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, Aug. 1989.

[139] C. Chnafa, S. Mendez, and F. Nicoud, "Image-based large-eddy simulation in a realistic left heart," *Computers & Fluids*, vol. 94, pp. 173–187, May 2014.

[140] R. Faludi *et al.*, "Left ventricular flow patterns in healthy subjects and patients with prosthetic mitral valves: An in vivo study using echocardiographic particle image velocimetry," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 139, no. 6, pp. 1501–1510, Jun. 2010.

[141] T. G. Grossmann, U. J. Komorowska, J. Latz, and C.-B. Schönlieb, "Can Physics-Informed Neural Networks beat the Finite Element Method?" Feb. 2023, arXiv:2302.04107 [math.NA].

[142] H. Wang, S. Gao, M. Mozumi, M. Omura, R. Nagaoka, and H. Hasegawa, "Preliminary investigation on clutter filtering based on deep learning," *Japanese Journal of Applied Physics*, vol. 60, p. SDDE21, May 2021.

[143] G. E. Spoorthi, R. K. Sai Subrahmanyam Gorthi, and S. Gorthi, "Phasenet 2.0: Phase unwrapping of noisy data based on deep learning approach," *IEEE Trans. Image Process.*, vol. 29, pp. 4862–4872, 2020.

[144] L. J. Krovetz and S. D. Goldbloom, "Frequency Content of Intravascular and Intracardiac Pressures and Their Time Derivatives," *IEEE Transactions on Biomedical Engineering*, vol. BME-21, no. 6, pp. 498–501, Nov. 1974.

[145] A. Ramalli, A. Rodriguez-Molares, J. Avdal, J. D'hooge, and L. Løvstakken, "High-Frame-Rate Color Doppler Echocardiography: A Quantitative Comparison of Different Approaches," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 5, pp. 923–933, May 2020.

[146] J. Puig *et al.*, "Boosting Cardiac Color Doppler Frame Rates with Deep Learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024, doi: 10.1109/TUFFC.2024.3424549.

[147] A. Ramalli *et al.*, "Real-Time High-Frame-Rate Cardiac B-Mode and Tissue Doppler Imaging Based on Multiline Transmission and Multiline Acquisition," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 11, pp. 2030–2041, Nov. 2018.

[148] M. Guerra *et al.*, "Improvement in left intraventricular pressure gradients after aortic valve replacement in aortic stenosis patients," *Experimental Physiology*, vol. 102, no. 4, pp. 411–421, 2017.

[149] R. Yotti *et al.*, "A Noninvasive Method for Assessing Impaired Diastolic Suction in Patients With Dilated Cardiomyopathy," *Circulation*, vol. 112, no. 19, pp. 2921–2929, Nov. 2005.

[150] T. Ohara *et al.*, "Loss of Adrenergic Augmentation of Diastolic Intra-LV Pressure Difference in Patients With Diastolic Dysfunction: Evaluation by Color M-Mode Echocardiography," *JACC: Cardiovascular Imaging*, vol. 5, no. 9, pp. 861–870, Sep. 2012.

[151] Y. Zhong *et al.*, "Left ventricular diastolic pressure gradient and outcome in advanced chronic kidney disease patients with preserved ejection fraction," *The International Journal of Cardiovascular Imaging*, vol. 37, no. 9, pp. 2663–2673, Sep. 2021.

[152] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 618–626.

# FOLIO ADMINISTRATIF

## THESE DE L'INSA LYON, MEMBRE DE L'UNIVERSITE DE LYON

NOM : LING

DATE de SOUTENANCE : 07/11/2024

Prénoms : Hang Jung

TITRE : Écoulement intraventriculaire en échocardiographie Doppler avec réseaux de neurones fondés sur la physique

NATURE : Doctorat

Numéro d'ordre : 2024ISAL0087

Ecole doctorale : 162 Mécanique, Énergétique, Génie civil, Acoustique (MEGA)

Spécialité : Acoustique

RESUME : Le cœur, en tant qu'organe central du système cardiovasculaire, est responsable de pomper le sang vers les cellules et tissus de l'organisme. L'évaluation de la santé cardiaque est cruciale pour la prévention des maladies cardiovasculaires. Pour cela, l'échocardiographie est couramment utilisée en routine clinique à cause de sa portabilité et de son coût abordable. Les examens échocardiographiques évaluent la fonction systolique et diastolique du cœur, mais les mesures de la fonction diastolique peuvent parfois donner des résultats de diagnostic discordants. Explorer des biomarqueurs alternatifs, comme le flux sanguin intracardiaque, pourrait améliorer la précision de la quantification de la fonction diastolique.

La cartographie du flux vectoriel intraventriculaire ou "intraventricular vector flow mapping" (iVFM) est une technique qui reconstruit le flux sanguin vectoriel à partir des champs scalaires fournis par l'échocardiographie Doppler couleur, mais elle nécessite des étapes de prétraitement chronophages. Dans cette thèse, nous avons utilisé l'apprentissage profond (DL) pour automatiser ces étapes, y compris la segmentation du ventricule gauche et la correction des artefacts de repliement de phase ou l'aliasing. Nous avons également développé des méthodes basées sur les réseaux de neurones fondés sur la physique pour reconstruire l'écoulement vectoriel intraventriculaire, montrant que ces approches peuvent améliorer l'efficacité et la précision de l'iVFM.

L'automatisation complète du pipeline d'iVFM à l'aide de réseaux de neurones, de la segmentation à la reconstruction du flux vectoriel, améliore la fiabilité de l'iVFM. La prochaine étape serait d'appliquer cet outil en milieu clinique pour explorer et extraire de nouveaux biomarqueurs basés sur le flux, ce qui pourrait bénéficier à la détection précoce des maladies cardiovasculaires.

MOTS-CLÉS : Apprentissage profond, dealiasing, Doppler couleur, échocardiographie, flux cardiaque, réseaux de neurones guidés par la physique, réseaux de neurones informés par la physique, segmentation, ultrasons, imagerie du flux vectoriel

Laboratoire (s) de recherche : Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé (CREATIS)

Directeur de thèse :
Damien Garcia, DR, INSERM, Directeur de thèse
Olivier Bernard, PR, INSA Lyon, Co-directeur de thèse
Pierre-Yves Courand, PU-PH, Hospices Civils de Lyon, Co-directeur de thèse

Présidente de jury : Irène Vignon-Clementel, DR, INRIA

Composition du jury :
Lasse Løvstakken, PR, NTNU Norvège, Rapporteur
Nadjia Kachenoura, DR, INSERM, Rapporteure
Maxime Sermesant, DR, INRIA, Examinateur
Olivier Villemain, PH, CHU Bordeaux, Examinateur