



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Development and Evaluation of an Automated Intraoperative CT–US Registration Pipeline for Endovascular Navigation

TESI DI LAUREA MAGISTRALE IN
BIOMEDICAL ENGINEERING - INGEGNERIA BIOMEDICA

Author: **Tommaso Vantini**

Student ID: 245487

Advisor: Prof. Emiliano Votta

Co-advisors: Veronica Ruozzi

Academic Year: 2024-25

Abstract

Image-guided navigation is central to minimally invasive cardiovascular interventions. Fluoroscopy provides real-time device visualization but lacks soft-tissue information and involves ionizing radiation, whereas ultrasound (US) provides anatomical detail without radiation but suffers from noise and limited field of view. Reliable alignment between pre-operative CT models and intraoperative US remains a key challenge for simulation-assisted navigation. This thesis presents a CT-US registration framework for updating patient-specific vascular geometries using robot-tracked US acquisitions. The pipeline integrates automatic vessel segmentation from 2D US via convolutional neural networks (CNN), 3D surface reconstruction from tracked sweeps, and a two-stage registration strategy introducing a symmetry-aware rigid alignment based on Iterative Closest Point (ICP) with deformable refinement using Coherent Point Drift (CPD). The framework is evaluated *in vitro* on custom vascular phantoms reproducing artificial and patient-specific geometries, assessing rigid alignment accuracy, initialization sensitivity and the contribution of deformable refinement. The proposed rigid strategy improves robustness to initialization over single-start ICP, while deformable refinement consistently reduces residual alignment errors, achieving millimetric accuracy within intraoperative time and non-invasiveness constraints. Since registration validation is limited to phantom settings, segmentation — a key pipeline component — is analysed to characterize variability affecting the overall framework. Multiple architectures are compared across phantom and clinical datasets to assess how dataset heterogeneity, acquisition variability and vessel type affect segmentation reliability within the pipeline. Simpler CNN models provide more consistent boundary delineation across heterogeneous and data-limited conditions, reducing variability that would otherwise propagate to registration. These results demonstrate the feasibility of combining US-based segmentation with multimodal point cloud registration into a coherent workflow for simulation-driven endovascular navigation.

Keywords: Endovascular navigation, CT-US registration, Point cloud registration, Ultrasound segmentation, Deep Learning, Vascular phantoms

Abstract in lingua italiana

La navigazione guidata da immagini è centrale nelle procedure cardiovascolari mini-invasive. La fluoroscopia consente la visualizzazione in tempo reale dei dispositivi ma non fornisce informazioni sui tessuti molli ed espone a radiazioni ionizzanti, mentre l'ecografia (US) è affetta da rumore e campo visivo limitato. L'allineamento affidabile tra modelli TC preoperatori e dati US intraoperatori rappresenta quindi una sfida cruciale per la navigazione endovascolare. Questa tesi propone un framework di registrazione TC-US per l'aggiornamento di geometrie vascolari paziente-specifiche mediante acquisizioni US tracciate da robot. La pipeline integra la segmentazione automatica del vaso in immagini US 2D tramite reti neurali convoluzionali (CNN), la ricostruzione 3D della superficie da sweep tracciati e una registrazione in due fasi che combina un allineamento rigido basato su Iterative Closest Point (ICP) con una successiva fase deformabile mediante Coherent Point Drift (CPD). La valutazione sperimentale *in vitro* su fantocci vascolari, riproducenti geometrie artificiali e paziente-specifiche, analizza accuratezza dell'allineamento rigido, sensibilità all'inizializzazione e contributo della fase deformabile. La strategia rigida proposta migliora la robustezza rispetto a ICP *single-start*, mentre CPD ne riduce sistematicamente l'errore residuo, raggiungendo accuratezza millimetrica compatibile con vincoli intraoperatori di tempo e non invasività. Poiché la registrazione è testabile solo in contesti controllati su fantoccio, la segmentazione — componente chiave della pipeline — è analizzata anche su dataset clinici per caratterizzare le principali fonti di variabilità che possono influenzare l'intero sistema. Diverse architetture sono confrontate per valutare l'effetto di eterogeneità del dataset e variabilità di acquisizione sull'affidabilità della segmentazione. Modelli CNN più semplici mostrano maggiore stabilità in condizioni eterogenee e a limitata disponibilità di dati, limitando la propagazione della variabilità alla registrazione. Nel complesso, i risultati dimostrano la fattibilità dell'integrazione tra segmentazione US e registrazione multimodale di point cloud in un workflow coerente per la navigazione endovascolare assistita da simulazione.

Parole chiave: Navigazione endovascolare, Registrazione TC-US, Registrazione di point cloud, Segmentazione ecografica, Deep Learning, Fantocci vascolari

Contents

Abstract	i
Abstract in lingua italiana	iii
Contents	v
1 Introduction	1
1.1 Cardiovascular diseases	1
1.2 Treatment approaches	1
1.2.1 Traditional surgery	1
1.2.2 Interventional cardiology	1
1.2.3 Endovascular navigation	4
1.3 Imaging systems	5
1.3.1 Fluoroscopy	6
1.3.2 Ultrasound-based techniques	7
1.3.3 Mixed Reality integration	9
1.4 Research framework for endovascular navigation	10
1.4.1 The ARTERY project	10
1.4.2 Finite element modeling of catheter–vessel interaction	11
1.5 Aim of the thesis	13
2 State of the art	15
2.1 Image registration methods	15
2.1.1 Problem formulation and classification	16
2.1.2 Optimization-based registration	17
2.1.3 Learning-based registration	20
2.1.4 Hybrid approaches	21
2.2 Image segmentation	22
2.2.1 Classical segmentation	23

2.2.2	Learning-based segmentation	23
2.2.3	US image segmentation challenges	25
2.2.4	Limited data and generalization.	26
2.3	Vascular phantoms for in-vitro testing	29
2.3.1	Design requirements for vascular phantoms	29
2.3.2	Thin-walled vascular phantoms	30
2.3.3	Wall-less vascular phantoms	31
2.3.4	Tissue-mimicking materials	32
3	Materials and methods	35
3.1	Methodological Pipeline for CT-US	
	Vascular Registration	36
3.1.1	System architecture and coordinate definition	36
3.1.2	Preoperative data preparation	39
3.1.3	Online intraoperative pipeline	40
3.2	Automatic ultrasound image segmentation	46
3.2.1	Datasets	47
3.2.2	Dataset preparation	48
3.2.3	Model design and training	50
3.2.4	Model aggregation and calibration	53
3.2.5	Inference and evaluation	54
3.3	In-vitro experiments for registration evaluation	55
3.3.1	Vascular phantom	55
3.3.2	Data acquisition protocol	57
3.3.3	Registration experiments	58
3.4	In-vivo evaluation of the segmentation network	60
3.4.1	Segmentation experiments	61
4	Results and discussion	65
4.1	In-vitro registration results	65
4.1.1	Reconstruction consistency	65
4.1.2	Rigid registration robustness	66
4.1.3	Registration accuracy	69
4.1.4	Computational performance	71
4.2	Segmentation results	73
4.2.1	Model comparison across datasets	73
4.2.2	Influence of dataset characteristics	75
4.2.3	Vessel type and domain aggregation analysis	78

5	Conclusions	83
5.1	Summary of the work.	83
5.2	Limitations	84
5.3	Future developments	84
	 Bibliography	 87
 A	 Evaluation metrics	 97
A.1	Segmentation metrics	97
A.2	Registration metrics	99
 B	 Algorithm Parameters and Code Settings	 101
	 List of Figures	 103
	 List of Tables	 107

1 | Introduction

1.1. Cardiovascular diseases

According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the leading cause of death worldwide, accounting for approximately 32% of all global deaths [77, 78]. Among them, ischemic heart disease and stroke are the most prevalent conditions, imposing a significant clinical, social and economic burden. Despite notable advances in prevention and clinical care, the incidence of CVDs is expected to further increase in the coming decades due to the growing prevalence of cardiovascular risk factors and population aging [34, 47]. This has stimulated ongoing research into innovative diagnostic and therapeutic technologies aimed at improving early detection, enhancing treatment efficacy and ultimately reducing the morbidity, mortality and societal costs associated with these conditions.

1.2. Treatment approaches

1.2.1. Traditional surgery

Conventional open-heart surgery has long represented the standard treatment for several cardiovascular pathologies. These procedures typically require median sternotomy and cardiopulmonary bypass, allowing direct access to the heart and major vessels at the cost of high invasiveness. Although effective, open surgical approaches are associated with prolonged recovery times and increased perioperative risk. In particular, elderly or fragile patients may not tolerate the stresses of bypass, the trauma of a major thoracic incision or the effects of total and prolonged sedation.

1.2.2. Interventional cardiology

Interventional cardiology has emerged as a minimally invasive alternative to conventional surgery for the treatment of a wide range of cardiovascular diseases, including stenosis, valvulopathies, dissections or aneurysms. These procedures are performed through per-

cutaneous vascular access, typically via the femoral artery or vein, and rely on catheter-based systems to deliver therapeutic devices to the target anatomical site (Figure 1.1). Since these procedures preserve native tissue integrity, patients benefit from shorter recovery times (often in a day-hospital setting), lower infection rate and lower overall costs compared to open-heart surgery.

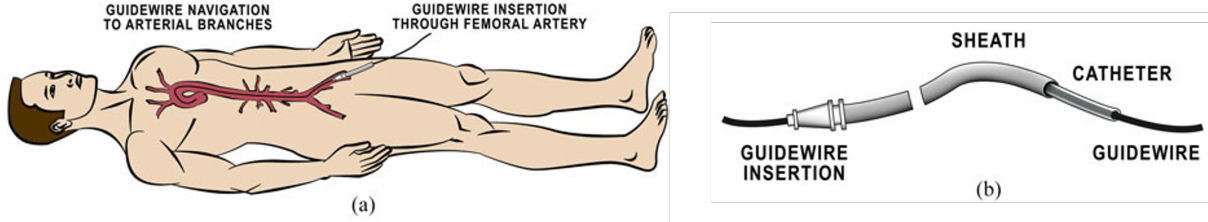


Figure 1.1: Left (a): Example of endovascular catheter placement via the femoral artery. Right (b): Typical structure of a delivery system for minimally invasive procedures, including catheter, guidewire and sheath. Adapted from [23].

Many devices are currently available for catheter-based cardiovascular interventions, including balloons, stents, valve prostheses, embolization coils and ablation catheters. These systems are delivered through long sheaths and guidewires and are designed to navigate complex vascular pathways while enabling controlled and precise deployment at the target site (Figure 1.2).

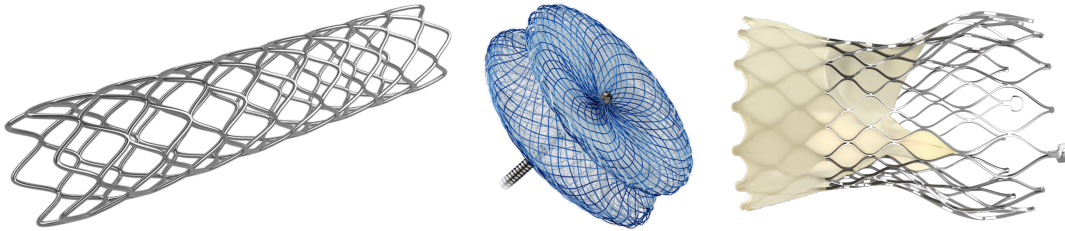


Figure 1.2: Some examples of catheter-based devices used in interventional cardiology. From left to right: stent, atrial appendage occluder and prosthetic valve.

Minimally invasive treatment is the only option to intervene in high-risk patients. Moreover, although open-heart surgery remains the gold standard, recent clinical studies have demonstrated that minimally invasive transcatheter procedures can achieve clinical outcomes comparable to conventional surgery, while significantly reducing procedural invasiveness and recovery time [9, 45]. These advantages have contributed to the rapid adoption of transcatheter techniques such as Transcatheter Aortic Valve Implant (TAVI) and Transcatheter Edge-to-Edge Repair (TEER) in clinical practice, also in patients with

low or mid surgical risk. Indeed, in the last decade transcatheter procedures are becoming prevalent with respect to traditional surgery (Figure 1.3).

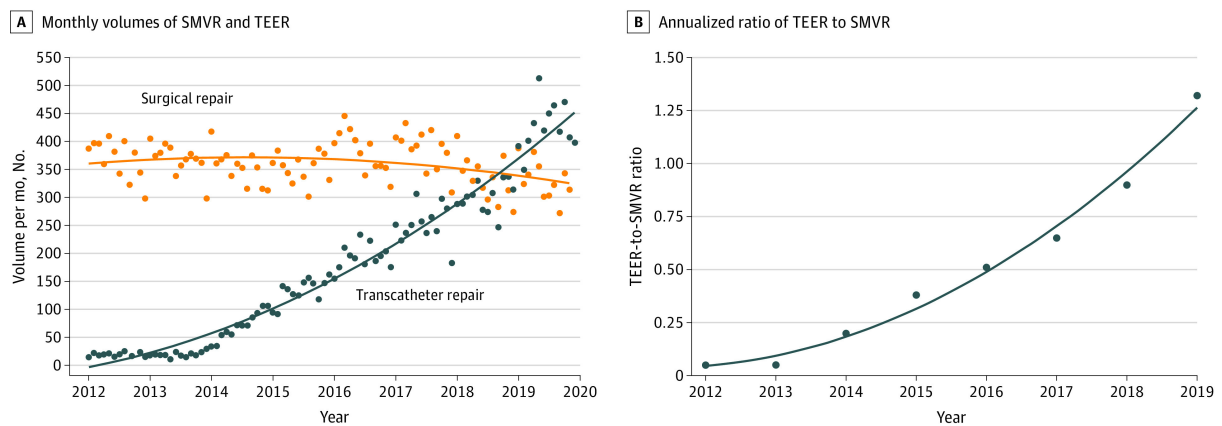


Figure 1.3: Trends in TEER vs. traditional surgical mitral valve repair (SMVR) in the USA, highlighting the increasing relevance of minimally invasive approaches in cardiology [82].

Despite their clinical success, these procedures remain technically demanding because operators must steer long and flexible devices within patient-specific anatomies under indirect intraoperative feedback. For this reason, endovascular navigation constitutes a central challenge in transcatheter interventions.

Examples of interventional procedures

Transcatheter Edge-to-Edge Repair. TEER is a minimally invasive procedure to treat mitral regurgitation by approximating the mitral valve leaflets to enhance coaptation and reduce regurgitation using a transcatheter clip device. The procedure is typically performed via a femoral venous approach. The device is navigated through the femoral vein and the inferior vena cava into the right atrium. A crucial part of the procedure is the trans-septal puncture, to cross the interatrial septum. From the left atrium, the catheter can be precisely maneuvered into position above the mitral valve. The device is then deployed to clasp the leaflets, creating a double-orifice valve structure that restores the mitral valve function.

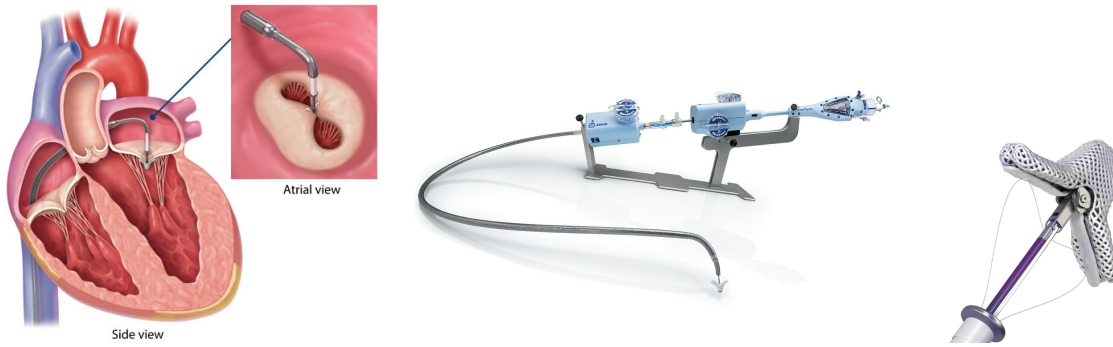


Figure 1.4: The MitraClip™ system (Abbott Vascular, Santa Clara, California). Left: side and atrial views of the clipping phase. Center: delivery system, including catheter and steering mechanism. Right: close-up view of the clip in the open configuration.

Transcatheter Aortic Valve Implantation. TAVI is a percutaneous approach for treating severe aortic stenosis by deploying a prosthetic valve across the native aortic valve. In this case, the femoral artery access allows the navigation through the aorta up to the heart. After the correct positioning, the system deploys a collapsible valve mounted on a catheter and expanded at the valve site to restore effective outflow.

1.2.3. Endovascular navigation

Endovascular navigation refers to the controlled advancement and steering of flexible instruments (e.g., guidewires and delivery systems) within the vascular network. Devices are advanced over long and tortuous anatomical pathways until the target site is reached. Successful navigation requires maintaining an accurate understanding of the instrument configuration and its spatial relationship with patient-specific anatomy, while ensuring safe interaction with surrounding tissues. Although transcatheter interventions offer clear clinical advantages, endovascular navigation remains technically demanding and represents one of the main sources of procedural complexity.

Dependence on imaging modalities. Since interventional cardiology procedures are performed without opening the chest cavity, the operator lacks direct visual feedback of the anatomical structures they are working on. As a result, the procedure relies heavily on imaging technologies to provide real-time guidance. Commonly used modalities include fluoroscopy, intravascular ultrasound (IVUS) and transesophageal echocardiography (TEE). However, these methods present trade-offs in terms of space and time resolution, invasiveness, radiation exposure and operator dependence. Inadequate imaging can compromise procedural accuracy, device positioning and clinical outcomes.

Steep learning curve. The manipulation of endovascular instruments, such as catheters or guidewires, requires a high manual expertise and spatial awareness. Operators must rely on indirect visual and tactile cues to navigate complex vascular pathways. This skill set is difficult to acquire and demands extensive hands-on training. Many high-fidelity surgical simulators and phantoms have been developed, offering clinicians a safe environment to practice. These platforms aim to enhance hand-eye coordination and the interpretation of haptic feedback under realistic physiological conditions.

Risk of tissue damage. During catheter navigation through the vascular system and inside the cardiac chambers, there is an inherent risk of damaging soft tissues. Excessive mechanical force and prolonged frictional contact with vessel walls can result in endothelial trauma, dissection or perforation. These injuries may have severe clinical consequences, including stroke, arrhythmia and vascular rupture. Modern robotic catheters and force-sensing tools attempt to mitigate these risks, but real-time monitoring and force control remain suboptimal in many settings.

Overall, these limitations highlight that safe and effective endovascular navigation critically depends on intraoperative guidance. For this reason, imaging systems play a central role in modern transcatheter procedures, providing real-time information about both device motion and anatomical context.

1.3. Imaging systems

In interventional cardiology, intraoperative visualization plays a crucial role in supporting endovascular navigation and ensuring procedural safety, since transcatheter interventions are performed without direct access to the operative field. Effective intraoperative visualization should provide complementary information on both anatomy and device motion. Indeed, it should:

- enable the identification of relevant cardiovascular structures, such as vessels, cardiac chambers and valves;
- allow continuous tracking of catheters, guidewires and deployed devices during navigation and deployment;
- achieve an adequate balance between spatial resolution, temporal resolution and ease of interpretation.

No single imaging modality can fully satisfy all these requirements. Each technique presents intrinsic limitations related to resolution, field of view, invasiveness, radiation

exposure or operator dependence. For this reason, intraoperative visualization in interventional cardiology typically relies on the combined use of multiple imaging systems, each contributing complementary information.

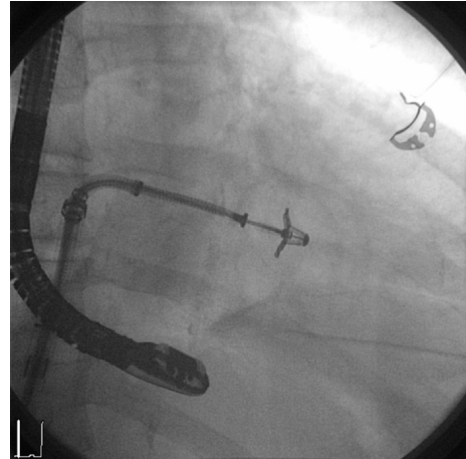
1.3.1. Fluoroscopy

Fluoroscopy is a radiologic imaging technique that provides real-time two-dimensional visualization of radiopaque structures by transmitting X-rays through the body onto a fluorescent detector. In interventional cardiology, fluoroscopy represents the primary modality for intraoperative guidance and is routinely used to track the position and orientation of devices.

The imaging system typically consists of an X-ray source and a detector mounted on a movable C-arm, which can be translated and rotated to acquire projections from different viewpoints (Figure 1.5a). This configuration allows the operator to dynamically adjust the imaging geometry according to procedural needs. Fluoroscopy is particularly effective at visualizing metallic components, making it indispensable for monitoring the advancement and deployment of interventional devices during endovascular navigation (Figure 1.5b). The injection of radiopaque contrast agents temporarily enhances the visualization of vessels and cardiac chambers.



(a) Fluoroscopy setup



(b) Intraoperative fluoroscopic image

Figure 1.5: Left (a): A typical fluoroscopy setup, including the C-arm (with X-ray source and detector) and the display monitor (*GE Healthcare*). Right (b): Fluoroscopic image acquired during a TEER procedure, showing the delivery catheter, the MitraClip™ device and the RT3DTEE probe [4].

Despite its widespread adoption, fluoroscopy presents relevant limitations. The most critical drawback is the exposure to ionizing radiation for both patients and medical staff [10]. Minimally invasive procedures typically last around 20 minutes or longer, during which operators may receive a mean effective dose of approximately $1.2 \mu\text{Sv}$ via femoral access and $2.3 \mu\text{Sv}$ via radial access [27, 65]. Although strict radiation-protection protocols are implemented in the interventional room, cumulative exposure remains a concern, as it may increase the long-term risk of cancer or cataract formation [41].

In addition, fluoroscopy provides only two-dimensional projections and lacks intrinsic depth information. As a result, the true three-dimensional spatial relationship between interventional devices and surrounding anatomical structures cannot be directly inferred and must be mentally reconstructed by the operator. Finally, the use of iodinated contrast agents may introduce procedure-related risks, particularly in patients with impaired renal function.

1.3.2. Ultrasound-based techniques

Ultrasound imaging represents a valuable complementary modality to fluoroscopy in interventional cardiology. Unlike X-ray-based techniques, ultrasound does not involve ionizing radiation and provides direct visualization of soft tissues, enabling the assessment of anatomical structures that are poorly visible or completely invisible under fluoroscopy. For these reasons, ultrasound is increasingly adopted to improve intraoperative guidance, reduce radiation exposure and enhance procedural safety.

Ultrasound imaging is based on the interaction between high-frequency mechanical waves and biological tissues. An echographic system emits acoustic pulses and records the echoes reflected by tissue interfaces with different acoustic impedances. Structures with higher echogenicity appear brighter, whereas fluid-filled regions appear anechoic. This imaging principle allows real-time visualization of cardiovascular anatomy and device-tissue interactions, offering information that complements the device-centric view provided by fluoroscopy.

In clinical practice, real-time three-dimensional transesophageal echocardiography (RT3DTEE) and intravascular ultrasound (IVUS) are particularly relevant for intraoperative guidance.

Real-time 3D transesophageal echocardiography. RT3DTEE is an echocardiographic technique that enables real-time volumetric visualization of cardiac structures, but it is not used to monitor vessels. The ultrasound probe is introduced orally and advanced into the esophagus, where it lies in close proximity to the heart (Figure 1.6, left).

This configuration minimizes acoustic interference from surrounding tissues and allows high-quality imaging of cardiac chambers and valves.

Compared with transthoracic echocardiography, the esophageal approach places the transducer nearer to the structures of interest, with fewer intermediate tissues (such as bones) that could introduce artifacts or impede the visualization. This technique is very useful because it provides clear, accurate and continuous images of the heart structures under manipulation. Moreover, the acquired data are processed to render a real-time 3D image, often enhanced with color coding to improve depth perception (Figure 1.6, right). Despite being non-ionizing, the esophageal probe may cause patient discomfort, especially during prolonged procedures performed under conscious sedation.

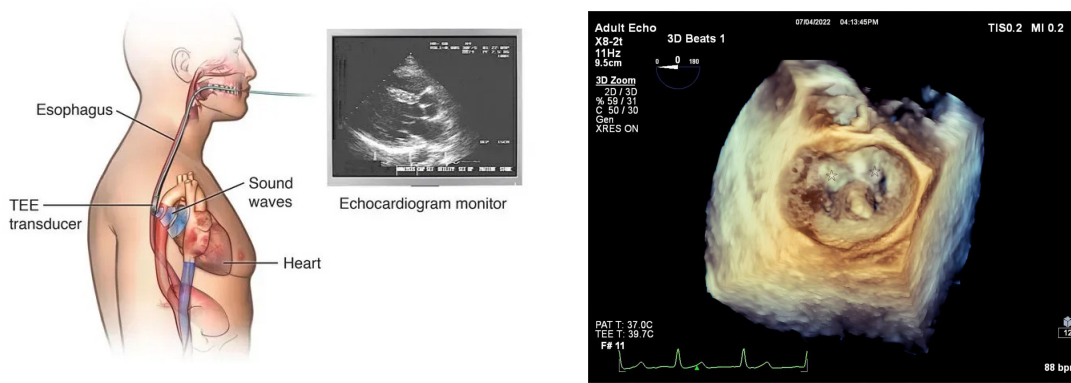


Figure 1.6: Real-time 3-dimensional transesophageal echocardiography (RT3DTEE). Left: Schematic representation of the probe's position in the esophagus relative to the heart. Right: Real-time 3D rendering of the mitral valve acquired using RT3DTEE, showing detailed anatomical features.

Intravascular ultrasound. IVUS is an ultrasound-based imaging technique in which a miniaturized transducer is mounted on the tip of a catheter and advanced inside the vessel lumen. It provides high-resolution cross-sectional images of the vessel wall and surrounding structures, enabling detailed assessment of vascular morphology, lumen diameter and plaque composition. Thanks to its intraluminal perspective, IVUS offers information that cannot be obtained from external imaging modalities. However, its field of view is limited to local vessel segments, and its use requires additional intravascular instrumentation. As a result, IVUS is selectively employed in specific procedural contexts rather than as a universal guidance modality.

1.3.3. Mixed Reality integration

While conventional imaging modalities provide high-resolution 2D or 3D images, these data are typically visualized on flat monitors placed outside the operative field (Figure 1.7, left). As a consequence, clinicians are required to continuously shift their attention between the patient and external displays and to mentally reconstruct complex spatial relationships between interventional devices and anatomical structures.

A first step toward improved spatial understanding consists in combining multiple sources of information, such as electromagnetic tracking of the catheter and preoperative three-dimensional reconstructions of the vasculature, into a unified digital representation displayed on a screen (Figure 1.7, center). Although this approach could enable interactive visualization of the intraoperative scenario, the spatial separation between the physical environment and the digital model still limits intuitive perception and hand–eye coordination.

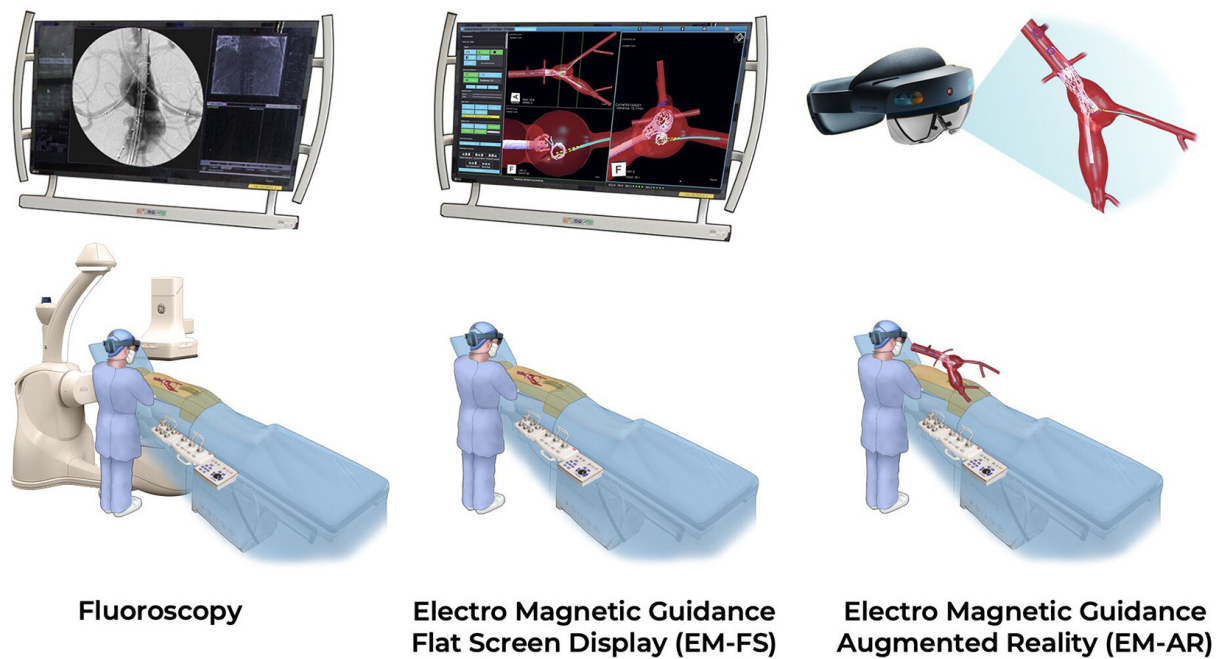


Figure 1.7: Comparison of navigation techniques for endovascular interventions. Left: Conventional 2D fluoroscopy displayed on a flat screen. Center: Catheter tracking combined with a virtual 3D vascular model. Right: Integration of catheter tracking and 3D vascular model in a mixed reality environment.

Mixed Reality (MR) technologies have been proposed as an advanced visualization paradigm to overcome these limitations. By means of head-mounted displays (HMDs), MR systems superimpose interactive holographic models directly onto the operator’s view of the pa-

tient and surrounding environment, while preserving natural vision for safety reasons. This allows digital anatomies to be perceived in situ, spatially aligned with the physical world.

Modern MR platforms integrate depth sensors, inertial measurement units and optical tracking systems to enable real-time environment mapping, spatial anchoring and gesture-based interaction. In this way, volumetric anatomical models, catheter trajectories and navigation cues can dynamically respond to both patient anatomy and operator input. Beyond procedural guidance, MR has also demonstrated benefits in anatomical education, clinician–patient communication and remote collaboration among geographically distributed teams [32]. However, several challenges remain open, including accurate registration between virtual and physical entities, robust real-time tracking and seamless integration into existing clinical workflows.

1.4. Research framework for endovascular navigation

1.4.1. The ARTERY project

In recent years, research efforts have focused on developing integrated approaches to support endovascular navigation during minimally invasive cardiovascular procedures, aiming to improve operator awareness and safety while reducing radiation exposure.

ARTERY (*Autonomous Robotics for Transcatheter dEliveRy sYstems*) was a European research initiative coordinated by Politecnico di Milano, involving academic and industrial partners working in robotics, biomedical engineering and cardiovascular surgery. The project aimed at developing advanced robotic and imaging technologies to assist clinicians during minimally invasive transcatheter cardiovascular procedures. By integrating robotic control, advanced imaging and physical simulations, the project aimed to improve operator awareness and safety, while reducing radiation exposure.

To achieve these objectives, the project combined several technologies within a single integrated framework:

- *Sensorized catheters* that can continuously track their own shape and position within the cardiovascular system [30]
- *Shared-autonomy robotic control*, enabling semi-autonomous navigation of the catheter through complex anatomical pathways [83]
- *Real-time finite element simulations*, which model the mechanical interaction between the catheter and the vessel walls [21, 62]

- *Mixed Reality visualization*, allowing the operator to observe in real time the catheter's position and its interaction with the cardiovascular structures [80]

The ARTERY project provides the broader research framework within which the present thesis is developed, ensuring conceptual and methodological continuity with ongoing work in the hosting laboratory.

1.4.2. Finite element modeling of catheter–vessel interaction

As discussed in *Ruozzi and Vassena, MSc Thesis* (Politecnico di Milano, 2021 [62]), the integration of a real-time finite element simulation within the navigation workflow could substantially improve the safety and control of minimally invasive cardiovascular procedures. By coupling the physical motion of a sensorized catheter with a continuously updated digital model (Figure 1.8), the operator can visualize the deformation of the vessel during the intervention and anticipate possible critical interactions. In this way, simulation becomes not only a visualization tool, but also a mean to interpret sensor data and provide predictive feedback to the robotic controller.

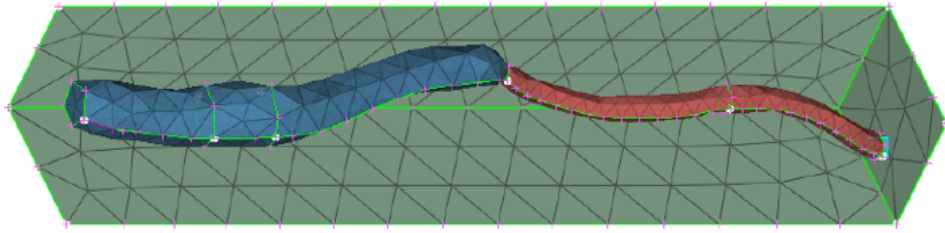


Figure 1.8: Patient-specific vascular model, including the femoral vein (red), a segment of the inferior vena cava (blue) and a bounding box (green) modelling the mechanical effect of surrounding tissues. Adapted from [62].

In this context, the reliability of real-time simulations strongly depends on the accuracy of the underlying finite element model. Several aspects play a critical role in determining the realism and predictive capability of the simulation.

Geometrical modeling. The geometrical model should be patient-specific, reconstructed from pre-operative imaging data to represent a realistic vascular anatomy. However, the pre-operative geometry does not necessarily coincide with the intraoperative configuration. Soft tissues are deformable and subject to physiological variability, and their shape may change over time due to posture, loading conditions and surrounding tissue interaction. Moreover, patient positioning during the intervention differs from that during

pre-operative imaging acquisition: while imaging is often performed with the patient in a supine position, during endovascular procedures the limb used for vascular access is commonly flexed and externally rotated to facilitate catheter insertion. These factors contribute to discrepancies between pre-operative and intraoperative geometries and represent a major challenge for accurate simulation.

Boundary conditions and catheter–vessel interaction. Boundary conditions play a crucial role in finite element simulations of endovascular procedures, especially at the interface between the catheter and the vessel wall. In the proposed framework [62], the effect of surrounding tissues is modeled through a bounding box whose material properties emulate the mechanical constraint imposed by the anatomical environment. While boundary conditions applied to the outer surfaces of this domain are important, the most critical aspect lies at the endovascular interface, where contact interactions between the catheter and the vessel govern the deformation and displacement of the vascular wall. Accurately modeling this interaction requires a continuous exchange of information between the physical space (the operating room, in which the catheter is manipulated) and the digital space (where the finite element model is solved). Establishing this link ensures that imposed displacements and contact constraints in the simulation correspond to the real configuration of the device, enabling a realistic reproduction of catheter–vessel interaction.

Material properties. Accurate modeling of the mechanical behavior of vascular tissues requires appropriate estimation of material properties. In *Dodesini, MSc Thesis* (Politecnico di Milano, 2024 [21]), a sensitivity analysis on the Young’s modulus (E) and the Poisson’s ratio (ν) demonstrated that variations in these parameters can significantly affect simulation outcomes, in particular the prediction of contact forces.

1.5. Aim of the thesis

As outlined in the previous sections, simulation-assisted approaches for endovascular navigation require a reliable correspondence between the physical intraoperative scenario and its digital representation. In a clinical context, this correspondence must be accurate, robust and compatible with intraoperative time constraints, while avoiding reliance on anatomical landmarks that may be difficult to identify in practice.

The primary aim of this thesis is to advance an existing research framework toward practical applicability by developing an intraoperative registration pipeline capable of aligning and updating a patient-specific vascular model using non-invasive imaging data. The proposed approach operates without explicit anatomical landmarks and aims to balance accuracy and computational efficiency.

To achieve this goal, the thesis addresses two closely related objectives:

1. the development of an **intraoperative registration pipeline** that aligns pre-operative vascular models with intraoperative data acquired using external ultrasound imaging, including both rigid registration for spatial alignment and deformable registration to adapt the geometry of a finite element vessel model;
2. the development of a **fast and automatic vessel segmentation method** based on deep learning, enabling the extraction of vascular geometry from ultrasound images in an accurate and operator-independent manner, suitable for intraoperative use.

The proposed approaches are evaluated through a twofold validation strategy. The segmentation network is tested on publicly available datasets as well as on data acquired in-house, while the registration pipeline is experimentally assessed in vitro on vascular phantoms.

2 | State of the art

This chapter reviews the state of the art of the methodological components underlying this thesis, with a focus on techniques relevant to multimodal vascular imaging and intraoperative scenarios.

The chapter is structured around three complementary research areas. First, **image registration methods** are discussed, with particular emphasis on approaches developed for medical applications and on strategies suitable for real-time or near real-time alignment in vascular contexts. Second, **image segmentation techniques** are reviewed, focusing on learning-based methods that enable robust extraction of anatomical structures from challenging clinical images. Finally, **vascular phantoms** are introduced as experimental tools for in-vitro validation, outlining common design principles and material choices used to test and benchmark imaging and registration pipelines.

2.1. Image registration methods

Image registration is the process of aligning, in a common coordinate system, two or more images acquired at different times, from different viewpoints or using different imaging modalities. As illustrated in Figure 2.1, registration can be formulated as the estimation of a spatial transformation that maps a moving image onto a fixed reference. In medical applications, registration plays a fundamental role in image-guided interventions, where preoperative information must be accurately aligned with intraoperative data.

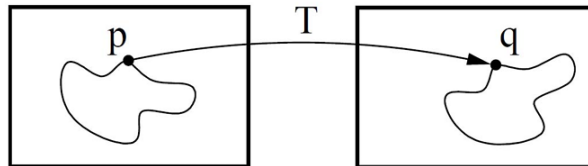


Figure 2.1: Conceptual example of image registration [37]. A transformation T is estimated to map a moving object (point p) onto a fixed reference (point q).

Accurate image registration remains a challenging task, since factors like resolution, noise

and contrast can impact the alignment performance. To address these issues, numerous methods have been proposed, from classical optimization-based approaches to more recent data-driven frameworks. Each method offers specific advantages and limitations depending on the data characteristics and the clinical context.

2.1.1. Problem formulation and classification

Image registration methods can be categorized according to several complementary criteria, which reflect both the characteristics of the data and the intended clinical application.

Image dimensionality. When images share the same dimensionality (2D–2D or 3D–3D), alignment is generally more straightforward, although differences in acquisition parameters can still pose challenges. By contrast, registration between datasets of different dimensionalities (e.g., 3D CT and 2D US) often requires dedicated strategies, including projection models or intermediate representations, to bridge the gap between spatial domains.

Imaging modality. The relationship between intensity values across images strongly influences the choice of similarity measures. In monomodal registration both images come from the same modality and therefore intensity values can be directly compared. Instead, in multimodal registration, voxel values represent different physical quantities (e.g., tissue density for CT versus acoustic impedance for US), which complicates direct comparison and motivates the use of modality-independent similarity metrics.

Alignment strategy. Feature-based approaches rely on identifiable structures such as landmarks or contours and are often more robust in multimodal settings. Intensity-based approaches, instead, exploit the entire image information by comparing intensity distributions, achieving high precision when image quality is adequate. Hybrid strategies combining both are increasingly common to balance robustness and accuracy.

Methodological approach. Image registration can be approached either as an optimization problem, where a transformation is iteratively adjusted to minimize a dissimilarity measure, or through learning-based methods, where a model directly predicts the transformation from input images. While the former has long been the standard, the latter enables near real-time registration. More recently, hybrid methods combining both paradigms have been proposed to merge their complementary strengths.

Transformation model. The transformation defines the degrees of freedom allowed during alignment (Figure 2.2). Rigid transformations (translations and rotations) are appropriate for non-deformable structures. Affine models add scaling and shearing to account for global differences. Deformable models allow for local displacements, typically represented by parametric functions (e.g., B-splines) or dense fields regularized for smoothness and anatomical plausibility.

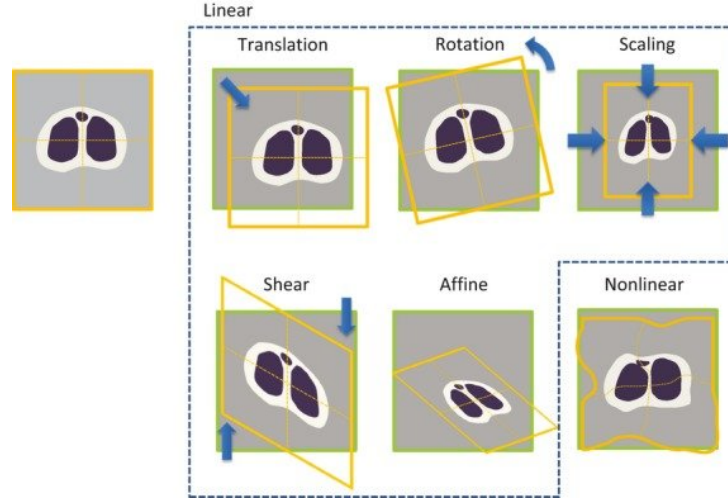


Figure 2.2: Examples of transformation models used in image registration, ranging from rigid transformations to nonlinear deformations [72].

2.1.2. Optimization-based registration

Traditional image registration methods are commonly formulated as optimization problems. The objective is to find the transformation T_θ , parameterized by θ , that aligns the moving image $M(x)$ with the reference image $R(x)$, both defined over a spatial domain $\Omega \subset \mathbb{R}^n$. In its general form, adaptable to different transformation models, the problem can be expressed as:

$$\theta^* = \arg \min_{\theta} [D(R, M \circ T_\theta) + \lambda S(T_\theta)], \quad (2.1)$$

where T_θ is the transformation parameterized by θ , $D(\cdot, \cdot)$ is a dissimilarity measure quantifying image misalignment, $S(T_\theta)$ is a regularization term and λ controls the trade-off between data fidelity and regularity.

The components involved in this optimization process are illustrated in Figure 2.3. The optimization is typically performed with gradient-based or derivative-free algorithms, depending on the differentiability of D and the dimensionality of the parameter space.

Multi-resolution strategies are often employed to improve convergence and robustness, starting from coarse representations and progressively refining the solution.

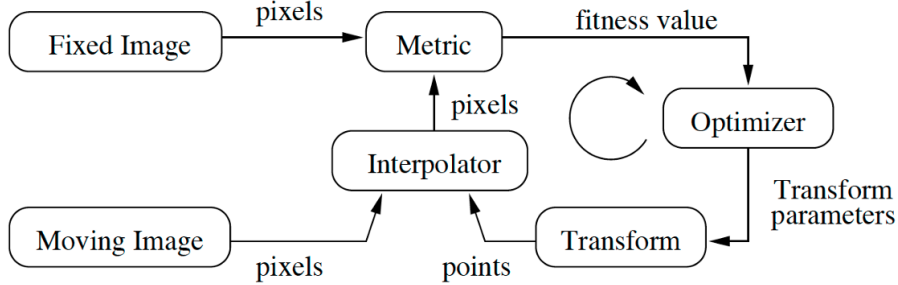


Figure 2.3: Schematic overview of a classical optimization-based image registration framework [37]. The moving image is transformed into the reference coordinate system (*Transform*), interpolated at non-grid locations (*Interpolator*) and compared to the fixed image through a similarity metric (*Metric*). An *Optimizer* iteratively updates the transformation parameters to minimize the overall cost function

Intensity-based methods

Intensity-based registration methods operate directly on voxel or pixel intensities, without requiring explicit feature extraction. By exploiting the full image information, these approaches can achieve high alignment accuracy when image quality is sufficient and a reasonable initialization is available. However, they are often computationally demanding and sensitive to noise, artifacts and local intensity inconsistencies.

In monomodal registration, direct intensity comparisons are appropriate. Common similarity measures include the *Sum of Squared Differences* (SSD) and the *Normalized Cross-Correlation* (NCC). Instead, in multimodal registration statistical dependence between the modalities is exploited through information-theoretic measures such as *Mutual Information* (MI) and its normalized form (NMI). Variants such as *Local Mutual Information* (LMI) or *Normalized Gradient Fields* (NGF) have been developed to improve robustness, particularly when local contrast or structural boundaries carry more information than absolute intensity values.

Practical implementations of intensity-based registration are available in widely used open-source frameworks such as *Elastix* [42, 67], which supports multiple transformation models, similarity measures and optimization strategies.

Feature-based methods

Feature-based registration relies on identifiable structures extracted from the images, such as landmarks, contours or surfaces, and aligns datasets by minimizing distances between corresponding features. By operating on sparse and geometrically meaningful representations, these approaches are often more robust in multimodal scenarios.

Landmark-based registration. When corresponding landmarks r_i and m_i can be identified in reference and moving images, the transformation parameters are estimated by minimizing:

$$E(\theta) = \sum_{i=1}^N \|r_i - T_\theta(m_i)\|^2. \quad (2.2)$$

This approach is computationally efficient but heavily depends on the accuracy of landmark detection, which can be challenging in medical images due to noise and low contrast.

Geometric point-cloud registration. When anatomical structures can be represented as point clouds or surfaces, registration can be performed directly on these geometries rather than on voxel intensities. By relying on spatial geometry rather than image texture, this approach avoids modality-dependent artifacts and reduces sensitivity to contrast or noise.

- **Iterative Closest Point (ICP).** Among geometric methods, the *Iterative Closest Point* (ICP) algorithm [12] was one of the first ones to be implemented and it is still the base of many modern approaches. ICP seeks the rigid transformation that minimizes the distance between two point clouds or surfaces. At each iteration, the algorithm performs two steps: (i) correspondence estimation, by assigning to each point in the moving set its nearest neighbor in the reference; (ii) transformation update, typically solved via least-squares minimization of point-to-point or point-to-plane distances. The process repeats until convergence, when further iterations do not significantly reduce the alignment error. ICP is conceptually simple and computationally efficient, but its performance strongly depends on initialization and on the presence of outliers or partial overlaps, which can drive convergence toward local minima. Numerous variants have been proposed to improve robustness and convergence. These include point-to-plane formulations [16], robust approaches for partial overlap such as *Trimmed ICP* [17] and probabilistic extensions such as *Generalized-ICP* [66].
- **Coherent Point Drift (CPD).** To further enhance robustness and allow for non-rigid alignment, probabilistic formulations have been developed. A representative

example is the *Coherent Point Drift* (CPD) algorithm [49], which interprets one point cloud as centroids of a Gaussian Mixture Model (GMM) and estimates the transformation by maximizing the likelihood of the other set under this model. The resulting expectation–maximization framework provides soft correspondences between points and includes a spatial regularization term that enforces local smoothness of the deformation field. This formulation enables CPD to handle noise, missing data and moderate non-rigid deformations while producing smooth and physically plausible transformations. For these reasons, CPD and its extensions (e.g., Bayesian [31] or B-spline [61] variants) are widely adopted in medical applications involving point-based deformable registration.

2.1.3. Learning-based registration

In recent years, deep learning has emerged as an alternative paradigm for image registration. Instead of solving a new optimization for every image pair, a neural network is trained to predict the spatial transformation directly. A network f_ϕ , parameterized by weights ϕ , estimates a displacement field $u(x)$:

$$u = f_\phi(R, M). \quad (2.3)$$

The training loss typically combines an image similarity term with a regularization term on the predicted deformation, in analogy to the classical formulation:

$$L(\phi) = D(R, M \circ (Id + f_\phi(R, M))) + \lambda S(f_\phi). \quad (2.4)$$

Supervised learning. When ground-truth deformations u^* are available, networks can be trained to predict u so that it matches the reference. This approach achieves high accuracy and stability but requires reliable annotated data, which are rarely available in medical imaging. Moreover, retraining for new modalities or scanners demands additional annotations, limiting generalization.

Synthetic supervision. To address the lack of ground-truth data, several frameworks exploit synthetic data generation. For instance, realistic US-like images can be simulated from segmented CT volumes using acoustic or ray-tracing models, enabling supervised learning with consistent anatomical labels. For example, *Velikova et al., 2024* [75] exploited an intermediate representation (IR) space between CT and US images to bridge

the inter-modality gap. The method was shown to outperform conventional supervised baselines in aortic segmentation tasks.

Unsupervised learning. In the absence of ground truth, registration networks can be trained by directly minimizing image dissimilarity between a reference R and a warped moving image $M \circ (Id + u)$, combined with deformation regularization. These methods are more flexible but rely heavily on the choice of similarity metric. Balakrishnan et al., 2019 [8] proposed *VoxelMorph*, an encoder-decoder CNN that predicts dense displacement fields from image pairs. The framework employs similarity measures such as NCC or MI, together with regularization constraints, to produce accurate and diffeomorphic transformations. It achieves accuracy comparable to classical tools such as *Elastix* while reducing runtime from minutes to milliseconds.

2.1.4. Hybrid approaches

Hybrid registration approaches aim to combine the interpretability and robustness of classical optimization-based methods with the efficiency of deep learning. Rather than replacing traditional pipelines entirely, these methods integrate learning-based components into specific stages of the registration process, such as similarity evaluation, initialization or anatomical constraint. This integration is particularly relevant in multimodal and intra-operative scenarios, where image quality, noise and anatomical variability can challenge purely intensity-based or purely learning-based solutions. By exploiting complementary strengths, hybrid frameworks seek to improve robustness and convergence behavior while retaining computational efficiency.

Similarity learning. One line of research focuses on learning modality-invariant similarity measures that can be embedded within classical optimization-based registration frameworks. Instead of directly predicting transformations, neural networks are trained to evaluate similarity or correspondence between image regions across modalities, replacing or augmenting handcrafted metrics such as MI. For example, *Grzech et al., 2022* [29] presented a variational Bayesian model in which the similarity metric is optimized jointly with the registration model. This not only enhances alignment accuracy compared to fixed similarity functions but also provides voxel-wise uncertainty estimation of the transformation.

Segmentation-guided registration. Another prominent hybrid strategy leverages anatomical segmentation to guide or constrain the registration process. Segmentation

masks can be used to weight similarity terms, restrict optimization to anatomically relevant regions or generate geometric representations such as landmarks and surface point clouds. By incorporating prior anatomical knowledge, these approaches improve robustness and reduce the influence of irrelevant or noisy image regions. In intraoperative contexts, real-time segmentation networks, often based on U-Net architectures and their variants, are increasingly employed to extract anatomical structures from ultrasound data. The resulting segmentations can be used to initialize registration, enforce anatomical consistency, or drive subsequent geometric alignment stages.

Azampour et al., 2024 [7] proposed a CT-to-US registration pipeline that combines segmentation and neural point-cloud matching. The method extracts spinal bone surfaces from both modalities using a U-Net and registers them through a deep network constrained by anatomy-aware losses, including rigidity and biomechanical consistency. These constraints enforce rigid motion of bones while allowing for realistic deformation of surrounding soft tissues, resulting in improved robustness to initialization and anatomically consistent registration outcomes (Figure 2.4).

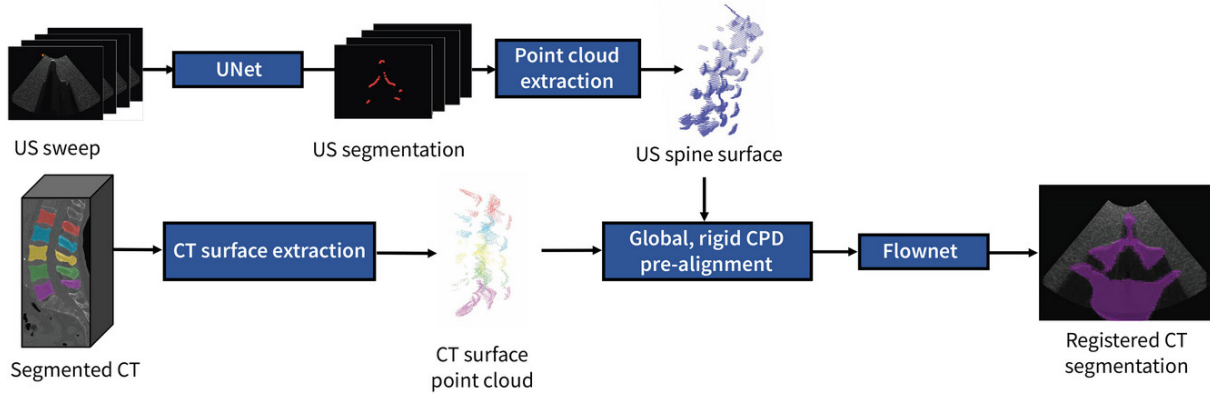


Figure 2.4: Example of a segmentation-guided registration pipeline [7]. Ultrasound volumes are segmented with a U-Net to extract a surface point cloud, which is used to pre-align a CT-derived point cloud through a global rigid CPD step. A subsequent network-based refinement (*FlowNet*) produces the final registered CT segmentation.

2.2. Image segmentation

Image segmentation is the process of dividing an image into distinct regions in order to identify structures or areas of interest. In medical imaging, segmentation is commonly used to isolate anatomical structures, support quantitative analysis and simplify the interpretation of complex image data. In the context of image-guided interventions,

segmentation plays a particularly important role, as it allows the extraction of anatomical information from intraoperative imaging modalities, which often suffer from limited field of view, noise, and low contrast. In such scenarios, reliable segmentation is essential to derive geometric representations that can be exploited for navigation and registration purposes [19, 84]. An example of a semantic, multi-class medical image segmentation is shown in Figure 2.5.

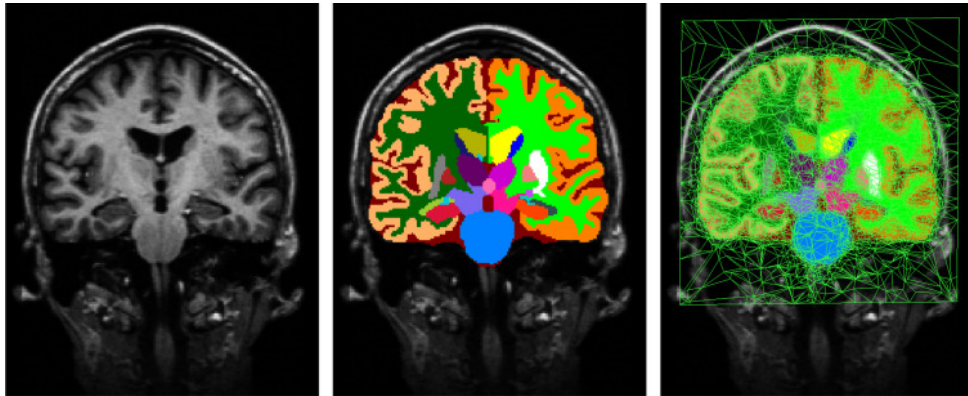


Figure 2.5: Example of multi-class anatomical segmentation of brain MRI, obtained using a parametric Bayesian modeling approach [54].

2.2.1. Classical segmentation

Early segmentation techniques relied on handcrafted algorithms exploiting low-level image features such as intensity, gradients and texture. Common techniques include threshold-based methods, edge detection and region-based algorithms, which segment images by applying explicit rules or heuristics derived from image statistics. While these methods are computationally efficient and relatively simple to implement, their performance is often sensitive to noise, image quality and parameter selection. As a result, classical segmentation techniques tend to struggle in medical imaging scenarios characterized by low contrast, artifacts and high inter-patient variability [19, 84].

Due to these limitations, classical segmentation approaches have gradually been replaced by robust and automatic learning-based techniques, particularly in applications involving complex anatomy and challenging imaging conditions.

2.2.2. Learning-based segmentation

To overcome the limitations of handcrafted techniques, learning-based segmentation methods have become the dominant approach in medical imaging. These methods aim to learn

relevant feature representations directly from data, enabling robust performance across varying image quality, noise levels and anatomical variability.

By leveraging large annotated datasets, learning-based models can capture complex appearance patterns that are difficult to encode explicitly, making them particularly well suited for challenging modalities such as ultrasound.

Convolutional neural networks

Among learning-based approaches, convolutional neural networks (CNNs) have emerged as the most widely adopted models for medical image segmentation. CNNs employ convolutional filters to extract hierarchical feature representations while preserving the spatial structure of the input image. This property is essential for segmentation tasks, where accurate localization of anatomical boundaries is required in addition to semantic discrimination [46].

Through successive convolutional and pooling or strided convolution layers, CNNs progressively increase their receptive field, allowing the integration of contextual information over larger image regions. Despite their effectiveness, maintaining precise spatial localization while exploiting global context remains a key challenge, particularly when segmenting thin or poorly contrasted anatomical structures [14, 46].

Encoder-decoder architectures

For image segmentation tasks, the most commonly adopted CNN architectures are based on an encoder–decoder structure. These architectures consist of a contracting path (encoder), which progressively extracts high-level features while reducing spatial resolution, and an expanding path (decoder), which reconstructs a dense segmentation map at the original image resolution.

A widely used family of encoder–decoder architectures is the U-Net [14, 59] (Figure 2.6), which introduced skip connections between corresponding encoder and decoder layers. These connections allow the network to combine high-level contextual information with fine-grained spatial details, significantly improving segmentation accuracy, particularly for small or thin anatomical structures. Due to its effectiveness and relatively simple design, the U-Net has become a reference architecture in medical image segmentation and has been successfully applied across different imaging modalities.

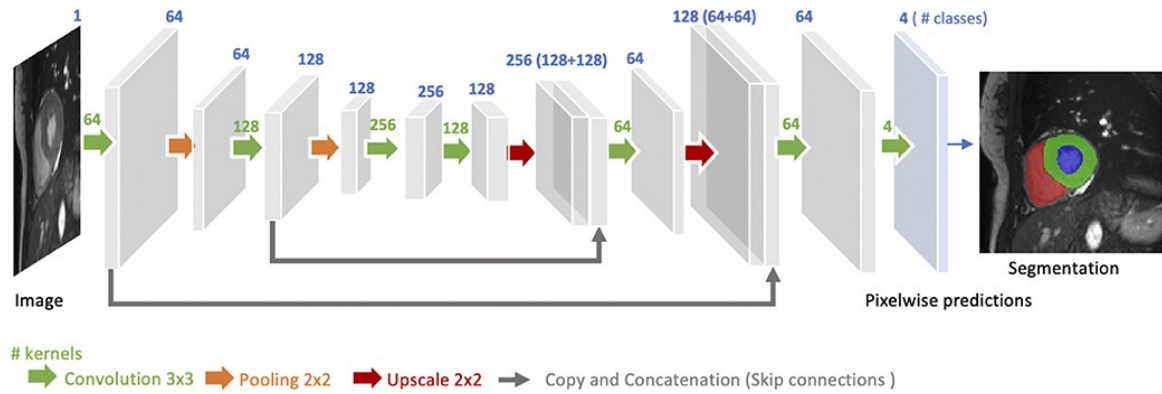


Figure 2.6: Schematic representation of a U-Net encoder–decoder architecture for multi-class cardiac chambers segmentation [14].

Several variants of the U-Net architecture have been proposed to address specific limitations of the original design in challenging segmentation scenarios.

Oktay et al., 2018 [52] developed the *Attention U-Net* architecture. It introduces attention gates within the skip connections to selectively emphasize anatomically relevant regions while suppressing irrelevant background features, improving segmentation accuracy in cases where target structures occupy a small portion of the image or exhibit low contrast.

Another example is the *U-Net3+*, by *Huang et al., 2020* [33]. It extends the encoder–decoder paradigm by incorporating full-scale skip connections that aggregate feature maps from multiple resolution levels at each decoding stage. This design enhances multi-scale feature fusion and improves boundary delineation, particularly in tasks requiring precise localization across varying spatial scales.

Despite these architectural differences, such variants retain the core encoder–decoder principle of U-Net while targeting improved robustness and accuracy in complex medical imaging applications.

2.2.3. US image segmentation challenges

Ultrasound imaging poses specific challenges for automatic segmentation due to its physical acquisition principles and operator-dependent nature. Common sources of degradation include speckle noise, low signal-to-noise ratio, reverberation artifacts, intensity inhomogeneities and acoustic shadowing (Figure 2.7). These factors often obscure anatomical boundaries and reduce image contrast, making the reliable delineation of structures particularly difficult [55].

In intraoperative settings, these challenges are further amplified by time constraints, lim-

ited field of view, and variability in probe positioning and acquisition parameters. As a result, segmentation models must operate robustly under non-ideal and rapidly changing imaging conditions.

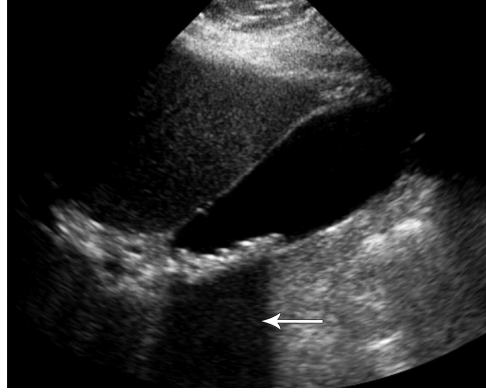


Figure 2.7: Example of the acoustic shadowing artifact in a liver ultrasound image [50].

2.2.4. Limited data and generalization.

From a learning perspective, ultrasound segmentation is often performed under severe practical constraints. Annotated datasets are typically limited in size due to the cost and expertise required for manual labeling. Moreover, strong class imbalance is common, as target anatomical structures frequently occupy a small fraction of the image compared to the background.

In addition, ultrasound data exhibit substantial variability across patients, operators, probes, and acquisition settings. This domain variability increases the risk of overfitting and limits the generalization capability of segmentation models trained on narrowly distributed datasets. To address these issues, robust regularization strategies are required at both the data level (e.g., augmentation) and model level (e.g., dropout).

Data augmentation. Data augmentation aims to improve generalization by exposing the model to plausible variations of the input during training, effectively increasing the diversity of the training distribution without collecting new labels [68]. In medical imaging, standard spatial augmentations include random rotations, translations, scaling, elastic deformations and cropping; intensity augmentations include brightness/contrast shifts and additive noise (Figure 2.8). In practice, augmentation acts as a strong prior: it encourages the network to learn representations that are invariant to nuisance factors (probe angle, gain, mild motion) while remaining sensitive to the target anatomy [53, 68].

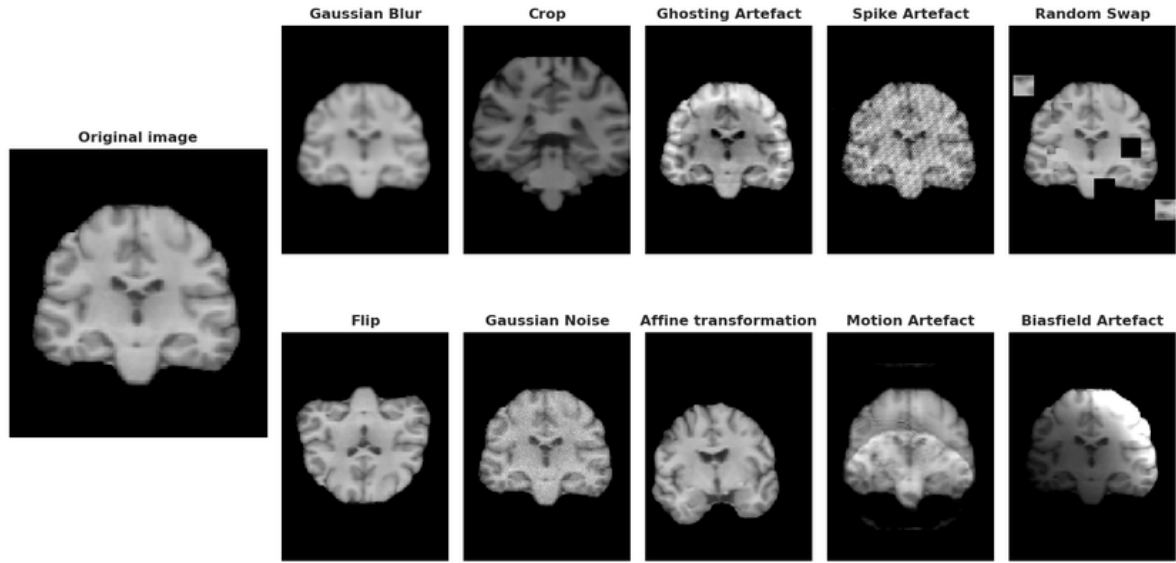


Figure 2.8: Examples of data augmentation strategies commonly applied during training, including spatial transformations and intensity variations [22].

Dropout for regularization and uncertainty estimation. Dropout was originally introduced as a regularization technique that reduces co-adaptation of network features by randomly deactivating units during training, as shown in Figure 2.9. It often improves generalization in over-parameterized models.

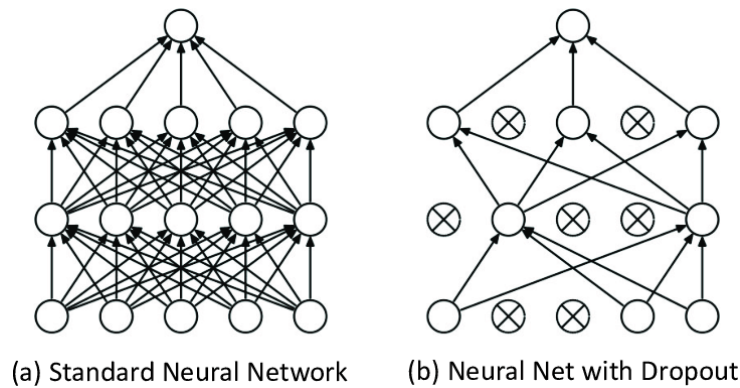


Figure 2.9: Schematic representation of the dropout regularization strategy [58].

Beyond its use in training, dropout can be kept active at inference time to perform multiple stochastic forward passes, a technique commonly referred to as *Monte Carlo dropout*. Under suitable assumptions, this procedure provides an approximation to Bayesian inference in deep neural networks and enables empirical estimation of predictive uncertainty [26]. In ultrasound segmentation, this is useful because not all pixels are equally reliable:

ultrasound artifacts can induce to ambiguous boundaries prediction. By sampling the model multiple times with dropout enabled, per-pixel uncertainty maps can be computed (e.g., predictive variance or entropy) alongside the mean prediction. Pixel-wise uncertainty maps derived from Monte Carlo sampling can support quality control, conservative decision-making or downstream filtering of reconstructed geometries.

Loss functions for imbalanced segmentation. A central design choice in segmentation is the loss function, which shapes the optimization landscape and determines how the model trades off false positives and false negatives. A common baseline for pixel-wise classification is Binary Cross-Entropy (BCE):

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (2.5)$$

where $y_i \in \{0, 1\}$ denotes the ground-truth label and $p_i \in [0, 1]$ the predicted probability at pixel i . While BCE provides stable gradients and effective optimization in balanced settings, it tends to be dominated by background pixels under strong class imbalance, which is common in vascular and organ segmentation tasks.

To address this limitation, overlap-based loss functions have been widely adopted. A common formulation is the Dice loss, defined as:

$$\mathcal{L}_{\text{Dice}} = 1 - \text{DSC}, \quad (2.6)$$

where the Dice Similarity Coefficient (DSC) (see appendix A.1) measures the overlap between predicted and ground-truth regions. Dice-based losses directly optimize region overlap and have been shown to improve performance in highly imbalanced scenarios [46].

More flexible formulations have been proposed to further control the balance between false positives and false negatives. Among these, the Focal-Tversky loss builds on the Tversky Index (TI) (see appendix A.1) and introduces a focusing mechanism to emphasize hard-to-segment pixels:

$$\mathcal{L}_{\text{FT}} = (1 - \text{TI})^\gamma, \quad \gamma > 1. \quad (2.7)$$

By increasing the contribution of misclassified and boundary pixels, the Focal-Tversky loss improves robustness in the presence of extreme class imbalance and thin anatomical structures, and has demonstrated superior performance in challenging medical segmentation tasks [1].

In practice, pixel-wise losses such as BCE are often combined with overlap-based terms

to stabilize optimization while promoting region-level consistency.

Cross-validation for model selection. When annotated datasets are limited, performance estimates obtained from a single train–validation split may be unstable and optimistic. Cross-validation (CV) mitigates this issue by repeating training and evaluation across multiple data splits and averaging the results, providing a more reliable estimate of generalization performance [73]. K-fold CV is commonly used: the data are partitioned into K folds; each fold acts as a validation set once while the remaining folds form the training set.

In medical imaging, careful CV design is essential to avoid data leakage. In particular, data should be split at the patient level rather than at the frame level, as samples from the same patient are often highly correlated. Failure to do so can lead to overly optimistic performance estimates and misleading conclusions [57, 74].

2.3. Vascular phantoms for in-vitro testing

Vascular phantoms are physical models used to reproduce selected features of blood vessels in a simplified and controllable way. In biomedical research, they are commonly adopted to support *in-vitro* experiments, providing a simpler, cheaper and practical alternative to *in-vivo* and *ex-vivo* tests. Their use allows experimental conditions to be defined in advance and maintained consistently, making vascular phantoms suitable for systematic testing and comparison of different methods.

In the context of image-guided interventions and image processing pipelines, vascular phantoms play a crucial role in method development and validation. They allow for systematic evaluation of imaging quality, segmentation robustness, and registration accuracy without the ethical, logistical, and variability constraints associated with clinical data.

Depending on the intended application, vascular phantoms can be designed to emphasize geometric realism, mechanical compliance, imaging compatibility, or flow dynamics. As a result, a wide variety of phantom designs has been reported in the literature, each tailored to specific experimental requirements.

2.3.1. Design requirements for vascular phantoms

To effectively support experimental studies, vascular phantoms are typically designed to satisfy a limited set of application-dependent requirements, including:

- accurate reproduction of **vascular geometry**,

- appropriate **mechanical response** of the vessel or surrounding material,
- compatibility with the targeted **imaging modality**,
- suitability for controlled **flow** or **loading conditions**.

The relative importance of each requirement depends strongly on the intended use case. Imaging-oriented studies prioritize geometric fidelity and imaging contrast, whereas mechanical or hemodynamic investigations require realistic material properties and stable integration into flow loops. Consequently, phantom design involves a trade-off between realism, durability, fabrication complexity and experimental control.

2.3.2. Thin-walled vascular phantoms

Thin-walled vascular phantoms (Figure 2.10) consist of hollow structures in which the vessel is represented by a physical wall with limited thickness. They are commonly fabricated using elastomeric materials, which provide long-term mechanical stability and a deformable mechanical response under loading. These phantoms are typically integrated into in-vitro flow loops, where controlled pulsatile pressure or flow conditions can be applied to reproduce physiological loading scenarios.

For example, *Yazdi et al., 2019* [81] developed a silicone-based compliant arterial phantom and employed it in a mock circulatory loop with pulsatile flow to investigate hemodynamic behavior and wall motion, demonstrating the suitability of thin-walled phantoms for fluid-dynamic and compliance studies.

One of the main advantages of thin-walled phantoms is the possibility to directly observe wall deformation, both visually and through optical or imaging-based techniques, due to the transparency of the materials often used. This feature makes them particularly suitable for mechanical testing, such as compliance evaluation and dynamic response analysis under pulsatile flow. In addition, these phantoms are generally durable and maintain their geometrical and mechanical properties over long periods, allowing repeated experimental use without significant degradation.



Figure 2.10: Example of a thin-walled silicone vascular phantom reproducing a patient-specific pathological arterial geometry [18]. The phantom was developed for experimental evaluation and training in catheter-based endovascular procedures.

2.3.3. Wall-less vascular phantoms

Wall-less vascular phantoms (Figure 2.11) are characterized by channels embedded within a continuous block of material, where the vessel lumen is defined as a cavity rather than by a physical wall. They are commonly fabricated using soft tissue-mimicking materials that can be cast around sacrificial molds. This fabrication approach is relatively simple and enables the creation of complex vascular geometries with good reproducibility. The main advantage of wall-less phantoms lies in their high compatibility with medical imaging modalities. The absence of a distinct vessel wall reduces imaging artefacts and improves contrast at the lumen–background interface, making these phantoms especially suitable for imaging-based validation studies and segmentation or registration benchmarks.

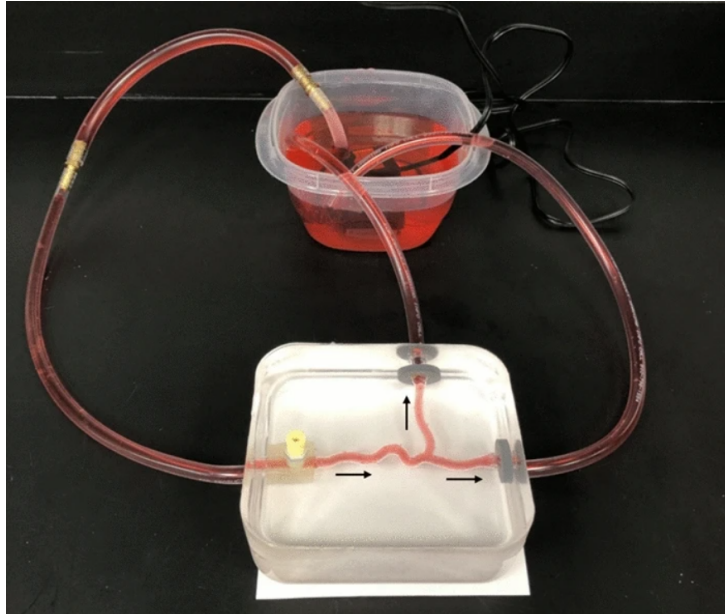


Figure 2.11: Wall-less phantom consisting of a bifurcating arterial geometry embedded in a gelatin-based tissue-mimicking matrix and integrated into a closed-loop flow circuit [43].

However, wall-less phantoms present several limitations. The materials are generally fragile, prone to cracking and poorly suited to withstand mechanical loads or pulsatile conditions for a prolonged period of time. As a result, compliance testing and dynamic mechanical studies are more difficult. Furthermore, these materials may undergo swelling or deformation over time, leading to reduced geometric stability and a shorter usable lifetime. The removal of sacrificial cores in complex geometries can also be challenging, further limiting their applicability.

2.3.4. Tissue-mimicking materials

The selection of tissue-mimicking materials plays a crucial role in vascular phantom fabrication and is closely linked to the intended experimental application. Rather than being inherently associated with a specific phantom category, material choice reflects a compromise between mechanical behavior, imaging compatibility and fabrication constraints. Elastomeric polymers (e.g. silicone) are commonly adopted when mechanical stability and durability are required. Water-rich polymer gels [25] (e.g. gelatine, agar, PVA) are often preferred for imaging-oriented applications due to their acoustic similarity to soft tissues. Oil-based polymer gels [13] (e.g. ballistic gel, *gel-wax*) represent an alternative solution, offering different mechanical properties and handling characteristics.

In addition to the base materials, various additives are frequently introduced to tailor specific properties. Scattering particles or contrast agents can be incorporated to improve ultrasound visibility, while fillers or cross-linkers may be used to adjust stiffness, attenuation or long-term stability. As illustrated in Figure 2.12, different material compositions result in distinct ultrasound echogenicity profiles, which directly affect image contrast and vessel visibility. These differences are particularly relevant in ultrasound-based experiments, where material-dependent imaging characteristics can influence segmentation accuracy and registration performance.

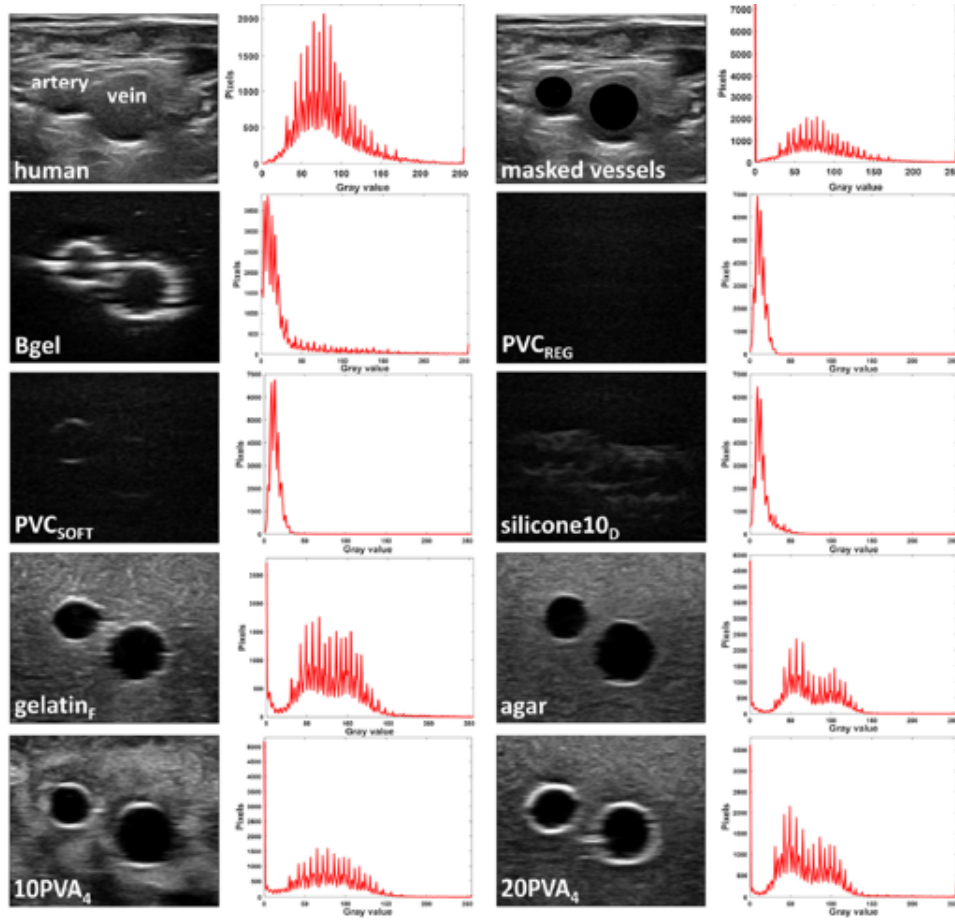


Figure 2.12: Ultrasound images and corresponding gray-level intensity histograms of different common tissue-mimicking materials [35].

3 | Materials and methods

This chapter describes the materials and methods adopted to develop and assess an intraoperative CT–US registration framework for vascular applications. The proposed approach is structured as a modular pipeline that integrates preoperative data preparation, real-time intraoperative processing and automatic ultrasound image analysis.

The first part of the chapter presents the methodological components of the framework. It introduces the intraoperative registration pipeline, detailing both the preoperative and intraoperative stages and outlining the overall structure of the proposed workflow. A dedicated section describes the automatic ultrasound vessel segmentation module based on convolutional neural networks. For the sake of conciseness, the subsections focus on the rationale and structure of the implemented methodologies, while the complete list of parameter values used in the study is reported in Appendix B.

The computational tools implemented in the previous part are tested through two sets of experiments. First, in-vitro laboratory experiments are conducted to assess the accuracy and robustness of the CT–US registration pipeline under controlled and repeatable conditions. Subsequently, the automatic US vessel segmentation network is evaluated on in-vivo data to investigate its applicability in clinically realistic scenarios.

All experiments were conducted on a laptop workstation equipped with an Intel® Core™ i7-13620H CPU (13th generation, 2.40 GHz), 16 GB RAM (5200 MT/s), and an NVIDIA GeForce RTX 4060 Laptop GPU with 8 GB dedicated memory, running a 64-bit Windows operating system. The CPU was primarily used for geometric processing tasks, whereas the GPU was employed for neural network training and inference through PyTorch with CUDA acceleration.

3.1. Methodological Pipeline for CT–US Vascular Registration

The method is designed to perform the automatic registration between a patient-specific preoperative CT model and intraoperative US data. The process starts from a simple US sweep, collected together with the corresponding poses provided by a robotic arm. This acquisition is completely non-invasive and is carried out just before the intervention, while the patient is already positioned on the operating table.

The registration must be accurate enough to support surgical navigation while remaining fast and robust, as the entire process needs to be compatible with the constraints of an operating room environment. For this reason, all steps of the intraoperative pipeline are designed to be computationally efficient and to minimize user interaction.

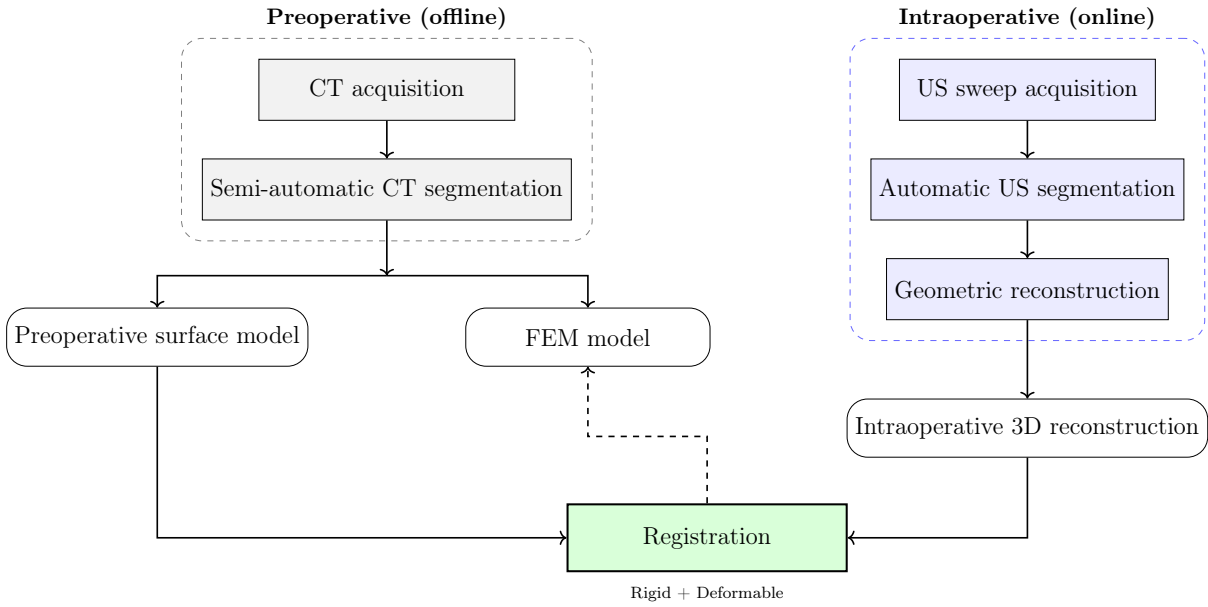


Figure 3.1: Overview of the proposed framework. By registering the preoperative model with the intraoperative reconstruction, the proposed framework aligns the associated reference frames and estimates the transformations required to update the patient-specific finite element model (FEM).

3.1.1. System architecture and coordinate definition

Proposed interventional workflow

The proposed framework is designed to integrate into the clinical interventional workflow, minimizing additional procedural complexity and avoiding disruption of standard

clinical practice. As shown in Figure 3.1, the overall workflow is organized into an offline preoperative phase and an online intraoperative phase. In the preoperative phase, a patient-specific anatomical model is generated from a clinical CT scan. This step is performed offline, prior to the intervention, and does not introduce additional burden during the procedure. During the intraoperative phase, after patient positioning and prior to the interventional task, a short freehand ultrasound sweep is acquired using a robot-mounted probe. The acquisition is non-invasive and requires no manual annotations or user-driven initialization. Ultrasound images are automatically segmented, reconstructed into a 3D representation and registered to the preoperative model without operator intervention.

The resulting CT–US alignment provides a spatial relationship between preoperative and intraoperative data, enabling the update of the patient-specific biomechanical model and supporting downstream navigation or simulation tasks.

Reference frames and transformations

To enable consistent data acquisition, a system of reference frames must be defined (see Figure 3.2). The world frame corresponds to the robot base, which remains fixed. From this reference, all other subsystems are described in a common spatial framework.

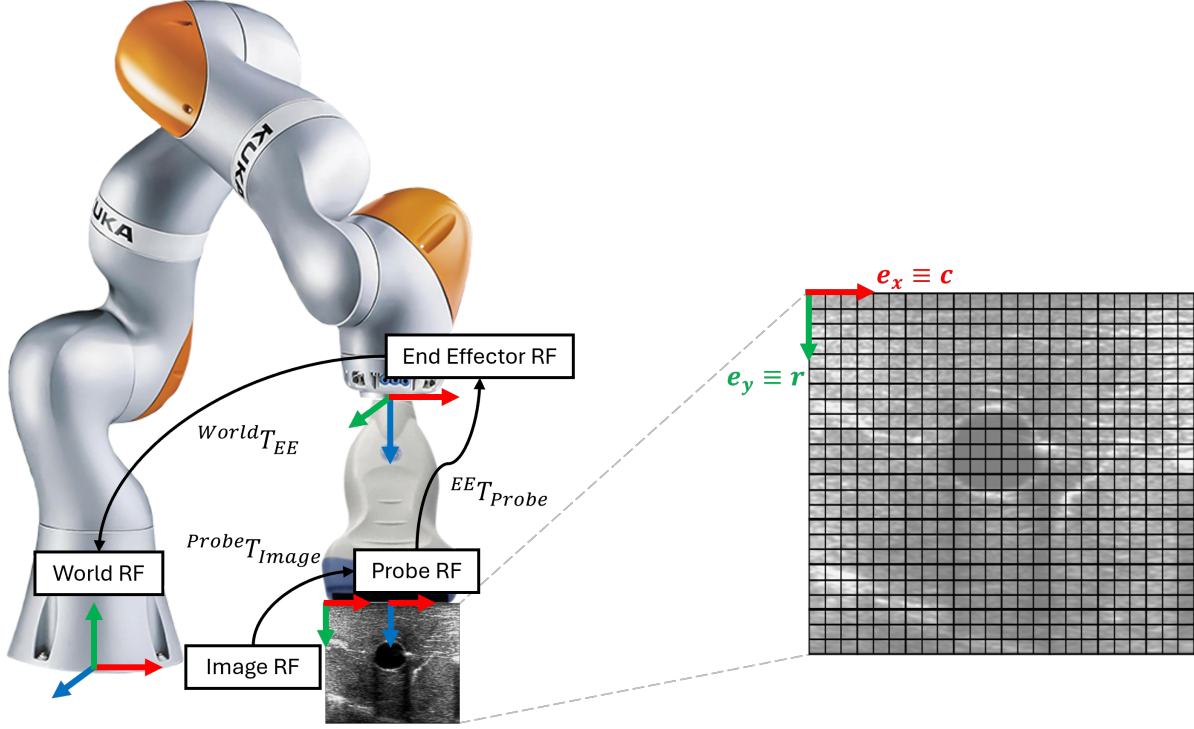


Figure 3.2: Reference frames and transformation chain relating the robot base, end effector, ultrasound probe and image coordinate system. Image axes follow the conventional pixel-based notation.

Rigid transformations are represented in homogeneous form, following the notation ${}^A T_B$, which maps coordinates from frame B to frame A . The relevant frames in the context of this thesis are the **World** reference frame (W), the **End effector** reference frame (E), the **Probe** reference frame (P) and the **Image** reference frame (I). The image reference frame is defined with its origin in the top-left corner, the x-axis oriented rightward, and the y-axis oriented downward, following the conventional image coordinate notation.

The main transform chain is:

$${}^W \mathbf{T}_I = {}^W \mathbf{T}_E \circ {}^E \mathbf{T}_P \circ {}^P \mathbf{T}_I \quad (3.1)$$

where:

- ${}^W \mathbf{T}_I$ is the transformation mapping image pixel coordinates into the physical world;
- ${}^W \mathbf{T}_E$ is obtained directly from the robot controller (*pose messages* in ROS);
- ${}^E \mathbf{T}_P$ is constant, determined from probe space calibration (already performed);
- ${}^P \mathbf{T}_I$ is the fixed geometric mapping from image coordinates to the probe reference

frame.

3.1.2. Preoperative data preparation

The preoperative pipeline is performed offline, before the surgical procedure. Its goal is to generate a geometrical model of the anatomical structure of interest. This stage requires the availability of a preoperative 3D clinical image of the patient. In many cases, a CT scan is already part of the standard diagnostic workflow, so it can be exploited without modifying the clinical routine.

Data acquisition and segmentation. The patient-specific geometry is reconstructed starting from a clinical CT scan provided by IRCCS Ospedale San Raffaele. The target vessel is automatically segmented using the *TotalSegmentator* extension [76] in 3D Slicer [24, 40] (version 5.8.1), and the result is then manually refined to ensure anatomical consistency.

The segmented volume is post-processed in MeshMixer [5, 64] (version 3.5.474) to remove residual artefacts and to obtain a clean 3D model of the vessel. Then, in ParaView [2, 6] (version 5.13.2), the volume is clipped at two positions along the vessel in order to isolate the region of interest, obtaining sharp and well-defined inlet and outlet cross-sections. This operation ensures well-defined open boundaries, which are required for subsequent centerline extraction and for the robustness of the registration pipeline.

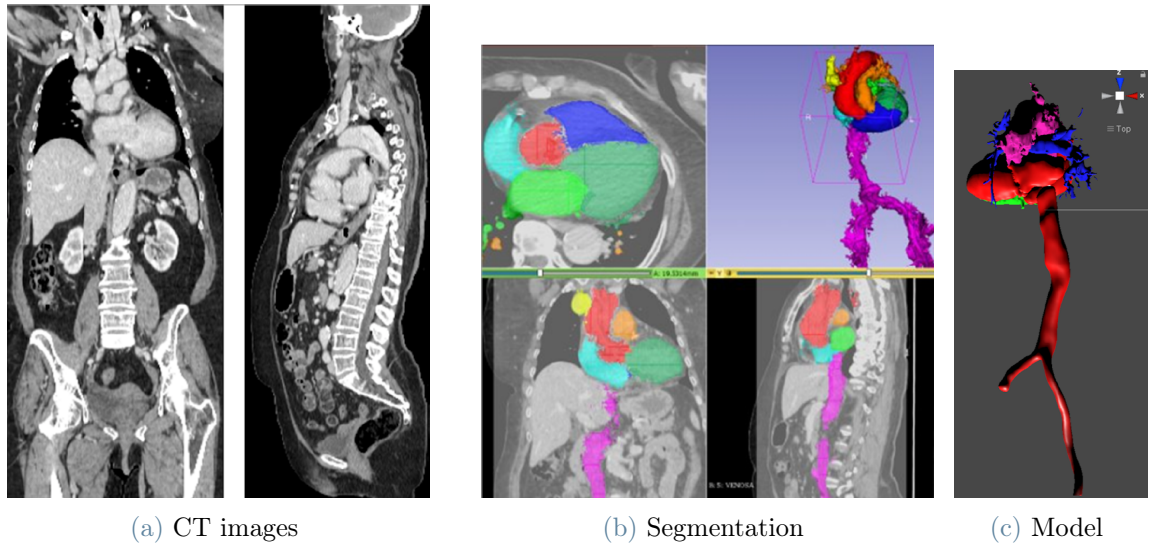


Figure 3.3: The model of all the biological structures involved in MitraClipTM procedure is obtained from the segmentation of CT preoperative images [62].

Centerline extraction. The centerline of the vessel was extracted using a custom Python script integrating the *Vascular Modeling Toolkit* (VMTK). The algorithm computes the path connecting the inlet and outlet sections through the lumen center, with a mild smoothing applied to obtain a continuous and noise-free curve. The extracted centerline is used in the registration process, allowing the initial alignment between the CT model and the intraoperative data using a reduced number of representative points.

Point cloud extraction. Starting from the volumetric segmentation-derived surface model, the geometry is converted into a point cloud representation. Surface points (and associated normals) are extracted to obtain a discrete representation of the anatomy, suitable for efficient registration with intraoperative data.

3.1.3. Online intraoperative pipeline

Intraoperative data acquisition and processing

In the proposed pipeline, intraoperative data are acquired using an ultrasound system whose probe is mounted on the end-effector of a robotic arm. The robot provides both the position and orientation of the probe during the scanning procedure, enabling the reconstruction of space calibrated US images.

The acquisition is managed through a ROS 2 interface (*rclpy* library), where two distinct topics are published: one with the US images and the other with the end-effector poses. Due to the different acquisition frequencies, image–pose synchronization is needed. To this aim, pose interpolation is carried out along the spatial trajectory of the end-effector. At this stage, the poses refer to the end-effector with respect to the world reference frame (${}^W\mathbf{T}_E$). The pose data, originally represented as translations and quaternions, are converted into 4×4 homogeneous transformation matrices. The composition with a constant calibration matrix (${}^E\mathbf{T}_I = {}^E\mathbf{T}_P \circ {}^P\mathbf{T}_I$) allows to determine the pose of each US image in the world coordinate system (${}^W\mathbf{T}_I$).

When native DICOM export or spatial metadata are not available (e.g., in proprietary ultrasound systems), frames can be acquired via a frame grabber capturing the monitor output. In such cases, a cropping operation is applied to isolate the effective ultrasound field of view from the full-screen image prior to further processing. Cropping and basic image handling operations were performed using OpenCV.

If in-plane spatial resolution (pixel spacing) is not provided, it can be estimated from the known probe field of view (FOV). Assuming isotropic pixel geometry, the pixel spacing is obtained by dividing the probe FOV (in mm) by the number of pixels spanning the probe

direction in the acquired frame. In the laboratory setup adopted in this work, the probe field of view was 40 mm over 251 pixels, resulting in an estimated in-plane pixel spacing of approximately 0.1594 mm.

Let s_x and s_y denote the pixel spacing along the image axes. Given the image pose $\mathbf{t}_{Image}$ and the orthonormal basis vectors $(\mathbf{e}_x, \mathbf{e}_y)$ defining the image plane orientation, using the same notation shown in Figure 3.2, the mapping from image coordinates (r, c) to the world coordinate system is defined as:

$$\mathbf{P}_{World}(r, c) = \mathbf{t}_{Image} + cs_x\mathbf{e}_x + rs_y\mathbf{e}_y \quad (3.2)$$

Automatic segmentation

The automatic segmentation module integrated in the intraoperative pipeline is designed to process 2D US images in order to generate a binary mask of the target vessel (see Figure 3.4). The workflow consists of three main stages: (i) pre-processing of the input image, (ii) segmentation by a convolutional neural network (CNN) and (iii) post-processing of the predicted mask.

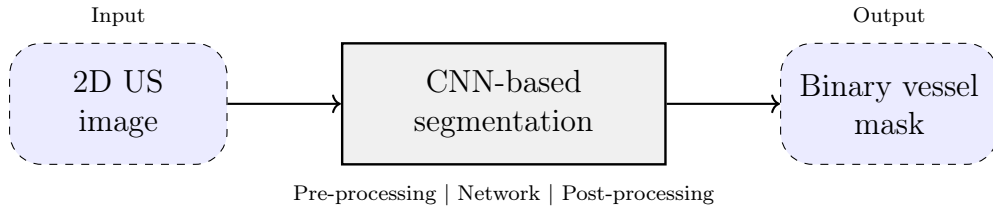


Figure 3.4: Automatic segmentation module integrated in the intraoperative pipeline.

All models were implemented in PyTorch and trained/inferred on GPU using CUDA acceleration.

Pre-processing. US images often present variability in resolution, contrast and noise level, mainly due to speckle and differences among acquisition systems and settings. Before being passed to the network, each image undergoes a standardized pre-processing procedure to ensure consistent input quality and stable predictions, to reduce the influence of acquisition variability on model performance. The pre-processing is the same described in Section 3.2.2.

US segmentation network. The US segmentation step is performed by a convolutional neural network (CNN) that automatically detects the vessel region within each

ultrasound frame. The network follows a typical encoder–decoder structure, allowing the extraction of relevant spatial and semantic features in the encoding path and reconstructs a dense prediction map in the decoding path. The network output consists of logits, representing the raw confidence values for each pixel. These logits are then converted into calibrated probabilities through affine temperature scaling, ensuring a more reliable and interpretable probability map. Finally, a hysteresis thresholding procedure is applied to obtain a binary segmentation. Further details about the network architecture, training strategy and calibration procedure are provided in Sections 3.2.3 and 3.2.4.

Post-processing. After the segmentation network produces the binary mask of the vessel, a post-processing stage refines the result to obtain a clean and coherent segmentation. This step includes a combination of morphological operations aimed at improving spatial continuity, removing small artefacts and ensuring a consistent vessel shape. A more detailed description of the post-processing procedures is provided in Section 3.2.5. The final binary mask represents the segmented cross-section of the vessel and serves as the geometric input for the subsequent 3D reconstruction and registration steps.

Geometric reconstruction

After automatic segmentation, each ultrasound frame is associated with a binary vessel mask and with a calibrated pose in the world reference frame. As illustrated in Figure 3.5, geometric reconstruction aims at converting the sequence of 2D segmented cross-sections into consistent 3D representations of the vessel: a surface point cloud, a watertight surface mesh and an estimate of the vessel centerline.

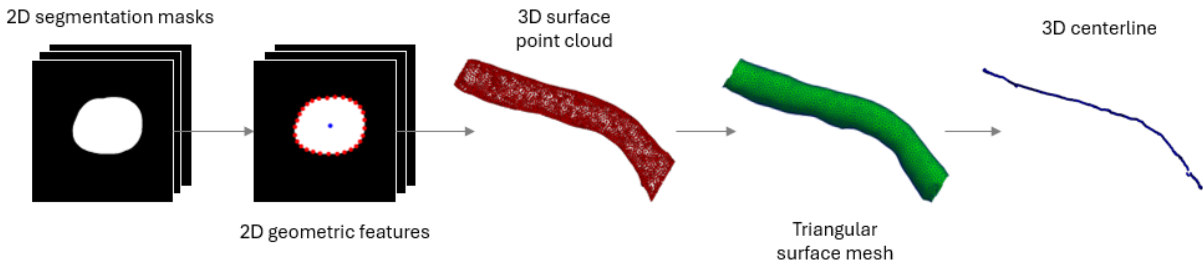


Figure 3.5: Overview of the geometric reconstruction process.

Contour points extraction. For each segmented frame, the vessel boundary is extracted from the binary mask by identifying the outer contour. The resulting 2D polyline is optionally resampled at a fixed spatial step to obtain a uniform point distribution across frames. Each contour point (r, c) is then mapped to 3D coordinates using the

known mapping between the image and the world reference frame (Equation 3.2).

Point cloud fusion and processing. All 3D contour profiles are aggregated to form a raw surface point cloud. Since intraoperative acquisitions can contain irregular sampling density and occasional spurious points, the surface point cloud undergoes a processing step to improve geometric consistency. In particular, point density is regularized through voxel-based downsampling (with *voxel size* of 1.25 mm). A radius-based neighborhood definition is adopted to both remove isolated outliers and define local neighborhoods for subsequent Taubin smoothing [71]. Normals are estimated after processing, producing a cleaned surface representation suitable for next steps. Point cloud processing operations (downsampling, outlier removal and normal estimation) were implemented with Open3D.

Surface mesh reconstruction. A triangulated surface mesh is reconstructed from the preprocessed oriented point cloud using Poisson surface reconstruction [11, 38] (Open3D implementation). This approach is well suited for freehand ultrasound acquisitions because it provides a smooth, watertight approximation of the underlying surface and naturally integrates surface normals. The reconstruction is followed by an internal trimming step that removes low-density regions, reducing surface artefacts generated by sparsely sampled areas. Inlet and outlet boundaries are removed, obtaining an open lateral surface of the vessel.

Remeshing and filtering. The raw Poisson mesh is further processed to obtain a well-conditioned triangulation for subsequent geometric operations. First, an isotropic remeshing stage is applied (PyMeshLab), including hole closing to handle small discontinuities. The remeshing target edge length is chosen equal to the voxel size (1.25 mm) adopted for point cloud downsampling, to achieve a consistent resolution across representations. Then, a bilateral smoothing filter [20] is applied to reduce high-frequency artefacts while preserving geometric features; a boundary-aware strategy is adopted to avoid excessive shrinkage near open boundaries. The resulting processed mesh is the reference surface representation used in the following registration stages.

Centerline extraction. An estimate of the vessel centerline is extracted from the reconstructed surface mesh, following a similar approach as in Section 3.1.2 (*Centerline Extraction*). Two seed points are automatically selected from the acquisition by taking the mask centroids of the first and last frames providing a valid contour, thereby defining proximal and distal endpoints. Given the surface mesh and the endpoint seeds, the centerline is computed using a VMTK-based algorithm and subsequently resampled at a

fixed spatial step. The uniform 2 mm resampling enforces a constant spatial resolution; therefore, the number of points depends on vessel length, while the inter-point spacing remains fixed.

Rigid registration

Rigid registration estimates a global rigid transformation between two 3D representations of the vessel. In the proposed framework, rigid alignment is implemented as a single registration module applied twice in the pipeline: first on centerlines to obtain a stable initialization, and then on surface point clouds to refine the alignment using full geometric detail.

Rigid registration module. Given a source point set and a target point set, the rigid module (Figure 3.6) estimates the transformation $\mathbf{T}$ minimizing the alignment error under nearest-neighbor correspondences. The procedure consists of two sequential stages:

- **Initial optimization.** Both point sets are first centered at their centroids. Then, an initial rigid transform is estimated by minimizing the Root Mean Square Error (RMSE) between the transformed source points and their closest target points using a BFGS optimizer (Broyden–Fletcher–Goldfarb–Shanno) in a 6-DOF parameterization (3 Euler angles + translation). The standard SciPy BFGS implementation is used, relying on its default gradient-based convergence criterion, with a maximum number of iterations set to 100.
- **Multi-start ICP refinement.** Starting from the BFGS solution, a set of 8 candidate initializations is generated to handle typical ambiguities of quasi-cylindrical anatomies. The main axis of the source set is estimated via Principal Component Analysis (PCA), and multiple rotations around this axis are tested (e.g., quarter-turn symmetries). In addition, a flip candidate is included to account for possible inlet–outlet inversions. Each candidate is refined independently using point-to-plane ICP [12, 44] (Open3D), and the solution showing the minimum RMSE is selected as final rigid alignment.

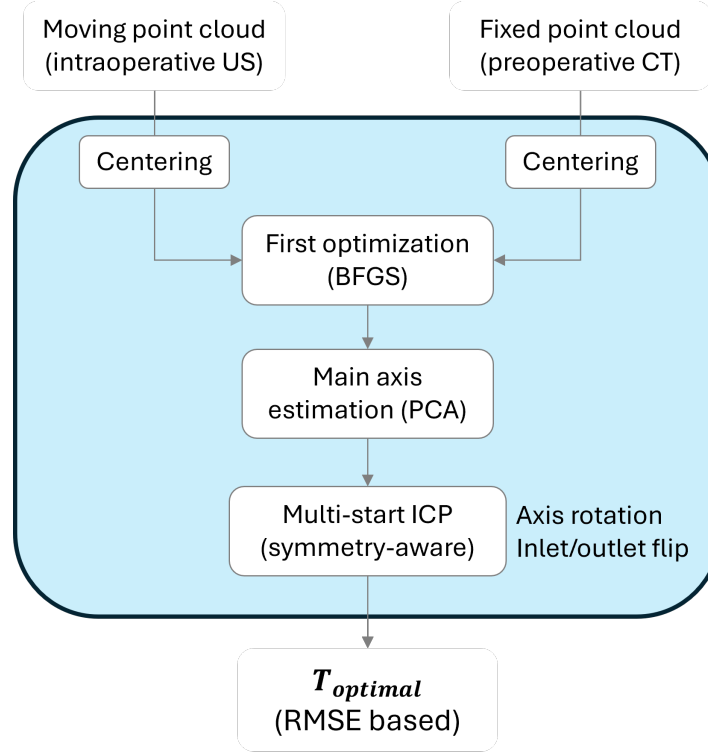


Figure 3.6: Rigid registration strategy scheme based on BFGS initialization and symmetry-aware multi-start ICP refinement.

Two-stage strategy. The rigid registration module is used twice, with different geometric inputs:

1. **Centerline-based rigid alignment.** The first application aligns the preoperative and intraoperative centerlines. Since centerlines provide a compact and stable descriptor of vessel path, this step yields a robust global initialization even when surface data are noisy or partially observed.
2. **Surface-based rigid refinement.** The centerline-derived transformation $T_{\text{Centerline}}$ is then applied to the preoperative surface point cloud, providing a physically meaningful initialization for the surface alignment. Starting from this initialization, the rigid registration module (BFGS + multi-start ICP) is executed on the surface point clouds, producing the final rigid transformation $T_{\text{Rigid}} = T_{\text{Surface}} \circ T_{\text{Centerline}}$.

The resulting rigid transformation provides a geometrically consistent initialization for the subsequent deformable registration stage.

Deformable registration

After rigid alignment, a deformable registration step is applied to compensate for residual local discrepancies between the preoperative CT-derived model and the intraoperative ultrasound reconstruction. These discrepancies are mainly due to vessel deformation induced by probe pressure, patient positioning and physiological variability. In the proposed framework, deformable registration is performed using an existing PyTorch-based implementation of the Coherent Point Drift (CPD) algorithm in its non-rigid formulation (*torchcpd*).

The non-rigid CPD parameters, namely the Gaussian kernel width β and the regularization weight λ , were selected through preliminary empirical tuning. In the CPD formulation, β controls the spatial smoothness of the deformation field, while λ balances data fidelity and regularization [49]. Values reported in the literature depend on the spatial scale and sampling density and typically lie in the range of a few units. In this work, $\beta = 5$ and $\lambda = 8$ were adopted, which fall within commonly used ranges and are consistent with the millimetric resolution of the reconstructed surfaces.

The source point set corresponds to the rigidly transformed preoperative surface, while the target point set is given by the intraoperative surface reconstruction. The algorithm estimates a smooth deformation field that improves surface correspondence while preserving the global vessel topology. Only a single deformable registration step is performed, and no additional anatomical constraints are introduced. The resulting deformation is used to refine the alignment prior to updating the patient-specific finite element model.

3.2. Automatic ultrasound image segmentation

This section describes the automatic ultrasound image segmentation module used within the proposed intraoperative framework. The goal of the segmentation network is to extract a reliable binary representation of the target vessel from 2D ultrasound images, which is subsequently exploited for geometric reconstruction and registration.

The segmentation pipeline is designed to operate robustly under challenging ultrasound imaging conditions, including speckle noise, variable contrast and limited field of view. To this end, the proposed approach combines standardized image preprocessing, data augmentation, supervised training of a convolutional neural network and a dedicated calibration and post-processing strategy.

3.2.1. Datasets

In this work, four US datasets characterized by different acquisition conditions, anatomical targets and levels of variability (see Figure 3.7) were used. The datasets include both phantom and *in-vivo* data and are selected to progressively increase the complexity of the segmentation task, from controlled *in-vitro* conditions to heterogeneous clinical scenarios. Manual binary annotations for the two in-house datasets (*Phantom* and *San Raffaele*) were performed by the author using *3D Slicer* (version 5.8.1).

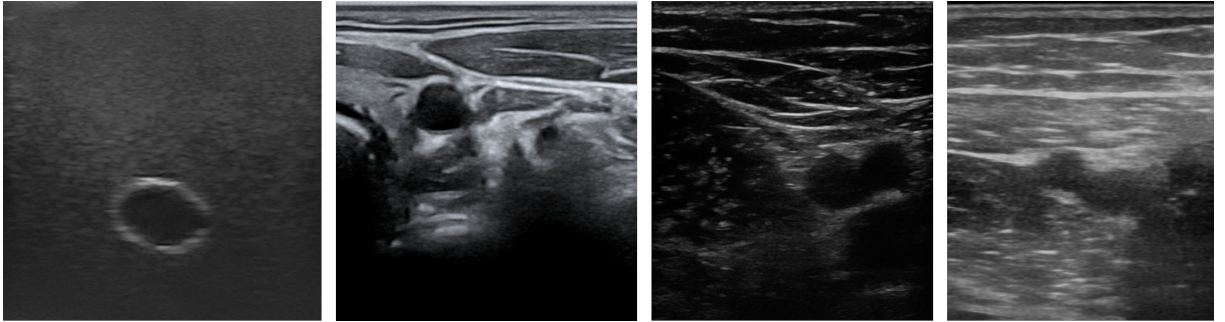


Figure 3.7: Examples of US images from different datasets. From left to right: *Phantom*, *Carotid*, *San Raffaele* and *Mus-V*.

Phantom dataset. This dataset consists of US acquisitions performed on the vascular phantoms described in Section 3.3.1. Images were acquired using a *Telemed MicrUS EXT-1H* ultrasound system with a linear probe, operating in B-mode. The dataset includes approximately 11,000 images collected from 20 independent acquisitions performed on phantoms with different lumen geometries.

Public carotid artery dataset. This publicly available dataset [48] consists of short-axis US images of the common carotid artery acquired *in-vivo*. It includes 1,100 images collected from 11 acquisitions using a *Mindray UMT-500Plus* ultrasound system. Corresponding expert segmentation masks were produced by a trained technician and verified by a clinical expert prior to publication.

San Raffaele femoral vein dataset. This dataset consists of *in-vivo* ultrasound images of the femoral vein acquired at IRCCS Ospedale San Raffaele using a *Vivid iQ* ultrasound system (GE Healthcare). The dataset includes 3,494 images collected from 35 independent acquisitions.

Mus-V dataset. The Mus-V dataset [28] is a heterogeneous *in-vivo* ultrasound dataset comprising 3,114 images collected from 105 acquisitions using an *Angell Pioneer H20* ul-

trasound scanner. The dataset includes both femoral and neck acquisitions and provides multiclass annotations distinguishing arterial and venous structures. Since the proposed network performs binary segmentation, the Mus-V dataset was reformulated into separate binary segmentation tasks. The underlying images remain unchanged across configurations, while only the target class definition differs. In particular, two configurations were derived:

- *Mus-V Artery*, where arterial structures are considered foreground and all remaining pixels are treated as background;
- *Mus-V Vein*, where venous structures are considered foreground.

3.2.2. Dataset preparation

Dataset split

A careful data splitting strategy was required to obtain reliable performance estimates and to avoid data leakage, particularly in ultrasound imaging, where consecutive frames acquired during the same scan are strongly correlated.

For this reason, dataset splitting was performed at the *acquisition level*. All images belonging to the same acquisition were always assigned to the same subset, ensuring that no highly correlated samples from the same scan appear simultaneously in training, validation or test sets. This constraint prevents overly optimistic performance estimates that may arise when temporally adjacent frames are treated as independent samples. In this study, datasets were partitioned into 70% training, 15% validation and 15% test sets.

Dataset pruning

Ultrasound datasets may contain a high degree of redundancy, especially when acquisitions consist of long sequences of frames collected under quasi-static conditions. In such cases, consecutive images can be nearly identical and strongly correlated, providing limited additional information to the learning process. Training on highly redundant data may increase computational cost and promote overfitting to acquisition-specific appearance patterns rather than meaningful anatomical variability. To mitigate this effect, a dataset pruning strategy was adopted exclusively for the San Raffaele dataset, which was acquired in vivo using cine-loop US sequences and is therefore characterized by intrinsically limited intra-acquisition variability. The goal of pruning is to select a representative subset of images that preserves the variability of each acquisition while discarding highly redundant samples.

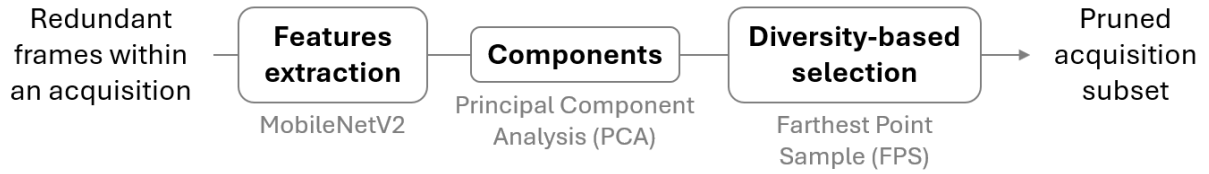


Figure 3.8: Dataset pruning pipeline: feature embedding, PCA projection and Farthest Point Sampling selection.

The pruning procedure (Figure 3.8) operates by embedding individual frames into a compact feature space using a lightweight convolutional feature extractor (*MobileNetV2*). Principal Component Analysis (PCA) is applied to capture the dominant modes of variation within each acquisition. A subset of frames is finally selected by maximizing coverage of the feature space, using a Farthest Point Sampling strategy (FPS) to retain samples that are maximally dissimilar from one another. Specifically, frames are iteratively selected until the subset explains at least 99% of the total variance of the acquisition in PCA space, while enforcing a minimum of 45 images per acquisition to prevent excessive pruning in highly redundant sequences. Feature embedding and pruning were implemented in PyTorch using a pretrained *MobileNetV2* backbone.

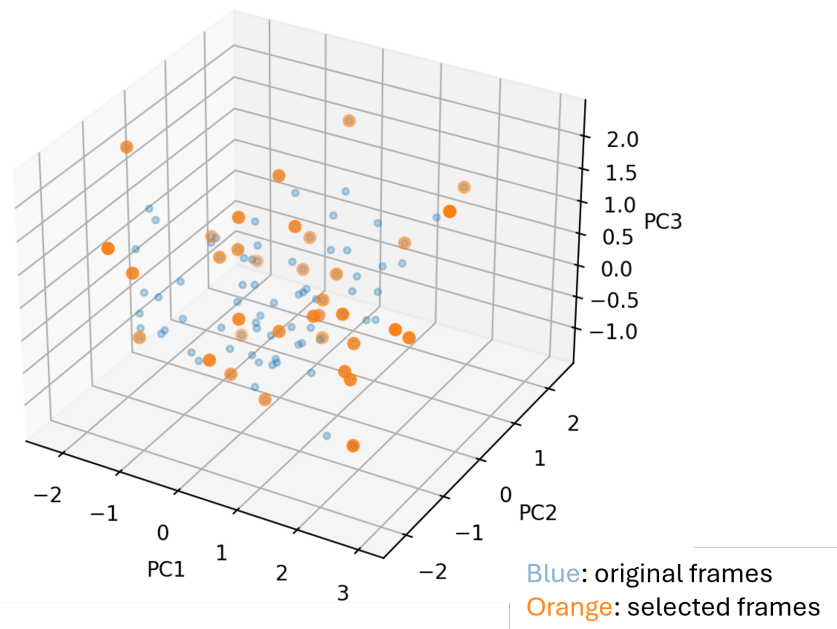


Figure 3.9: Example of pruning outcome on a single acquisition, showing reduced redundancy while preserving variability. PC1, PC2, PC3 represent the three principal components obtained by PCA.

This process is applied independently to each acquisition and allows reducing redundancy while preserving appearance diversity. An example for a single acquisition is shown in Figure 3.9. The resulting pruned dataset, reduced from 3494 to 1575 frames, provides a more balanced and informative training set, without altering the acquisition-level structure required for proper dataset splitting.

Image pre-processing

Ultrasound images may exhibit significant variability in resolution, contrast and noise level, mainly due to speckle and differences in acquisition settings and imaging systems. Before being passed to the segmentation network, each image undergoes a standardized pre-processing procedure, implemented in OpenCV, to ensure consistent input quality and stable predictions.

The pre-processing pipeline includes:

- conversion to single-channel grayscale (8-bit) and linear intensity normalization to $[0, 1]$ without additional clipping or saturation;
- resizing and padding performed jointly to obtain a fixed resolution of $H \times W$ (in this study: 256×256) while preserving the original aspect ratio and avoiding geometric distortions of anatomical structures;
- linear normalization of intensity values to the range $[0, 1]$, reducing variability due to acquisition parameters and probe settings.

Since padded regions do not correspond to meaningful image content, a binary *validity mask* is associated with each image to identify pixels belonging to the original field of view. This mask is propagated through the pre-processing pipeline and is subsequently used during both training and evaluation, ensuring that loss computation and performance metrics are restricted to valid image regions only.

3.2.3. Model design and training

Model design

CNN architectures. Three convolutional neural network architectures were evaluated in this work (see Section 2.2.2): a standard U-Net model, an attention-based variant of U-Net and a multi-scale aggregation architecture. These models were selected to assess the impact of different mechanisms for feature propagation and fusion on ultrasound vessel segmentation.

All networks follow an encoder–decoder design with three downsampling stages and three corresponding skip connections between encoder and decoder. The baseline U-Net uses two consecutive 3×3 convolutional layers per stage (with batch normalization and ReLU), with a progressive increase in feature channels (32–64–128) and a 256-channel bottleneck. The Attention U-Net has the same overall structure, but each skip connection is modulated by an attention gate: encoder features are weighted by a learned attention map computed from the decoder features at the same scale, so that irrelevant responses are suppressed before fusion. Finally, the multi-scale aggregation model combines information from multiple resolutions at each decoder stage by rescaling and concatenating encoder and decoder feature maps, and it employs auxiliary outputs during training (deep supervision) to stabilize learning.

Loss function. The model optimization was based on a composite loss combining a *Binary Cross-Entropy* (BCE) term and a *Focal-Tversky* (FT) term, selected to mitigate the strong class imbalance between vessel and background pixels and to improve overlap and boundary consistency (see Section 2.2.4). The relative contribution of the two terms is controlled by a weighting factor $\lambda \in [0, 1]$:

$$\mathcal{L}_{\text{total}} = (1 - \lambda)\mathcal{L}_{\text{BCE}} + \lambda\mathcal{L}_{\text{FT}}, \quad (3.3)$$

In this study, λ was set to 0.75, assigning greater weight to the Focal-Tversky term. This value was selected empirically based on preliminary experiments, as the network exhibited a tendency toward under-segmentation; increasing the contribution of the Focal-Tversky component helped mitigate false negatives without introducing excessive false positives. Loss computation is restricted to valid image regions by exploiting the validity mask introduced during pre-processing, so that padded areas do not contribute to the optimization.

Regularization strategy. Regularization is achieved through a combination of data-level and model-level strategies (see Section 2.2.4). First, on-the-fly data augmentation is applied during training to improve robustness to the variability typical of ultrasound acquisitions. In particular, standard geometric transformations (random flips, rotations, scaling and elastic deformations) and intensity-based perturbations (contrast and brightness variations) are applied consistently to images and corresponding masks. Second, dropout is used (with probability $p = 0.1$) as a model-level regularizer to reduce overfitting. Additional regularization is provided by weight decay in the optimizer and by early stopping during training.

Training and validation strategy

Training was performed in PyTorch with CUDA acceleration, using mixed-precision to reduce memory usage. The complete set of settings and parameters is reported in Appendix B.

Weights initialization and optimization. Since no pretrained weights are available, all convolutional layers are initialized using Kaiming Normal initialization to preserve stable activation variance across layers. Optimization is performed using the AdamW optimizer with an initial learning rate of 1×10^{-3} . A cosine annealing schedule progressively reduces the learning rate over epochs, promoting smooth convergence. Mixed-precision training is employed to accelerate computation and reduce GPU memory consumption.

Early stopping. Training is monitored on a validation set at the end of each epoch using the same preprocessing pipeline adopted for training, without data augmentation. Training is performed for a maximum of 50 epochs and is terminated early if the validation loss does not improve for 10 consecutive epochs (patience criterion).

Cross-validation protocol. To ensure robust performance estimation and reduce dependence on a specific train/validation split, a k -fold cross-validation scheme is adopted. For each fold, the model is re-initialized and trained independently on a different combination of data partitions. The best-performing checkpoint from each fold is saved and later used for model fusion. In the context of this work, a 5-fold CV strategy is employed. An overview of the full training and validation workflow is reported in Fig. 3.10. Model fusion and calibration are detailed in Section 3.2.4.

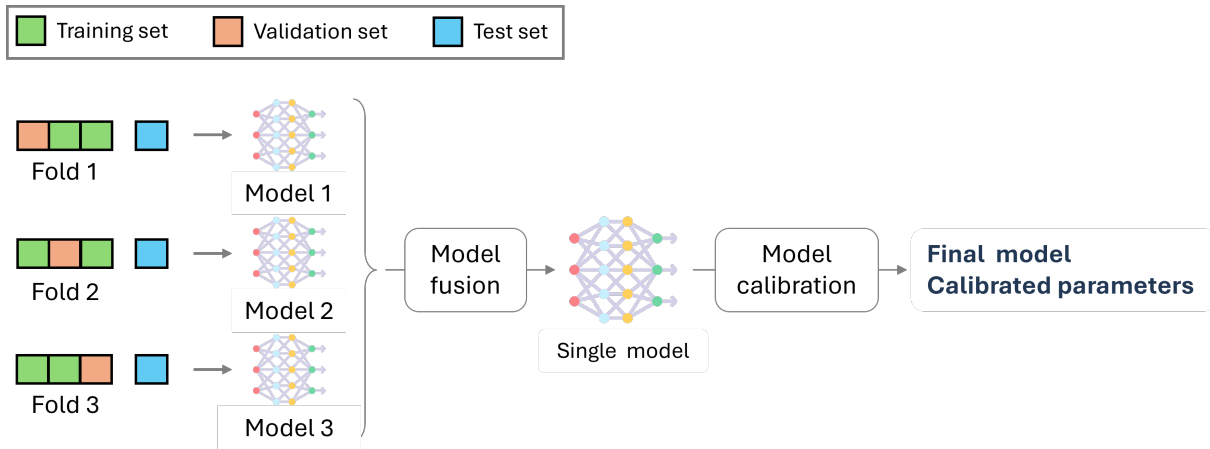


Figure 3.10: K-fold cross-validation workflow with model fusion and post-hoc calibration to obtain a single calibrated model. Example of a 3-fold CV scheme.

3.2.4. Model aggregation and calibration

Model fusion

The cross-validation protocol yields $k = 5$ independently trained models, each corresponding to a different train/validation split. While ensembling all models at inference time can improve robustness, it is not compatible with the near real-time requirements of the intraoperative pipeline. Therefore, a single consolidated model is obtained by parameter averaging using a *weighted model soup* strategy, inspired by *Wortsman et al* [79].

Let $\theta_1, \dots, \theta_k$ be the parameters of the best checkpoint from each fold. The fused parameters are computed as a convex combination:

$$\theta_{\text{fused}} = \sum_{i=1}^k w_i \theta_i, \quad \sum_{i=1}^k w_i = 1, \quad w_i \geq 0. \quad (3.4)$$

The weights w_i are optimized on a global validation subset by minimizing the segmentation loss. After fusion, Batch Normalization statistics are re-estimated by forwarding training data through the network without gradient updates. The resulting model is the final configuration used for calibration and testing.

Calibration and thresholding

A possible problem in DL-based image segmentation consists in the possibility to have models that achieve good accuracy, e.g. in terms of DSC and HD95, but with output probabilities that are poorly calibrated, especially under strong class imbalance. To address this potential issue, a post-hoc calibration and thresholding step was applied to the fused model.

Affine Temperature Scaling (ATS). Calibration was performed using an affine temperature scaling applied to the logits before the sigmoid activation:

$$\tilde{p} = \sigma(az + b), \quad (3.5)$$

where (a, b) are learned scalar parameters. The parameters were estimated on training and validation data by minimizing the binary cross-entropy loss and then fixed for all subsequent inference runs.

Hysteresis thresholding. After calibration, the probability map is converted into a binary mask using hysteresis thresholding with thresholds $(t_{\text{high}}, t_{\text{low}})$. Pixels with $\tilde{p} \geq$

t_{high} are immediately classified as part of the vessel region, while those with $\tilde{p} < t_{\text{low}}$ are discarded. Pixels whose probability falls between the two thresholds are included only if they are spatially connected to high-confidence regions. This procedure improves spatial coherence, suppresses isolated false positives and preserves weak but anatomically consistent boundaries. Optimal thresholds were selected empirically on the validation set by minimizing the 95th percentile Hausdorff distance (HD95) and then applied uniformly to all test and inference images.

3.2.5. Inference and evaluation

Mask post-processing

After network inference, a post-processing stage (OpenCV) is applied to obtain a single, clean binary vessel mask suitable for geometric reconstruction and registration.

- **Morphological filtering.** The binary mask is refined via disk-shaped morphological closing to seal small boundary gaps, followed by hole filling. A subsequent disk-shaped opening removes small spurious structures and regularizes the mask shape.
- **Single-component selection.** Connected components are extracted from the cleaned mask. If multiple candidates are present, the mask required by the pipeline is obtained by selecting the component maximizing a confidence–area score,

$$s(\Omega) = (\text{mean}_{x \in \Omega} p(x)) |\Omega|, \quad (3.6)$$

where $p(x)$ is the predicted probability map and $|\Omega|$ is the component area.

- **Boundary smoothing.** Finally, the selected component is optionally smoothed by filtering its contour in polar coordinates $r(\theta)$ with a 1D Gaussian filter to reduce high-frequency artefacts while preserving the overall vessel shape.

Testing

The final model is evaluated on a held-out test set, fully independent from training and validation data. For each test image, inference follows the same deterministic pipeline adopted during training. Images are pre-processed according to the standard procedure (see Section 3.2.2) and passed through the network to obtain pixel-wise logits. These are converted into calibrated probabilities by applying the sigmoid activation with the fixed affine temperature scaling parameters estimated during calibration (see Section 3.2.4).

The probability maps are then binarized using the two hysteresis thresholds selected on the validation set and post-processed as described in (see Section 3.2.5), yielding the final binary vessel masks. Segmentation performance is assessed on the test set using pixel-wise, region-based and boundary-based metrics, computed exclusively on valid image regions (see Appendix A.1).

3.3. In-vitro experiments for registration evaluation

This section describes the experimental assessment of the proposed registration pipeline using vascular phantoms in a laboratory setting. The objective of these experiments is to quantify the registration accuracy achieved by the pipeline when applied to realistic ultrasound acquisitions performed on known geometries.

3.3.1. Vascular phantom

Vascular phantoms are developed to provide a controlled in-vitro environment for the experimental evaluation of the CT-US registration pipeline. The phantoms are designed to reproduce realistic vascular geometries while ensuring good ultrasound visibility and geometric stability during repeated acquisitions.

Three different lumen geometries were considered in this work:

- an *Artificial* geometry, a tortuous tubular structure with variable cross-section;
- a patient-specific femoral *Vein* segment;
- a patient-specific femoral *Artery* segment.

The two anatomical geometries were extracted from the same preoperative CT dataset employed in the registration pipeline, ensuring consistency between the digital preoperative model and the physical phantom used during in-vitro experiments.

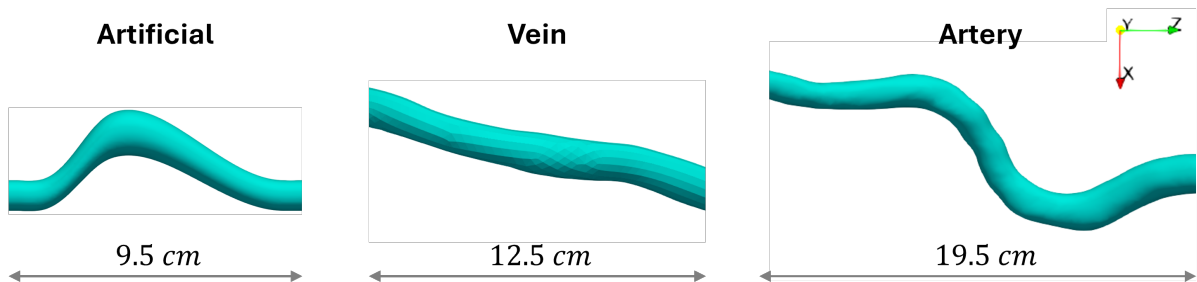


Figure 3.11: Representation of the three vessel geometries used to build the phantoms.

Material. The phantoms were fabricated using *gel-wax*, a commercially available polymer-based gel commonly adopted for ultrasound phantoms due to its good acoustic properties and ease of processing. Gel-wax provides sufficient echogenicity to clearly visualize vessel boundaries in B-mode ultrasound images and exhibits soft mechanical behavior compatible with tissue-mimicking applications.

To improve the stability and handling of the material, additives were mixed into the base gel-wax. In particular, paraffin wax was added at approximately 7% by weight to increase mechanical robustness and reduce fragility, while talcum powder was added at approximately 1% by weight to enhance ultrasound scattering and improve image contrast. This composition provided a suitable trade-off between echogenicity, deformability and durability for repeated imaging experiments.

Phantom fabrication. The phantoms were manufactured through a casting process based on a custom mold, designed using *Autodesk Fusion* (vers. 2605.1.52). The mold consisted of two main components:

- an external rigid shell, defining the outer shape of the phantom;
- an internal sacrificial core, defining the vessel lumen geometry.

The external shell was 3D-printed in PLA and assembled from multiple parts to facilitate demolding. The internal core was 3D-printed in water-soluble PVA, allowing its removal after casting without mechanically damaging the phantom. The mold components were prepared and sliced using *Ultimaker Cura* (vers. 5.10.0) and fabricated on a fused deposition modeling 3D printer (*Ultimaker 3*).

During fabrication, the gel-wax mixture was heated in a water bath until fully molten (approximately at 70 °C) and then poured in a single casting step into the assembled mold containing the PVA core. A single continuous pour was adopted to ensure material homogeneity, as gel-wax does not undergo self-polymerization and sequential castings would result in mechanically distinct layers. After pouring, the phantom was allowed to cool and solidify at room temperature; complete solidification required several hours, depending on the total volume of material. Once fully solidified, the phantom was immersed in warm water to dissolve the sacrificial PVA core, leaving a hollow lumen with the desired geometry. The dissolution time depended on the vessel diameter, length, and access openings, as well as on the printing parameters of the PVA core (e.g., infill density), and typically ranged from several hours to longer durations for smaller or more complex geometries. This process resulted in soft vascular phantoms with smooth internal surfaces and well-defined lumen shapes, suitable for ultrasound imaging and registration

experiments.

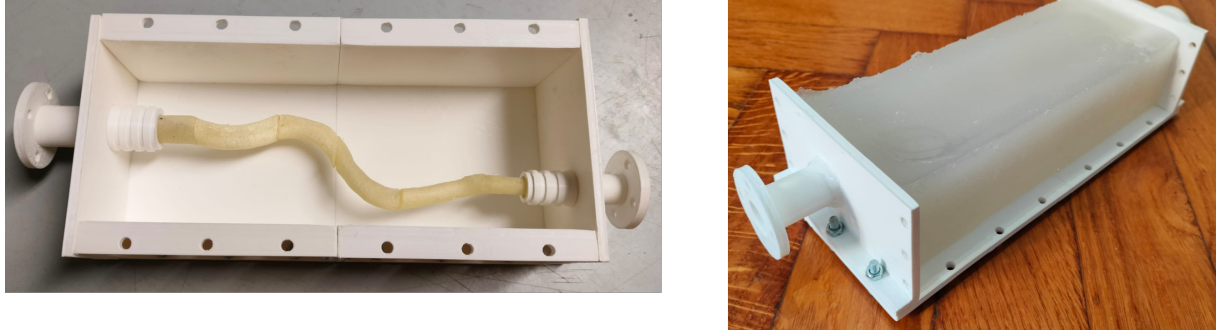


Figure 3.12: Femoral artery phantom fabrication using a PLA external mold, a sacrificial PVA core and gel-wax casting.

3.3.2. Data acquisition protocol

In-vitro data acquisition was performed in a laboratory environment using the vascular phantoms described in the previous section. The experimental setup was designed to reproduce a simplified but realistic ultrasound acquisition scenario, while ensuring controlled and repeatable conditions across different experiments. The setup consisted of three main components:

- the **vascular phantom**, positioned on a rigid support;
- the **ultrasound system** (*Telemed MicrUS EXT-1H*), operating in B-mode with a linear probe;
- the **robotic arm** (*7-DoF KUKA LBR iiwa*) used to manipulate the ultrasound probe.

The robotic arm was used as a tracking device, providing the pose of the ultrasound probe during acquisition, while probe motion was guided manually.



Figure 3.13: *In-vitro* US acquisition setup with probe pose tracking via robotic arm.

During data acquisition, the phantom remained stationary, mimicking the presence of a patient limb resting on the operating table. For each experiment, a sequence of ultrasound images was acquired by performing a freehand sweep along the longitudinal direction of the phantom. The sweep direction was approximately aligned with the vessel centerline, resulting in a sequence of cross-sectional views of the lumen, corresponding to short-axis vessel images.

3.3.3. Registration experiments

Each experiment follows the processing workflow described in Section 3.1. The experiments are designed to evaluate the performance of the proposed CT-US registration pipeline along four complementary aspects: reconstruction consistency, rigid registration robustness, registration accuracy and computational performance. Code registration parameters are listed in Table B.1 (Appendix B), while the evaluation metrics adopted are defined in Appendix A.2.

Three different vascular phantom geometries (*Artificial*, *Vein* and *Artery*) were considered. For each phantom, 8 independent ultrasound sweeps were acquired under identical experimental conditions and reconstructed using the complete pipeline described in Sec-

tion 3.1.

Reconstruction consistency

Reconstruction consistency was evaluated to quantify the variability introduced by repeated acquisitions of the same physical phantom. To assess repeatability, all pairwise combinations of the reconstructed surfaces were compared. Since the phantom was manually repositioned between acquisitions, a preliminary rigid alignment was performed prior to metrics computation in order to remove global translations and rotations due to physical repositioning. Distances were computed between the final processed surface meshes produced by the pipeline. Symmetric surface distance metrics (HD95, MSD and Chamfer distance) were adopted, as repeated acquisitions of the same phantom are expected to cover the same anatomical region. For each geometry, this procedure results in 28 pairwise comparisons, which are used to characterize reconstruction repeatability under variations in probe motion and acquisition conditions.

Rigid registration robustness

The robustness of the rigid alignment stage was evaluated by comparing the proposed multi-start strategy with a conventional single-start ICP baseline. Robustness was assessed by systematically perturbing the initial alignment of the intraoperative reconstruction through controlled rotational misalignments and additive noise.

Rotational perturbations were applied independently around the three coordinate axes using the discrete set $\{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$. All possible combinations were considered, resulting in 512 rotational configurations. No translational perturbation was introduced. In addition, zero-mean gaussian noise was added to the reconstructed point cloud with standard deviation $\sigma \in \{0.0, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0\}$ mm, emulating acquisition variability and segmentation uncertainty. For each combination of rotation and noise level, rigid registration was performed using both methods. The resulting alignment was evaluated in terms of RMSE and *orientation error* ($\Delta\theta$).

The reference alignment is defined as the rigid transformation computed by the proposed multi-start strategy between the original (non-perturbed) intraoperative reconstruction and the fixed surface. For each artificially perturbed configuration, the expected reference rotation is derived from this reference transformation. The orientation error $\Delta\theta$ is then computed as the angular difference between the estimated rotation and the expected rotation associated with the perturbed configuration. An alignment was considered successful when $\Delta\theta < 10^\circ$, ensuring consistency with the reference orientation.

The reference alignment was visually verified to correspond to a correct geometric overlap between the intraoperative reconstruction and the fixed surface. Empirically, configurations with low $\Delta\theta$ consistently showed proper surface alignment, while high $\Delta\theta$ values were associated with convergence to incorrect local minima (e.g., flipped or misoriented solutions). This confirms the suitability of $\Delta\theta$ as a reliability indicator in this setting.

Registration accuracy

Registration accuracy was evaluated by computing geometric distance-based metrics between the intraoperative reconstruction (moving surface) and the corresponding CAD phantom model derived from the preoperative mesh (fixed reference).

Since the intraoperative reconstruction typically covers only a portion of the vessel surface, distances were computed from the moving surface to the fixed reference (moving→fixed). Symmetric distances were avoided, as partial surface coverage would make them difficult to interpret and potentially inflate the error. Accuracy was quantified using RMSE and HD95. RMSE captures the average geometric discrepancy, while HD95 reflects residual localized mismatches. Metrics were evaluated after rigid alignment and after deformable registration, allowing the contribution of the non-rigid refinement step to be assessed relative to the rigidly aligned configuration.

Computational performance

The computational performance of the registration pipeline was evaluated by measuring the execution time of its main processing stages, as defined in Section 3.1. For each acquisition, the runtime of every pipeline component, together with the total execution time, was recorded. The reported execution time reflects the full end-to-end processing required to obtain the final aligned reconstruction.

Median execution times across repeated acquisitions were used to summarize performance for each phantom geometry. This analysis provides an estimate of the overall workflow duration and allows assessment of the practical feasibility of the proposed framework within intraoperative time constraints.

3.4. In-vivo evaluation of the segmentation network

This section presents the experimental evaluation of the automatic vessel segmentation network on ultrasound data acquired under both controlled and clinically realistic conditions. In particular, segmentation performance is analyzed on a phantom ultrasound

dataset, used within the in-vitro registration pipeline, as well as on multiple in-vivo datasets acquired on real vascular anatomies.

Phantom ultrasound images provide a simplified and controlled scenario, characterized by homogeneous background, high vessel-to-background contrast and limited anatomical variability. These data are employed to validate the basic behavior of the segmentation network and for the in-vitro registration experiments. In contrast, in-vivo ultrasound images are affected by significantly more challenging acquisition conditions, including strong speckle noise, variable contrast, surrounding anatomical structures and motion-related artefacts, which increase the complexity of the segmentation task.

The experiments presented in this section are designed to assess the segmentation network’s generalization capability and potential applicability in clinical scenarios. To this end, multiple datasets acquired with different imaging systems, anatomical targets and levels of variability are considered.

3.4.1. Segmentation experiments

The experimental protocol was designed to evaluate the performance and generalization capability of the proposed neural network architectures under different data conditions. All experiments were conducted using the same training and inference pipeline described in Section 3.2, and evaluation was performed using the metrics defined in Appendix A.1. The segmentation parameters adopted are listed in Table B.2 (Appendix B).

The following experiments are designed to assess both architectural variability and dataset-dependent generalization. For each dataset or dataset configuration, the network was trained and evaluated independently using acquisition-level splits (to avoid frame-level leakage) within the adopted cross-validation framework. Segmentation performance was primarily assessed using the Dice Similarity Coefficient (DSC), together with complementary metrics such as the 95th percentile Hausdorff Distance (HD95) reported in Appendix A.1. While DSC captures global overlap, HD95 provides additional insight into boundary accuracy, which is particularly relevant for thin structures such as vessels.

Model comparison across datasets

Three encoder–decoder architectures (U-Net, Attention U-Net, and U-Net 3+) were compared to assess the impact of architectural complexity on segmentation performance across datasets with different sizes and variability levels. All models were trained and evaluated under identical conditions.

Influence of dataset characteristics

This experiment investigates how segmentation performance is affected by (i) acquisition-level variability and (ii) dataset size, using the San Raffaele femoral vein dataset as reference. Three configurations were defined in order to disentangle these effects:

- **Single-session configuration:** a homogeneous subset including acquisitions performed within the same clinical session, using identical ultrasound settings and acquisition protocol. This configuration minimizes inter-acquisition variability while preserving the original acquisition structure.
- **Full configuration:** the complete dataset, including acquisitions obtained in different anatomical regions of the limb and acquired with varying ultrasound parameters. This setting introduces increased heterogeneity in anatomical appearance and imaging conditions.
- **Pruned configuration:** a reduced version of the *Full* dataset obtained through a pruning strategy aimed at removing highly redundant frames while preserving acquisition-level variability.

The comparison between the *Single-session* and *Full* configurations allows assessment of the impact of acquisition heterogeneity on segmentation performance. In contrast, the comparison between the *Full* and *Pruned* configurations isolates the effect of dataset size while maintaining similar variability across acquisitions. This experimental design enables evaluation of whether segmentation performance is primarily influenced by increased heterogeneity across acquisitions or by the total number of available training samples.

Vessel type and domain aggregation analysis

This experiment investigates segmentation performance with respect to two factors: (i) vessel type (artery vs vein) and (ii) domain aggregation (single anatomical district vs aggregated multi-district data). The Mus-V dataset includes both femoral and neck acquisitions. To enable a controlled analysis, six configurations were defined as summarized in Table 3.1.

Table 3.1: Mus-V dataset configurations used for vessel-type and domain aggregation analysis.

	Femoral district	Neck district	Aggregated
Artery	Mus-V Femoral Artery	Mus-V Carotid Artery	Mus-V Artery
Vein	Mus-V Femoral Vein	Mus-V Jugular Vein	Mus-V Vein

Single-district configurations allow comparison between arteries and veins within the same anatomical region, minimizing cross-domain variability. Aggregated configurations combine femoral and neck data for the same vessel type, enabling analysis of how domain aggregation affects segmentation performance.

4 | Results and discussion

This chapter presents and discusses the results obtained from the experimental evaluation described in Chapter ???. Results are organized according to the two main experimental components: (i) *in-vitro* registration experiments and (ii) segmentation experiments on phantom and *in-vivo* ultrasound data.

4.1. In-vitro registration results

4.1.1. Reconstruction consistency

Reconstruction consistency was evaluated as described in Section 3.3.3, following the procedure outlined in Section 3.1. Results are reported separately for the three phantom geometries.

Table 4.1: Reconstruction repeatability across phantom geometries. Median and interquartile range (IQR) are reported.

Geometry	HD95 [mm]	Chamfer [mm ²]	MSD [mm]
Artificial	2.126 [1.927–2.321]	1.395 [1.193–1.507]	0.947 [0.926–1.009]
Vein	2.070 [1.874–2.315]	1.209 [0.964–1.308]	0.872 [0.822–0.921]
Artery	1.989 [1.681–2.396]	1.442 [0.991–2.449]	0.853 [0.784–0.970]

The quantitative results summarized in Table 4.1 show a bounded variability across repeated reconstructions for all geometries. Median values are comparable across phantoms, while differences are mainly reflected in the spread of the distributions. Overall, the artificial and venous phantoms show similar repeatability. In particular, the artificial phantom presents the most compact distributions for MSD, while the venous geometry shows slightly wider dispersion on HD95 and Chamfer distance, with comparable median values. The arterial phantom exhibits the largest variability, most noticeably in Chamfer distance and also in HD95/MSD, indicating a higher sensitivity to acquisition and

segmentation variability in this geometry.

This behaviour is further illustrated in Figure 4.1, where median values remain stable and variability is mainly reflected by the interquartile range rather than by outlier-dominated behaviour.

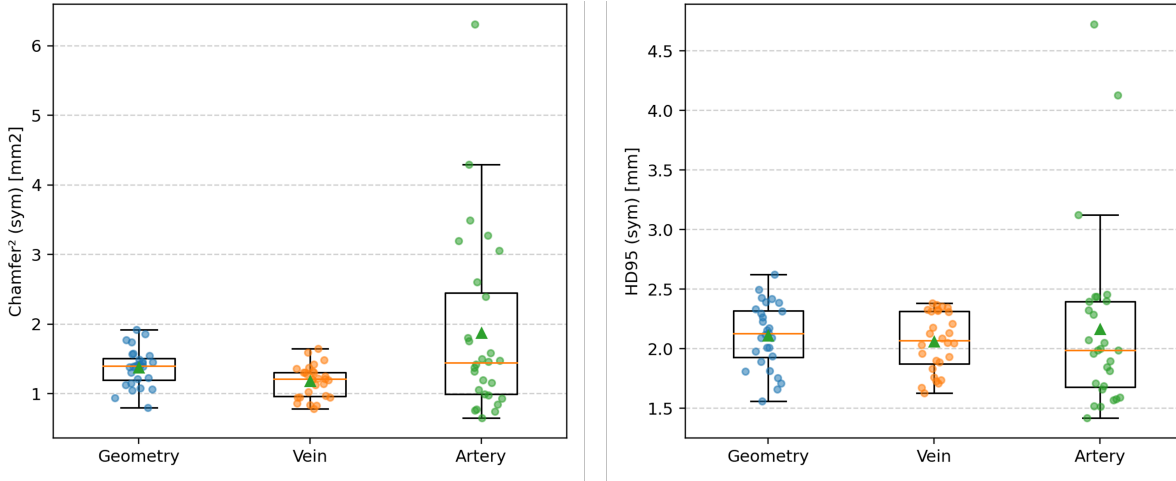


Figure 4.1: Distribution of Chamfer distance and HD95 across repeated reconstructions for each phantom geometry. Boxplots highlight comparable medians with geometry-dependent dispersion.

The increased dispersion observed for the arterial geometry is consistent with its higher geometric complexity, characterized by the presence of locally narrower vessel segments combined with increased curvature and irregular shape. These factors can amplify the impact of local reconstruction inaccuracies, particularly on tail-based metrics such as HD95.

Overall, these results suggest that the reconstruction pipeline is repeatable, with geometry-dependent variability that remains within a limited range. This supports the use of the reconstructed surfaces for subsequent registration and deformation analysis.

4.1.2. Rigid registration robustness

Rigid registration robustness was evaluated as described in Section 3.3.3, by analysing convergence behaviour under increasing rotational perturbations and noise levels.

Figure 4.2 provides qualitative examples comparing the single-start ICP baseline method and the proposed multi-start strategy for the three phantom geometries. For each phantom, one representative failure case of the baseline method and one successful case of the proposed method are shown. These examples illustrate typical incorrect convergences of

the baseline method under unfavorable initialization, while the proposed strategy is able to recover the correct orientation in the same conditions. Note that not all baseline runs fail, and not all proposed runs succeed.

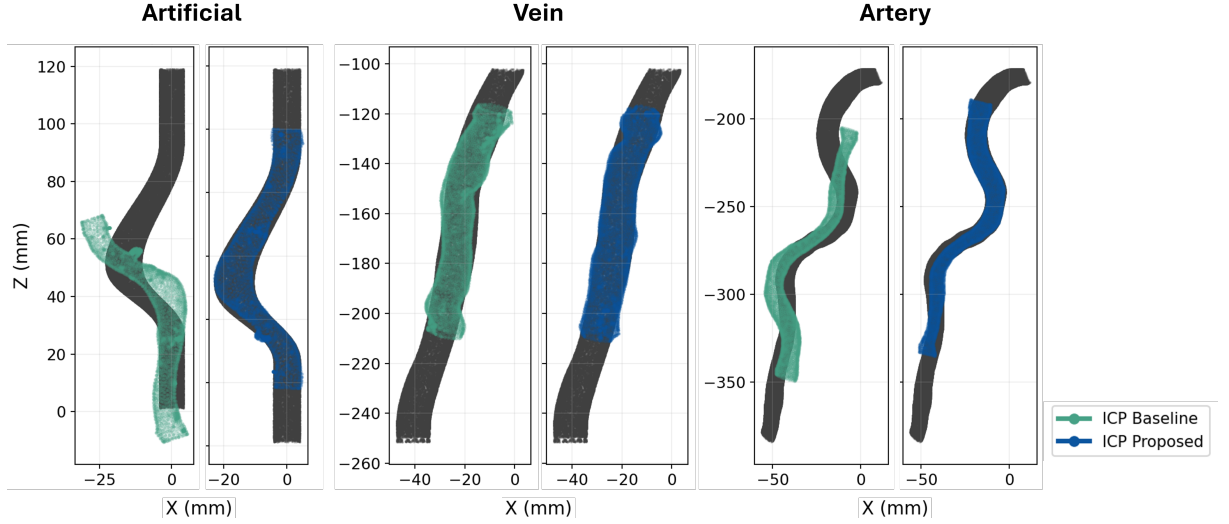


Figure 4.2: Representative examples of rigid registration under challenging initialization conditions. For each phantom geometry, one failure case of the single-start ICP baseline is compared with a successful alignment obtained using the proposed multi-start strategy.

A quantitative analysis of robustness is reported in Figure 4.3. For each phantom, the success rate is shown as a function of initial rotation magnitude and noise level. The success rate as a function of initial rotation highlights the main limitation of the baseline approach. The proposed multi-start strategy consistently maintains success rates above 80% across the entire rotational range for all phantom geometries. In contrast, the baseline method exhibits severe drops for specific rotational configurations, with success rates decreasing to approximately 10% in the venous phantom and remaining well below 50% in other unfavorable orientations. This behaviour indicates frequent convergence to incorrect local minima when the initialization is far from the correct basin of attraction.

Conversely, the success rate as a function of noise shows minimal sensitivity for both methods within the tested range ($\sigma \leq 2$ mm). However, the absolute performance levels remain markedly different. The baseline method remains approximately constant around 30–40% across all phantoms and noise levels, whereas the multi-start strategy consistently achieves success rates above 90%, reaching nearly 100% for the arterial phantom. This indicates that convergence robustness is primarily limited by rotational initialization rather than by moderate geometric noise, and confirms the effectiveness of the proposed multi-start strategy in expanding the basin of convergence.

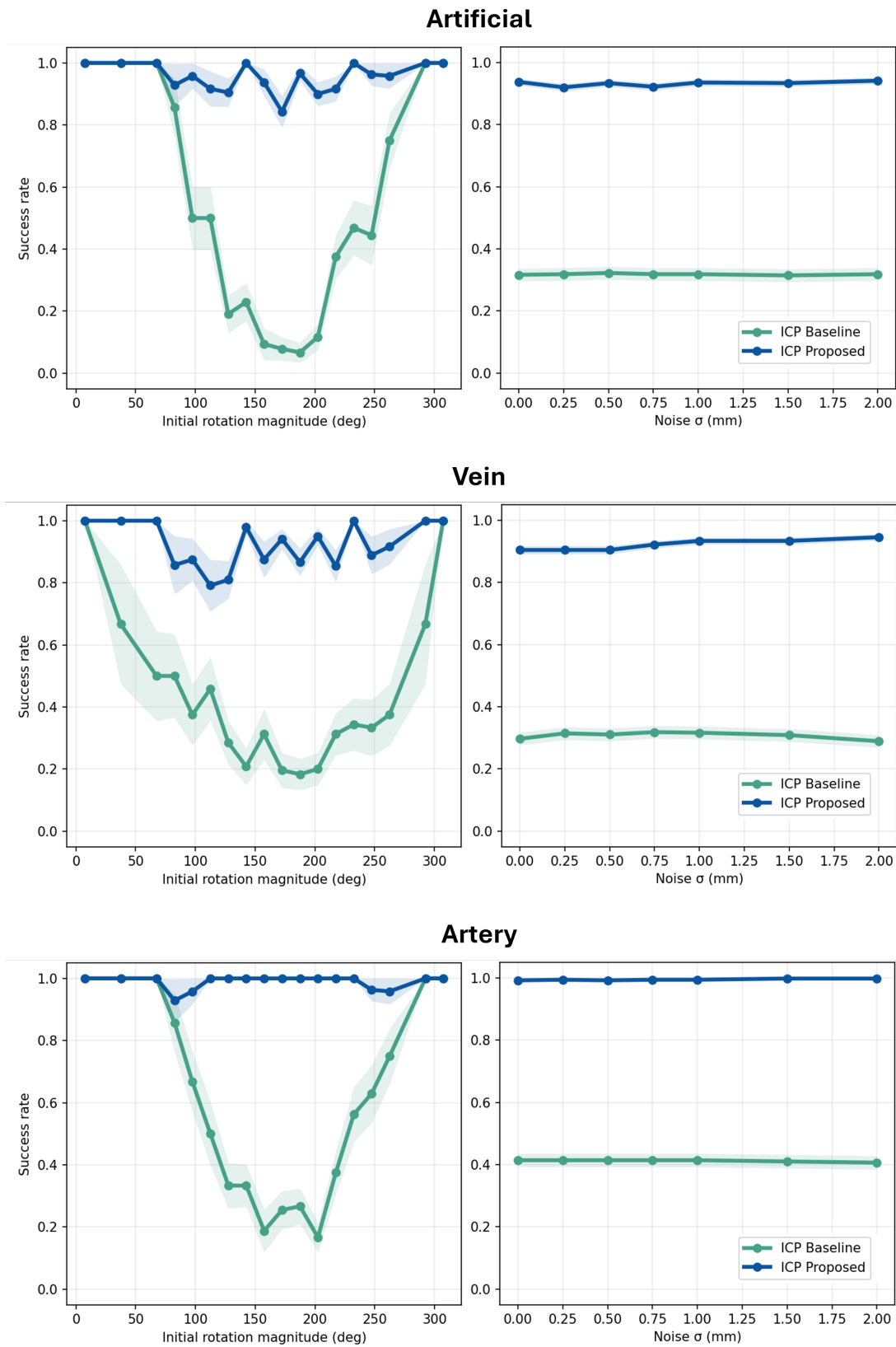


Figure 4.3: Success rate of rigid registration for each phantom geometry. Left column: success rate as a function of initial rotation magnitude. Right column: success rate as a function of gaussian noise level.

4.1.3. Registration accuracy

Figure 4.4 reports the registration accuracy obtained after rigid alignment and after deformable registration for the three phantom geometries. Results are shown as vertical bar plots for RMSE and HD95, allowing a direct comparison between the two stages of the pipeline. For all phantoms, deformable registration consistently improves alignment accuracy with respect to rigid registration. Both RMSE and HD95 are reduced in every case, with RMSE median values of 0.24, 1.15, and 0.54 mm for the Artificial, Vein, and Artery phantom, respectively. This suggests a systematic refinement rather than isolated corrections. The magnitude of the improvement is particularly pronounced for the Artificial and Artery geometries. For the Artificial phantom, RMSE and HD95 are reduced by approximately 80% and 84%, respectively. Similarly, for the Artery phantom, the reductions amount to approximately 62% for RMSE and 69% for HD95. For the venous geometry, the gain is more moderate (about 20%) but remains consistent across both metrics.

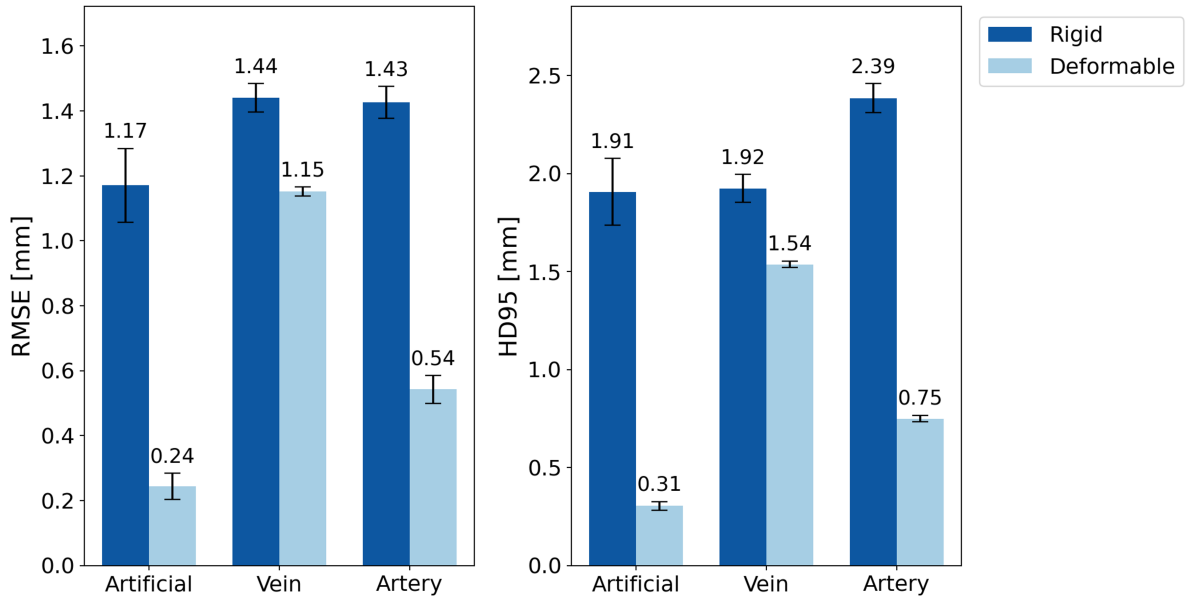


Figure 4.4: Registration accuracy after rigid and deformable alignment for the three phantom geometries. RMSE and HD95 are computed from the intraoperative reconstruction to the CAD reference model (moving→fixed). Deformable registration consistently reduces both average and tail errors.

The quantitative reduction in both RMSE and HD95 is consistent with the improved surface overlap observed in the qualitative comparison exemplified in Figure 4.5 for the arterial phantom. After rigid alignment, visible residual mismatches remain between

the intraoperative reconstruction and the CAD model. Following deformable refinement, surface correspondence is visibly improved and residual gaps are reduced.

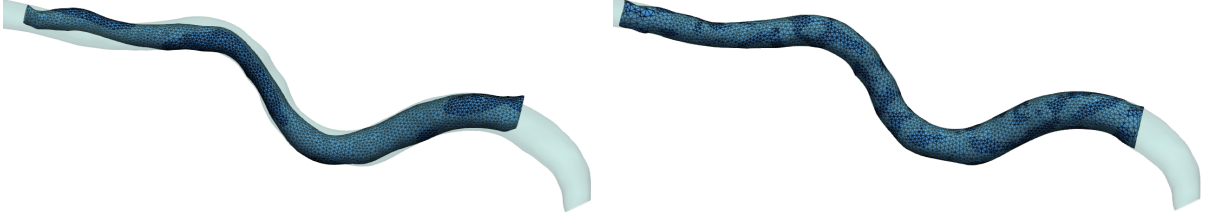


Figure 4.5: Qualitative comparison of surface alignment after rigid (left) and deformable (right) registration for the Artery phantom. The CAD phantom model (fixed geometry) is shown in light blue transparency, while the intraoperative reconstruction (moving geometry) is displayed in blue with visible mesh edges.

Although deformable registration improves alignment in all cases, the venous phantom remains the most challenging geometry after non-rigid refinement. The residual error appears systematic rather than driven by isolated failures, as dispersion across repetitions is comparable to the other phantoms. In the absence of an independent ground truth, this residual discrepancy cannot be attributed exclusively to the registration algorithm. It may reflect persistent geometric differences between the reconstructed surface and the CAD model, potentially related to phantom manufacturing tolerances, minor material deformations, or acquisition and segmentation biases.

Figure 4.6 illustrates the spatial distribution of the estimated deformation for the Artery phantom. Larger displacement magnitudes are predominantly observed near the extremities of the reconstructed surface. This pattern is consistently observed across multiple acquisitions. Several factors may contribute to this behaviour. First, the phantom is connected to external adapters at its extremities, which may introduce local geometric irregularities or reduced structural rigidity. Second, preprocessing and filtering operations are inherently less stable near boundary regions, potentially amplifying residual discrepancies. Third, since the intraoperative reconstruction covers only a partial portion of the CAD model, the CPD-based deformable registration may compensate for incomplete longitudinal coverage, resulting in slightly increased deformation near the ends. Although regularization constraints are applied to limit excessive warping, a residual boundary effect may still be present. However, no spatially incoherent or unrealistic deformation patterns are observed, and the estimated displacement field remains smooth and anatomically plausible.



Figure 4.6: Estimated displacement field introduced by the non-rigid registration exemplified for the Artery phantom. Points of the intraoperative reconstruction are colored according to displacement magnitude (in mm), illustrating the spatial distribution of the applied correction.

Geometric registration approaches reported in the literature typically achieve residual errors of the order of a few millimetres. Mean target registration errors of approximately 4–5 mm after rigid alignment, reduced to about 1.7–2.1 mm after deformable refinement, have been reported in surgical image-guidance settings [56]. Similarly, locally rigid, vessel-based registration methods have shown residual errors of approximately 0.7 mm in phantom experiments and mean values of about 3–4 mm in vivo [69].

In this context, the proposed deformable registration achieves millimetric-level accuracy, with RMSE and HD95 values consistently close to 1 mm or even smaller, depending on the phantom geometry. These results indicate effective compensation of residual geometric discrepancies that cannot be resolved by rigid alignment alone.

4.1.4. Computational performance

Figure 4.7 shows the computational cost of the full pipeline, decomposed into its main processing stages. Reported values correspond to median execution times across acquisitions for each phantom geometry. Total execution time increases from approximately 43 s for the artificial phantom to 70 s for the arterial case. This variation is primarily related to differences in spatial extent rather than geometric complexity. Larger phantoms require longer ultrasound sweeps and generate denser initial point clouds, directly increasing acquisition and preprocessing time.

Ultrasound acquisition represents the dominant component of the total runtime across all geometries. After voxel-based downsampling, point cloud density becomes more uniform,

resulting in relatively limited variations in the subsequent processing stages. Compared to acquisition and preprocessing, rigid and deformable registration introduce just a small overhead (about 6–7% of the total time), confirming that registration itself does not represent the computational bottleneck of the pipeline.

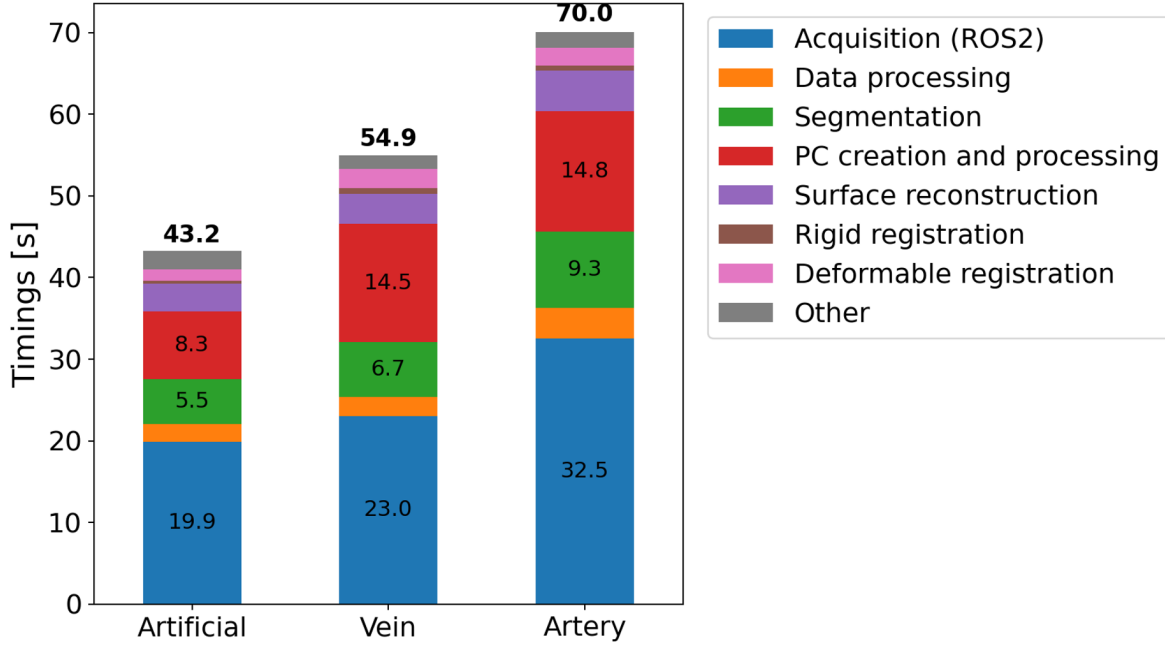


Figure 4.7: Breakdown of median computational time across the main stages of the registration pipeline for the three phantom geometries. Ultrasound acquisition represents the dominant contribution, while rigid and deformable registration account for a limited fraction of the total runtime.

Reported runtimes for geometric multimodal registration pipelines range from a few seconds for rigid alignment to several minutes when including deformable refinement [56]. In comparison, the full acquisition-to-registration workflow of the proposed framework is completed within approximately 40–70 s.

It is important to note that the full processing time occurs during the preparatory phase, while clinical staff set up the operative environment. Therefore, total runtimes on the order of one minute remain compatible with typical intraoperative workflows and do not require real-time execution during the interventional procedure. Overall, these results indicate that the proposed framework achieves accurate registration with a computational cost that remains stable across geometries of different size, with acquisition time being the primary determinant of total runtime.

4.2. Segmentation results

4.2.1. Model comparison across datasets

This section reports the results of the model comparison experiment described in Section 3.4.1. For each dataset, models were trained independently and evaluated on a held-out test set of images. Mean DSC and HD95 values obtained by each architecture are summarized in Tables 4.2 and 4.3.

Table 4.2: Mean DSC obtained by each architecture across datasets.

Dataset	U-Net	Attention U-Net	U-Net3+
Phantom	94.2%	94.0%	86.7%
Carotid	94.7%	94.6%	93.7%
San Raffaele	73.5%	80.6%	61.4%
Mus-V Artery	90.7%	91.2%	89.8%
Mus-V Vein	72.1%	69.0%	58.4%

Table 4.3: Mean HD95 obtained by each architecture across datasets. HD95 values are reported in pixel units.

Dataset	U-Net	Attention U-Net	U-Net3+
Phantom	4.57	4.84	17.19
Carotid	2.42	2.23	2.99
San Raffaele	18.89	10.12	22.40
Mus-V Artery	3.44	3.12	3.62
Mus-V Vein	12.28	13.94	20.29

A clear separation in performance levels can be observed across experimental conditions. All models achieved high overlap ($\text{DSC} \geq 93.7\%$) and low boundary errors ($\text{HD95} \leq 5$ pixels) on the Carotid set, and similarly high performance was observed on the Phantom set for U-Net and Attention U-Net ($\text{DSC} \approx 94\%$ with $\text{HD95} \leq 5$ pixels), whereas U-Net3+ showed a clear degradation ($\text{DSC} = 86.7\%$ and $\text{HD95} = 17.19$ pixels). On the remaining datasets, performance was worse and depended on the model architecture. On the Mus-V Artery subset, results were only slightly inferior to those obtained on Phantom and Carotid, with DSC values around 89.8–91.2% and HD95 between 3 and 4 pixels; among

the evaluated architectures, the Attention U-Net achieved the best overall performance (DSC = 91.2%, HD95 = 3.12 pixels). Performance degraded more substantially on the San Raffaele dataset, where DSC remained below 81% (61.4–80.6%) and HD95 was always above 10 pixels (10.12–22.40); in this case, the Attention U-Net provided the best results (DSC = 80.6%, HD95 = 10.12 pixels). The most challenging scenario was observed for the Mus-V Vein dataset, where DSC decreased further (58.4–72.1%) and HD95 increased beyond 12 pixels (12.28–20.29). In this setting, U-Net achieved the best performance (DSC = 72.1%, HD95 = 12.28 pixels), indicating that the optimal architecture varies across datasets and vessel types.

From an architectural perspective, the results indicate that performance rankings vary depending on the specific configuration. Attention U-Net achieves the best results in the Mus-V Artery subset (DSC = 91.2%, HD95 = 3.12 pixels) and also provides the highest accuracy in the San Raffaele dataset (DSC = 80.6%, HD95 = 10.12 pixels). In both cases, it outperforms the other architectures in terms of both overlap and boundary error. U-Net achieves the best performance in the Mus-V Vein configuration (DSC = 72.1%, HD95 = 12.28 pixels), exceeding both Attention U-Net and U-Net3+ in this setting. On Phantom and Carotid, U-Net and Attention U-Net show comparable behaviour, with no substantial difference in performance. In contrast, U-Net3+ does not exhibit systematic advantages across tasks. Although architecturally more complex, it shows reduced robustness in several configurations (e.g., marked boundary error increase on Phantom and performance degradation in heterogeneous in-vivo datasets), suggesting that increased architectural capacity does not necessarily translate into improved segmentation accuracy when training data are limited or when acquisition variability is high.

Beyond anatomical differences, these trends may also reflect intrinsic characteristics of B-mode ultrasound imaging. In ultrasound, vessel boundaries are not defined by sharp and stable intensity discontinuities but rather by gradual echogenic transitions embedded in speckle-dominated backgrounds. As a consequence, boundary localization can be inherently sensitive to acquisition-dependent factors such as probe orientation, local contrast and surrounding tissue appearance. This intrinsic ambiguity may contribute to the increased dispersion observed in HD95 under heterogeneous in-vivo conditions, even when DSC values remain moderately high. While no direct causal relationship can be established, this interpretation is consistent with the variability observed across experimental settings from an imaging perspective.

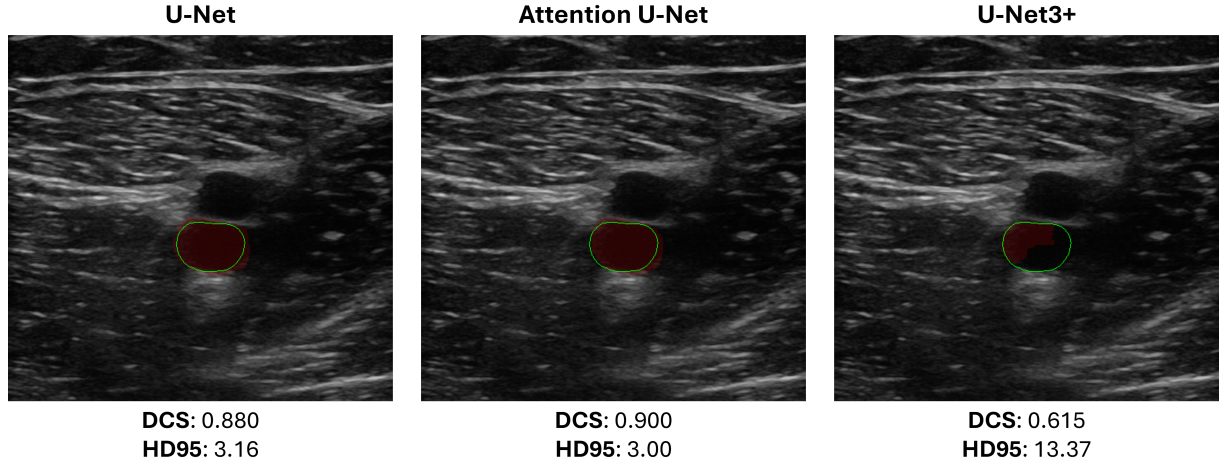


Figure 4.8: Qualitative comparison of the three architectures on a *San Raffaele* frame. Note the partial collapse of the segmentation in U-Net3+.

Reported Dice scores in literature for femoral vessel segmentation in B-mode ultrasound typically range between 0.80 and 0.92 for arteries [36, 60], with slightly lower values for veins [15]. On phantom-based acquisitions, segmentation performance can also reach Dice values of about 0.9 [3]. The results obtained in this work fall within these ranges. However, the comparison across datasets highlights that maintaining segmentation robustness under increasing anatomical and acquisition heterogeneity remains a primary challenge in ultrasound-based vessel analysis.

4.2.2. Influence of dataset characteristics

This section reports the results of the experiment described in Section 3.4.1. The analysis focuses on the three configurations of the *San Raffaele* dataset: *Single-session*, *Full* and *Pruned*. Figure 4.9 shows the DSC comparison across configurations, while Figure 4.10 reports the corresponding HD95 values.

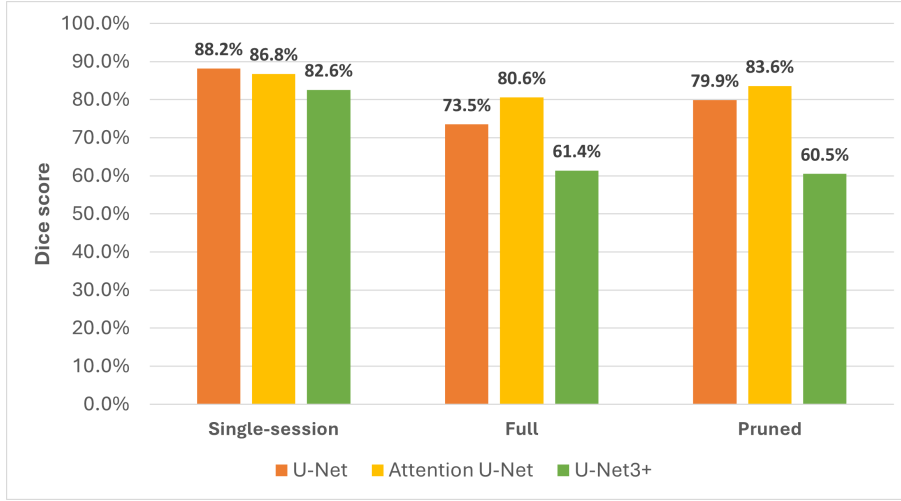


Figure 4.9: Mean DSC comparison across *Single-session*, *Full* and *Pruned* dataset configurations.

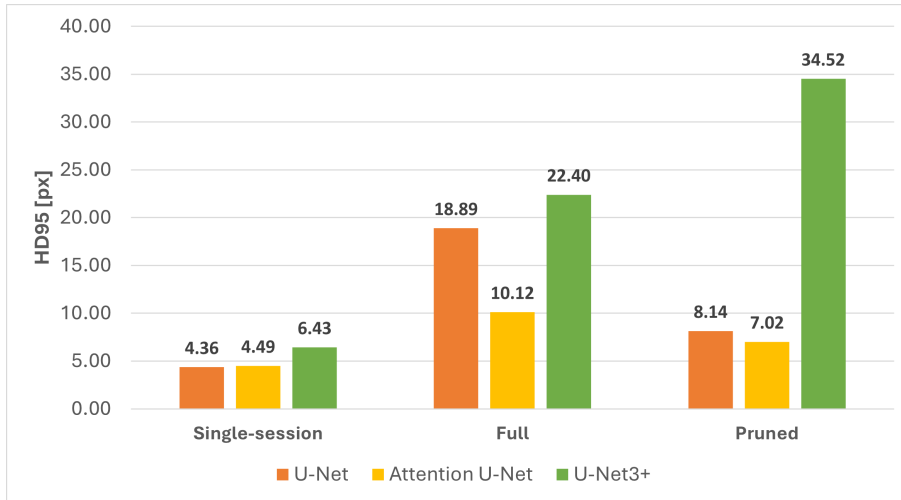


Figure 4.10: Mean HD95 comparison across the *Single-session*, *Full* and *Pruned* configurations of the San Raffaele dataset. Values are expressed in pixel units.

Segmentation performance is highest in the *Single-session* configuration for all architectures. When the dataset is extended to the *Full* configuration, which includes increased variability across acquisitions and patients, DSC decreases and HD95 increases consistently across models. This confirms that introducing additional heterogeneity makes the segmentation task more challenging, even when the total number of training images increases.

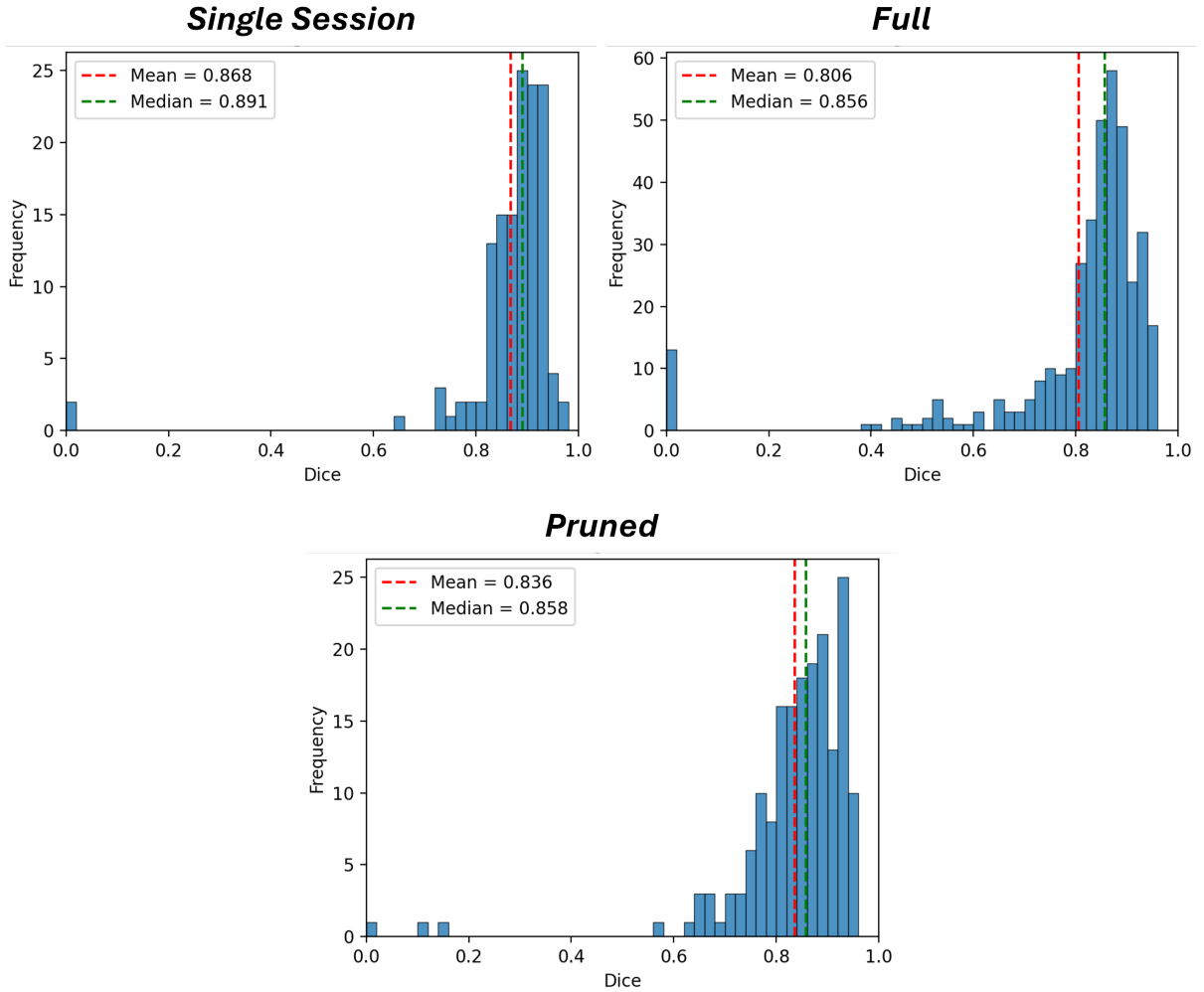


Figure 4.11: Distribution of Dice values for the *Single-session*, *Full* and *Pruned* configurations of the San Raffaele dataset, using the Attention U-Net.

To better understand the nature of this degradation, the DSC histogram distributions are shown in Figure 4.11, for the Attention U-Net. A key observation is that the drop in performance in the *Full* configuration is not uniform across all frames. While the median DSC remains relatively high, the mean DSC decreases substantially due to the presence of a significant number of low-DSC outliers, including near-zero cases. A similar pattern can be observed for HD95: the median remains within a moderate range, while the mean increases due to a long tail of large boundary errors. This behaviour indicates that the performance reduction in the *Full* configuration is driven by a subset of particularly challenging samples rather than by a global shift affecting all images equally. An inspection of the failing cases confirms that low-DSC and high-HD95 outliers originate from a limited number of acquisitions. Therefore, the degradation is primarily acquisition-level rather than frame-level: when an acquisition is challenging, for most of its frame segmentation

is not sufficiently reliable. This is coherent with the strong correlation among frames within the same acquisition, which share the same probe positioning, anatomical context and imaging conditions. As a consequence, the heavy-tailed distributions observed in Figure 4.11 reflect clustered acquisition-specific failures rather than uniformly degraded segmentation performance across the entire dataset.

Comparing the *Full* and *Pruned* configurations allows for isolating of the effect of dataset size while preserving acquisition-level variability. U-Net and Attention U-Net show a slight improvement after pruning, suggesting that removing highly redundant frames does not reduce the effective information content of the dataset and may limit the over-representation of highly similar samples. The distributions further indicate that pruning primarily reduces the weight of extreme outliers, with limited impact on the median performance, which remains comparable to the *Full* configuration. In contrast, U-Net 3+ remains more sensitive to reduced training data and does not benefit from pruning to the same extent.

Overall, these results indicate that segmentation performance is more strongly affected by acquisition-level variability than by moderate reductions in dataset size, provided that inter-acquisition diversity is preserved.

4.2.3. Vessel type and domain aggregation analysis

This section reports the results of the experiment described in Section 3.4.1. The analysis is structured along two complementary aspects: (i) the intrinsic difference between arterial and venous segmentation under controlled conditions and (ii) the effect of cross-district aggregation within each vessel type.

Vessel type comparison under controlled conditions

Figure 4.12 reports the DSC obtained on single-district configurations as well as in the aggregated setting, separated by vessel type. Within the same anatomical region and acquisition setting, arterial segmentation consistently achieves slightly higher performance than venous segmentation across architectures. Although the absolute DSC gap remains moderate, the trend is stable across all tested models, indicating a systematic performance difference rather than architecture-specific behaviour.



Figure 4.12: DSC comparison between arterial and venous segmentation across architectures.

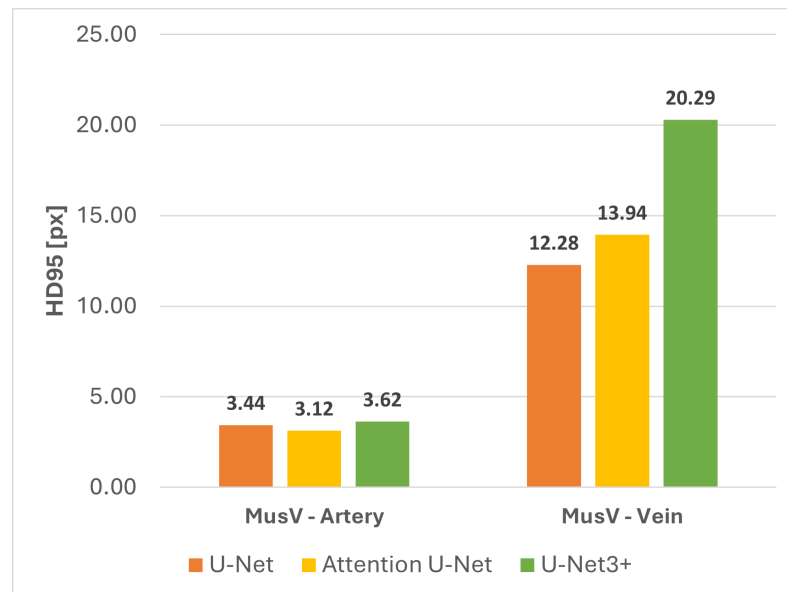


Figure 4.13: HD95 comparison between arterial and venous segmentation. Values are reported in pixel units.

To complement the overlap-based analysis, boundary accuracy was further evaluated using HD95, as shown in Figure 4.13. The trends observed for HD95 are consistent with those obtained for DSC, while performance differences between vessel types appear more pronounced. This suggests that the discrepancy between arterial and venous segmentation primarily manifests at the boundary level rather than in global overlap alone. Since HD95 is more sensitive to localized contour deviations, the larger separation observed for veins indicates increased boundary variability under otherwise controlled conditions.

To provide a structural interpretation of this behaviour, a geometric analysis of the manual annotations was performed. For each acquisition, circularity and foreground prevalence were computed and are reported in Figure 4.14.

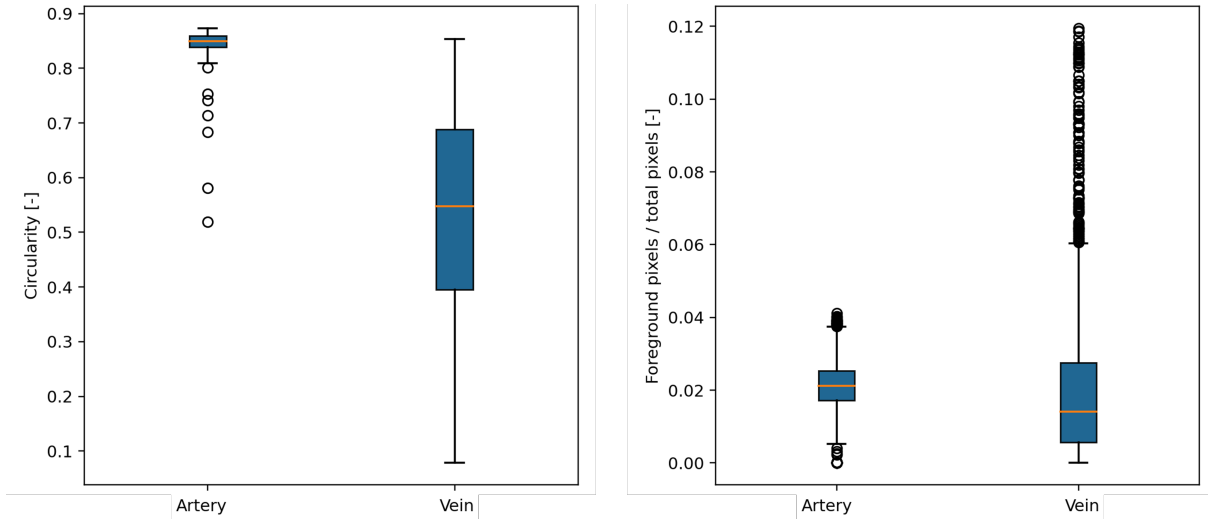


Figure 4.14: Distribution of circularity and foreground prevalence across MusV configurations. In the box plots, the orange horizontal line indicates the median value; blue boxes indicate the interquartile range; whiskers indicate the 5th and 95th percentile. Outliers are indicated by empty circles.

Arteries exhibit higher circularity values with lower dispersion, indicating a more regular and geometrically stable morphology across frames. In contrast, venous structures show substantially larger variability in both circularity and foreground prevalence. This broader geometric dispersion is consistent with the increased variability observed in HD95 for venous segmentation. Greater shape irregularity and more variable foreground prevalence may contribute to increased boundary instability, as irregular contours and fluctuating class proportions can make precise delineation more sensitive to local prediction errors [39, 51, 63, 70]. While this does not establish a direct causal relationship, the geometric analysis provides a coherent explanation for the systematic performance gap observed

between vessel types under controlled acquisition settings.

Effect of domain aggregation within each vessel type

The impact of aggregating femoral and neck data is also shown in Figures 4.12 and 4.13. For arteries, the aggregated configuration achieves performance comparable to or slightly higher than single-district training. This suggests that arterial morphology remains sufficiently consistent across anatomical regions for the model to benefit from the increased number of training samples without introducing substantial additional variability. For veins, the opposite trend is observed. The aggregated venous configuration shows a clear performance degradation compared to single-district training, and this effect is more evident in HD95 than in DSC. The stronger separation in HD95 indicates that domain aggregation primarily affects boundary precision rather than overall overlap. In parallel, the geometric descriptors in Figure 4.14 show a wider dispersion of circularity and foreground prevalence across venous configurations, suggesting increased intra-class variability when combining districts. Such variability may challenge the model’s ability to maintain stable boundary delineation across domains [39, 70], even though the total number of training samples increases.

Overall, these results indicate that vessel type and domain aggregation interact in a non-uniform manner. Arterial segmentation appears relatively robust to cross-district aggregation, whereas venous segmentation shows greater sensitivity to anatomical and morphological heterogeneity. The combined metric and geometric analyses consistently support this interpretation.

5 | Conclusions

5.1. Summary of the work.

This thesis presented an end-to-end framework for CT-US registration aimed at supporting model-based intraoperative visualization and guidance in catheter-based interventions. The proposed pipeline integrates automatic vessel segmentation from 2D ultrasound images, 3D geometric reconstruction of the vascular surface, a robust rigid alignment strategy and a final non-rigid refinement step to compensate for residual local discrepancies between preoperative and intraoperative geometries.

Since the pipeline could not be tested in vivo, ad hoc in vitro experiments were run. These allowed for integrating the different computational tools and testing their communication. Also, in vitro tests demonstrated that the pipeline is capable of achieving accurate surface alignment under controlled conditions. The proposed multi-start rigid registration strategy improved robustness to initialization compared to a conventional single-start baseline, particularly in the presence of rotational ambiguities and partial geometric overlap. Computational analysis showed that the overall runtime remains compatible with intraoperative workflows, with data acquisition and preprocessing representing the most time-consuming stages, while the registration steps introduce limited additional overhead. However, in vitro tests did not allow to understand if the U-Net models selected in this work could be successfully applied to the segmentation of real-world B-mode US images. Thus, the automated segmentation component of the pipeline was tested also on images acquired in vivo on arteries and on veins. These further tests showed stable performance across different network architectures, with simpler models achieving competitive accuracy in data-limited settings. The experimental analysis also highlighted the impact of dataset variability and pruning strategies on generalization performance.

Overall, the presented framework demonstrates the feasibility of integrating segmentation, reconstruction and registration into a coherent pipeline suitable for intraoperative use, while identifying specific limitations and directions for future improvement.

5.2. Limitations

Despite the promising results obtained in this work, several limitations should be acknowledged.

Data availability. The *in-vivo* segmentation experiments rely in part on a dataset acquired at a single clinical center, whose size is limited as compared to large-scale medical imaging datasets. While this dataset is representative of realistic clinical conditions, its limited size may affect the robustness and generalization capability of the trained models. Larger and more diverse datasets would be required to draw stronger conclusions on segmentation performance across different patient populations and acquisition settings.

Experimental setup constraints. The experimental validation of the registration pipeline was limited to *in-vitro* conditions. This constraint is primarily due to hardware availability: the laboratory setup provided robotic tracking but only a basic ultrasound system, whereas clinical acquisitions were performed with a high-end scanner without access to probe pose information. As a result, end-to-end *in-vivo* validation of the registration pipeline was not possible within the scope of this work.

Simplified anatomical and physiological conditions. *In-vitro* experiments were conducted on static phantoms, which do not reproduce physiological effects such as arterial pulsatility or patient motion. While this simplification allows for controlled evaluation of the pipeline, it does not capture dynamic effects that may influence reconstruction and registration accuracy in real clinical scenarios.

Limited anatomical extent of registration. The proposed registration strategy focuses on a single vascular segment, such as the femoral vein or artery. While this choice is appropriate for an initial validation, it does not account for global anatomical variations or posture-induced deformations that may occur along the full vascular path during an intervention. As a consequence, the impact of distal registration errors on more proximal (and critical) anatomical regions was not investigated.

5.3. Future developments

Several directions can be followed to extend and improve the proposed framework, both on the registration and segmentation components.

***In-vivo* registration validation.** A primary future development consists in validating the proposed registration pipeline directly *in vivo*. As an intermediate step, experiments on healthy volunteers could be considered, combining pre-operative MRI (to avoid unnecessary ionizing radiation) with laboratory-controlled US acquisitions. In this setting, the robotic platform could be used to acquire controlled B-mode sequences, enabling the evaluation of registration performance under physiological conditions not reproducible in static phantoms, such as pulsatility or limb motion.

Dynamic effects and pulsatility. For arterial structures, vessel pulsatility and motion may introduce geometric inconsistencies along ultrasound sweeps. Future developments could incorporate temporal modeling, ECG-based synchronization strategies or correction methods to compensate for these effects and improve reconstruction consistency.

Registration of extended vascular models. Future work should address the registration of longer vascular segments, including proximal anatomical landmarks such as the inferior vena cava. The availability of multiple anchoring regions would improve robustness to global posture changes and mitigate error amplification when mapping distal alignments to proximal anatomical targets.

Estimation of patient-specific mechanical properties. Although not addressed in this thesis, the estimation of vascular and surrounding tissue mechanical properties is essential for realistic real-time biomechanical simulations aimed to computing the interaction between the catheter and the vessel. Future research could investigate non-invasive estimation strategies, such as probe-based compression tests or approaches correlating vessel deformation with known pressure conditions, to improve the reliability of real-time FEM-based models.

Extensions of the segmentation framework. Several extensions can be explored for the segmentation component. These include the evaluation of alternative loss functions and dedicated ablation studies to better understand the impact of training choices on model performance. In addition, multi-class segmentation strategies jointly segmenting arteries and veins could be investigated, potentially leveraging anatomical context and spatial relationships to improve robustness and consistency of vessel predictions.

Bibliography

- [1] N. Abraham and N. M. Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. 2018. doi: 10.48550/arXiv.1810.07842. URL <https://arxiv.org/abs/1810.07842>.
- [2] J. Ahrens, B. Geveci, and C. Law. Paraview: An end-user tool for large data visualization. In C. D. Hansen and C. R. Johnson, editors, *Visualization Handbook*. Elsevier, Amsterdam, NL, 2005. ISBN 9780123875822.
- [3] L. Al-Zogbi et al. Robotic ultrasound-guided femoral artery reconstruction of anatomically-representative phantoms. 2025.
- [4] E. J. Armstrong, R. Low, and J. Rogers. *EVEREST II Functional Mitral Regurgitation*, pages 357–375. Springer London, London, 2013. ISBN 978-1-4471-4294-2. doi: 10.1007/978-1-4471-4294-2_22. URL https://doi.org/10.1007/978-1-4471-4294-2_22.
- [5] Autodesk, Inc. Autodesk Meshmixer, 2020. URL <http://www.meshmixer.com>. 3D mesh editing and processing software. RRID: SCR_015736.
- [6] U. Ayachit. *The ParaView Guide: A Parallel Visualization Application*. Kitware Inc., 2015. ISBN 9781930934306.
- [7] M. F. Azampour, M. Tirindelli, J. Lameski, M. Gafencu, E. Tagliabue, E. Fatemizadeh, I. Hacıhaliloglu, and N. Navab. Anatomy-aware computed tomography-to-ultrasound spine registration. *Medical Physics*, 51(3):2044–2056, 2024. doi: <https://doi.org/10.1002/mp.16731>. URL <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16731>.
- [8] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019. doi: 10.1109/TMI.2019.2897538. URL <https://pubmed.ncbi.nlm.nih.gov/30716034/>.
- [9] S. Baldus, T. Doenst, R. Pfister, J. Gummert, M. Kessler, P. Boekstegers, E. Lubos,

- J. Schröder, H. Thiele, T. Walther, M. Kelm, J. Hausleiter, I. Eitel, U. Fischer-Rasokat, A. Bufe, A. Schmeisser, H. Ince, P. Lurz, R. S. von Bardeleben, C. Hagl, T. Noack, S. Reith, H. Beucher, H. Reichenspurner, W. Rottbauer, P. C. Schulze, W. Müller, J. Frank, M. Hellmich, T. Wahlers, and V. Rudolph. Transcatheter repair versus mitral-valve surgery for secondary mitral regurgitation. *New England Journal of Medicine*, 391(19):1787–1798, 2024. doi: 10.1056/NEJMoa2408739. URL <https://www.nejm.org/doi/full/10.1056/NEJMoa2408739>.
- [10] G. Bartal, E. Vano, G. Paulo, and D. Miller. Management of patient and staff radiation dose in interventional radiology: Current concepts. *Cardiovascular and interventional radiology*, 37, 07 2013. doi: 10.1007/s00270-013-0685-0.
- [11] M. Berger, A. Tagliasacchi, L. M. Seversky, P. Alliez, J. A. Levine, A. Sharf, and C. T. Silva. A survey of surface reconstruction from point clouds. *Computer Graphics Forum*, 36(1):301–329, 2017. doi: 10.1111/cgf.12802.
- [12] P. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. doi: 10.1109/34.121791.
- [13] L. C. Cabrelli et al. Oil-based gel phantom for ultrasound and optical imaging. In *Proceedings of SPIE*, 2015. doi: 10.1117/12.2187916.
- [14] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert. Deep learning for cardiac image segmentation: A review. *Frontiers in Cardiovascular Medicine*, 7:25, 2020. ISSN 2297-055X. doi: 10.3389/fcvm.2020.00025. URL <https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2020.00025/full>.
- [15] P.-W. Chen et al. Deep learning model for diagnosis of venous thrombosis from lower extremity peripheral ultrasound imaging. *iScience*, 2024. doi: 10.1016/j.isci.2024.111318.
- [16] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992. doi: 10.1016/0262-8856(92)90066-C.
- [17] D. Chetverikov, D. Stepanov, and P. Krsek. Robust euclidean alignment of 3d point sets: The trimmed iterative closest point algorithm. *Image and Vision Computing*, 23(3):299–309, 2005. doi: 10.1016/j.imavis.2004.05.009.
- [18] J. Coles-Black, D. Bolton, and J. Chuen. Accessing 3d printed vascular phantoms for

- procedural simulation. *Frontiers in Surgery*, 7:626212, 2021. doi: 10.3389/fsurg.2020.626212. URL <https://www.frontiersin.org/articles/10.3389/fsurg.2020.626212>.
- [19] M. P. Deshmukh and U. Bhosle. A survey of image registration. *International Journal of Image Processing (IJIP)*, 5(3):245–269, 2011. URL <https://www.cscjournals.org/manuscript/Journals/IJIP/Volume5/Issue3/IJIP-364.pdf>.
- [20] J. Digne and C. de Franchis. The bilateral filter for point clouds. *Image Processing On Line*, 7:278–287, 2017. doi: 10.5201/ipol.2017.179.
- [21] C. Dodesini. Verso la validazione di simulazioni real-time dell’interazione vasocateretere in procedure di cardiologia interventistica: caratterizzazione delle proprietà meccaniche di materiali per fantocci in vitro e analisi del loro effetto sui risultati numerici. Master’s thesis, Politecnico di Milano, April 2024.
- [22] B. Dufumier, P. Gori, I. Battaglia, J. Victor, A. Grigis, and E. Duchesnay. Benchmarking cnn on 3d anatomical brain mri: Architectures, data augmentation and deep ensemble learning. 06 2021. doi: 10.48550/arXiv.2106.01132.
- [23] S. Estrada, C. Duran, D. Schulz, J. Bismuth, M. Byrne, and M. O’Malley. Smoothness of surgical tool tip motion correlates to skill in endovascular tasks. *IEEE Transactions on Human-Machine Systems*, 46:1–13, 10 2016. doi: 10.1109/THMS.2016.2545247.
- [24] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. M. Fennessy, M. Sonka, J. V. Miller, S. Pieper, and R. Kikinis. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341, 2012. doi: 10.1016/j.mri.2012.05.001.
- [25] K. Funamoto, O. Yamashita, and T. Hayase. Poly(vinyl alcohol) gel ultrasound phantom with durability and visibility of internal flow. *Journal of Medical Ultrasonics (2001)*, 42(1):17–23, 2014. doi: 10.1007/s10396-014-0560-x.
- [26] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. 48:1050–1059, 2016. URL <https://proceedings.mlr.press/v48/gal16.html>.
- [27] V. García-Vázquez, F. von Haxthausen, S. Jäckle, C. Schumann, I. Kuhlemann, J. Bouchagiar, A.-C. Höfer, F. Matysiak, G. Hüttmann, J. P. Goltz, M. Kleemann, F. Ernst, and M. Horn. Navigation and visualisation with hololens in endovascular

- aortic repair. *Innovative Surgical Sciences*, 3(3):167–177, 2018. doi: doi:10.1515/iss-2018-2001. URL <https://doi.org/10.1515/iss-2018-2001>.
- [28] Y. Geng, G. Meng, M. Chen, G. Cao, M. Zhao, J. Zhao, and H. Liu. Force sensing guided artery-vein segmentation via sequential ultrasound images. In M. G. Linguraru, Q. Dou, A. Feragen, S. Giannarou, B. Glocker, K. Lekadir, and J. A. Schnabel, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024: 27th International Conference, Marrakesh, Morocco, October 6–10, 2024, Proceedings, Part IV*, volume 15004 of *Lecture Notes in Computer Science*, pages 656–666. Springer, 2024. doi: 10.1007/978-3-031-72083-3_61. URL https://doi.org/10.1007/978-3-031-72083-3_61.
- [29] D. Grzech, M. F. Azampour, B. Glocker, J. Schnabel, N. Navab, B. Kainz, and L. L. Folgoc. A variational bayesian method for similarity learning in non-rigid image registration. pages 119–128, 2022. doi: 10.1109/CVPR52688.2022.00022.
- [30] X. T. Ha, D. Wu, M. Ourak, G. Borghesan, A. Menciassi, and E. V. Poorten. Sensor fusion for shape reconstruction using electromagnetic tracking sensors and multi-core optical fiber. *IEEE Robotics and Automation Letters*, 8(7):4076–4083, 2023. doi: 10.1109/LRA.2023.3280456.
- [31] O. Hirose. A bayesian formulation of coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11):2269–2281, 2017. doi: 10.1109/TPAMI.2016.2625800.
- [32] H.-z. Hu, X.-b. Feng, Z.-w. Shao, M. Xie, S. Xu, X.-h. Wu, and Z.-w. Ye. Application and prospect of mixed reality technology in medical field. *Current medical science*, 39:1–6, 2019.
- [33] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu. Unet 3+: A full-scale connected unet for medical image segmentation. *arXiv preprint arXiv:2004.08790*, 2020. doi: 10.48550/arXiv.2004.08790.
- [34] Istituto Superiore di Sanità (ISS) - Progetto Cuore. CuoreData – Dati su rischio cardiovascolare. <https://www.cuore.iss.it/indagini/CuoreData>, 2025. Accessed: 2025-07-10.
- [35] R. Jafary, S. Armstrong, T. Byrne, A. Stephens, V. Pellegrino, and S. D. Gregory. Fabrication and characterization of tissue-mimicking phantoms for ultrasound-guided cannulation training. *ASAIO Journal*, 68(7):940–948, 2022. doi: 10.1097/MAT.0000000000001593. URL <https://doi.org/10.1097/MAT.0000000000001593>.

- [36] B. Jiang et al. Automatic vessel segmentation with deep spatiotemporal context learning. 2021.
- [37] H. J. Johnson, M. M. McCormick, L. Ibáñez, and the Insight Software Consortium. *The ITK Software Guide: Design and Functionality (Books 1 & 2)*. Kitware, Inc., fourth edition, 2024. URL <https://itk.org/ItkSoftwareGuide.pdf>. Accessed: 25 Oct 2025.
- [38] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. 2006. doi: 10.2312/SGP/SGP06/061-070.
- [39] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed. Boundary loss for highly unbalanced segmentation. In *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning (MIDL 2019)*, volume 102 of *Proceedings of Machine Learning Research*, pages 285–296. PMLR, 2019. URL <https://proceedings.mlr.press/v102/kervadec19a.html>.
- [40] R. Kikinis, S. D. Pieper, and K. Vosburgh. 3d slicer: A platform for subject-specific image analysis, visualization, and clinical support. In F. A. Jolesz, editor, *Intra-operative Imaging and Image-Guided Therapy*, volume 3(19). Springer, 2014. ISBN 9781461476566.
- [41] L. W. Klein, D. L. Miller, S. Balter, W. Laskey, D. Haines, A. Norbash, M. A. Mauro, and J. A. Goldstein. Occupational health hazards in the interventional laboratory: Time for a safer environment. *Radiology*, 250(2):538–544, 2009. doi: 10.1148/radiol.2502082558. URL <https://doi.org/10.1148/radiol.2502082558>. PMID: 19188321.
- [42] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, Jan. 2010. doi: 10.1109/TMI.2009.2035616.
- [43] M. E. Laughlin, S. E. Stephens, J. A. Hestekin, and M. O. Jensen. Development of custom wall-less cardiovascular flow phantoms with tissue-mimicking gel. *Cardiovascular Engineering and Technology*, 13(1):1–13, 2022. doi: 10.1007/s13239-021-00546-7. URL <https://doi.org/10.1007/s13239-021-00546-7>.
- [44] M. Lyu, J. Yang, Z. Qi, R. Xu, and J. Liu. Rigid pairwise 3d point cloud registration: A survey. *Pattern Recognition*, 151:110408, 2024. doi: 10.1016/j.patcog.2024.110408.
- [45] M. J. Mack, M. B. Leon, V. H. Thourani, R. Makkar, S. K. Kodali, M. Russo, S. R. Kapadia, S. C. Malaisrie, D. J. Cohen, P. Pibarot, J. Leipsic, R. T. Hahn,

- P. Blanke, M. R. Williams, J. M. McCabe, D. L. Brown, V. Babaliaros, S. Goldman, W. Y. Szeto, P. Genereux, A. Pershad, S. J. Pocock, M. C. Alu, J. G. Webb, and C. R. Smith. Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients. *New England Journal of Medicine*, 380(18):1695–1705, 2019. doi: 10.1056/NEJMoa1814052. URL <https://www.nejm.org/doi/full/10.1056/NEJMoa1814052>.
- [46] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016. doi: 10.48550/arXiv.1606.04797. URL <https://arxiv.org/abs/1606.04797>.
- [47] Ministero della Salute. *Relazione sullo Stato Sanitario del Paese 2017-2021*. Ministero della Salute, Direzione generale della digitalizzazione, del sistema informativo sanitario e della statistica, Roma, 2022. URL https://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?lingua=italiano&id=3080. Accessed: 2025-07-10.
- [48] A. Momot. Common carotid artery ultrasound images, 2022. URL <https://data.mendeley.com/datasets/d4xt63mgjm/1>.
- [49] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010. doi: 10.1109/TPAMI.2010.46.
- [50] A. Oglat, M. Alshipili, M. Sayah, and M. S. Ahmad. Artifacts in diagnostic ultrasonography. *Journal for Vascular Ultrasound*, 05 2020. doi: 10.1177/1544316720923937.
- [51] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O’Regan, B. Kainz, B. Glocker, and D. Rueckert. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, 2018. doi: 10.1109/TMI.2017.2743464.
- [52] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. doi: 10.48550/arXiv.1804.03999.
- [53] L. Perez and J. Wang. The effectiveness of data augmentation in image classification using deep learning. 2017. doi: 10.48550/arXiv.1712.04621. URL <https://arxiv.org/abs/1712.04621>.

- [54] O. Puonti, J. E. Iglesias, and K. Van Leemput. Fast and sequence-adaptive whole-brain segmentation using parametric bayesian modeling. *NeuroImage*, 143:235–249, 2016. doi: 10.1016/j.neuroimage.2016.09.011. URL <https://doi.org/10.1016/j.neuroimage.2016.09.011>.
- [55] M. M. Quien and M. Saric. Ultrasound imaging artifacts: How to recognize them and how to avoid them. *Echocardiography*, 35(9):1388–1401, 2018. doi: 10.1111/echo.14116. URL <https://onlinelibrary.wiley.com/doi/10.1111/echo.14116>.
- [56] S. Reaungamornrat et al. Deformable image registration for cone-beam ct guided transoral robotic base of tongue surgery. *Physics in Medicine and Biology*, 58(14):4953–4979, 2013.
- [57] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J. R. Weir-McCall, Z. Teng, D. Gopalan, D. Rueckert, and T. Schaeffter. Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. *Nature Machine Intelligence*, 3:199–217, 2021. doi: 10.1038/s42256-021-00307-0. URL <https://www.nature.com/articles/s42256-021-00307-0>.
- [58] G. Roffo. *Ranking to Learn and Learning to Rank: On the Role of Ranking in Pattern Recognition Applications*. PhD thesis, 05 2017.
- [59] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015. doi: 10.48550/arXiv.1505.04597.
- [60] A. G. Roy et al. Ultrasound-guided femoral vessel segmentation using deep learning. In *Proceedings of MICCAI Workshop / Medical Imaging Conference*, 2018.
- [61] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 1679 of *Lecture Notes in Computer Science*, pages 712–721. Springer, 1999. doi: 10.1007/10704282_78.
- [62] V. Ruozzi and P. Vassena. Integrazione di simulazioni real time e realtà aumentata per il monitoraggio dell’interazione cateterevaso in procedure di cardiologia interventistica. Master’s thesis, Politecnico di Milano, April 2022.
- [63] S. S. M. Salehi, D. Erdogmus, and A. Gholipour. Tversky loss function for image

- segmentation using 3d fully convolutional deep networks. In *Machine Learning in Medical Imaging (MLMI 2017)*, volume 10541 of *Lecture Notes in Computer Science*, pages 379–387. Springer, Cham, 2017. doi: 10.1007/978-3-319-67389-9_44.
- [64] R. Schmidt and K. Singh. meshmixer: an interface for rapid mesh composition. In *ACM SIGGRAPH 2010 Talks*, SIGGRAPH '10. ACM, 2010. doi: 10.1145/1837026.1837034.
- [65] A. Sciahbasi, E. Frigoli, A. Sarandrea, M. Rothenbühler, P. Calabrò, A. Lupi, F. Tomassini, B. Cortese, S. Rigattieri, E. Cerrato, D. Zavalloni, A. Zingarelli, P. Calabria, P. Rubartelli, G. Sardella, M. Tebaldi, S. Windecker, P. Jüni, D. Heg, and M. Valgimigli. Radiation exposure and vascular access in acute coronary syndromes: The rad-matrix trial. *Journal of the American College of Cardiology*, 69(20):2530–2537, 2017. ISSN 0735-1097. doi: <https://doi.org/10.1016/j.jacc.2017.03.018>. URL <https://www.sciencedirect.com/science/article/pii/S0735109717360424>.
- [66] A. Segal, D. Haehnel, and S. Thrun. Generalized-icp. In *Proceedings of Robotics: Science and Systems (RSS)*, Seattle, USA, 2009.
- [67] D. P. Shamonin, E. E. Bron, B. P. F. Lelieveldt, M. Smits, S. Klein, and M. Staring. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer’s disease. *Frontiers in Neuroinformatics*, 7:50, Jan. 2014. doi: 10.3389/fninf.2013.00050.
- [68] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6:60, 2019. doi: 10.1186/s40537-019-0197-0. URL <https://link.springer.com/article/10.1186/s40537-019-0197-0>.
- [69] Y. Song, Y. Hu, D. Wu, et al. Locally rigid, vessel-based registration for laparoscopic liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 10(12):1951–1961, 2015.
- [70] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA/ML-CDS 2017)*, volume 10553 of *Lecture Notes in Computer Science*, pages 240–248. Springer, Cham, 2017. doi: 10.1007/978-3-319-67558-9_28.
- [71] G. Taubin. A signal processing approach to fair surface design. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*, pages 351–358. ACM, 1995. doi: 10.1145/218380.218473.

- [72] S. Uchida. Image processing and recognition for biological images. *Development, growth differentiation*, 55, 04 2013. doi: 10.1111/dgd.12054.
- [73] S. Varma and R. Simon. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7:91, 2006. doi: 10.1186/1471-2105-7-91. URL <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-91>.
- [74] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145:166–179, 2017. doi: 10.1016/j.neuroimage.2016.10.038. URL <https://www.sciencedirect.com/science/article/pii/S105381191630595X>.
- [75] Y. Velikova, W. Simson, M. F. Azampour, P. Paprottka, and N. Navab. Cactus: Common anatomical ct-us space for us examinations. *International Journal of Computer Assisted Radiology and Surgery*, 19:861–869, 2024. doi: 10.1007/s11548-024-03060-y. URL <https://doi.org/10.1007/s11548-024-03060-y>.
- [76] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth. Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. doi: 10.1148/ryai.230024. URL <https://doi.org/10.1148/ryai.230024>.
- [77] World Health Organization. Cardiovascular Diseases (CVDs): Fact sheet. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), June 2021. Accessed: 2025-07-10.
- [78] World Health Organization. *World health statistics 2025: Monitoring health for the SDGs, Sustainable Development Goals*. World Health Organization, Geneva, 2025. ISBN 978-92-4-011049-6. URL <https://www.who.int/publications/i/item/9789240110496>. Available under CC BY-NC-SA 3.0 IGO. Accessed: 2025-07-10.
- [79] M. Wortsman, G. Gur-Ari, O. Sharir, L. Berrada, H. Larochelle, R. Gontijo-Lopes, A. Morcos, and Y. Carmon. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International Conference on Machine Learning (ICML)*, 2022. URL <https://proceedings.mlr.press/v162/wortsman22a.html>.
- [80] D. Wu, Z. Li, M. H. D. Ansari, X. T. Ha, M. Ourak, J. Dankelman, A. Menci-

- assi, E. De Momi, and E. V. Poorten. Comparative Analysis of Interactive Modalities for Intuitive Endovascular Interventions . *IEEE Transactions on Visualization & Computer Graphics*, 31(02):1371–1388, Feb. 2025. ISSN 1941-0506. doi: 10.1109/TVCG.2024.3362628. URL <https://doi.ieeecomputersociety.org/10.1109/TVCG.2024.3362628>.
- [81] S. G. Yazdi, L. Huetter, P. D. Docherty, P. N. Williamson, D. Clucas, M. Jermy, and P. H. Geoghegan. A novel fabrication method for compliant silicone phantoms of arterial geometry for use in particle image velocimetry of haemodynamics. *Applied Sciences*, 9(18), 2019. doi: 10.3390/app9183811. URL <https://www.mdpi.com/2076-3417/9/18/3811>.
- [82] M. N. Young, S. Kearing, M. A. Albaghdadi, A. Latib, and A. Iribarne. Trends in transcatheter vs surgical mitral valve repair among medicare beneficiaries, 2012 to 2019. *JAMA Cardiology*, 7(7):770–772, 07 2022. ISSN 2380-6583. doi: 10.1001/jamacardio.2022.0775. URL <https://doi.org/10.1001/jamacardio.2022.0775>.
- [83] X. Zhang, A. Sridhar, X. T. Ha, S. Z. Mehdi, A. Fortuna, M. Magro, A. Peloso, A. Bicchi, M. Ourak, A. Aliverti, E. Votta, E. Vander Poorten, and E. De Momi. Path tracking control of a steerable catheter in transcatheter cardiology interventions. *International Journal of Computer Assisted Radiology and Surgery*, 19:757–766, 2024. doi: 10.1007/s11548-024-03069-3. URL <https://doi.org/10.1007/s11548-024-03069-3>.
- [84] B. Zitová and J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21(11):977–1000, 2003. doi: 10.1016/S0262-8856(03)00137-9.

A | Evaluation metrics

This appendix reports the formal definitions of all evaluation metrics employed in this work. Metrics are grouped into two macro-categories, reflecting the two main stages of the pipeline: segmentation and registration.

A.1. Segmentation metrics

Let $y_i \in \{0, 1\}$ denote the ground-truth label of pixel i and $p_i \in \{0, 1\}$ the corresponding predicted binary label. True positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) are defined in the standard way. All segmentation metrics are computed exclusively over valid image regions, as defined by the validity mask introduced during pre-processing.

Region-based overlap metrics

Dice similarity coefficient (DSC). The Dice coefficient measures the spatial overlap between predicted and ground-truth regions and is widely adopted in medical image segmentation.

$$\text{Dice} = \frac{2 \text{ TP}}{2 \text{ TP} + \text{FP} + \text{FN}} \quad (\text{A.1})$$

Tversky index. The Tversky index generalizes Dice and IoU by allowing different weights for false positives and false negatives, and is particularly suited to imbalanced segmentation problems.

$$\text{TI} = \frac{\text{TP}}{\text{TP} + \alpha \text{FP} + \beta \text{FN}} \quad (\text{A.2})$$

Boundary-based metrics

Let ∂S_p and ∂S_g denote the sets of boundary points of the predicted and ground-truth segmentation masks, respectively.

Boundary F1-score (bF1). Boundary precision and recall are computed by verifying whether boundary points of one mask lie within a fixed tolerance distance r from the boundary of the other mask. The bF1-score evaluates contour alignment independently of region overlap.

$$\text{bF1} = 2 \cdot \frac{\text{bPrecision} \cdot \text{bRecall}}{\text{bPrecision} + \text{bRecall}} \quad (\text{A.3})$$

95th percentile Hausdorff Distance (HD95). HD95 captures worst-case boundary discrepancies while remaining robust to outliers.

$$\text{HD95} = \max \left\{ \text{percentile}_{95}(d(x, \partial S_g))_{x \in \partial S_p}, \text{percentile}_{95}(d(y, \partial S_p))_{y \in \partial S_g} \right\} \quad (\text{A.4})$$

Calibration metrics

Brier score. The Brier score measures the mean squared error between predicted probabilities and ground-truth labels, capturing both calibration and discrimination quality.

$$\text{Brier} = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2 \quad (\text{A.5})$$

Expected Calibration Error (ECE). Predicted probabilities are grouped into M confidence bins $\{B_m\}$. The Expected Calibration Error is defined as:

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{N} |\text{acc}(B_m) - \text{conf}(B_m)| \quad (\text{A.6})$$

where $\text{acc}(B_m)$ denotes empirical accuracy and $\text{conf}(B_m)$ the average confidence in bin B_m .

A.2. Registration metrics

Let $\mathcal{P} = \{p_i\}_{i=1}^N$ and $\mathcal{Q} = \{q_j\}_{j=1}^M$ denote two point sets representing the registered geometry and the reference anatomy, respectively.

Distance-based accuracy metrics

Root Mean Square Error (RMSE). RMSE measures the average Euclidean distance between corresponding points after registration, penalizing large alignment errors.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|p_i - q_i\|^2} \quad (\text{A.7})$$

Median distance error. The median distance provides a robust estimate of typical registration accuracy, reducing sensitivity to outliers.

$$\text{Median} = \text{median}_i (\|p_i - q_i\|) \quad (\text{A.8})$$

Set-to-set distance metrics

Squared Chamfer Distance (CD). The squared Chamfer distance measures the symmetric proximity between two point sets without requiring explicit correspondences.

$$\text{CD}(\mathcal{P}, \mathcal{Q}) = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \min_{q \in \mathcal{Q}} \|p - q\|^2 + \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \min_{p \in \mathcal{P}} \|q - p\|^2 \quad (\text{A.9})$$

95th percentile Hausdorff Distance (HD95). HD95 evaluates worst-case surface misalignment while limiting sensitivity to isolated outliers.

$$\text{HD95} = \max \left\{ \text{percentile}_{95}(d(p, \mathcal{Q}))_{p \in \mathcal{P}}, \text{percentile}_{95}(d(q, \mathcal{P}))_{q \in \mathcal{Q}} \right\} \quad (\text{A.10})$$

B | Algorithm Parameters and Code Settings

This appendix lists the fixed configuration parameters used in the proposed segmentation and registration pipelines. Run-dependent quantities estimated during calibration or optimization (e.g., temperature scaling coefficients or adaptive thresholds) are not reported, as they vary across runs.

Table B.1: Fixed parameters used for geometric reconstruction and registration.

Parameter	Value	Meaning / notes
2D mask refinement and contour smoothing		
Closing kernel size	5 [px]	Morphological closing applied before contour extraction.
Opening kernel size	5 [px]	Morphological opening applied before contour extraction.
Radial smoothing σ	7.0	Gaussian smoothing of the radial contour profile.
Point cloud preprocessing		
Filter radius	5.0 [mm]	Spatial radius used for local point filtering.
Neighbors (k)	20	Number of neighbors used in local neighborhood computations.
Alpha	0.7	Strength factor of the iterative point cloud smoothing operator.
Lambda	0.5	Smoothing parameter of the iterative filtering scheme.
Mu	-0.51	Counter-smoothing parameter limiting shrinkage effects.
Iterations	40	Number of preprocessing iterations.
Voxel size	1.5 [mm]	Voxel size used for PC downsampling.
Surface reconstruction and remeshing		
Poisson depth	11	Octree depth controlling surface reconstruction resolution.
Density quantile	0.01	Quantile threshold used for density-based trimming.
Remesh target edge	1.5 [mm]	Target edge length for isotropic remeshing.
Remesh iterations	6	Number of remeshing refinement steps.
Feature angle	30.0 [°]	Angle threshold used for feature preservation.
Boundary-aware mesh smoothing		
Bilateral iterations	20	Number of bilateral smoothing iterations.
σ_{spatial}	4.5	Spatial bandwidth controlling neighborhood influence.
σ_{normal}	0.35	Normal bandwidth controlling feature preservation.
Freeze band	6.0 [mm]	Boundary band excluded from smoothing.

Continued on next page

Parameter	Value	Meaning / notes
Max step	1.125 [mm]	Maximum vertex displacement per iteration.
Deformable registration (CPD)		
CPD β	5.0	Gaussian kernel width controlling deformation smoothness.
CPD λ	8.0	Regularization weight penalizing non-smooth deformations.

Table B.2: Fixed parameters used for the ultrasound segmentation pipeline (training and inference).

Parameter	Value	Meaning / notes
Input data and splitting		
Image size	256	Input images resized to 256×256 .
Number of splits	5	Number of folds used for cross-validation.
Train/Val/Test split	(0.7, 0.15, 0.15)	Dataset partition ratios for random splitting.
Seed	42	Random seed used to ensure reproducible dataset splits.
Network architecture and regularization		
Initial features	12	Number of feature maps in the first encoder layer.
Dropout probability	0.1	Dropout rate used as regularization during training.
Training setup		
Train batch size	10	Batch size used during training and validation.
Initial learning rate	1×10^{-3}	Initial learning rate of the optimizer.
Weight decay	1×10^{-4}	Weight decay coefficient used in AdamW.
Max epochs	50	Maximum number of training epochs.
Early patience	10	Number of consecutive epochs without improvement before stopping.
Early min delta	1×10^{-5}	Minimum improvement required to reset patience.
Loss function		
BCE positive weight	2.0	Positive-class weight in the binary cross-entropy loss.
FT coefficient	0.75	Relative weight of the Focal-Tversky loss term.
Tversky α	0.25	False-positive penalty in the Tversky index.
Tversky β	0.75	False-negative penalty in the Tversky index.
Focal γ	1.33	Focusing parameter emphasizing hard examples.
Mask post-processing		
Closing kernel size	5 [px]	Morphological closing kernel size applied to binary masks.
Opening kernel size	5 [px]	Morphological opening kernel size applied to binary masks.

List of Figures

1.1	Scheme of transcatheter delivery system	2
1.2	Some examples of catheter-based devices used in interventional cardiology. From left to right: stent, atrial appendage occluder and prosthetic valve. . .	2
1.3	Trends in TEER vs. traditional surgical mitral valve repair (SMVR) in the USA, highlighting the increasing relevance of minimally invasive approaches in cardiology [82].	3
1.4	The MitraClip™ system (<i>Abbott Vascular, Santa Clara, California</i>). Left: side and atrial views of the clipping phase Center: delivery system, includ- ing catheter and steering mechanism. Right: close-up view of the clip in the open configuration.	4
1.5	Fluoroscopy image	6
1.6	Real-time 3-dimensional transesophageal echocardiography (RT3DTEE) . .	8
1.7	Comparison of navigation techniques for endovascular interventions	9
1.8	Patient-specific vascular model, including the femoral vein (red), a segment of the inferior vena cava (blue) and a bounding box (green) modelling the mechanical effect of surrounding tissues. Adapted from [62].	11
2.1	Conceptual example of image registration [37]. A transformation T is es- timated to map a moving object (point p) onto a fixed reference (point q).	15
2.2	Examples of transformation models used in image registration, ranging from rigid transformations to nonlinear deformations [72].	17
2.3	Schematic overview of a classical optimization-based image registration framework [37]. The moving image is transformed into the reference coordi- nate system (<i>Transform</i>), interpolated at non-grid locations (<i>Interpolator</i>) and compared to the fixed image through a similarity metric (<i>Metric</i>). An <i>Optimizer</i> iteratively updates the transformation parameters to minimize the overall cost function	18

2.4	Example of a segmentation-guided registration pipeline [7]. Ultrasound volumes are segmented with a U-Net to extract a surface point cloud, which is used to pre-align a CT-derived point cloud through a global rigid CPD step. A subsequent network-based refinement (<i>FlowNet</i>) produces the final registered CT segmentation.	22
2.5	Example of multi-class anatomical segmentation of brain MRI, obtained using a parametric Bayesian modeling approach [54].	23
2.6	Schematic representation of a U-Net encoder-decoder architecture for multiclass cardiac chambers segmentation [14].	25
2.7	Example of the acoustic shadowing artifact in a liver ultrasound image [50].	26
2.8	Examples of data augmentation strategies commonly applied during training, including spatial transformations and intensity variations [22].	27
2.9	Schematic representation of the dropout regularization strategy [58].	27
2.10	Example of a thin-walled silicone vascular phantom reproducing a patient-specific pathological arterial geometry [18]. The phantom was developed for experimental evaluation and training in catheter-based endovascular procedures.	31
2.11	Wall-less phantom consisting of a bifurcating arterial geometry embedded in a gelatin-based tissue-mimicking matrix and integrated into a closed-loop flow circuit [43].	32
2.12	Ultrasound images and corresponding gray-level intensity histograms of different common tissue-mimicking materials [35].	33
3.1	Overview of the proposed framework. By registering the preoperative model with the intraoperative reconstruction, the proposed framework aligns the associated reference frames and estimates the transformations required to update the patient-specific finite element model (FEM).	36
3.2	Reference frames and transformation chain relating the robot base, end effector, ultrasound probe and image coordinate system. Image axes follow the conventional pixel-based notation.	38
3.3	The model of all the biological structures involved in MitraClip TM procedure is obtained from the segmentation of CT preoperative images [62]. . .	39
3.4	Automatic segmentation module integrated in the intraoperative pipeline. .	41
3.5	Overview of the geometric reconstruction process.	42
3.6	Rigid registration strategy scheme based on BFGS initialization and symmetry-aware multi-start ICP refinement.	45

3.7	Examples of US images from different datasets. From left to right: <i>Phantom</i> , <i>Carotid</i> , <i>San Raffaele</i> and <i>Mus-V</i>	47
3.8	Dataset pruning pipeline: feature embedding, PCA projection and Farthest Point Sampling selection.	49
3.9	Example of pruning outcome on a single acquisition, showing reduced redundancy while preserving variability. PC1, PC2, PC3 represent the three principal components obtained by PCA.	49
3.10	K-fold cross-validation workflow with model fusion and post-hoc calibration to obtain a single calibrated model. Example of a 3-fold CV scheme. . . .	52
3.11	Representation of the three vessel geometries used to build the phantoms. .	55
3.12	Femoral artery phantom fabrication using a PLA external mold, a sacrificial PVA core and gel-wax casting.	57
3.13	<i>In-vitro</i> US acquisition setup with probe pose tracking via robotic arm. . .	58
4.1	Distribution of Chamfer distance and HD95 across repeated reconstructions for each phantom geometry. Boxplots highlight comparable medians with geometry-dependent dispersion.	66
4.2	Representative examples of rigid registration under challenging initialization conditions. For each phantom geometry, one failure case of the single-start ICP baseline is compared with a successful alignment obtained using the proposed multi-start strategy.	67
4.3	Success rate of rigid registration for each phantom geometry. Left column: success rate as a function of initial rotation magnitude. Right column: success rate as a function of gaussian noise level.	68
4.4	Registration accuracy after rigid and deformable alignment for the three phantom geometries. RMSE and HD95 are computed from the intraoperative reconstruction to the CAD reference model (moving→fixed). Deformable registration consistently reduces both average and tail errors. . .	69
4.5	Qualitative comparison of surface alignment after rigid (left) and deformable (right) registration for the Artery phantom. The CAD phantom model (fixed geometry) is shown in light blue transparency, while the intraoperative reconstruction (moving geometry) is displayed in blue with visible mesh edges.	70
4.6	Estimated displacement field introduced by the non-rigid registration exemplified for the Artery phantom. Points of the intraoperative reconstruction are colored according to displacement magnitude (in mm), illustrating the spatial distribution of the applied correction.	71

4.7	Breakdown of median computational time across the main stages of the registration pipeline for the three phantom geometries. Ultrasound acquisition represents the dominant contribution, while rigid and deformable registration account for a limited fraction of the total runtime.	72
4.8	Qualitative comparison of the three architectures on a <i>San Raffaele</i> frame. Note the partial collapse of the segmentation in U-Net3+.	75
4.9	Mean DSC comparison across <i>Single-session</i> , <i>Full</i> and <i>Pruned</i> dataset configurations.	76
4.10	Mean HD95 comparison across the <i>Single-session</i> , <i>Full</i> and <i>Pruned</i> configurations of the San Raffaele dataset. Values are expressed in pixel units. .	76
4.11	Distribution of Dice values for the <i>Single-session</i> , <i>Full</i> and <i>Pruned</i> configurations of the San Raffaele dataset, using the Attention U-Net.	77
4.12	DSC comparison between arterial and venous segmentation across architectures.	79
4.13	HD95 comparison between arterial and venous segmentation. Values are reported in pixel units.	79
4.14	Distribution of circularity and foreground prevalence across MusV configurations. In the box plots, the orange horizontal line indicates the median value; blue boxes indicate the interquartile range; whiskers indicate the 5th and 95th percentile. Outliers are indicated by empty circles.	80

List of Tables

- 3.1 Mus-V dataset configurations used for vessel-type and domain aggregation analysis. 63
- 4.1 Reconstruction repeatability across phantom geometries. Median and interquartile range (IQR) are reported. 65
- 4.2 Mean DSC obtained by each architecture across datasets. 73
- 4.3 Mean HD95 obtained by each architecture across datasets. HD95 values are reported in pixel units. 73
- B.1 Fixed parameters used for geometric reconstruction and registration. . . . 101
- B.2 Fixed parameters used for the ultrasound segmentation pipeline (training and inference). 102

