Medical Image Analysis 113 (2026) 104139

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

PIVOTS: Aligning unseen structures using preoperative to intraoperative volume-to-surface registration for liver navigation ☆

Peng Liu [a,b,c,d,*,1], Bianca Güttner [a,b,c,d,1], Yutong Su [a,b,c,d], Chenyang Li [a,b,c,d], Jinjing Xu [a,b,c,d], Mingyang Liu [g], Zhe Min [g], Andrey Zhylka [h], Jasper Smit [h], Karin Olthof [h], Matteo Fusaglia [h], Rudi Apolle [i,c,b,d], Matthias Miederer [i,c,b,d], Laura Frohneberg [c], Carina Riediger [c], Jürgen Weitz [c], Fiona Kolbinger [e,j], Stefanie Speidel [a,b,c,d,f], Micha Pfeiffer [a,b,c,d]

[a] Department of Translational Surgical Oncology, National Center for Tumor Diseases (NCT), NCT/UCC Dresden, a partnership between DKFZ, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, and Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany, Fetscherstrasse 74/PF 64, Dresden, 01307, Saxony, Germany
[b] German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg, 69120, Baden-Württemberg, Germany
[c] Faculty of Medicine and University Hospital Carl Gustav Carus, Dresden, 01307, Saxony, Germany
[d] Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Saxony, Germany
[e] Department of Visceral, Thoracic and Vascular Surgery, University Hospital and Faculty of Medicine Carl Gustav Carus, TUD Dresden University of Technology, Dresden, 01307, Saxony, Germany
[f] Centre for Tactile Internet with Human-in-the-Loop, TU Dresden, Dresden, Saxony, Germany
[g] School of Control Science and Engineering, Shandong University, Jinan, Shandong, China
[h] Surgical Department, The Netherlands Cancer Institute, Amsterdam, Netherlands
[i] Translational Imaging in Oncology, National Center for Tumor Diseases, Fetscherstrasse 74/PF 64, Dresden, 01307, Saxony, Germany
[j] Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA

ARTICLE INFO

Dataset link: https://github.com/pengliu-nct/PIVOTS

Keywords:
Intraoperative
Liver navigation
Point clouds
Nonrigid
Registration
Volume-to-surface

ABSTRACT

Non-rigid registration is essential for augmented reality-guided laparoscopic liver surgery, as it enables the fusion of preoperative information such as tumor location and vascular structures into the limited intraoperative view, thereby enhancing surgical navigation. A prerequisite is the accurate prediction of intraoperative liver deformation, which remains highly challenging due to factors such as large deformation caused by pneumoperitoneum, respiration and tool interaction as well as noisy intraoperative data, and limited field of view due to occlusion and constrained camera movement. To address these challenges, we introduce **PIVOTS**, a **P**reoperative to **I**ntraoperative **VO**lume-**T**o-**S**urface registration neural network that directly takes point clouds as input for deformation prediction. The geometric feature extraction encoder allows multi-resolution feature extraction, and the decoder, comprising inter-modality cross attention modules, enables information exchange between pre- and intraoperative features and accurate multi-level displacement prediction. We train the neural network on a large synthetic dataset created using a biomechanical simulation pipeline that explicitly targets the mentioned intraoperative challenges and validate its performance on both synthetic and real datasets. Results demonstrate superior registration performance of our method compared to baseline methods, exhibiting strong robustness against high amounts of noise, large deformation, and various levels of intraoperative visibility. The network is fast enough to run multiple times per second and directly generalizes to new patients without retraining. We publish training and test sets as evaluation benchmarks in an effort to contribute to the development of more robust liver registration methods based on volume-to-surface data. Code, docker container and datasets are available at https://github.com/pengliu-nct/PIVOTS.

☆ Given their role as Associate Editor, Stefanie Speidel had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor.

* Corresponding author at: Department of Translational Surgical Oncology, National Center for Tumor Diseases (NCT), NCT/UCC Dresden, a partnership between DKFZ, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, and Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany, Fetscherstrasse 74/PF 64, Dresden, 01307, Saxony, Germany.
E-mail address: peng.liu@nct-dresden.de (P. Liu).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.media.2026.104139
Received 10 June 2025; Received in revised form 6 April 2026; Accepted 20 May 2026
Available online 26 May 2026
1361-8415/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

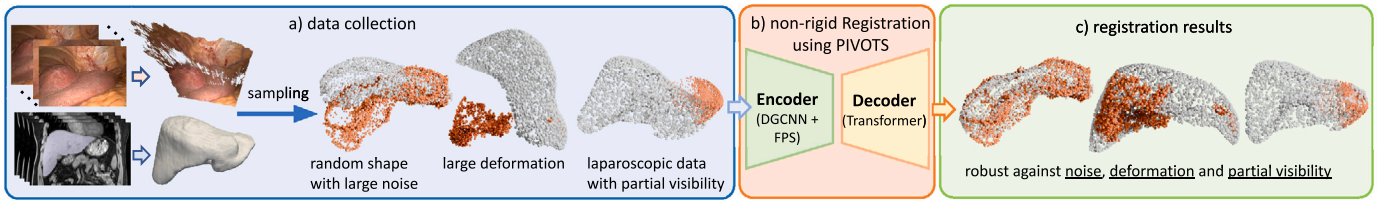


Fig. 1. Registration with our method. First, the intraoperative surface (orange) is reconstructed from laparoscopic video and the preoperative volume (gray) is segmented from MRI, and the surface and volume are subsampled as point clouds, depicted in (a). These point clouds are then fed into PIVOTS for deformation prediction in (b). Finally, the registration results in (c) show the robustness of PIVOTS against substantial noise, large deformation and partial visibility.

1. Introduction

Laparoscopic liver surgery (LLS) is a minimally invasive approach for hepatic resection that utilizes small abdominal incisions, specialized instruments, and high-resolution camera systems. During LLS, surgeons strategically insert trocars into the abdominal cavity, through which they introduce surgical instruments and a laparoscope for real-time imaging, allowing them to navigate and resect the targeted liver tissue. Compared to open surgery, LLS offers decreased blood loss, postoperative pain and infectious complications, shorter hospital stays, and faster recovery, while matching the oncological results with respect to margin accuracy and recurrence rates (Novitsky et al., 2004; Abu Hilal et al., 2018; Amodu et al., 2022; Nguyen et al., 2011; Ratti et al., 2018).

Despite these advantages, such minimally invasive procedures also introduce multiple challenges. Intraoperative views only show the tissue surface that presents scarce distinct texture features. They are often restricted by occlusions from adjacent tissues or instruments and confined operating spaces, and may therefore only show a small part of the surface without salient landmarks like the ridge line, especially at later stages of a procedure. The image can additionally be degraded by camera artifacts and light reflections. At the same time, preoperative MRI or CT scans provide a comprehensive overview of the abdominal anatomy, detailing critical internal anatomical structures, such as vessels and tumors. Each patient presents a different setting due to morphological variations (Singh and Rabi, 2019; Srimani and Saha, 2020; Santhosh et al., 2024) or due to large tumors distorting the natural shape of the organ. Besides instrument interaction, during surgery, the liver undergoes significant deformation due to respiratory motion, cardiac pulsation, pneumoperitoneum pressure, the round ligament stabilizing the left lobe and contact with other organs. The exact forces applied to the organ by these conditions remain unknown. Similarly, the tissue properties of a patient's liver, e.g. stiffness, are unknown prior to surgery while they hold a big influence on the intraoperative deformation and show great variability across the population and across measurement techniques and conditions (Mattei and Ahluwalia, 2016; Marchesseau et al., 2017; Murad Gutiérrez and Romero Enciso, 2018; Lemine et al., 2024): Healthy liver tissue is soft and pliable with a Young's modulus below 2.5 kPa, becoming much stiffer with comorbidities like liver cirrhosis or fatty liver disease, with ranges up to 16 kPa in non-invasive elastography. Furthermore, the organ's volume may change between the diagnostic preoperative scan and the time of the operation (Leung et al., 1986a). These unknowns and mismatches between the data collection modalities impede the direct transfer of preoperative information to intraoperative settings.

However, accurate information transfer is critical for patient safety, since liver surgery involves complex and variable vascular anatomy (hepatic arteries, portal and hepatic veins) and bile duct anatomy, the identification of which can be challenging, particularly in the limited view setting of LLS. Injuries to these structures can lead to serious complications like bleeding, reduced blood supply to the remaining liver, and, potentially, liver failure. Therefore, image-guided laparoscopic navigation systems with augmented reality (AR) are designed to facilitate orientation. Such a navigation system must satisfy several key requirements: It must (i) ensure an accurate alignment of pre- and intraoperative structures, reliably identifying the visible surface part, (ii) infer a precise, detailed and sufficiently smooth online deformation of critical structures (such as vessels and small metastases) from the visible parts of the organ, (iii) demonstrate a high robustness against intraoperative conditions like noise and (iv) require limited manual intervention while (v) running on available hardware at very high speeds. To meet these demands, a navigation system must rely on a registration algorithm that is both precise and computationally efficient.

In a typical intraoperative pipeline 3D–3D registration methods utilize the preoperative liver volume and align it to an intraoperative 3D surface. This intraoperative target surface can be constructed using various modalities, such as *Structure from Motion* (SfM) or *Simultaneous Localization and Mapping* (SLAM) methods (Docea et al., 2021, 2022) or by tracing the surface with tracked tools (Heiselman et al., 2024). Such reconstructed surfaces are typically small with little overlap with the preoperative full liver volume, and are subject to different types of noise compared to the segmented preoperative 3D liver surfaces. Noise properties differ between reconstruction pipelines based on the choice of endoscope, calibration method, segmentation method, error accumulation due to heating, tracking and mapping method, amongst others.

Rigid 3D–3D registration algorithms then aim to find the optimal rigid transformation between the pre- and intraoperative surfaces (Z. Yang et al., 2023). Building on this initial alignment, non-rigid 3D–3D registration methods estimate a deformation field for the full liver volume, including internal vessels and tumors. Applying this deformation to every point of the preoperative volume results in an estimated up-to-date intraoperative state of the liver which can be used to reveal the hidden internal structures to surgeons via AR (Pfeiffer et al., 2020). While methods exist that estimate the displacement exclusively on the surface of the preoperative data, this is not sufficient for our use case since a meaningful internal prediction is needed to reliably update the position of internal anatomical structures with a high distance to the liver surface. Surface-to-surface methods in this case would require an additional method to calculate the internal displacement.

Many current non-rigid registration methods fail to translate to clinical use-cases due to various challenges, including high sensitivity to noise, their inability to generalize to new patients without re-parameterization or re-training and performance degradation when dealing with large deformations. Many existing methods are not real-time capable - and those that are tend to lose in accuracy and physical plausibility (Pfeiffer et al., 2020; Heiselman et al., 2024). Our goal is a network that is applicable for translation. We therefore follow above system requirements as closely as possible and target a fully automatic method without interactive guidance which only relies on data that is collected before and during surgery by default, and can be used throughout a procedure, including limited views. We aim for one-shot prediction of volumetric displacements to avoid slow iterations or subsequent interpolation methods.

In line with this, we propose a novel neural network (Fig. 1) that yields better registration results with higher robustness against the mentioned challenges. It estimates plausible liver deformation from partial views of the organ surface while implicitly inferring unknown

boundary conditions and tissue properties. The architecture's full potential is achieved by a targeted, challenge-driven design, enabled by custom synthetic training data that incorporate these adverse conditions early and enforce sufficiently expressive components. Our contributions can be summarized as follows:

- We propose the Point-based **Preoperative to Intraoperative Volume to Surface registration (PIVOTS)**, a new attention-based neural network for the volume-to-surface formulation of deformable liver registration, which **generalizes to unseen patients on the fly** and can handle **larger tissue deformations** than previous learning-based methods. It outperforms current state-of-the-art methods while being **about two orders of magnitude faster** at about 4.3 fps. At its core, allowing additional attention of the target surface to the surface volume guides the receptive field along the intraoperative surface, which stabilizes the implicit correspondence and displacement estimation for challenging scenarios. Code and weights for the network are made publicly available.
- We build upon a **soft-tissue simulation pipeline for registration pair creation** to introduce the network to difficult conditions early on and to mitigate the common issue of lacking training and evaluation data. To this end, we generate a dataset with almost 200k highly variable training samples grounded in biomechanics, substantially expanding over previous approaches for synthetic training regimes. Synthetic training and test dataset containing 200k/3k samples are made publicly available.
- Our **extensive evaluations** show PIVOTS' high registration performance, especially strong robustness against substantial amounts of noise and various amounts of intraoperative visibility. We show **one-shot generalization** to phantom, human breathing motion and real laparoscopic data without re-training, and find that in our experiments, across datasets, native surface-to-surface methods cannot solve the volume-to-surface problem correctly, even when re-trained on our data.
- For thorough evaluation despite the lack of ground-truth human data in liver registration tasks, we curate a **healthy human liver breathing motion (HHLBM)** dataset from dedicated MRI acquisitions in different poses. By localizing structures inside the livers, we can benchmark our registration method on real human deformation data quantitatively. The dataset is also made publicly available.

Our paper is organized in four main sections: we introduce related work in 3D registration in Section 2, followed by the elaboration of the proposed non-rigid neural network and synthetic scene generation in Section 3. For evaluation, we show a comparison with 3D–3D SOTA methods (Section 4.3) and a quantitative evaluation on a recent intraoperative dataset (Section 4.3.1), detailed studies of sensitivity to different deformation levels (Section 4.4), noise levels (Section 4.5), and visible surface percentages (Section 4.6), followed by network ablations (Section 4.7). We conclude with a discussion of strengths and remaining limitations of our method in Section 5.

2. Related work

Registration algorithms, which are the core component in laparoscopic navigation systems, can be broadly divided into 3D–3D and 3D–2D registration based on the input modalities. State-of-the-art 3D–3D registration methods can be further categorized into two primary types: (1) traditional approaches, and (2) learning-based methods.

Among traditional registration methods, feature-based methods detect keypoints (Sipiran and Bustos, 2011; Streiff et al., 2021; Jin et al., 2024), which are described with either local or global point descriptors (Rusu et al., 2009; Guo et al., 2013; Zhao et al., 2022). Explicit correspondences are then established through feature matching (Rusu et al., 2008; Guo et al., 2013; Rodola et al., 2013), or voting mechanisms for correspondence optimization (Tombari and Di Stefano, 2010; J. Yang et al., 2023). However, in the context of liver registration, explicit feature detection often fails because the liver surface contains many featureless or very similar regions. Furthermore, due to the multi-modal nature of the task (preoperative data is from CT or MRI, intraoperative data is from an endoscope or other tracking method) the noise, artifacts and even shapes vary widely between the source and target, making it difficult to establish correspondences between the two meshes.

In contrast to methods relying on one-shot correspondence, traditional iterative techniques such as ICP (Besl and McKay, 1992) and its variants (e.g. Go-ICP Yang et al., 2015, non-rigid ICP Amberg et al., 2007) optimize both point correspondences and transformation. Another family of methods relies on implicit correspondences, for instance, CPD (Myronenko and Song, 2010; Hirose, 2022) models the source point cloud as the centroids of Gaussians with equal isotropic covariance matrices in a Gaussian Mixture Model. These iterative models can be computational expensive and slow. Moreover, they focus on surface-to-surface registration, ignoring the volumetric nature and biomechanical properties of an organ. However, the idea of letting the correspondence search and the deformation estimation step work in tandem is an intelligent idea, which we also build implicitly into our method.

To integrate biomechanical properties, liver navigation methods often employ biomechanical models for soft-tissue simulation under boundary conditions to infer intraoperative deformation, often using Finite Element Methods (Özgür et al., 2018; Suwelack et al., 2014; Yang et al., 2024). In theory, these methods model material behavior in a physically accurate way. However, in practice they usually require many simplifications which negate this advantage. Furthermore, the parameterization of their boundary conditions can require a lot of effort, they are relatively difficult to parallelize, they often suffer from long run-times and can have difficulties adapting to larger deformations.

In light of the marked success of deep learning-based methods, data-driven 3D registration enables end-to-end frameworks, offering computational efficiency and robustness. According to different representations of 3D point clouds and deformation, these learning-based methods can be divided into grid-based, graph-based and point-based methods.

Grid-based methods can leverage highly optimized convolutional operators, but require the data to be represented in a regular grid structure. Methods such as 3DMatch (Zeng et al., 2017) and PerfectMatch (Gojcic et al., 2019) convert local patches into a volumetric representation of truncated distance field and employ a 3D convolutional neural network to learn local descriptors. V2S-Net (Pfeiffer et al., 2020) transforms preoperative liver volume and intraoperative partial surface into signed distance field and unsigned distance field, respectively, while Cue-Net (Liu et al., 2024a) integrates user inputs into the voxel grid as auxiliary information to guide the registration. Both train 3D CNNs to predict a displacement between pre- and intraoperative data for every voxel. Similarly, Chen et al. (2024) focuses on voxels surrounding the liver volume and surfaces, but uses a transformer architecture to learn deformation. Voxel representation suffers from several limitations, such as reduced flexibility, since all inputs need to be converted to a fixed voxel grid, and high memory consumption, as every voxel is processed regardless of its proximity to the target. Furthermore, voxel-based methods can lead to grid patterns in the estimated displacement, which produces unrealistic deformations when applied to the organ meshes (Heiselman et al., 2024). Due to these shortcomings, we build PIVOTS as a point-cloud processor which can work with continuous point positions and does not force input data into a grid structure.

Point clouds present challenges for neural networks due to their irregularity, unordered structure, and varying density. However, with the advent of robust backbones for point cloud analysis, learning on

point clouds directly becomes increasingly efficient. For example, architectures such as PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), PointTransformer (Zhao et al., 2021), and KPConv (Thomas et al., 2019) facilitate the development of point cloud registration (Z. Yang et al., 2023; Huang et al., 2020; Sarode et al., 2019; Shi et al., 2021; Yew and Lee, 2022; Qin et al., 2023a; Li and Harada, 2022a).

Harnessing PointNet (Qi et al., 2017a) to extract global features, PointNetLK (Aoki et al., 2019) modifies the Lucas & Kanade algorithm with global feature differences for iterative alignment, whereas PCR-Net (Sarode et al., 2019) directly concatenates global features of source and target to regress poses. More recently, LiverMatch (Z. Yang et al., 2023), RegTR (Yew and Lee, 2022), GeoTransformer (Qin et al., 2023a) and Lepard (Li and Harada, 2022a) utilize KPConv (Thomas et al., 2019) for grid-downsampling. Combined with point feature extraction and self- and cross-attention for feature propagation, the final correspondences and registration results are yielded according to various regression designs. RegTR (Yew and Lee, 2022) directly regresses the transformed point clouds from the conditioned features via dual MLP decoders. GeoTransformer (Qin et al., 2023a) adopts a local-to-global scheme, where transformation candidates are first generated between patches with conditioned features, then final transformation is the optimal candidate on the global point clouds. Lepard (Li and Harada, 2022a) integrates positional encoding into a two-block transformer architecture to incrementally achieve robust alignment, suitable for both rigid and non-rigid cases. In contrast, NDP (Li and Harada, 2022b) employs positional encoding with progressively increasing frequencies to address non-rigid registration by shifting from stiff and global alignment to soft and local adjustment. Building on these methods, our PIVOTS network utilizes attention mechanisms, positional encoding, and a hierarchical encoding- and decoding-structure as these are powerful tools for point-cloud processing.

Another family of 3D learning is Graph-based methods, beginning with the introduction of graph convolutional networks (Scarselli et al., 2008) and later incorporating convolutional networks (Duvenaud et al., 2015). Dynamic Graph CNN (DGCNN) (Wang et al., 2019) applies dynamic edge convolution to capture both local and global features in point clouds. DeepClosestPoint (Wang and Solomon, 2019) exploits DGCNN to extract local features for correspondence prediction with an attention module, followed by a differentiable SVD layer for transformation calculation. Predator (Huang et al., 2021) first extracts keypoint features using KPConv, then applies dynamic edge convolution and cross-attention to predict an overlap score between the two groups of superpoints, which are decoded into correspondences by a linear decoder. GraphSCNet (Qin et al., 2023b) constructs a deformation graph over the source point cloud and employs a graph-based local spatial consistency measure to aggregate and embed correspondences, thereby effectively filtering out outliers for robust non-rigid registration. Similarly, GCNNet (Zhu et al., 2022) utilizes GNN to enhance semantic encoding between local regions combined with geometry encodings constructed from the k-NN for multi-level feature extraction. We find that the DGCNN module is highly effective and works well as a point-cloud encoder in our network.

However, while these point- and graph-based methods perform increasingly well on surface-to-surface registration tasks, they usually do not model volume elements. Thus, they cannot be used directly to estimate the deformation of *internal* vessel-trees and tumors as is required in liver navigation systems. Furthermore, any information can only travel along the edges of the surface of the mesh or must jump over the large internal gap from one side of the mesh to the other. This may stop the method from learning a physically accurate soft-tissue deformation model – in which forces and displacement information must travel through the organ – and might make it difficult for the model to preserve volume. Our problem formulation thus requires internal volume points at every step. This allows information to travel through the organ and for the network to directly output a displacement field for the full volume. This also allows supervising the network training at every internal volume point if the ground truth displacement field is known.

Some registration methods ensure physical accuracy of displacement fields by using a physically informed loss term during training or by explicitly building biomechanical models into their networks. An example is the recently published BiomPINN-PBMs (Han et al., 2025) algorithm, which guides the registration method with an integrated finite element method. In contrast, PIVOTS learns any physical properties only implicitly from its training data.

The registration problem in liver navigation has also been approached using a 3D–2D formulation where the target are 2D laparoscopic images with hand-crafted or automatically detected features. Given that the segmented source liver model is textureless, these methods often rely on global geometrical features such as the liver ridge, falciform ligament and silhouette based on camera views.

Works such as Robu et al. (2018), Koo et al. (2017), Labrunie et al. (2024) manage to extract corresponding landmarks on 2D images and minimize the discrepancies between 3D and 2D landmarks, achieving rigid or non-rigid registration. Alternatively, approaches like Collins et al. (2016) enrich the source model with real texture by first aligning intraoperative images with the 3D model, followed by deformable registration based on the deviation in textures and boundaries. The P2ILF challenge (Ali et al., 2025) benchmarks 3D–2D registration by providing a dataset with manually labeled camera poses, where camera pose optimization with differential rendering demonstrates superior performance. Additionally, Mhiri et al. (2025) proposed a learning-based method that combines 3D synthetic deformation and 2D landmarks to predict intraoperative deformation, yet only patient-specific. The Self-P2IR method (Zhou et al., 2025) takes this one step further and trains in a self-supervised fashion on real data. While much progress has been made in 3D–2D registration in recent years, current methods are not yet robust enough to work reliably during surgery. A major of the 3D–2D formulation is that a single view often does not capture enough of the organ. Espinel et al. (2021) tries to solve this issue by using multiple views, but this may become difficult when the organ deforms between views. In practice, 3D–2D methods depend strongly on a well-chosen camera position which shows large parts of the liver. However, commonly used landmarks for 3D–2D alignment (ridge line, ligaments, liver contour) will quickly go out of view as soon as the camera is moved closer to the liver — as is usually done as soon as liver resection starts. Thus, 3D–2D registration methods may mainly be useful for initial alignment during the early phases of the surgery. Additionally, the methods may suffer from ambiguity issues inherent to the under-constrained formulation of a 3D–2D registration problem. The use of 3D intraoperative point clouds may circumvent these problems, because it is naturally suited for SLAM and Structure-from-motion methods which can be used to create an up-to-date model of the intraoperative scene as the registration target. Due to these considerations, we have decided to build PIVOTS as a 3D–3D volume-to-surface registration network.

We note that some of the methods mentioned in this section were not built specifically for liver registration, and they have to be re-trained or re-parameterized to work in this context. However, no available real or phantom dataset contains enough registration pairs for such training, since ground truth registration data with information about the deformation of internal structures is very difficult and time-consuming to obtain. Phantom datasets provide the highest number of samples, with a prominent example being the 112 registration pairs in the SparseDataChallenge dataset (Heiselman et al., 2024). The 3D-IRCADb-02 dataset contains a single in-vivo sample of liver deformation due to breathing motion (IRCAD, 2022). Intraoperatively, the DePOLL porcine dataset with 12 deformed states of one organ (Modrzejewski et al., 2019) and a clinical dataset with four patients (Rabbani et al., 2022) are available. Most of these datasets have only sparse or no landmarks available, which make it difficult to train methods since a dense displacement field is needed for training the volume-to-surface registration task.

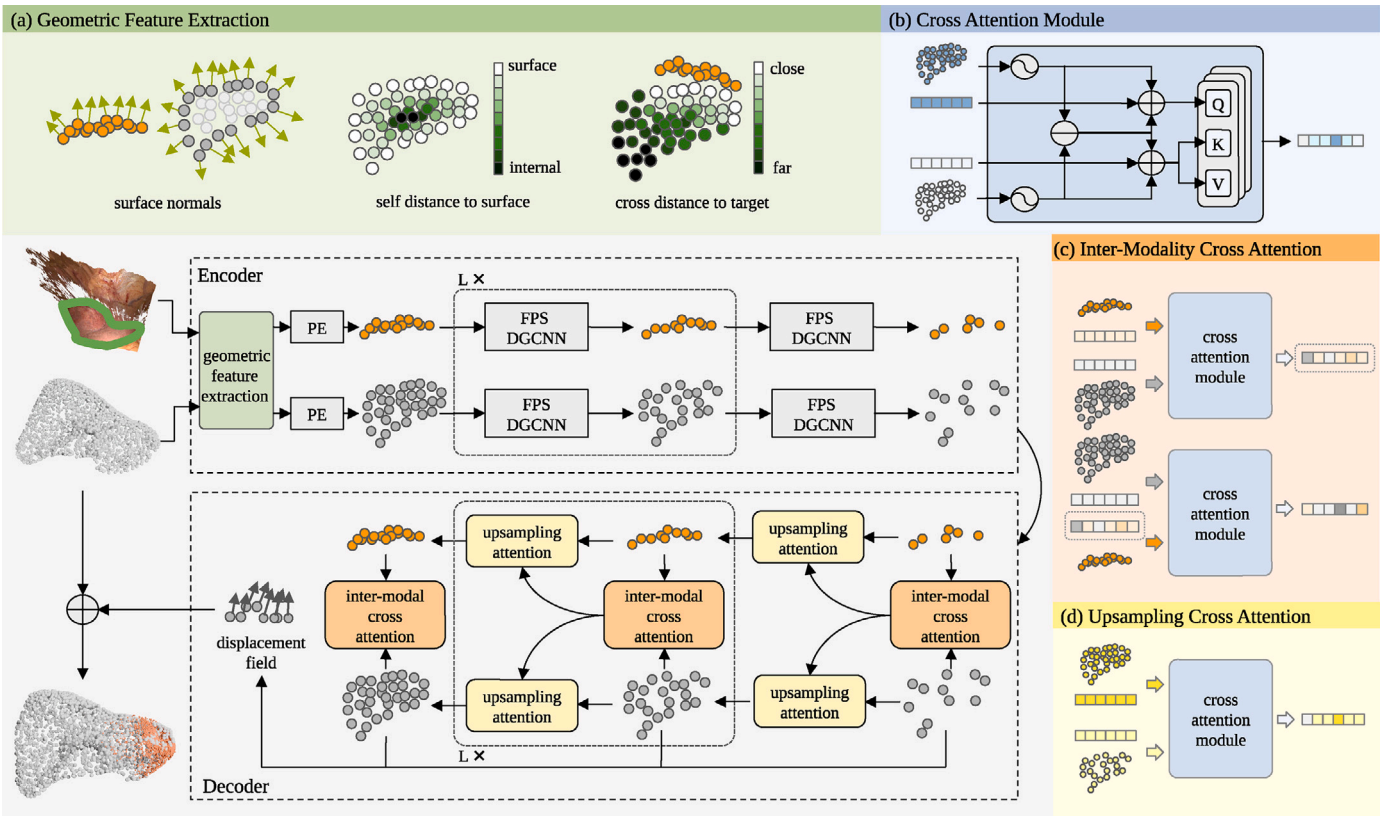


Fig. 2. Architecture of the proposed registration network. (a) For the intraoperative target points as well as the surface points of the preoperative point cloud, we encode the surface normals (see Section 3.2). For the preoperative source points we additionally encode the smallest distance to the mesh surface as well as the distance to the nearest intraoperative surface point. (b) The cross attention module based on our relative point attention is a core component of the network (described in Section 3.3.1). (c) Building upon this attention module, the Inter-Modality Cross Attention combines information from pre- and intraoperative point clouds (see Section 3.3.2). (d) The point attention module is also used for the Upsampling Cross Attention module, which propagates features from lower resolutions of the pre- and intraoperative points to higher resolutions (see Section 3.3.3). The final network architecture (left) partially resembles a U-Net, with information first being condensed to a lower resolution and then iteratively refined back to the original resolution.

It has been shown that registration networks can successfully transfer from synthetic training data to applications in the real domain Pfeiffer et al. (2019, 2020), Ali et al. (2025), Mendizabal et al. (2019). The resulting models tend to be fast at inference time (Pfeiffer et al., 2019; Mendizabal et al., 2019), and can be used for a volume-to-surface registration task (Pfeiffer et al., 2020). To fully leverage ground-truth synthetic training data, we generate a very large, novel synthetic training dataset with varying geometries, large deformations, controllable amounts of surface visibility and strong surface noise. We then build a novel neural network architecture which is specifically designed to handle the 3D–3D volume-to-surface task in liver registration.

3. Methods

Given a preoperative liver volume from MRI/CT, represented by point cloud $\mathbf{V} = \{\mathbf{v}_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}$, and an intraoperative partial liver surface reconstructed from laparoscopic video, noted as point cloud $\mathbf{S} = \{\mathbf{s}_j \in \mathbb{R}^3 \mid j = 1, \dots, M\}$, a non-rigid 3D–3D registration algorithm aims to find a transformation $\mathcal{T} : \mathbb{R}^3 \to \mathbb{R}^3$ that deforms the source $\mathbf{V}$ to align with the fixed target $\mathbf{S}$. The deformed $\tilde{\mathbf{V}} = \mathcal{T}(\mathbf{V})$ should reflect the current status of the intraoperative liver. In our case, this transformation $\mathcal{T}$ can be represented by a 3D displacement field $\boldsymbol{\Phi} \in \mathbb{R}^{N\times 3}$, i.e. $\mathcal{T}(\mathbf{V}) = \mathbf{V} + \boldsymbol{\Phi}$. In this formulation of the problem, the displacement field contains a value for each surface point *and* internal point in $\mathbf{V}$ and thus represents the full volume displacement. Our goal is to build a deep network **PIVOTS** which can predict $\boldsymbol{\Phi}$ for all points in $\mathbf{V}$ directly from the geometry $\mathbf{V}$ and $\mathbf{S}$, i.e.

$$\boldsymbol{\Phi} = \text{PIVOTS}(\mathbf{V}, \mathbf{S}). \tag{1}$$

During computation, the points in $\mathbf{V}$ and $\mathbf{S}$ are not moved. After the network has computed $\boldsymbol{\Phi}$, this displacement can be applied to the points in $\mathbf{V}$ to align V to the surface $\mathbf{S}$.

Our method adopts a dual-stream encoder–decoder architecture. At the encoder branch, we use farthest point sampling (FPS) and DGCNN (Wang et al., 2019) to iteratively downsample the preoperative and intraoperative point clouds and propagate features at each level. The decoder uses these features to predict the displacement field at each level, using cross-attention to merge information from $\mathbf{V}$ and $\mathbf{S}$ and propagate the merged information between layers. The architecture overview is shown in Fig. 2 and code is available at https://github.com/pengliu-nct/PIVOTS. The following will describe all components of the architecture in detail.

3.1. Point cloud preprocessor

The original preoperative and intraoperative point clouds are first enhanced by preprocessing steps before being fed into the registration network for deformation prediction. This serves as a standardization scheme for data from various sources and creates additional geometric information which the network can act upon.

First, both pre- and intraoperative point clouds are re-sampled to have a common resolution of roughly 5 mm, and several pre-computed features are extracted to help the network understand the geometry of inputs. PIVOTS requires internal points in the preoperative data, to allow information to travel through the volume and to predict a displacement on. For the synthetic training data, an internal point is placed at every vertex of the tetrahedral volume elements. In the

case of other data sources, such as surfaces obtained from real patient data, such internal mesh elements may not exist. In this case, we distribute random internal points throughout the preoperative volume (at a similar density as the mesh elements in the training data). Then, we calculate various point features: *Point distance field,* F_d: Inspired by V2S-Net, for each internal point of $\mathbf{V}$, we compute its distance to the closest point on the surface of $\mathbf{V}$. Unlike V2S-Net, we additionally compute cross-modality closest distances: For each point in $\mathbf{S}$ the smallest distance to the surface of $\mathbf{V}$ is computed and similarly for each point in the preoperative volume $\mathbf{V}$, we compute the distance to the closest point in intraoperative surface S. *Surface normals,* F_n: for the surface points in $\mathbf{V}$ and $\mathbf{S}$, surface normals are computed to help the network identify the orientation of surfaces. *Sinusoidal positional encoding*: for pre- and intraoperative point clouds, sinusoidal positional encodings are computed as a fast and effective way to capture higher-frequency variations. We use 8 frequencies with the mappings: $\xi : \mathbb{R}^3 \rightarrow \mathbb{R}^{6\omega}$, e.g., positional encoding of preoperative points is $\xi_v = [sin(2^{\omega} v_i), cos(2^{\omega} v_i)]$, where $\omega \in (2^{-1}, 2, 4, 8, 16, 32,)$ is the list of encoding frequencies.

Moreover, to allow batch-processing, both pre- and intraoperative point clouds are further set to exactly n points by either subsampling randomly or by adding dummy points that are way outside the space of interest and ignored during farthest point sampling (FPS).

The final inputs to the network are: $\mathbf{V}_r = \{P_v, \xi_v, F_v\}$ and $\mathbf{S}_r = \{P_s, \xi_s, F_s\}$, where $P_v, P_s \in \mathbb{R}^{n\times 3}$, $\xi_v, \xi_s \in \mathbb{R}^{n\times(6\omega)}$ and $F_v, F_s \in \mathbb{R}^{n\times 5}$ (see Fig. 2(a)). We store all distances and locations in meters.

3.2. Encoder: Geometric feature extraction

While downsampling significantly reduces computational cost, it results in sparse representations where fine-grained local geometry is less explicit compared to the original volumes. To address this, we adopt the widely used DGCNN (Wang et al., 2019) to extract and propagate robust local geometry features between L layers in different resolutions obtained by farthest point sampling (FPS): $(P^{\ell+1}, \xi^{\ell+1}, F^{\ell+1}) = \Pi_{fps}(P^{\ell}, \xi^{\ell}, F^{\ell}, N^{\ell})$, where Π_{fps} refers to mapping of FPS function, N^{ℓ} is the number of points at ℓ-th layer (see a in Fig. 2). Features of next layer $F_v^{\ell+1}, F_s^{\ell+1}$ are updated using two-stream edge convolution for V_r and S_r separately due to the fact that the geometry information of pre- and intraoperative data is naturally different:

$$\tilde{F}^{\ell+1} = \max_{j\in\mathcal{K}(i)} \mathrm{ReLu}(\mathrm{MLP}(F_j^{\ell} - F_i^{\ell+1}, F_i^{\ell+1})) \tag{2}$$

where F^{ℓ} is features of either preoperative or intraoperative point clouds, $\mathcal{K}$ represents the K nearest neighbors from layer $\ell+1$ of point i in layer $\ell+1$, and features are aggregated by the multi-layer perception MLP, followed by ReLu and maxpool layers. $\tilde{F}^{\ell+1}$ is the aggregated features for the $(\ell+1)$-th layer.

The combination of FPS with DGCNN in the encoder allows fast downsampling while ensuring that the whole point cloud is covered despite subsequent local kernels. We decided against global operations on the full point cloud due to the effectiveness of the paradigm to re-use the same kernel with a fixed-size in CNNs.

3.3. Decoder: Displacement field prediction via point attention network

Pre- and intraoperative liver point clouds are orderless and can vary largely in density, shape, and appearance. Thus, it is substantial to design a mechanism that allows features to flow and propagate between the two point cloud pairs in the same or different resolutions. The attention mechanism demonstrates promising performance for point cloud analysis due to its invariance to point order and its ability to enable dynamic information interaction between two arbitrary point clouds. These make the attention mechanism a suitable and efficient analyzer in dealing with pre- and intraoperative point cloud pairs. Therefore, we proposed a novel *Displacement field predictor* consisting of *Inter-Modality Cross* and *Upsampling Cross Attention* modules using *Relative Point Attention* as the core, for pre- and intraoperative point cloud feature interaction and propagation. They are introduced in Sections 3.3.2 and 3.3.3.

3.3.1. Relative point attention module

We propose a *relative point attention* (see b in Fig. 2) for a cross attention module which is used in multiple parts of our decoder to compute new features of a *query* point $P_Q \in \mathbb{R}^{N\times 3}$ from its local neighborhood in the same – or a different – point cloud. We employ vector attention starting with finding the k nearest neighbors $P_K \in \mathbb{R}^{N\times k\times 3}$ of P_Q, which act as the *key* and *value* points. To each feature vector ($F_Q \in \mathbb{R}^{N\times f}$ and $F_{K,k} \in \mathbb{R}^{N\times k\times f}$), we concatenate the point's coordinates both in absolute form and relative to P_Q and P_K, as well as positional encodings ξ (sine and cosines at different frequencies) thereof. We then calculate the query Q, key K and value V vectors as:

$$Q = W_Q(\xi_Q \oplus (\xi_Q \ominus \xi_K) \oplus F_Q) \tag{3}$$

$$K = W_K(\xi_Q \oplus (\xi_Q \ominus \xi_K) \oplus F_{K,k}) \tag{4}$$

$$V = W_V(\xi_Q \oplus (\xi_Q \ominus \xi_K) \oplus F_{K,k}) \tag{5}$$

where $W_q, W_k, W_v \in \mathbb{R}^{c\times f}$ are linear projection matrices, resulting in $Q \in \mathbb{R}^{c\times N}$ and $K, V \in \mathbb{R}^{c\times N\times k}$, $\xi \in \mathbb{R}^{6\omega}$ is sinusoidal positional encoding, here we also include the positional encoding of the N query points to their k neighbors. F_Q is the point cloud features, and $F_{K,k}$ is the features of the k nearest neighbors of query points. $\oplus$ is a concatenation operation and $\ominus$ stands for the calculation of relative point positional encoding between P_Q and the k neighbors in P_K.

Then, attention weights are calculated using *local additive attention*:

$$att(Q, K) = Softmax\,(\mathrm{MLP}(\tanh(Q + K))) \tag{6}$$

where $att \in \mathbb{R}^{d\times N\times k}$ is the attention score of Q on K. For each key point, the weighted average of the value features V is computed, resulting in the new features $\tilde{F}_Q \in \mathbb{R}^{d\times N}$ for the query point Q. The query feature F_Q is then the combination of multi-heads of attention modules:

$$\tilde{F}_Q \leftarrow \mathrm{MultiHead}(Q, K, V) = \mathrm{MLP}(cat[att_i(Q, K) \cdot V]) \tag{7}$$

where $\tilde{F}_Q \in \mathbb{R}^{N\times d'}$ is the updated query features. In practice, we set the number of attention heads to four. Depending on whether the query and value points are taken from the same or different point clouds, this module can be used for self- or cross-attention.

3.3.2. Inter-modality cross attention: S2V & V2S

To predict an accurate alignment between the preoperative and intraoperative point clouds, we must allow information to flow between the two clouds. Therefore, based on the proposed local *relative point attention* module, we assemble an *inter-Modality cross attention* module (see c in Fig. 2) between preoperative volume points $\mathbf{V}_r^{\ell} = (P_v^{\ell}, \xi_v^{\ell}, F_v^{\ell})$ and intraoperative surface points $\mathbf{S}_r^{\ell} = (P_s^{\ell}, \xi_s^{\ell}, F_s^{\ell})$ at each resolution ℓ. Because the *relative point attention* modules attend to the *k nearest neighbors in the value point cloud* (independent of their actual distance to the query point), using it for cross attention can allow information to quickly hop larger spatial gaps, such as those that can occur between two strongly deformed meshes.

Our *inter-modality cross attention* module consists of two cross attention steps. Intuitively, because we aim to calculate the displacement for the preoperative points given the intraoperative surface points (which act as fixed target points), a cross attention module which lets the preoperative points $\mathbf{V}_r^{\ell}$ attend to the intraoperative surface points $\mathbf{S}_r^{\ell}$ (V2S) is a straightforward addition. However, we find that the network benefits from first enriching the intraoperative surface points with information about their closest preoperative point, and so our module first lets $\mathbf{S}_r^{\ell}$ to attend to $\mathbf{V}_r^{\ell}$ (S2V):

$$\tilde{F}_s^{\ell} \leftarrow \mathrm{MultiHead}(Q_s^{\ell}, K_v^{\ell}, V_v^{\ell}) \tag{8}$$

where Q_s^ℓ is the linearly transformed relative encodings ξ_v, ξ_s and surface features F_s^ℓ, and K_v^ℓ, V_v^ℓ come from relative encodings ξ_v, ξ_s and volume features F_v^ℓ (see Eq. (5)).

With the updated intraoperative surface features being informative of preoperative data, the second attention module is applied, aiming to convey such knowledge to the preoperative point (V2S):

$$\tilde{F}_v^\ell \leftarrow \text{MultiHead}(Q_v^\ell, \tilde{K}_s^\ell, \tilde{V}_s^\ell) \tag{9}$$

where $\tilde{K}_s^\ell, \tilde{V}_s^\ell$ stem from the updated surface features $\tilde{F}_s^\ell$, and $\tilde{F}_v^\ell$ is the updated preoperative features, being aware of the intraoperative points at the current resolution. At this point, information can flow back and forth between the point clouds, and can propagate between neighborhoods, and $\tilde{F}_v^\ell$ should contain all necessary information about deformation. Hence, we let the network predict a (downsampled) version of the displacement field Φ^ℓ for every level ℓ:

$$\Phi^\ell = \text{MLP}(\text{ReLu}(\tilde{F}_v^\ell)) \tag{10}$$

Such a design (**S** attends to **V**, then **V** attends to **S**) enables the information flow in the presence of larger deformation between pre- and intraoperative points, as all points from both sides gain the opportunity to update their geometric knowledge regarding the k nearest neighbors in their counterparts. We hypothesize that this bilateral instead of unilateral search reduces ambiguity by mutually aligning the feature space. In practice, this approach often focuses the receptive field of preoperative points onto larger intraoperative surface regions (see the Supplementary Material, Sec. 2). We demonstrate the effectiveness of the module for the full registration task in the ablation study in Section 4.7.

3.3.3. Upsampling cross attention

The abstract features calculated at lower resolutions must later be incorporated back into the features for the higher resolutions in the decoder. For this, *Upsampling Cross Attention* modules (see d in Fig. 2) are built to propagate information from lower layers to higher layers. Taking preoperative volume as an example:

$$\check{F}_v^{\ell-1} \leftarrow \text{MultiHead}(\tilde{Q}_v^{\ell-1}, \tilde{K}_v^\ell, \tilde{V}_v^\ell) \tag{11}$$

where $\tilde{K}_v^\ell, \tilde{V}_v^\ell$ are the projected encoding and feature vectors at the lth layer in lower resolution (see Eq. (8)), and $\tilde{Q}_v^{\ell-1}$ (the feature of (ℓ-1)-th layer in higher resolution, see Eq. (9)) is used as the query, which will be updated by cross attention to be $\check{F}_v^{\ell-1}$. Similar operations are applied between intraoperative points. In the overall flow, the deformation and upsampling cross attention are applied alternatingly according to the chosen number of layers L. $\check{F}_v^{\ell-1}$ and $\check{F}_s^{\ell-1}$ serve as input features to Eqs. (8) and (9) on the next higher resolution: $\check{F}_s^{\ell-1} \rightarrow Q_s^{\ell-1}$, $\check{F}_v^{\ell-1} \rightarrow K_v^{\ell-1}, V_v^{\ell-1}, Q_v^{\ell-1}$.

3.4. Supervision

The displacement field Φ_ℓ is predicted at each layer ℓ, so the total outputs of an L layers network is:

$$\Phi = \{\Phi^\ell \in \mathbb{R}^{N\times 3} | \ell = 1, \dots, L\} \tag{12}$$

where Φ^L is the predicted displacement field at the highest resolution.

We train the network in a supervised learning fashion on a synthetic dataset generated using a simulation pipeline (see Section 3.5). For each training sample, the ground truth displacement field Φ_{gt} is known, so the predicted displacement fields can be compared at each respective resolution using a mean squared error (MSE). The total loss function $\mathcal{L}$ is computed as a weighted sum of the losses:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_i^N \|\Phi_{gt,i} - \Phi_i\|_2^2 \tag{13}$$

$$\mathcal{L} = \sum_\ell w_\ell \mathcal{L}_{MSE,\ell} \tag{14}$$

where $\Phi_{gt,i}, \Phi_i \in \mathbb{R}^3$ are the predicted and ground truth displacement vectors of a preoperative point. The lower-level losses guide the network to output better features for displacement estimation at high resolution layers, speeding up convergence, while higher-level losses ensure the fidelity and accuracy of the final prediction (Pfeiffer et al., 2020).

3.5. Synthetic training data generation

The synthetic training data generation builds upon an in-house simulation pipeline. Our training data is designed towards providing the network with two core abilities: (i) generalizing to new patients without retraining and (ii) performing robustly in challenging intraoperative conditions.

Generalization is fostered with respect to organ shape, material and boundary conditions. Diverse, randomly generated organ-like shapes support the adaptation to the natural variation of patients' liver shapes. We try to address the unknown highly variable tissue properties by covering the full range of tissue material parameters. In order to approximate the unknown intraoperative forces which are extremely difficult to measure, like the effects of the pneumoperitoneum, we create a variety of scenarios with differences in magnitude as well as position and surface area which the initial forces act on.

The challenging intraoperative conditions are abstracted as partial views and unavoidable point cloud reconstruction noise. Since it is difficult to encompassingly characterize the different noise properties, we emulate it with a mixture of two noise types in varying strength such that the network can learn to mitigate their effects.

The following sections provide more detail about the training data generation. It starts with the scene setup, implementing the shape variability and placement of boundary conditions, followed by simulating the material's behavior under the specified forces, extracting a part of the deformed surface and adding noise.

3.5.1. Shape variability

A scene comprises of a deformable object and its surroundings. The liver is represented by a random organ-like shape, created according to a previously published algorithm (Pfeiffer, 2022): A sphere is subjected to several extrusions and subtractions, followed by smoothing. Convex parts reflect the anterior side while concave parts mimic the concavities around the hepatic portal on the posterior side. This results in a shape of roughly the dimensions $d_x = d_y = d_z \sim U(100, 300)$ mm, an example is illustrated in Fig. 3. Fig. 1 in the Supplementary Material shows the variability of the created shapes alongside two exemplary real livers from the AMOS dataset (Ji et al., 2022), which display the natural variation. In order to create space within the synthetic abdominal cavity, 0 to 3 more organs of the size $d_x = d_y = d_z \sim U(50, 200)$ mm are created additionally. A convex hull spans around these objects with a minimum distance according to $\sim U(0, 50)$ mm, expanded up to 20 mm of Perlin noise along its outward normal, corresponding to a simplified abdominal wall. One or two random paths of up to 100 mm along the deformable organ and extruded towards the abdominal wall imitate ligaments. Finally, a fixed boundary is placed by expanding from a random face of the deformable organ until 2 to 5% of the surface is covered. A comprehensive list of parameters for the scene generation can be found in Table 1 in the Supplementary Material.

3.5.2. Deformation and dense displacement field

Subsequently, the created scene objects are equipped with physical properties. Trying to mimic high deformations of liver tissue implies that high strains are to be expected, the deformable organ is therefore modeled as a hyperelastic and nonlinear material with a Young's modulus $E \sim U(3, 30)$ kPa and a Poisson's ratio $\nu \sim U(0.45, 0.48)$ informed by literature values (Glozman and Azhari, 2010; Mattei and Ahluwalia, 2016; Marchesseau et al., 2017; Estermann et al., 2020; Lemine et al., 2024), sampled independently in order to cover the

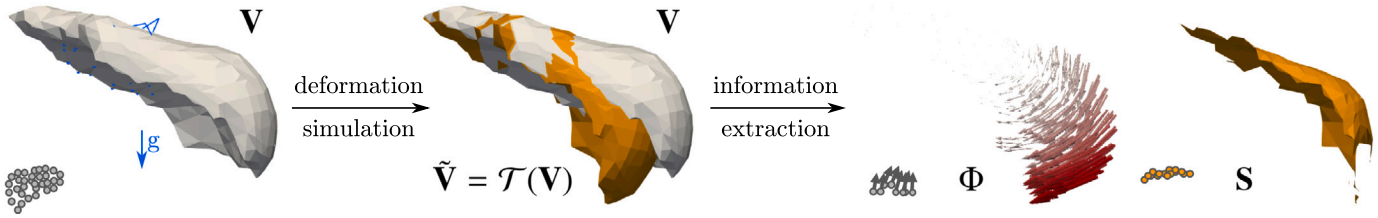


Fig. 3. Training data generation. From left to right: Random boundary conditions (blue), including gravity, fixed boundaries (dots) and linear spring boundaries (lines), are placed around a randomly generated organ-like shape with random material properties, representing the preoperative volume $\mathbf{V}$. Simulation under these conditions leads to a deformed configuration $\tilde{\mathbf{V}}$ (orange) which models the deformation encountered during surgery, e.g. due to tool-tissue interaction. Finally, the ground truth dense displacement field $\boldsymbol{\Phi}$ (magnitude-colored glyphs) and a partial surface with added noise (orange), mimicking an intraoperative point cloud $\mathbf{S}$ reconstructed from the laparoscopic video, can be extracted as training inputs.

parameter space when creating a sufficiently high number of samples. For computer-assisted surgery, material behavior at low strain rates is relevant, however, in order to present the network with as many configurations as possible, the parameter range is chosen very generously. Hyperelasticity is realized using the isotropic compressible Neohookean formulation by Bonet and Wood (2008) in SOFA's implementation:

$$\Psi = \frac{\mu}{2}(I_1 - 3) - \mu \ln J + \frac{\lambda}{2}(\ln J)^2$$
$$\text{with} \quad \mu = \frac{E}{2(1+\nu)}, \ \lambda = \frac{E}{3(1-2\nu)},$$

where Ψ denotes the strain energy, I_1 the first invariant of the right Cauchy–Green tensor and J the Jacobian. For simplification, the material is assumed to be homogeneous and isotropic (Marchesseau et al., 2017) with a mass density of $\rho \sim U(1.05, 1.09)\,\mathrm{g/cm^3}$ (Lemine et al., 2024). Since we do not consider time series but snapshots of deformation states, we also exclude viscosity and porosity which are mostly considered for time and frequency domain responses.

The previously created surface is processed into a volumetric tetrahedral mesh using GMSH (Geuzaine and Remacle, 2009) (mesh generation) and the built-in Netgen interface (optimization) (Schöberl et al., 2024), symbolizing the still undeformed, preoperative volume $\mathbf{V}$. A maximum mesh element size of 1 cm presents a trade-off between accuracy and runtime.

In order to create diverse deformation scenarios we employ three kinds of boundary conditions: Gravity, fixed boundaries and linear elastic springs (blue elements on the left of Fig. 3). The positions of the latter two have been determined in the scene generation step. The fixed boundaries are enforced as a Dirichlet zero-displacement boundary constraint on the deformable organ's surface. The spring boundaries between the abdominal wall and the deformable organ's surface can only move on the organ side. Defined in a state outside of equilibrium, they serve two purposes: In a stretched state, they will pull on the deformable organ, reflecting anatomical ligaments of different strength. In a compressed state, they represent a very simplified version of other organs pushing against the liver. The springs' initial state is determined by their current length $l = l_0 + \delta l$, the rest length factor $c_0 = \frac{l_0}{l} \sim U(0.9, 1.1)$ to show the deviation from the relaxed state, and the stiffness $k \sim U(100, 300)\frac{\mathrm{N}}{\mathrm{m}}, k = \text{const}$. Each spring i in a 10 cm ligament object lig is assigned the same k_{lig} and $c_{0,lig}$, with different $\delta l_{lig,i}$ based on the slightly different l due to the initial placement. Each spring's initial force corresponds to $F_{lig,i} = -k_{lig} \cdot \delta l_{lig,i}$. It follows that $c_0 < 1$ corresponds to a stretched spring (pulling), $c_0 > 1$ to a compressed string (pushing).

The simulation is set up using the SOFA framework (Faure et al., 2012) with an implicit Euler time stepping scheme (step size 50 ms) and a Conjugate Gradient linear system solver. The simulation is run until equilibrium is reached, but in order to obtain higher deformations for training the network, we export the timestep with the highest deformation instead of the final state, therefore including transient dynamic states of the tissue in the training. The difference between the undeformed state $\mathbf{V}$ (gray) and deformed state $\tilde{\mathbf{V}} = \mathcal{T}(\mathbf{V}) = \mathbf{V} +$ $\boldsymbol{\Phi}$ (orange) is illustrated in Fig. 3, second from the left. Since point correspondences between the undeformed and the deformed mesh stay intact, the ground truth dense displacement field $\boldsymbol{\Phi}$ is easily computed, as shown in red color and magnitude scaled glyphs in Fig. 3, second from the right.

3.5.3. Intraoperative surfaces

In order to recreate the partial and noisy information of the intraoperatively reconstructed point cloud, a part of the surface of the deformed mesh is extracted and perturbed at random (Pfeiffer, 2022). Starting from a randomly selected mesh vertex as a reference, a randomly weighted sum of the geodesic distance, the angle between the normal vectors and a Poisson noise value is used to select the points to remain. After increasing the number of points on this partial surface by subdivision, an offset is sampled for each direction independently from separate Perlin noise distributions sharing an amplitude A_P, using the point positions as an input $\varepsilon_{x_i} \sim A_P \cdot \mathrm{Perlin}(x_1 \cdot f_{x_i}, x_2 \cdot f_{x_i}, x_3 \cdot f_{x_i}), i \in 1, 2, 3$. Additionally, a Gaussian noise contribution is sampled from $\mathcal{N}(x_i, \sigma)$ for $i \in \{1, 2, 3\}$ and added to the coordinates, followed by a randomized sparsification step to remove few larger areas rather than many small ones. The result is an emulated noisy partial surface $\mathbf{S}$, visualized in orange in Fig. 3 (right).

In this manner, after discarding invalid configurations, we create 193,563 samples for training. A full overview of the parameters used in the data generation pipeline can be found in Table 1 in the Supplementary Material.

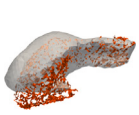
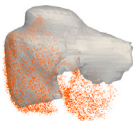
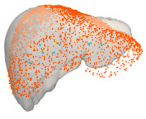
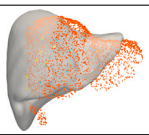
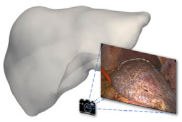
3.6. Implementation and training

PIVOTS is implemented in Pytorch and trained on an Nvidia A100 GPU with batch size 20 with a one-cycle learning-rate scheduler (Smith and Topin, 2019) and AdamW optimizer. We use Optuna (Akiba et al., 2019) to fine-tune our hyperparameters with a smaller subsection of the training dataset and then train with the full training dataset and the most promising parameters for 100 epochs. We use $k = 35$ for all neighborhoods in the point attention modules, $e = 29$ for all attention embedding vectors, and all input point clouds are standardized to contain $n = 2500$ points. We build PIVOTS with six down- and upsampling layers ($L = 6$) with $[8, 35, 92, 144, 239, 321]$ points for each downsampled layer respectively, and point feature vectors of length [200, 150, 110, 80, 60, 50] during the downsampling and upsampling stages of the network. While 2500 points may appear few considering the liver size, a reasonably high L was required in order to get adequate performance, limiting the choice of n in light of the speed requirement.

4. Evaluation

We evaluate the proposed network on five datasets, including two synthetic datasets, a phantom dataset, and two real liver datasets, which are introduced in Section 4.1. Extensive experiments are conducted to assess the generalizability to various data sources (see Section 4.3 and Section 4.3.1), robustness against high deformation (see

Table 1
Comparison of evaluation datasets.

	Synthetic	AMOS	Phantom	HHLBM	Laparoscopic	P2I-LReg
Visualization preoperative intraoperative landmarks						
size	3008	690 (×18 noise)	6 (×20 targets)	38 (×20 targets)	6	92
shape	synthetic	real	phantom	real	real	real
deformation	synthetic	synthetic	real (ti)	real (bm, p)	real (bm, hb, pp)	real (bm, hb, pp)
metrics	MED	MED	TRE	TRE	–	CD/Dice
PRD (mm)	32.14	32.04	17.35	25.78	–	–

Abbreviations: bm - breathing motion, hb - heartbeat, p - positioning, pp - pneumoperitoneum, ti - tool interaction.

Section 4.4), noise (see Section 4.5) and varying amounts of surface visibility (see Section 4.6). The ablation study in Section 4.7 highlights the contribution of the different components in our architecture. In section Section 4.8 we explore the physical plausiblility of the network's output. Visual examples of the registration quality in comparison to other methods are shown in Section 4.9.

4.1. Datasets

The test datasets were chosen to expose different aspects of PIVOTS:

- the **synthetic** test set presents the smallest step to something unseen,
- with **AMOS** we assess generalization to controlled setups of liver geometries and varying levels of known intraoperative challenges,
- the **Phantom** dataset allows to explore performance under local tool-induced deformations,
- **HHLBM** introduces more global in-vivo deformations due to pose and breathing motion,
- and the **P2I-LReg** and **Laparoscopic** datasets challenge it with real deformation and intraoperative point clouds with real characteristics.

Table 1 shows an overview of the dataset characteristics.

The **Synthetic** evaluation dataset is generated using the same simulation pipeline described in Section 3.5 with the same hyperparameters, resulting in a similar deformation and noise distribution, but the samples are unseen during training. This dataset contains 3008 samples.

The **AMOS** evaluation dataset is a semi-synthetic dataset, where the liver models are reconstructed from labeled real abdominal organ CT scans from the AMOS dataset (Ji et al., 2022) (for details on the data we refer the reader to the original publication) and the deformation is synthetically produced by the same simulation pipeline. 360 livers are selected, which corresponds to the number of scans with publicly available labels, and 3 different sets of boundary conditions are applied to each of the real liver models. Randomly selected noise levels are applied to the random partial surfaces as described above.

For a thorough evaluation of our method, we additionally create a noise level benchmark on this data, $\text{AMOS}_{\text{noise}}$. For this, in order to imitate a real intraoperative surface more closely, we change the partial surface extraction method of the data generation pipeline to mimic a laparoscopic camera view, and fix a set of constant noise amplitude combinations. The camera view is created by selecting a random position for the camera that remains outside the mesh with a distance of 50 to 200 mm throughout the simulation, setting its focus to a random point of the point cloud and extracting the whole part of the surface that lies within the field of view. The selected fixed noise levels include Perlin noise ranging from $A_P = 0$ mm to $A_P = 15$ mm with steps of 3 mm and Gaussian noise from $\sigma = 0$ mm to $\sigma = 5$ mm with steps of 2.5 mm. Fig. 5 illustrates their effects on the intraoperative surface. This results in 683 samples, each one containing 19 synthetic intraoperative partial surfaces: a random surface with randomly selected noise and one camera view surface for each of the 18 fixed noise levels. An overview of the parameters used to generate the AMOS datasets can be found in Table 2 in the Supplementary Material.

The **Phantom** dataset is acquired from a Kyoto Kagaku abdominal phantom (Zhylka et al., 2024) for controlled experiments on real data with complete knowledge of internal deformation.

A series of six deformations were applied to the liver using a fake MRI-compatible intraoperative US probe. For each compressed state, an MRI scan was performed (plus one for the uncompressed state). In each case a T2 image was acquired with TE=75 ms, TR=5.2 s, voxel size=0.88 × 0.88 × 3 mm^3 on Philips Ingenia 3.0 T. The phantom was filled with water to avoid artifacts during scanning. To keep the US probe in a fixed position during MRI acquisition, a specialized construction was designed consisting of a rail, fake probe and a holder (see Fig. 5 in the Supplementary Material). All scans were co-registered rigidly via the basin. We marked the same eight landmarks at clearly identifiable positions inside the liver scans in 3D Slicer, using the vessel bifurcations as main guidance. After segmentation of the MRI scans, a partial view of the anterior side of the liver was extracted for each of the six compressed states by simulating the field-of-view of a simple camera. We increased this field of view iteratively (resulting in 20 partial surfaces for each compressed state) and used the uncompressed volume as the preoperative. As PIVOTS requires internal volume points to compute a displacement, additional random support points are added in the preoperative volume as described in the data preprocessing in Section 3.1. The network uses them to compute a displacement field which is later interpolated onto the preoperative landmarks. Afterwards, the landmarks can be displaced according to these interpolated values and the TRE can be calculated on the landmarks.

The **Healthy Human Liver Breathing Motion (HHLBM)** dataset contains real in vivo liver deformation of healthy subjects during breathing without external disturbance. It enables the evaluation of the proposed network on real human liver deformation. For each healthy subject, data were collected by an MRI scan under three controlled variables: respiratory state (inhaled or exhaled), body posture (supine or prone), and acquisition day (first or second), resulting in a total of eight scan combinations per subject. Images were acquired using a VIBE sequence (T1) with multiple echo acquisitions according to the Dixon method on a Siemens Biograph mMr (3.0 T), voxel size 1.39 × 1.39 × 1.5 mm^3. Three annotators are involved in the segmentation process, which consist of one biomedical engineer and two experienced computer scientists. To facilitate the time-consuming manual segmentation process, nnUnetV2 (Isensee et al., 2021) trained on five public liver datasets (LiverHccSeg, Atlas, AMOS, TotalSegmentator MRI, and CHAOS) generates pre-segmentation masks, where only T1-Weighted scans are included to correspond to the focus of the HHLBM dataset. The masks are further refined by the annotators to be the ground truth liver volume. 20 partial surfaces are then extracted from the volumes using the same strategy as for the phantom dataset to reflect intraoperative observations. For landmarks independent of the surface, since no dense correspondence information is available, two

annotators search for clearly visible vessel bifurcations and mark their positions in the eight scans. The mean position of the vessel bifurcations is considered the ground truth position, and the mean deviation between landmarks from the two annotators is 2.76 mm. Scans with low quality are excluded from the experiments, such as strong artifacts due to the subject moving during the scanning process. For evaluating registration methods, one scan from each patient, the inhaled state in the prone position on the first acquisition day, was designated as the "preoperative" volume $\mathbf{V}$. The partial surfaces from the other scans of the same patient were chosen as the "intraoperative" partial surfaces $\mathbf{S}$, the displacement vectors of the corresponding internal landmarks act as a surrogate for the ground-truth internal displacement field. Additional support points are added in the same way as in the phantom dataset. In total, data from seven subjects was included in the dataset. Overall, this resulted in 38 valid samples for this dataset. The data was collected after obtaining ethics approval (TU Dresden, approval number BO-EK-174042023) and with informed consent from every subject.

The **P2I-LReg** (Zhou et al., 2025) is a public in vivo registration dataset containing laparoscopic liver videos of 21 patients. It comprises 346 keyframes annotated with liver segmentation masks and calibrated camera intrinsics, 92 of which are used for qualitative evaluation.

The **Laparoscopic** Liver dataset comprises six videos collected from real surgery, i.e., six intraoperative surfaces. The preoperative liver models are segmented from CT scans, and the intraoperative 3D scene is reconstructed from a stereoscopic video recorded from an endoscope during the exploration phase at the beginning of a surgery. Since fully automatic 3D reconstruction methods still struggle with the difficult intropeative lighting conditions, texture-less surfaces, breathing and heartbeat motion as well as limited visibility, a semi-automatic approach was used to create the intraoperative target surfaces. First, the FoundationStereo method (Wen et al., 2025) was used for stereo-reconstruction. From the result, overlapping point clouds were hand-picked and aligned using manually placed point correspondences in the CloudCompare Software. Afterwards, points not belonging to the liver surface were discarded. The result is a point cloud containing a partial view of the intraoperative liver, containing noise from the stereo reconstruction as well as alignment errors due to deformation (See Fig. 11). Patient consent was obtained and the data was collected after approval by the ethics committee (TU Dresden, BO-EK-137042018).

4.2. Evaluation metrics

Based on different characteristics of the test sets, we adopt different evaluation metrics. For synthetic and AMOS datasets, ground truth displacement fields are available. Therefore, the mean Euclidean distances (MED) are calculated as the dense registration error metric:

$$MED = \frac{1}{N} \sum \left\| \Phi_{pred} - \Phi_{gt} \right\|_2 \tag{15}$$

In contrast, phantom and HHLBM datasets, no dense correspondence information is available, so we employ target registration errors (TRE) to evaluate the model performance:

$$TRE = \frac{1}{N} \sum \left\| \mathcal{T}(L_P) - L_I \right\|_2 \tag{16}$$

where L_p and L_I are the sparse landmarks in pre- and intraoperative liver volume, $\mathcal{T}(L_P)$ denotes the deformed preoperative landmarks by the estimated displacement field. Both pre- and intraoperative landmarks are not included in the input points but only used for evaluation. TRE presents the deviation between the predicted position and actual positions of intraoperative landmarks.

We also defined the initial displacement for all datasets as PRD (Pre-Registration Displacement). For the synthetic and AMOS datasets with ground truth known, PRD is defined as:

$$PRD = \frac{1}{N} \sum \| \Phi_{gt} \|_2 \tag{17}$$

And for Phantom and HHLBM datasets with sparse landmarks:

$$PRD = \frac{1}{N} \sum \| L_I - L_P \|_2 \tag{18}$$

For the P2I-LReg dataset used in qualitative evaluation, we calculate the Chamfer Distance (CD) shipped with the dataset for benchmarking:

$$CD = \frac{1}{N_1} \sum_{x \in \tilde{\mathbf{V}}} \min_{y \in \mathbf{S}} \|x - y\|_2 + \frac{1}{N_2} \sum_{y \in \mathbf{S}} \min_{x \in \mathbf{S}} \|y - x\|_2 \tag{19}$$

where $\tilde{\mathbf{V}}$ and $\mathbf{S}$ are the deform liver point cloud and the reconstructed intraoperative point cloud. To measure the similarity between annotated liver masks and projected deformed models in 2D space, the Dice coefficient is calculated:

$$\text{Dice}(\tilde{\mathbf{V}}, \mathcal{M}) = \frac{2 \cdot |\Pi(\tilde{\mathbf{V}}) \cap \mathcal{M}|}{|\Pi(\tilde{\mathbf{V}})| + |\mathcal{M}|} \tag{20}$$

where Π is the project operation and $\mathcal{M}$ is the segmented liver mask on the 2D image.

In order to assess the physical consistency of the predicted displacement field, we employ the Jacobian determinant J that has been introduced as a registration metric by Heiselman et al. (2024):

$$J = \det \mathbf{J} = \det(\nabla \mathbf{u} + \mathbf{I}), \tag{21}$$

with the displacement $\mathbf{u}$ and identity $\mathbf{I}$. It can be interpreted as a ratio of local volume change between the undeformed and deformed state. For synthetic datasets, the reference displacement field is used to calculate $J_{\text{diff},t} = J_{\text{pred},t} - J_{\text{gt},t}$, where t indexes the tetrahedral mesh elements. $J_{\text{diff},t} < 0$ indicates that an element has been compressed more than in the reference data, $J_{\text{diff},t} > 0$ shows exaggerated extension. To avoid error cancellation between elements, we summarize the values over each mesh in a mean absolute error:

$$\bar{J}_{\text{diff}} = \frac{1}{T} \sum_t |J_{\text{diff},t}| \tag{22}$$

where T is the number of tetrahedra.

4.3. Comparison with state-of-the-art methods

Baselines: We select widely adopted registration methods from the general computer vision and liver registration communities as our baseline methods. They can be roughly categorized into rigid methods and non-rigid methods. ICP (Besl and McKay, 1992) serves as a rigid, non-learning baseline. GCNNet (Zhu et al., 2022) and LiverMatch (Z. Yang et al., 2023) are rigid learning-based approaches. Classical non-rigid, non-learning methods include CPD (Myronenko and Song, 2010) and HMM-CPD (Min et al., 2019, 2020), while PBSM (Suwelack et al., 2014) and BCF-FEM (Yang et al., 2024) represent biomechanics-based non-rigid methods. Learning-based non-rigid methods include Lepard (Li and Harada, 2022a), Robust-DefReg (Monji-Azad et al., 2024), LiverMatch+NDP (Li and Harada, 2022b), C2P-Net (Liu et al., 2024b), and PIVOTS. The hyperparameters of all methods are carefully optimized, and all learning-based methods are trained using the same synthetic training dataset as the proposed method. The initial rigid misalignment (PRD) of the inputs stems from the respective datasets and nonrigid methods receive no additional alignment preprocessing. For PBSM and BCF-FEM, the preoperative surfaces must be meshed with volume elements, which we achieve with the GMSH (Geuzaine and Remacle, 2009) software. Because the speed of these methods strongly depends on the number of elements in the meshes, we downsample surfaces to a maximum of roughly 5000 points.

The results of all methods are listed in Table 2, which presents a quantitative comparison of registration performance among a diverse set of baseline methods across four datasets: Synthetic, AMOS, Phantom, and HHLBM. As evaluation metric we use MED computed based on ground truth displacements (Synthetic and AMOS) and TRE calculated from sparse landmarks (Phantom and HHLBM), reported as mean ±

Table 2
Registration performance comparison between baseline methods across different datasets. Bold face represents the best mean registration error on each dataset, and underlined numbers stand for the second best. If a method increases the registration error compared to the dataset's PRD, we consider it failed, indicated by a gray font. The first three methods are rigid, the other eleven methods compute a non-rigid registration. The symbols further specify their categories as follows: □ Surface-to-Surface, ■ Volume-to-Surface, ○ learning-based, × not learning-based.

Method	Registration Error (mm, Mean ± SD)			
	Synthetic Samples: 3008 PRD: 32.14	**AMOS** Samples: 683 PRD: 32.04	**Phantom** Samples: 6 PRD: 17.35	**HHLBM** Samples: 38 PRD: 25.78
□ × ICP	73.37 ± 23.77	78.52 ± 11.84	25.46 ± 13.99	43.61 ± 52.26
□ ○ GCNNet	23.16 ± 13.30	26.03 ± 11.67	32.08 ± 10.10	31.20 ± 10.83
□ ○ LiverMatch	13.94 ± 13.08	14.57 ± 12.43	12.33 ± 3.95	13.45 ± 2.59
□ × CPD	48.68 ± 16.50	46.74 ± 10.84	36.41 ± 4.61	41.14 ± 6.45
□ × HMM-CPD	25.53 ± 15.70	26.31 ± 15.59	29.44 ± 3.09	39.40 ± 9.43
□ × HMM-CPD (aniso)	23.47 ± 15.27	26.08 ± 18.68	28.02 ± 3.02	25.29 ± 5.23
□ ○ Robust-DefReg	26.40 ± 16.10	29.07 ± 15.44	35.02 ± 1.45	42.87 ± 3.20
□ ○ LiverMatch + NDP	19.60 ± 18.08	16.60 ± 12.13	20.12 ± 2.92	22.34 ± 3.53
□ ○ Lepard	19.92 ± 14.73	21.81 ± 14.07	16.34 ± 7.35	19.28 ± 6.00
□ ○ C2P-Net	19.15 ± 13.73	19.17 ± 12.57	17.00 ± 1.07	22.21 ± 4.22
■ × PBSM	27.59 ± 29.02	30.28 ± 26.81	10.03 ± 6.02	8.59 ± 5.39
■ × BCF-FEM	21.90 ± 30.47	18.11 ± 22.49	9.02 ± 6.34	6.72 ± 2.88
■ ○ V2S-Net	6.36 ± 6.48	6.25 ± 6.04	9.03 ± 4.06	7.21 ± 1.86
■ ○ PIVOTS (ours)	**3.11 ± 3.38**	**4.06 ± 3.35**	**5.96 ± 1.47**	**6.38 ± 2.56**

standard deviation in millimeters. For reference, the pre-registration displacement (PRD) is included to indicate the initial misalignment magnitude in each dataset prior to registration.

Overall, the proposed method achieves the lowest mean registration error and Mean + SD across all datasets, outperforming both rigid and non-rigid, traditional and learning-based approaches. Notably, the performance gain is most pronounced in the Synthetic and AMOS datasets, where the average error (3.11 mm and 4.06 mm) is reduced by a substantial margin compared to the next-best methods. In particular, besides LiverMatch, V2S-Net, and our method, the other methods obtain much worse registration errors around 20 mm and much higher, although learning-based methods are trained on the same dataset and iterative methods are fine-tuned as much as possible. For example, ICP (73.37 mm and 78.52 mm), CPD (48.68 mm and 46.74 mm), and Robust-DefReg (26.40 mm and 29.07 mm) appear to struggle consistently. These results indicate that volume-to-surface registration is a challenging task and requires dedicated designs to tackle the high inconsistency between source and target and accurately infer the movement of internal structures. The comparison between the Synthetic and AMOS datasets also illustrates PIVOTS' capability to generalize from random shapes during training to real liver shapes at inference time, with an increase in MED from 3.11 ± 3.38 mm to 4.06 ± 3.35 mm and RMSE from 4.32 ± 4.35 mm to 4.96 ± 4.13 mm, respectively.

Furthermore, on datasets with real deformation, namely the Phantom and HHLBM datasets, our method similarly outperforms others with a registration error of 5.96 mm and 6.38 mm, respectively, suggesting higher robustness and ability of the method in dealing with real noise and deformation compared to baseline methods, e.g., V2S-Net (9.03 mm and 7.21 mm). Beyond absolute performance, the consistency and stability of each method across datasets with various noise and deformation is also critical and significant in practice to ensure safe navigation even when conditions in clinical application may differ. For instance, some methods such as Robust-DefReg and NDP show substantial variation in performance depending on the dataset, suggesting sensitivity to specific deformation characteristics or anatomical context. In contrast, our method demonstrates a stable and low error across all domains, reflecting effective generalization from synthetic to real anatomical deformation and across different evaluation protocols (dense field vs. sparse landmarks).

In summary, several insights can be obtained:

Volume-to-Surface Registration: For datasets with real deformation, only volume-to-surface baseline methods achieve a registration error lower than 1 cm, e.g. PBSM and V2S-Net, indicating the criticality of volume-to-surface methods in order to obtain high accuracy of the prediction for internal structures. Furthermore, deep learning methods originally designed for surface-to-surface registration, e.g. Lepard and C2P-Net, require dedicated modification to adapt to such a different task in liver navigation.

Necessity of Non-rigid Registration: Although LiverMatch achieves registration errors lower than the majority of non-rigid methods across all four datasets, none of the mean TREs undercut the 1 cm threshold, which is only achieved by four of the nonrigid methods.

Synthetic and Real Domain Gap: Our method obtained a mean registration error lower than 4 mm on datasets with synthetic deformation. However, the performance degrades on the other two real datasets by around 2 mm. This degradation hints at a noticeable gap between synthetic and real datasets, which could be attributed to different materials or properties of the added noise. An alternative influence could be the nature of the deformation: In the synthetic data, the organ can be stretched or folded. The main deformation in the Phantom dataset stems from a US probe causing a local perturbation and the in vivo deformation of the HHLBM dataset is limited by the undisturbed relation of the liver to the surrounding organs and tissues. Especially in later stages of a surgery, one would expect higher deformations than in both of those cases. Therefore, while the performance differences can inform room for improvement in the training data generation, an intraoperative dataset would be highly valuable in assessing the generalization ability to true intraoperative deformations in more detail.

Inference Speed: The methods based on biomechanical simulation (PBSM and BCF-FEM) obtain very good registration results on some samples, especially when the initial displacement (PRD) is low. However, they are orders of magnitudes slower and often take more than half a minute for a single sample as opposed to PIVOTS with 230 ms. We further investigate the influence of our choice of the number of input points n in the point cloud preprocessor on the inference speed and registration error in an ablation study in Section 4.7.

Table 3
Mean Dice scores, mean CD, and mean inference time on P2I-LReg.

Method	Dice (%) ↑	CD (mm) ↓	time (s) ↓	#samples
BCF-FEM	68.30 ± 8.81	4.88 ± 1.56	160.26 s	67
V2S-Net	72.02 ± 8.16	4.67 ± 1.48	0.13 s	92
Self-P2IR	72.99 ± 8.37	3.94 ± 1.37	3.97 s	92
Ours	73.69 ± 7.45	2.39 ± 1.52	0.23 s	92

4.3.1. Comparison on P2I-LReg

We employ the recent intraoperative dataset P2I-LReg in order to evaluate PIVOTS on a public real laparoscopic dataset quantitatively. As baselines we compare against V2S-Net and BCF-FEM as the best nonrigid learnable and traditional baseline methods from the previous

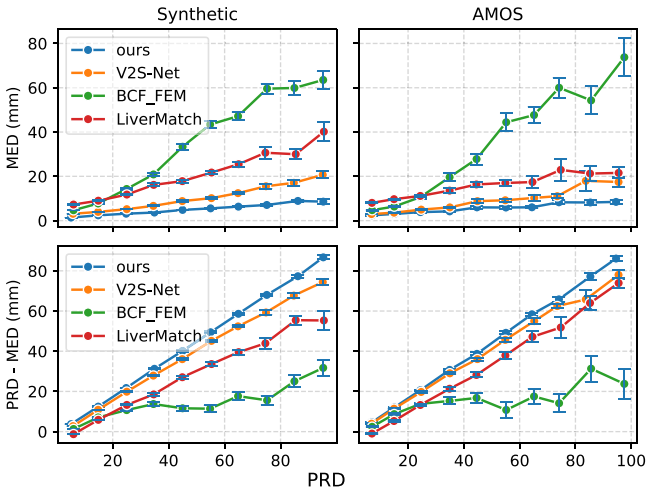


Fig. 4. Comparison of methods on various deformation levels (PRD) on the synthetic (left) and AMOS (right) datasets. Top row: Mean Euclidean distance (MED) versus PRD, demonstrating that our method exhibits the greatest robustness to initial misalignment. Bottom row: relative registration error (MED - PRD) versus PRD, showing that our approach achieves the most pronounced error reduction.

Table 4
Pearson and Spearson metrics for Synthetic and AMOS datasets.

Method	Synthetic		AMOS	
	Pearson	Spearman	Pearson	Spearman
BCF-FEM	0.437	0.557	0.359	0.561
LiverMatch	0.864	0.904	0.924	0.919
V2S-Net	0.982	0.985	0.981	0.985
ours	0.996	0.995	0.993	0.991

experiments, as well as against Self-P2IR (□ ○), which was published along with the dataset and is the current state-of-the-art in 3D/2D non-rigid registration. We report the mean Dice score and mean Chamfer Distance. Self-P2IR results are obtained using the official evaluation script, while PIVOTS, V2S-Net and BCF-FEM results are computed on rigidly aligned inputs using the Lepard model trained on P2I-LReg.

Quantitative results are summarized in Table 3: PIVOTS achieves the highest mean Dice score and the lowest mean Chamfer Distance. We chose to evaluate the small nominal differences using a paired sign test to show significance. Given the small sample size, a visible deviation from normality and asymmetry in the pairwise differences, this allows an assessment of directional differences without distributional assumptions. Bonferroni correction is applied to control for multiple comparisons. The difference in Chamfer Distance to the next best methods, Self-P2IR and V2S-Net, is statistically significant at a significance level of $\alpha = 0.05$ with p-values of 1.851×10^{-8} and 5.771×10^{-18}, respectively, and Self-P2IR also significantly improves over V2S-Net ($p=4.652\times10^{-9}$). The significance of differences in the Dice metric is strongly rejected with a consistent $p > 2.7$. Due to failures on several test cases arising from preprocessing requirements, a direct comparison with BCF-FEM should be interpreted with caution. However, a clear difference in performance is still evident. This shows that PIVOTS successfully generalizes to this intraoperative dataset while running around 16 times faster than Self-P2IR.

4.4. Deformation level experiment

Since the liver undergoes large deformation during surgery, e.g., due to tool manipulation, it is critical for the registration to reliably capture little as well as highly deformed states. We use the synthetic and AMOS test sets to quantify the influence of different deformation levels on the registration performance of representative methods from each group in Table 2 as visualized in Fig. 4, Also, we present the Pearson correlation coefficients in Table 4.

From Fig. 4, mean registration errors of our methods increase steadily with a relatively low slope, ending with errors below 10 mm for the largest deformation group (> 80 mm). In comparison, the other baseline methods present a more inclined trend with the increase of the deformation level, especially BCF-FEM, the mean error reaching more than 20 mm for a deformation larger than 40 mm. As a rigid registration method, LiverMatch surpasses BCF-FEM on both test sets and achieves similar errors as V2S-Net when presented with large deformations in the AMOS test set, with an MED of 20 mm.

The relative registration error is computed as $\Delta_{RE} = PRD - MED$, showing how much the registration error increases or decreases compared to the initial error prior to registration. Table 4 exhibits different dependencies of relative registration errors on initial error using Pearson correlation coefficients r and Spearman correlation coefficients ρ. Our method achieves near-perfect linear ($r = 0.996$) and rank-order ($\rho = 0.995$) correlations, indicating that larger initial misalignments are almost corrected by proportionally larger amounts. V2S-Net follows closely ($r = 0.981$, $\rho = 0.9850$), likewise demonstrating a very tight coupling between case difficulty and error reduction. In contrast, BCF-FEM shows a substantially weaker relationship ($r = 0.437$, $\rho = 0.557$), reflecting a more equilibrated strategy that yields more uniform improvements regardless of initial error magnitude. As a rigid alignment method, LiverMatch occupies an intermediate position ($r = 0.8643$, $\rho = 0.9043$), balancing targeted correction of challenging cases with non-negligible gains for easier ones. These results suggest that our method and V2S-Net yield proportional registration error correction, BCF-FEM presents global consistency regardless of the initial misalignment, and LiverMatch shows a comparable capability in error reduction. Overall, our methods exhibit superior performance regarding dealing with both small and large deformations.

4.5. Noise level experiments

In 3D–3D liver registration scenarios, the reconstructed liver surfaces using SfM or SLAM from the laparoscopic videos may be corrupted due to various noise or artifacts, substantially degrading the registration accuracy. To quantify the impact of various noise levels, we generated the $AMOS_{noise}$ dataset (see Section 4.1) with noisy partial surfaces as illustrated in Fig. 5.

We selected three representative methods with low registration error in Table 2 for this experiment, V2S-Net, BCF-FEM and LiverMatch.

Fig. 6 illustrates all registration results, organized by noise combinations, where the horizontal axis represents Perlin noise and the vertical direction corresponds to Gaussian noise. Overall, PIVOTS maintains a consistently low registration error across all noise combinations, showing strong robustness against various types and amounts of noise. On the other hand, V2S-Net and BCF-FEM present noticeable and rapid degradation with more noise, although the baseline methods work reasonably well with low amounts of noise. In contrast, the errors of PIVOTS show a much steadier deteriorating trend. In particular, the impact of Perlin noise is remarkably larger than Gaussian noise. This finding aligns well with the visualization in Fig. 6, where Gaussian noise makes the partial surface more fuzzy while the basic geometry remains, while on the other hand, target partial surfaces present a bumpy appearance under higher Perlin noise, potentially modifying the geometry and complicating the correspondence exploration for registration methods. These trends are also visible in the qualitative results, illustrated for a subset of the noise levels in Fig. 7: PIVOTS yields stable results with increasing amounts of Gaussian noise and slightly rising errors with higher Perlin noise amplitudes, but the results are qualitatively promising.

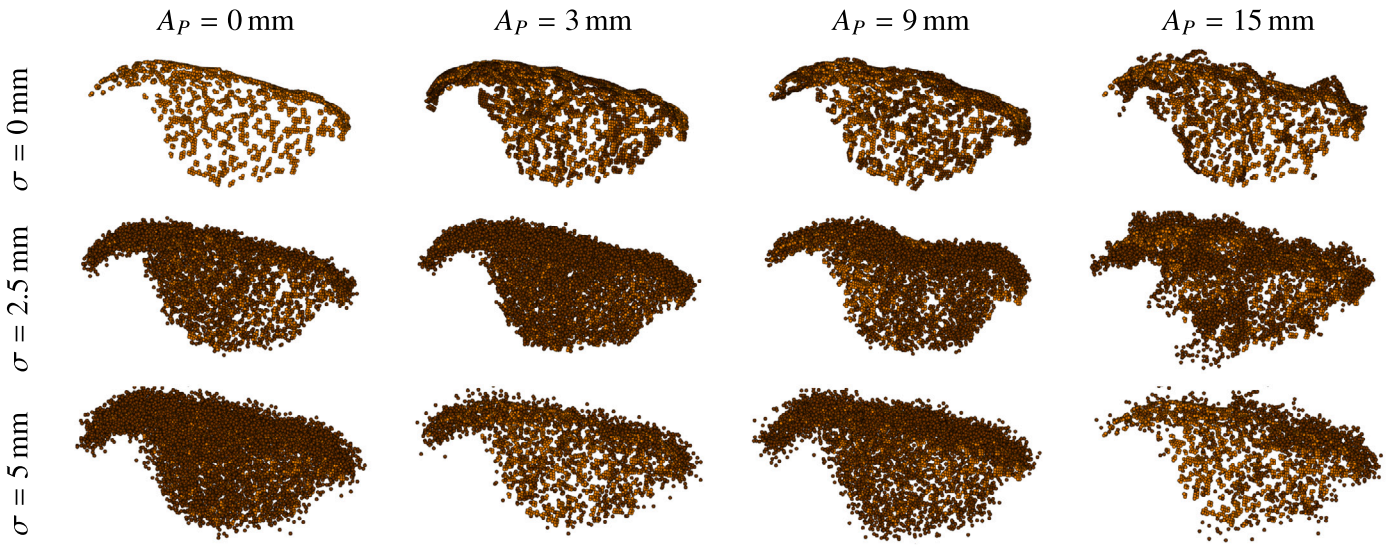


Fig. 5. Input for the noise level experiment. Example partial intraoperative surface **S** from the $\text{AMOS}_{\text{noise}}$ dataset (orange) with different combinations of Gaussian noise with deviation σ (rows) and Perlin noise with amplitude A_P (columns). Noise levels (brown) increase from left to right and from top to bottom. Gaussian noise blurs the intraoperative point cloud while Perlin noise causes nonuniform local distortions to the apparent surface shape. The randomized subsampling of the point cloud after adding noise additionally increases the difficulty.

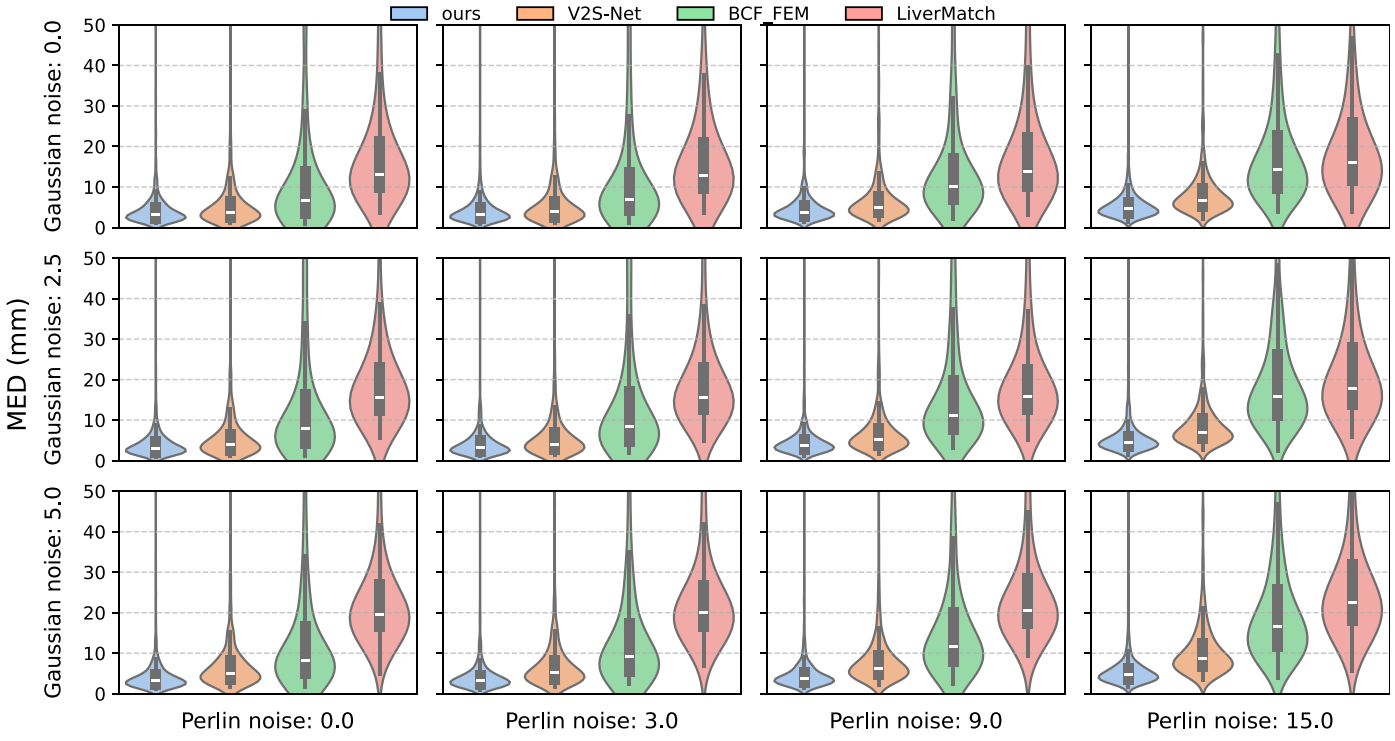


Fig. 6. Registration error distributions on various noise levels. Two types of noise are involved: Perlin noise (horizontal, range: $A_P \in \{0, 3, 6, 9, 12, 15\}$ mm) and Gaussian noise (vertical, range: $\sigma \in \{0, 2.5, 5.0\}$ mm). Each subplot exhibits the registration error distribution of four methods. Our method presents the most robustness against various amounts and types of noise.

Besides the mean error, the development of the error variability further underpins PIVOTS' strong noise handling capabilities. The top left plot (without noise) shows V2S-Net and PIVOTS clearly outperforming BCF-FEM and LiverMatch while on a similar level to each other. Despite BCF-FEM achieving a mean MED below 1 cm, the error distribution shows roughly a third of samples with an error above this value, making the method unreliable for a considerable subset of patients even in the ideal noise-free setting. With increasing noise levels, the error distribution spreads for all methods in general. However, PIVOTS still manages to contain the error for all but a small fraction of its predictions below the 1 cm threshold even at the highest noise level, giving it a strong edge over the V2S-Net baseline as well.

Additionally, this experiment allows the conclusion that PIVOTS is able to generalize between different surface extraction methods, as it was trained only on randomly extracted surfaces but $\text{AMOS}_{\text{noise}}$ contains camera-like partial surfaces, and the MED of this experiment remains comparable to the second column of Table 2.

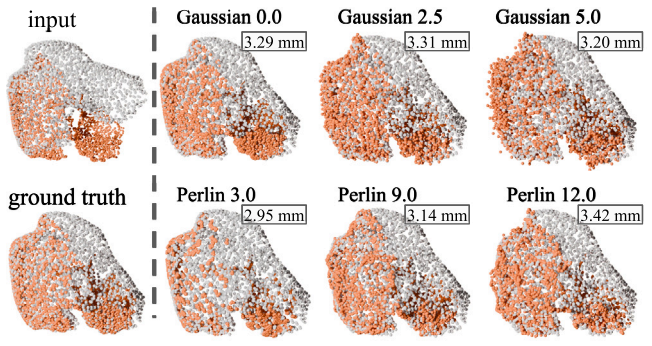


Fig. 7. Qualitative comparison between results of PIVOTS on intraoperative surfaces with various Gaussian noise (first row) and Perlin noise amplitudes (second row). The input and ground truth deformation are demonstrated on the left. PIVOTS exhibits strong robustness against various amounts of noise.

Table 5
TREs (mm) of visibility experiment on the Phantom and HHLBM datasets and average inference speed.

Dataset	Vis. Per. (%)	PBSM	BCF-FEM	V2S-Net	ours
Phantom	0–10	22.30 ± 6.83	10.22 ± 6.62	13.73 ± 4.88	**9.33 ± 4.23**
	10–20	14.84 ± 6.93	9.32 ± 6.06	9.88 ± 4.43	**6.71 ± 2.68**
	20–30	9.87 ± 5.17	8.89 ± 5.61	8.91 ± 4.18	**6.46 ± 1.47**
	30–40	10.18 ± 5.77	9.49 ± 6.57	9.36 ± 3.99	**6.35 ± 1.27**
HHLBM	0–10	20.56 ± 6.28	18.61 ± 5.67	**18.09 ± 7.29**	19.93 ± 9.92
	10–20	16.04 ± 5.78	13.03 ± 6.43	12.64 ± 6.78	**10.25 ± 3.63**
	20–30	11.01 ± 5.19	8.17 ± 3.72	8.06 ± 2.98	**6.95 ± 2.24**
	30–40	8.89 ± 3.79	7.49 ± 3.07	7.42 ± 2.86	**6.04 ± 2.09**
inference time		60.59 s	160.26 s	0.13 s	0.23 s

4.6. Intraoperative visibility experiments

Intraoperative liver surfaces are often only partially visible due to limited view angles and occlusion by instruments and adjacent organs, complicating the registration process. We carry out an experiment to examine the performance of the registration methods under various ranges of relative visibility, which is defined as $\alpha = \mathcal{A}(\mathbf{S})/\mathcal{A}(\partial(\mathbf{V}))$, where $\mathcal{A}(\cdot)$ is a function to calculate the area of a given surface, $\partial(\cdot)$ extracts the surface mesh from a volume, and the visibility $\alpha \in [0, 1]$. For this specific experiment we adopt the Phantom and HHLBM test sets. As described in Section 4.1, for each sample, both datasets contain 20 partial intraoperative surfaces which are extracted from the view of a virtual camera, observing the anterior side of the liver volumes and gradually moving away from the liver. The grouped results are presented in Table 5

For the Phantom dataset, at the most extreme occlusion (0–10% visibility), PBSM yields the largest TRE (22.30 ± 6.83 mm), whereas our method achieves the lowest TRE (9.33 ± 4.23 mm), improving upon V2S-Net by 32% and BCF-FEM by 9% of TRE. As the visibility increases, our method shows a declining trend and always obtains the best TREs around 6.5 mm with a notably lower variability, outperforming BCF-FEM and V2S-Net by $2-3$ mm. In comparison, the other methods present a fluctuating trend with the increase of visibility, stagnating around 9 mm TRE.

Regarding TREs on HHLBM dataset, under extreme occlusion (0–10%), V2S-Net and BCF-FEM outperform our method (18.09 mm and 18.61 mm vs. 19.93 mm for our method), suggesting that iterative optimization like BCF-FEM can assist exploration of a better solution and voxelized data like the inputs of V2S-Net favor sensing the ambient small geometry. In contrast, our method performs only one forward end-to-end inference and introduces no extra spatial information. Once over 10% visibility, the TRE of our method drops more quickly than the others and our method outperforms the baseline methods, with around 6 mm as the lowest error at the highest level of visibility. Overall, insights can be drawn from this experiment:

Table 6
Ablation study on the synthetic and partial $\text{AMOS}_{\text{noise}}$ ($A_P = 0$, $\sigma = 0$) datasets, MED (mm, Mean ± SD). All networks were trained on a third of the training data set, except for the last row which was copied from Table 2 for quick reference. The value for $\text{AMOS}_{\text{noise}}$ in the last row is from inference on all noise levels.

Architecture	Synthetic	$\text{AMOS}_{\text{noise}}$
shallower network ($L = 2$)	7.44 ± 6.40	8.67 ± 6.67
shallower network ($L = 4$)	5.09 ± 4.96	5.91 ± 5.16
w/o S2V&V2S cross attention	5.22 ± 4.74	5.69 ± 4.90
w/o pre-computed features	7.74 ± 6.91	8.10 ± 6.60
w/o relative coordinates in att.	4.75 ± 4.97	5.67 ± 4.41
full network	4.46 ± 4.25	4.93 ± 3.64
full network and full dataset	3.11 ± 3.38	4.06 ± 3.64

Sensitivity to visibility: All methods benefit from increased visibility (α), but PIVOTS exhibits the steepest decline in TREs, indicating superior exploitation of additional surface data.

Visibility Threshold: 10% visibility splits the performance of the methods on obtaining TRE lower than 1 cm. But above 10%, methods, especially PIVOTS, rapidly reduce TRE towards less than 7 mm. This indicates a critical visibility threshold that registration methods must aim to tackle for reliable accuracy in clinical setups.

4.7. Ablation studies

In this section, we explore alternative architectures to investigate the necessity of the components of our proposed method. To remain within a feasible time-limit, we train all models on a subset of roughly one-third of the full generated training set. The results are summarized in Table 6. Fig. 8 shows results for the synthetic test set grouped by different levels of deformation and visibility in order to visualize tendencies. Visualizations for the $\text{AMOS}_{\text{noise}}$ test set are attached in Supplementary Fig. 6.

Number of layers: We show that increasing network depth improves registration performance and enhances robustness to deformation and limited visibility. The diminishing returns on TRE and substantially higher training time place six layers as a practical trade-off for our setup.

Inter-Modality cross attention (S2V & V2S): We ablate the proposed inter-modality cross-attention module by replacing it with a simple implementation of point cross-attention (single flow from **V** to **S** without relative position and encodings). The module provides a clear advantage in challenging scenarios, behaving as intended under high deformation and extending to low-visibility settings. We attribute this to the ability of the bidirectional information exchange to cross large distances and cover wider surface areas. Notably, its removal yields performance comparable to reducing depth by two layers, indicating a substantial loss in expressivity. This suggests that inter-modality interaction and increased depth have partially similar effects in expanding the receptive field, although the former achieves this without additional layers or parameters.

Pre-computed Features: Removing the input features F_v, F_s has the second highest impact on performance and variability after the shallowest setup, with influence increasing under greater deformation. For reduced surface visibility their loss proves comparably detrimental to the loss of 4 layers of depth. The use of precomputed features, including the proposed cross-modality closest distances, appears to be highly effective for our architecture.

Relative coordinates and positional encoding: We ablate the Relative Point Attention module by using only semantic features for the cross attention. The explicit structural information in the original module appears to improve accuracy and increase robustness, which is especially prominent in challenging scenarios. It enables the source

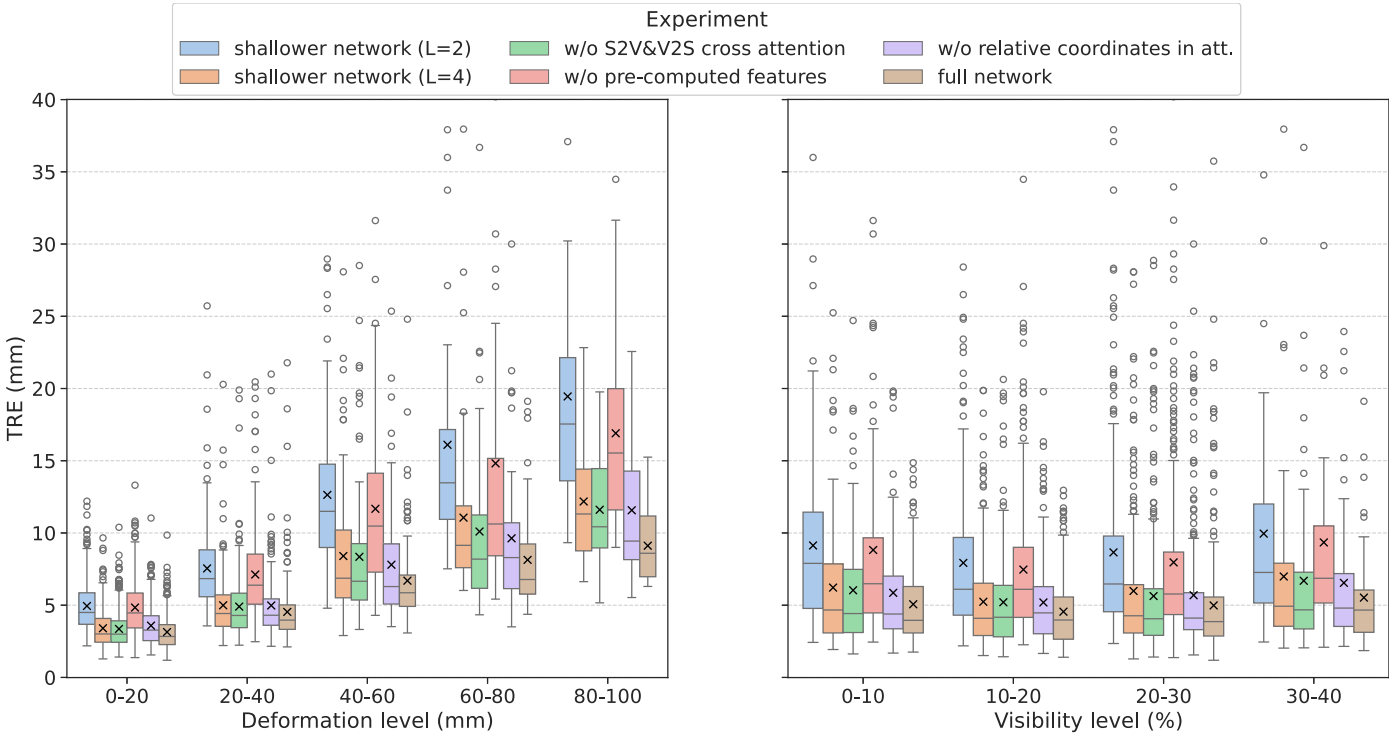


Fig. 8. Ablation results of PIVOTS on the synthetic test set, cropped to max. 40 mm TRE. We show the mean TRE per sample, with a stratification by amount of deformation on the left, and a distinction by amount of visible surface on the right. The mean per category is indicated by a cross. With higher deformation and lower visibility, the impact of various architectural components becomes more pronounced.

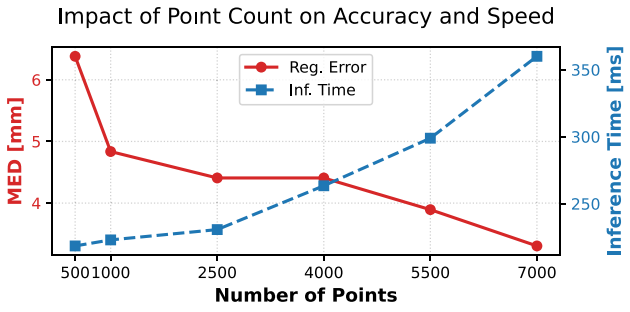


Fig. 9. Effects of varying the number of input points on inference time in relation to the mean registration error on the AMOS dataset. While the registration error decreases with a higher amount of points, inference time grows at an increasing rate.

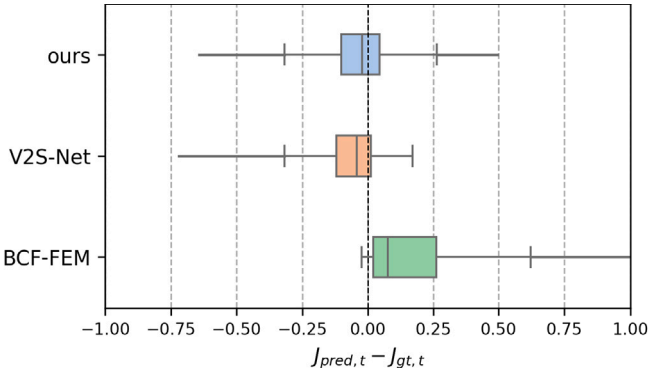


Fig. 10. Jacobian determinant error $J_{\text{diff},t}$ for inference of PIVOTS, V2S-Net and BCF-FEM on a subset of $\text{AMOS}_{\text{noise}}$ ($A_P = 0$, $\sigma = 0$), distribution on the level of tetrahedra. The distributions have been clipped at the 1% and 99% percentile for better visualization. The true maximum and minimum are reported in Table 7. Our method shows spurious compression ($J_{\text{diff},t} < 0$) as well as spurious extension ($J_{\text{diff},t} > 0$) while V2S-Net mostly overestimates compression and BCF-FEM almost exclusively overestimates in the tensile domain.

point to align with a target patch that is geometrically consistent with its local neighborhood, thereby acting as a soft spatial regularizer.

Robustness against noise: We find that robustness to noise is achieved across a range of architectural choices, with no single component clearly responsible, and emerges already at smaller training dataset sizes. A stratification plot is shown in Supplementary Fig. 7.

Inference speed and point density: In the baseline method comparison in Section 4.3 we note that they are considerably slower. For PIVOTS, inference speed as well as TRE depend on the number of input points n in the point cloud preprocessor. We further investigate this influence on the AMOS test dataset in Fig. 9. The MED decreases when the method has access to more points; logically, the denser the point cloud, the better the liver volume can be represented. This effect mainly shifts the location of the error distribution (see Fig. 8 in the Supplementary Material). However, with the increase of input point amounts, the mean inference time clearly increases. We adopted $n = 2500$ as a trade-off between registration accuracy and efficiency, but we note that PIVOTS with $n = 1000$ with an MED below 5 mm still outperforms the second-best V2S-Net with 6.25 mm in this test setting (AMOS column in Table 2).

4.8. Physical consistency

Fig. 10 shows the distribution of the Jacobian error on the level of the volume elements for inference of three methods on a noise-free subset of the $\text{AMOS}_{\text{noise}}$ dataset. Noticeably, their tendency is qualitatively different: Considering that the reference value of this

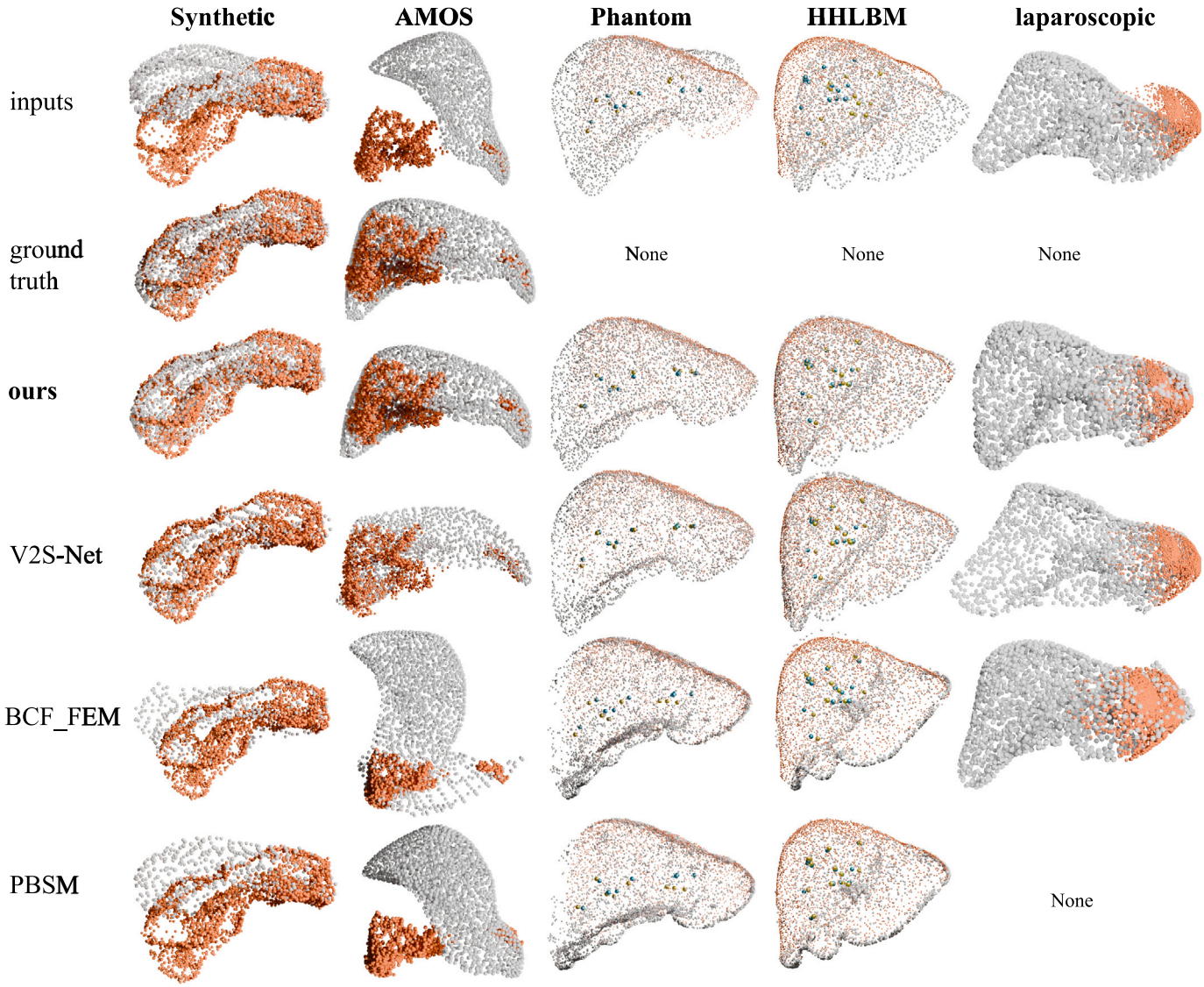


Fig. 11. Qualitative comparison on four evaluation datasets. Preoperative point clouds are depicted in gray •, and intraoperative points in orange •. For the Phantom and HHLBM datasets, preoperative landmarks are shown in yellow • and intraoperative ones in blue •. Ground truth information is only available for synthetic and AMOS datasets.

Table 7
Jacobian determinant errors.

Method	Min	$\bar{J}_{\text{diff}}$ (Mean + SD)	Max
ours	−28.79	0.12 ± 0.08	68.10
V2S-Net	−419.30	0.12 ± 0.11	423.00
BCF-FEM	−28.83	0.33 ± 0.69	52.66

metric in the simulation is very close to one, PIVOTS shows an almost balanced distribution around zero for the difference, indicating that it has learned to both extend and compress tissue. This is in line with its training on compressible material to reflect liver volume changes between preoperative acquisition day and the time of surgery (Leung et al., 1986b), however, this effect scales too strongly and does not appear spatially coherent for neighboring tetrahedra, not allowing deformations to propagate through the material. V2S-Net is biased towards material compression and has been shown to similarly exhibit localized centers of compression or extension (Heiselman et al., 2024). We attribute this weakness of the learning-based methods to the point-wise displacement predictions.

BCF-FEM consistently overestimates the local extension, producing a larger error in magnitude. This is likely due to the differences in material model, as BCF-FEM relies on a linear material which is only valid for small strains while the $\text{AMOS}_{\text{noise}}$ dataset was generated using a nonlinear material and explicitly provoking large deformations. Table 7 summarizes the absolute errors over the dataset.

In different ways, none of the methods yield a physically realistic result with respect to the local material deformation.

4.9. Qualitative results

We visualize registration results of various methods on the four 3D–3D test sets in Fig. 11. The first row shows the input point clouds to the method, where the difference of PRDs between the datasets is clearly visible, and the second row contains ground truth for the synthetic and AMOS datasets.

Synthetic data: For the synthetic and AMOS datasets, BCF-FEM and PBSM clearly struggle to accommodate large deformations: In the synthetic example, their output point clouds remain nearly in the initial pose, indicating a failure to drive the source towards the target geometry. On AMOS, PBSM leaves the left lobe essentially uncorrected,

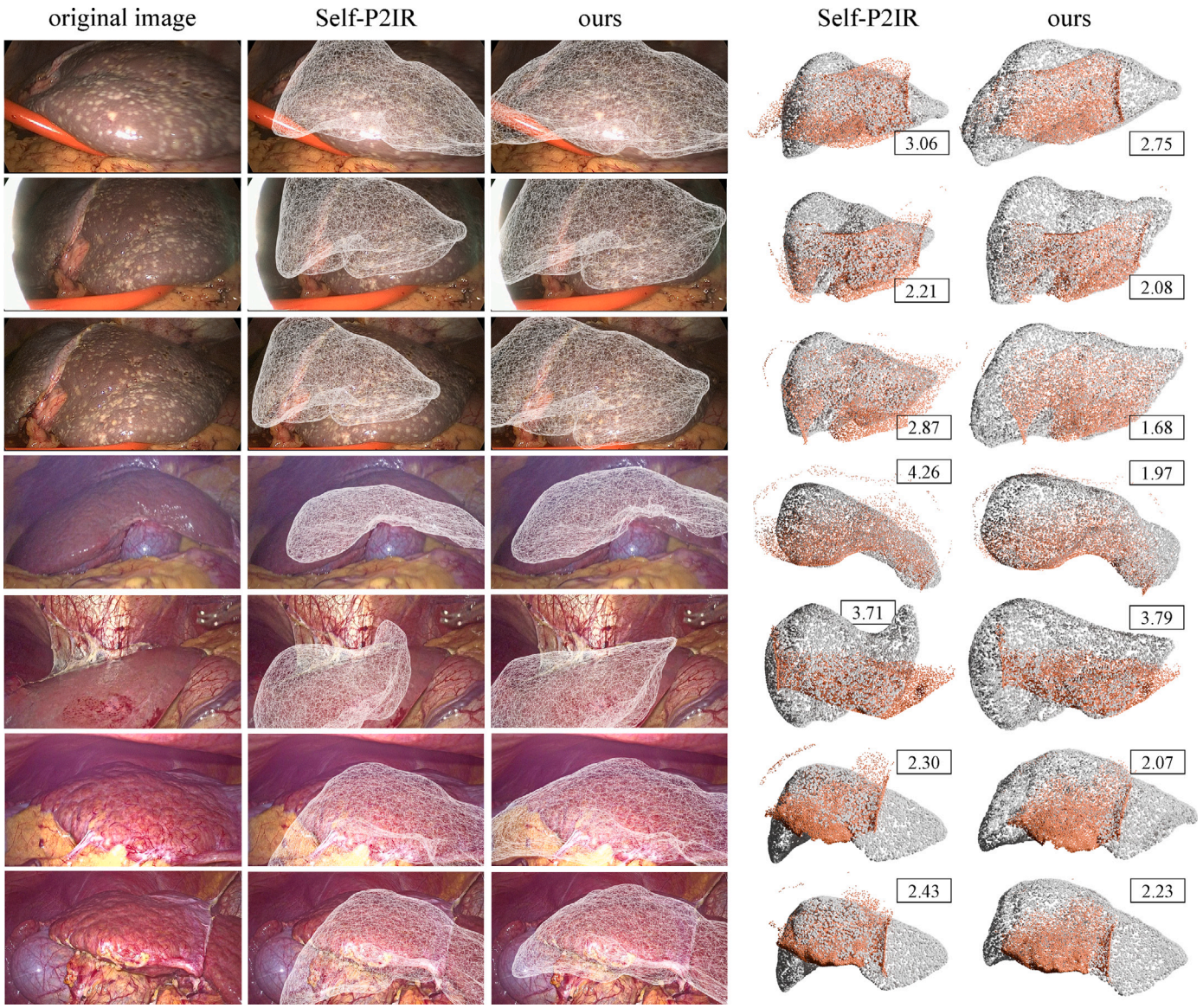


Fig. 12. Qualitative results of Self-P2IR and PIVOTS on P2I-LReg dataset. The first column shows original 2D frame, the next two columns present the projected deformed liver surface mesh as white wireframe, and the last two columns demonstrate the registration results of the aligned 3D preoperative liver point clouds (gray) and the intraoperative point clouds (orange). The 2D frames are from multiple patients.

and BCF-FEM creates a strong warp in the mesh due to too little identifiable overlap despite its strong physical constraints. By contrast, V2S-Net matches the ground-truth displacement well on the synthetic data, but on AMOS its deformation field exhibits abrupt changes. On the other hand, PIVOTS delivers much smoother and coherent displacement fields on both datasets and aligns the source point clouds well with the ground truth, providing an example for its superior quantitative performance.

Real data: For the Phantom and HHLBM samples, surface alignment improves for both BCF-FEM and PBSM relative to the synthetic test sets, aligning well with their gain in quantitative scores on real data in Table 2. Nonetheless, none of the two methods fully corrects the Phantom sample's left-lobe misalignment, and PBSM only slightly outperforms BCF-FEM on the right side. In the HHLBM example, PBSM yields the best result for the surface alignment and the landmarks, especially for the right lobe, and provides a more uniform misalignment vector distribution than V2S-Net. BCF-FEM achieves a good local fit at the center but remains the least effective overall, showing no clear deformation trend. While both V2S-Net and PIVOTS succeed in moving internal landmarks closer to their true positions, PIVOTS exhibits stronger results in both the center/left of the liver and the surface regions.

Intraoperative data: The Laparoscopic Liver dataset allows only qualitative comparison but provides a valuable real intraoperative setting. BCF-FEM yields a plausible close-to-rigid transformation for the majority of the surface but it locally deforms the left lobe of the preoperative surface unrealistically. On the other hand, managing the alignment of the surfaces in the visible region well, only leaving a gap on the leftmost lower part, V2S-Net introduces a strong deformation to the right lobe with a sharp transition without a basis in the intraoperative information. PIVOTS predicts the most coherent deformation without local distortions and matches the partial surface very closely. It also assumes a deformation of the right lobe without a clear indication in the data, however, the displacement is in line with the transformation towards the target area. Since PBSM requires a surface mesh of the intraoperative data and generating such a mesh from the noisy point cloud introduces many additional parameters, highlighting a shortcoming of the method, we exclude it from this experiment.

P2I-LReg dataset: A qualitative comparison between Self-P2IR and PIVOTS is shown in Fig. 12. Both methods produce coherent registra-

tions, but PIVOTS demonstrates a better understanding of the global intraoperative geometry, whereas Self-P2IR frequently misaligns the liver surface to sparse intraoperative points. Moreover, Self-P2IR often introduces compressions of the liver, while PIVOTS better preserves organ volume after deformation.

The qualitative results highlight the PIVOTS' ability to predict smooth displacement fields across various amounts of deformation, while V2S-Net struggles with abrupt field changes, the biomechanical methods fail to reliably establish correspondences in the presence of a high discrepancy between source and target, and the 3D–2D method tends to struggle with the global alignment.

5. Discussion and conclusion

5.1. Summary

Comparison with State-of-the-Art Methods: Our comparison experiment shows that our proposed network not only outperforms biomechanics-based methods like PBSM and BCF-FEM, but also existing deep learning-based methods, e.g. V2S-Net and Lepard, even after these methods are re-trained on the same training sets with hyperparameter tuning. Native surface-to-surface registration designs consistently struggle on the volume-to-surface problem and datasets. PIVOTS' attention mechanisms appear to be more powerful than the convolutions in V2S-Net on the unstructured point cloud data (at the cost of being roughly 2x slower), and our method is more robust to noise and produces fewer artifacts. Explicit biomechanical methods can compete on real data, however, in our experiments they struggle with large deformations and imperfect conditions, BCF-FEM ensures a spatially coherent displacement field but is prone to spuriously enlarging the liver volume due to simplified underlying assumptions, and both PBSM and BCF-FEM require a considerable amount of online parameter tuning in practice. For a favorable comparison we tuned the free parameters individually for each sample but in a translation setting, a solid solution for exploring the situation-specific (e.g. deformation level, in addition to patient-specific) parameter space would be needed. In comparison to the biomechanical methods PIVOTS reduces runtime by a factor of 260 (PBSM) and 700 (BCF-FEM). This speed in conjunction with competitive accuracy is underpinned by the choice of network depth, local kernel and neighborhood sizes.

Flexibility: Our method generalizes well across datasets in terms of (1) shape variety: the model trained on synthetic data can seamlessly transfer to inference on AMOS samples, and (2) real datasets: It maintains registration errors well below the 1 cm threshold on unseen real-world Phantom and HHLBM datasets and achieves a Chamfer Distance of 2.39 mm on the P2I-LReg dataset, all of which employ different surface acquisition methods. We conjecture that our simple downsampling strategy, despite being designed for high point density heterogeneity within a single point cloud, contributes to this strength.

Robustness: The robustness of PIVOTS extends to large deformations (0 mm–100 mm displacement before registration), where it achieves less than half the TRE of the next best method and three- to six-fold improvements over other baselines. Its noise resilience is equally impressive: While V2S-Net's error distribution broadens with increasing perturbation, PIVOTS preserves a tight error profile across both Gaussian and Perlin noise regimes. Furthermore, when faced with extremely limited fields of view (< 10% visible surface) on HHLBM samples, our method initially lags behind baseline methods, however, performance rapidly improves with just up to 20% visibility, achieving less than 1 cm errors across two real datasets.

PIVOTS also exhibits low variability across patients and experimental conditions, ensuring reliable performance not just on average but for every individual case, which additionally presents a critical requirement for clinical deployment.

Network components: The ablation study shows that several components contribute to the accuracy of the prediction: The use of precomputed features, including the (to our knowledge) newly introduced cross-modality closest distances, appears to be highly effective for our architecture. The S2V & V2S cross attention achieves expanding the receptive field without additional layers or parameters, giving PIVOTS an edge especially in high deformation scenarios. Adding relative coordinates to the point attention mostly decreases inter-sample variability.

We also examine the tradeoffs in our choices of network depth and input point density with regard to registration error and inference time.

5.2. Methodological insights

From these considerations, we derive the following core takeaways:

Pushing for generalization: Overall, the strong generalization observed in our model appears to arise from the interplay between the architecture and the training data rather than from any single dominant component. In the ablation study, we compare how performance degrades for both the synthetic test dataset (generic shapes) as well as the $\text{AMOS}_{\text{noise}}$ dataset (real liver shapes). For each ablated component, the performance tends to drop a similar percentage for both datasets and we do not observe catastrophic failure, which shows that none of these components is responsible for the generalization capabilities of PIVOTS on its own.

This suggests that explicitly exposing the network to the conditions it is expected to handle early on is key to achieving robustness. We used a staged development strategy, starting from simplified settings like shape encoding and progressively introducing more known challenging effects of the intraoperative scene using the training data to select generalization-capable components.

In our experience, a major enabler for the network's generalization to real liver shapes is the large amounts of diverse training data. Interestingly, both after training on the reduced ablation study samples and on the full dataset the performance on real livers tends to follow the performance on the synthetic validation dataset, i.e. generalization to our generic shapes goes hand-in-hand with generalization to real organ shapes. Our focus on quick inference speeds led us to build smaller models, which limits model capacity and may be another reason why we usually do not observe models that overfit to the synthetic data domain. It is also likely that the encoding–decoding structure with the low-resolution bottleneck acts as a regularization technique and enables generalization. This is backed by the fact that V2S-Net, which is based on a U-Net structure, can also generalize to unseen shapes. However, ablating this structure is difficult since it requires changing of the entire network architecture and quickly leads to a network that cannot learn the registration task at all.

Intrinsic dimensionality difference between volume and surface: Our main architectural contribution lies in the S2V & V2S intermodality cross attention module, more specifically, in enabling *bidirectional* attention flow between representations of different intrinsic dimensionality. The preoperative volume is inherently 3-dimensional, whereas the intraoperative surface is a 2-dimensional manifold within 3D space.

The ablation demonstrates that this bidirectional interaction makes the network more expressive, particularly in challenging settings with large deformations. Based on the receptive field study for a reduced 2D–1D setting (see the Supplementary Material, Section 2) we conjecture that this effect arises because the S2V pathway enables information to propagate across the surface in a lateral manner, effectively increasing the receptive field along the outside where correspondences are defined, while the V2S pathway anchors this information in the volumetric context. This cross-modal exchange allows the network to learn to guide its feature aggregation in a way that would be difficult to achieve within a single modality alone.

However, similar effects may in principle be approximated by deeper networks or larger kernels, suggesting that the benefit is not strictly unique but achieved here in a more parameter-efficient manner.

Importantly, this mechanism is tied to the difference in dimensionality: the observed behavior relies on the interaction between a 2D manifold and a 3D domain and would not arise in the same form if both inputs shared the same structure (e.g., volume–volume or surface–surface).

5.3. Limitations

Nevertheless, limitations still exist, hindering the translation into the operation room. Our method assumes the availability of a metric, dense intraoperative point cloud. However, real-time 3D reconstruction in the operating room remains challenging. Recent advances in intraoperative SLAM (Docea et al., 2021), Neural Radiance Fields (NeRFs) for dynamic scenes (Khojasteh et al., 2025), and learning-based deformation tracking (Gong et al., 2024) offer promising directions, but fully robust, low-latency reconstruction pipelines are not yet standard clinical tools.

Like other learning-based 3D–3D non-rigid registration approaches (e.g., V2S-Net), our method presumes an initial rigid alignment of preoperative and intraoperative data. This requirement can limit its applicability in extreme scenarios where only a wrong or no initial pose estimate is available. Future work could integrate a rigid registration module to enable fully end-to-end alignment from arbitrary initial poses.

Our evaluation of the displacement fields reveals that PIVOTS can output local mechanical inconsistencies within its smooth results. This likely arises from predicting the displacement field each point individually and using only implicit constraints on the physics and local coherence of the results via the dense displacement field from a biomechanical simulation as ground truth during training. PIVOTS learns any physical properties of the material implicitly through the synthetic training data. Addressing these inconsistencies while retaining fast inference remains open, especially given the high cost of explicit models. While physics-informed losses can be used to explicitly guide a network towards physical accuracy (Ragoza and Batmanghelich, 2023), in our experience their introduction can have a negative impact on the target registration error. This may be because the overall loss formulation becomes too strict; for example a loss that prohibits any stretching of volume may contradict the real test data where volume changes occur due to the diurnal digestive process. The implicit guidance through the training data has the advantage of being very flexible, and phenomena such as tissue cutting and inhomogeneous or compressible materials could be integrated without changing the network or loss functions. However, in the long term, it is possible that a combination of explicit and implicit physical guidance may become the most viable approach. Alternatively, future methods could incorporate partial elasticity information where available, infer it from dynamics on the fly, or output uncertainty estimates to better reflect regions of lower physical consistency.

We create the largest and most diverse training dataset to date (to our knowledge). Despite that, the noticeable sim-to-real gap calls for further improvements. This could be addressed by creating more realistic synthetic data modeling tool–tissue interaction, self-collision, heterogeneous material properties, volume changes due to pneumoperitoneum, and authentic surgical noise, and the effect of including each aspect would have to be evaluated. Training data efficiency could be increased by exploring adaptive sampling strategies, since in our case the ablation experiment shows that with one third of the training data, the performance loss remains below 2 mm. Gathering high-fidelity intraoperative datasets will be crucial for validating PIVOTS under true clinical conditions.

5.4. Implications for intraoperative pipelines

In our visibility experiments, all methods, including PIVOTS, struggle when less than 10% of the surface is visible. This finding indicates that pipelines for intraoperative surface acquisition must prioritize wide coverage: not only during initial mapping but also through continuous map maintenance, reintegration of temporarily occluded regions, and real-time update of moving structures. Without sufficient coverage, even the most advanced registration networks cannot compensate for missing geometry.

PIVOTS has strong potential to incorporate additional inputs to improve accuracy and possibly mitigate missing-geometry issues due to low visibility. In particular, laparoscopic ultrasound could serve as an automatic source of online refinement without manual intervention. The approach is well-suited to multi-modal extensions for two reasons: (1) its point-based formulation allows fine structures to be integrated more naturally than for voxel-based methods, and (2) the bidirectional attention between volume and surface handles differing dimensionalities effectively.

Our cross-dataset results indicate that switching to other acquisition devices should be feasible despite different characteristics, as they still produce surface-like data. Extensions to more challenging inputs such as segmented ultrasound slices (2D planes in 3D) or 1D structures remain more complex but are in principle compatible with the inter-modality paradigm. However, incorporating ultrasound also introduces additional alignment and calibration errors, which are often large and comparable to or exceeding current TRE levels.

Finally, the comparatively low point density required on the intraoperative surface allows freedom in the choice of faster, potentially lower resolution acquisition devices during surgery. However, even a perfect non-rigid registration of preoperative information onto the intraoperative situation can only visualize what is discernible in the preoperative scan, whereas especially small lesions are often first detected intraoperatively (Chu et al., 2023; Husarova et al., 2023). Therefore, advancing reliable intraoperative imaging remains indispensable for improved patient outcome.

5.5. Conclusion

We set out to develop a registration method that translates to real clinical use-cases by closely adhering to navigation system requirements for its design. To this end, we explicitly address key intraoperative challenges through a targeted, joint development of network architecture and high-fidelity training data. This results in a model comprising an encoder with DGCNN and FPS layers and a decoder with *Inter-Modality cross attention* and *upsampling cross attention* modules. To the best of our knowledge, this is the first application of inter-modality cross-attention to deformable liver volume-to-surface registration, where the additional S&V attention proves highly effective in handling substantial deformations.

Through extensive experiments, we demonstrate that the proposed network:

- (i) outperforms state-of-the-art baseline methods across different datasets,
- (ii) infers an accurate and smooth displacement field,
- (iii) displays substantial robustness against large deformation, considerable amounts of noise, and various levels of partial surface visibility,
- (iv) infers the displacement from point clouds that can be derived from standard clinical data directly and fully automatically for easier integration with surgical workflows,
- (v) achieves this level of accuracy at 4.3 fps.

These results render PIVOTS a strong candidate for a non-rigid registration method at the core of an intraoperative navigation system. In light of its generalizability across random shapes, we trust that the proposed method can be applied in deformation prediction of many other soft tissues, such as the prostate, lungs, etc.

Prospectively, we aim to refine the method to operate throughout the entire surgical procedure. Current limitations have largely been discussed above: Regarding requirement (ii), small local physical artifacts, with respect to (iii), issues with low visibility and the dependency on stereo reconstruction which is also affected by intraoperative image artifacts, and relating to (iv), the need of a fully automatic initial rigid registration.

Data scarcity remains a central bottleneck in developing nonrigid registration for laparoscopy. While we mitigate this through a large and highly diverse training dataset and a modular, flexible pipeline adaptable for various soft-tissue problems, simulation-based approaches inevitably face a sim-to-real gap. Beyond continued efforts to narrow this gap, we see systematic intraoperative validation data collection as the most critical lever to rigorously challenge these methods and move them towards clinical translation, both for liver surgery and other soft tissues.

Ultimately, this line of work aims to enable nonrigid registration as a *clinically viable* navigation tool in order to support intraoperative orientation and reduce cognitive load, thereby improving risk management regarding injury to critical internal structures.

CRediT authorship contribution statement

Peng Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Bianca Güttner:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation, Conceptualization. **Yutong Su:** Visualization, Methodology, Data curation. **Chenyang Li:** Writing – review & editing, Visualization. **Jinjing Xu:** Writing – review & editing, Visualization. **Mingyang Liu:** Validation, Methodology. **Zhe Min:** Validation, Methodology. **Andrey Zhylka:** Resources, Data curation. **Jasper Smit:** Resources, Data curation. **Karin Olthof:** Resources, Data curation. **Matteo Fusaglia:** Resources, Data curation. **Rudi Apolle:** Resources, Data curation. **Matthias Miederer:** Resources, Data curation. **Laura Frohneberg:** Resources, Data curation. **Carina Riediger:** Resources, Conceptualization. **Jürgen Weitz:** Resources, Conceptualization. **Fiona Kolbinger:** Writing – review & editing, Validation. **Stefanie Speidel:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Micha Pfeiffer:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the European Union through CloudSkin under grant agreement ID 101092646. It is also funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/2 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden, and through the Reinhart Koselleck-project – Project ID 560101272 – "SIMSURGE: Balancing the odds by simulating rare cases for surgical data science". Additionally, this work is co-funded by the project "Next Generation AI Computing (gAIn)," supported by the Bavarian Ministry of Science and the Arts and the Saxon Ministry for Science, Culture, and Tourism.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2026.104139.

Data availability

The code and dataset can be accessed from https://github.com/pengliu-nct/PIVOTS.

References

Abu Hilal, M., Aldrighetti, L., Dagher, I., Edwin, B., Troisi, R.I., Alikhanov, R., Aroori, S., Belli, G., Besselink, M., Briceno, J., Gayet, B., D'Hondt, M., Lesurtel, M., Menon, K., Lodge, P., Rotellar, F., Santoyo, J., Scatton, O., Soubrane, O., Sutcliffe, R., Van Dam, R., White, S., Halls, M.C., Cipriani, F., Van Der Poel, M., Ciria, R., Barkhatov, L., Gomez-Luque, Y., Ocana-Garcia, S., Cook, A., Buell, J., Clavien, P.-A., Dervenis, C., Fusai, G., Geller, D., Lang, H., Primrose, J., Taylor, M., Van Gulik, T., Wakabayashi, G., Asbun, H., Cherqui, D., 2018. The Southampton Consensus Guidelines for Laparoscopic Liver Surgery: From Indication to Implementation. Ann. Surg. 268 (1), 11–18. http://dx.doi.org/10.1097/SLA.0000000000002524, URL https://journals.lww.com/00000658-201807000-00003.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Ali, S., Espinel, Y., Jin, Y., Liu, P., Güttner, B., Zhang, X., Zhang, L., Dowrick, T., Clarkson, M.J., Xiao, S., et al., 2025. An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion from the MICCAI2022 challenge. Med. Image Anal. 99, 103371.

Amberg, B., Romdhani, S., Vetter, T., 2007. Optimal step nonrigid ICP algorithms for surface registration. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8.

Amodu, L.I., Howell, R.S., Daskalaki, D., Allendorf, J.D., 2022. Oncologic benefits of laparoscopic and minimally invasive surgery: a review of the literature. Ann. Laparosc. Endosc. Surg. 7, 5. http://dx.doi.org/10.21037/ales-21-19, URL https://ales.amegroups.com/article/view/7382/html.

Aoki, Y., Goforth, H., Srivatsan, R.A., Lucey, S., 2019. Pointnetlk: Robust & efficient point cloud registration using pointnet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7163–7172.

Besl, P., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14 (2), 239–256. http://dx.doi.org/10.1109/34.121791.

Bonet, J., Wood, R.D., 2008. Nonlinear Continuum Mechanics for Finite Element Analysis, second ed. Cambridge University Press, Cambridge.

Chen, Z., Zou, B., Kui, X., Shi, Y., Lv, D., Chen, L., 2024. Points of interest linear attention network for real-time non-rigid liver volume to surface registration. Med. Phys. 51 (8), 5351–5360.

Chu, K.-J., Kawaguchi, Y., Hasegawa, K., 2023. Current use of intraoperative ultrasound in modern liver surgery. Oncol. Transl. Med. 9 (4), 168–175. http://dx.doi.org/10.1097/ot9.0000000000000005, URL https://journals.lww.com/10.1097/ot9.0000000000000005.

Collins, T., Bartoli, A., Bourdel, N., Canis, M., 2016. Robust, real-time, dense and deformable 3D organ tracking in laparoscopic videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 404–412.

Docea, R., Pfeiffer, M., Bodenstedt, S., Kolbinger, F.R., Höller, L., Wittig, I., Hoffmann, R.-T., Troost, E.G.C., Riediger, C., Weitz, J., Speidel, S., 2021. Simultaneous localisation and mapping for laparoscopic liver navigation : a comparative evaluation study. In: Linte, C.A., Siewerdsen, J.H. (Eds.), Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling. 11598, SPIE International Society for Optics and Photonics, p. 115980B. http://dx.doi.org/10.1117/12.2582121.

Docea, R., Pfeiffer, M., Müller, J., Krug, K., Hardner, M., Riedel, P., Menzel, M., Kolbinger, F.R., Frohneberg, L., Weitz, J., Speidel, S., 2022. A laparoscopic liver navigation pipeline with minimal setup requirements. In: 2022 IEEE Biomedical Circuits and Systems Conference (BioCAS). pp. 578–582. http://dx.doi.org/10.1109/BioCAS54905.2022.9948587.

Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P., 2015. Convolutional networks on graphs for learning molecular fingerprints. Adv. Neural Inf. Process. Syst. 28.

Espinel, Y., Calvet, L., Botros, K., Buc, E., Tilmant, C., Bartoli, A., 2021. Using multiple images and contours for deformable 3d-2d registration of a preoperative ct in laparoscopic liver surgery. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24. Springer, pp. 657–666.

Estermann, S.-J., Pahr, D.H., Reisinger, A., 2020. Hyperelastic and viscoelastic characterization of hepatic tissue under uniaxial tension in time and frequency domain. J. Mech. Behav. Biomed. Mater. 112, 104038. http://dx.doi.org/10.1016/j.jmbbm.2020.104038, URL https://www.sciencedirect.com/science/article/pii/S1751616120305865?via%3Dihub.

Faure, F., Duriez, C., Delingette, H., Allard, J., Gilles, B., Marchesseau, S., Talbot, H., Courtecuisse, H., Bousquet, G., Peterlik, I., Cotin, S., 2012. SOFA: A Multi-Model Framework for Interactive Physical Simulation. In: Payan, Y. (Ed.), Soft Tissue Biomechanical Modeling for Computer Assisted Surgery. In: Studies in Mechanobiology, Tissue Engineering and Biomaterials, vol. 11, Springer, pp. 283–321. http://dx.doi.org/10.1007/8415_2012_125.

Geuzaine, C., Remacle, J.-F., 2009. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. Internat. J. Numer. Methods Engrg. 79 (11), 1309–1331. http://dx.doi.org/10.1002/nme.2579, Version 4.13.1 from 24 May 2024.

Glozman, T., Azhari, H., 2010. A Method for Characterization of Tissue Elastic Properties Combining Ultrasonic Computed Tomography With Elastography. J. Ultrasound Med. 29 (3), 387–398. http://dx.doi.org/10.7863/jum.2010.29.3.387, URL http://doi.wiley.com/10.7863/jum.2010.29.3.387.

Gojcic, Z., Zhou, C., Wegner, J.D., Wieser, A., 2019. The perfect match: 3d point cloud matching with smoothed densities. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5545–5554.

Gong, S., Long, Y., Chen, K., Liu, J., Xiao, Y., Cheng, A., Wang, Z., Dou, Q., 2024. Self-supervised cyclic diffeomorphic mapping for soft tissue deformation recovery in robotic surgery scenes. IEEE Trans. Med. Imaging PP, http://dx.doi.org/10.1109/TMI.2024.3439701.

Guo, Y., Sohel, F., Bennamoun, M., Lu, M., Wan, J., 2013. Rotational projection statistics for 3D local surface description and object recognition. Int. J. Comput. Vis. 105, 63–86.

Han, Z., Zhou, J., Pei, J., Qin, J., Fan, Y., Dou, Q., 2025. Toward reliable AR-guided surgical navigation: Interactive deformation modeling with data-driven biomechanics and prompts. IEEE Trans. Med. Imaging 44 (11), 4417–4431. http://dx.doi.org/10.1109/tmi.2025.3577759.

Heiselman, J.S., Collins, J.A., Ringel, M.J., Kingham, T.P., Jarnagin, W.R., Miga, M.I., 2024. The Image-to-Physical Liver Registration Sparse Data Challenge: comparison of state-of-the-art using a common dataset. J. Med. Imaging 11 (1), 015001. http://dx.doi.org/10.1117/1.JMI.11.1.015001.

Hirose, O., 2022. Geodesic-based bayesian coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. 45 (5), 5816–5832.

Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K., 2021. Predator: Registration of 3d point clouds with low overlap. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4267–4276.

Huang, X., Mei, G., Zhang, J., 2020. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11366–11374.

Husarova, T., MacCuaig, W.M., Dennahy, I.S., Sanderson, E.J., Edil, B.H., Jain, A., Bonds, M.M., McNally, M.W., Menclova, K., Pudil, J., Zaruba, P., Pohnan, R., Henson, C.E., Grizzle, W.E., McNally, L.R., 2023. Intraoperative Imaging in Hepatopancreatobiliary Surgery. Cancers 15 (14), 3694. http://dx.doi.org/10.3390/cancers15143694, URL https://www.mdpi.com/2072-6694/15/14/3694.

IRCAD, S., 2022. Respiratory cycle 3D-IRCADb-02. URL https://www.ircad.fr/research/data-sets/respiratory-cycle-3d-ircadb-02/.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211.

Ji, Y., Bai, H., Ge, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., et al., 2022. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. Adv. Neural Inf. Process. Syst. 35, 36722–36732.

Jin, H., Shen, Y., Lou, J., Zhou, K., Zheng, Y., 2024. KeypointDETR: An end-to-end 3D keypoint detector. In: European Conference on Computer Vision. Springer, pp. 374–390.

Khojasteh, S.B., Fuentes-Jimenez, D., Pizarro, D., Espinel, Y., Bartoli, A., 2025. MIS-NeRF: neural radiance fields in minimally-invasive surgery. Int. J. Comput. Assist. Radiol. Surg. http://dx.doi.org/10.1007/s11548-025-03429-7.

Koo, B., Özgür, E., Le Roy, B., Buc, E., Bartoli, A., 2017. Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 326–334.

Labrunie, M., Pizarro, D., Tilmant, C., Bartoli, A., 2024. Automatic 3D/2D deformable registration in minimally invasive liver resection using a mesh recovery network. In: Oguz, I., Noble, J., Li, X., Styner, M., Baumgartner, C., Rusu, M., Heinmann, T., Kontos, D., Landman, B., Dawant, B. (Eds.), Medical Imaging with Deep Learning. In: Proceedings of Machine Learning Research, vol. 227, PMLR, pp. 1104–1123, URL https://proceedings.mlr.press/v227/labrunie24a.html.

Lemine, A.S., Ahmad, Z., Al-Thani, N.J., Hasan, A., Bhadra, J., 2024. Mechanical properties of human hepatic tissues to develop liver-mimicking phantoms for medical applications. Biomech. Model. Mechanobiol. 23 (2), 373–396. http://dx.doi.org/10.1007/s10237-023-01785-4, URL https://link.springer.com/10.1007/s10237-023-01785-4.

Leung, N., Farrant, P., Peters, T., 1986a. Liver volume measurement by ultrasonography in normal subjects and alcoholic patients. J. Hepatol. 2 (2), 157–164. http://dx.doi.org/10.1016/S0168-8278(86)80074-5, URL https://www.sciencedirect.com/science/article/pii/S0168827886800745.

Leung, N., Farrant, P., Peters, T., 1986b. Liver volume measurement by ultrasonography in normal subjects and alcoholic patients. J. Hepatol. 2 (2), 157–164. http://dx.doi.org/10.1016/S0168-8278(86)80074-5, URL https://linkinghub.elsevier.com/retrieve/pii/S0168827886800745.

Li, Y., Harada, T., 2022a. Lepard: Learning partial point cloud matching in rigid and deformable scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5554–5564.

Li, Y., Harada, T., 2022b. Non-rigid point cloud registration with neural deformation pyramid. Adv. Neural Inf. Process. Syst. 35, 27757–27768.

Liu, P., Bodenstedt, S., Kolbinger, F., Riediger, C., Weitz, J., Speidel, S., Pfeiffer, M., 2024a. Leveraging expert knowledge for real-time online adaptation of intraoperative liver registration. In: International Workshop on Shape in Medical Imaging. Springer, pp. 137–148.

Liu, P., Golde, J., Morgenstern, J., Bodenstedt, S., Li, C., Hu, Y., Chen, Z., Koch, E., Neudert, M., Speidel, S., 2024b. Non-rigid point cloud registration for middle ear diagnostics with endoscopic optical coherence tomography. Int. J. Comput. Assist. Radiol. Surg. 19 (1), 139–145.

Marchesseau, S., Chatelin, S., Delingette, H., 2017. Nonlinear Biomechanical Model of the Liver. In: Biomechanics of Living Organs. Elsevier, pp. 243–265. http://dx.doi.org/10.1016/B978-0-12-804009-6.00011-0, URL https://linkinghub.elsevier.com/retrieve/pii/B9780128040096000110.

Mattei, G., Ahluwalia, A., 2016. Sample, testing and analysis variables affecting liver mechanical properties: A review. Acta Biomater. 45, 60–71. http://dx.doi.org/10.1016/j.actbio.2016.08.055, URL https://linkinghub.elsevier.com/retrieve/pii/S1742706116304603.

Mendizabal, A., Tagliabue, E., Brunet, J.-N., Dall'Alba, D., Fiorini, P., Cotin, S., 2019. Physics-based Deep Neural Network for Real-Time Lesion Tracking in Ultrasound-guided Breast Biopsy. In: Computational Biomechanics for Medicine XIV. Shenzhen, China, URL https://inria.hal.science/hal-02311277.

Mhiri, I., Pizarro, D., Bartoli, A., 2025. Neural patient-specific 3D–2D registration in laparoscopic liver resection. Int. J. Comput. Assist. Radiol. Surg. 20 (1), 57–64.

Min, Z., Liu, L., Meng, M.Q.-H., 2019. Generalized non-rigid point set registration with hybrid mixture models considering anisotropic positional uncertainties. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22. Springer, pp. 547–555.

Min, Z., Zhu, D., Ren, H., Meng, M.Q.-H., 2020. Feature-guided nonrigid 3-d point set registration framework for image-guided liver surgery: From isotropic positional noise to anisotropic positional noise. IEEE Trans. Autom. Sci. Eng. 18 (2), 471–483.

Modrzejewski, R., Collins, T., Seeliger, B., Bartoli, A., Hostettler, A., Marescaux, J., 2019. An in vivo porcine dataset and evaluation methodology to measure soft-body laparoscopic liver registration accuracy with an extended algorithm that handles collisions. Int. J. Comput. Assist. Radiol. Surg. 14 (7), 1237–1245. http://dx.doi.org/10.1007/s11548-019-02001-4, URL http://link.springer.com/10.1007/s11548-019-02001-4.

Monji-Azad, S., Kinz, M., Männel, D., Scherl, C., Hesser, J., 2024. Robust-DefReg: a robust coarse to fine non-rigid point cloud registration method based on graph convolutional neural networks. Meas. Sci. Technol. 36 (1), 015426.

Murad Gutiérrez, V., Romero Enciso, J., 2018. Liver elastography: What it is, how it is done, and how it is interpreted. RadiologíA (English Edition) 60 (3), 183–189. http://dx.doi.org/10.1016/j.rxeng.2018.03.003, URL https://linkinghub.elsevier.com/retrieve/pii/S217351071830020X.

Myronenko, A., Song, X., 2010. Point set registration: Coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. 32 (12), 2262–2275.

Nguyen, K.T., Marsh, J.W., Tsung, A., Steel, J.J.L., Gamblin, T.C., Geller, D.A., 2011. Comparative benefits of laparoscopic vs open hepatic resection: a critical appraisal. AArch Surg. 146 (3), 348–356.

Novitsky, Y.W., Litwin, D.E.M., Callery, M.P., 2004. The net immunologic advantage of laparoscopic surgery. Surg. Endosc. 18 (10), 1411–1419. http://dx.doi.org/10.1007/s00464-003-8275-x, URL http://link.springer.com/10.1007/s00464-003-8275-x.

Özgür, E., Koo, B., Le Roy, B., Buc, E., Bartoli, A., 2018. Preoperative liver registration for augmented monocular laparoscopy using backward–forward biomechanical simulation. Int. J. Comput. Assist. Radiol. Surg. 13, 1629–1640.

Pfeiffer, M., 2022. Non-Rigid Liver Registration for Laparoscopy using Data-Driven Biomechanical Models (Ph.D. thesis). Technische Universität Dresden, Dresden, Germany, URL https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-793702.

Pfeiffer, M., Riediger, C., Leger, S., Kühn, J.-P., Seppelt, D., Hoffmann, R.-T., Weitz, J., Speidel, S., 2020. Non-rigid volume to surface registration using a data-driven biomechanical model. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Springer International Publishing, Cham, pp. 724–734.

Pfeiffer, M., Riediger, C., Weitz, J., Speidel, S., 2019. Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks. Int. J. Comput. Assist. Radiol. Surg. 14 (7), 1147–1155. http://dx.doi.org/10.1007/s11548-019-01965-7, URL http://link.springer.com/10.1007/s11548-019-01965-7.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 652–660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: Advances in Neural Information Processing Systems, vol. 30.

Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Ilic, S., Hu, D., Xu, K., 2023a. Geotransformer: Fast and robust point cloud registration with geometric transformer. IEEE Trans. Pattern Anal. Mach. Intell. 45 (8), 9806–9821.

Qin, Z., Yu, H., Wang, C., Peng, Y., Xu, K., 2023b. Deep graph-based spatial consistency for robust non-rigid point cloud registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5394–5403.

Rabbani, N., Calvet, L., Espinel, Y., Le Roy, B., Ribeiro, M., Buc, E., Bartoli, A., 2022. A methodology and clinical dataset with ground-truth to evaluate registration accuracy quantitatively in computer-assisted Laparoscopic Liver Resection. Comput. Methods Biomech. Biomed. Eng.: Imaging & Vis. 10 (4), 441–450. http://dx.doi.org/10.1080/21681163.2021.1997642, URL https://www.tandfonline.com/doi/full/10.1080/21681163.2021.1997642.

Ragoza, M., Batmanghelich, K., 2023. Physics-informed neural networks for tissue elasticity reconstruction in magnetic resonance elastography. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Springer Nature Switzerland, Cham, pp. 333–343.

Ratti, F., Fiorentini, G., Cipriani, F., Catena, M., Paganelli, M., Aldrighetti, L., 2018. Laparoscopic vs open surgery for colorectal liver metastases. JAMA Surg. 153 (11), 1028–1035.

Robu, M.R., Ramalhinho, J., Thompson, S., Gurusamy, K., Davidson, B., Hawkes, D., Stoyanov, D., Clarkson, M.J., 2018. Global rigid registration of CT to video in laparoscopic liver surgery. Int. J. Comput. Assist. Radiol. Surg. 13, 947–956.

Rodola, E., Albarelli, A., Bergamasco, F., Torsello, A., 2013. A scale independent selection process for 3d object recognition in cluttered scenes. Int. J. Comput. Vis. 102, 129–145.

Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. In: 2009 IEEE International Conference on Robotics and Automation. IEEE, pp. 3212–3217.

Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M., 2008. Aligning point cloud views using persistent feature histograms. In: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 3384–3391.

Santhosh, I., Yadav, T., Garg, S., Kumar, K., Singh, H., Kurdia, K., 2024. Study of variations in the biliary anatomy of the liver (extra and intrahepatic biliary apparatus)-a study of 100 cadaveric liver dissection in a tertiary care center. HPB 26, S263–S264.

Sarode, V., Li, X., Goforth, H., Aoki, Y., Srivatsan, R.A., Lucey, S., Choset, H., 2019. Pcrnet: Point cloud registration network using pointnet encoding. arXiv preprint arXiv:1908.07906.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2008. The graph neural network model. IEEE Trans. Neural Netw. 20 (1), 61–80.

Schöberl, J., Lackner, C., Hochsteger, M., Dorigo, R., Hirnschall, S., Leonhartsberger, C., 2024. Netgen/NGSolve. svn revision 469 from 1 December 2011 with changes as stated in GMSH source. URL https://netgen-mesher.svn.sourceforge.net/svnroot/netgen-mesher. https://ngsolve.org/.

Shi, J., Wan, P., Chen, F., 2021. An unsupervised non-rigid registration network for fast medical shape alignment. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, pp. 1887–1890.

Singh, H.R., Rabi, S., 2019. Study of morphological variations of liver in human. Transl. Res. Anat. 14, 1–5.

Sipiran, I., Bustos, B., 2011. Harris 3D: a robust extension of the harris operator for interest point detection on 3D meshes. Vis. Comput. 27, 963–976.

Smith, L.N., Topin, N., 2019. Super-convergence: very fast training of neural networks using large learning rates. In: Pham, T. (Ed.), In: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, vol. 11006, SPIE International Society for Optics and Photonics, 1100612. http://dx.doi.org/10.1117/12.2520589.

Srimani, P., Saha, A., 2020. Liver morphology: anatomical study about the outer aspects. Surg. Radiol. Anat. 42, 1425–1434.

Streiff, D., Bernreiter, L., Tschopp, F., Fehr, M., Siegwart, R., 2021. 3D3l: Deep learned 3D keypoint detection and description for lidars. In: 2021 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 13064–13070.

Suwelack, S., Röhl, S., Bodenstedt, S., Reichard, D., Dillmann, R., dos Santos, T., Maier-Hein, L., Wagner, M., Wünscher, J., Kenngott, H., et al., 2014. Physics-based shape matching for intraoperative image guidance. Med. Phys. 41 (11), 111901.

Thomas, H., Qi, C.R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6411–6420.

Tombari, F., Di Stefano, L., 2010. Object recognition in 3d scenes with occlusions and clutter by hough voting. In: 2010 Fourth Pacific-Rim Symposium on Image and Video Technology. IEEE, pp. 349–355.

Wang, Y., Solomon, J.M., 2019. Deep closest point: Learning representations for point cloud registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3523–3532.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph cnn for learning on point clouds. ACM Trans. Graph. (Tog) 38 (5), 1–12.

Wen, B., Trepte, M., Aribido, J., Kautz, J., Gallo, O., Birchfield, S., 2025. FoundationStereo: Zero-shot stereo matching. CVPR.

Yang, J., Li, H., Campbell, D., Jia, Y., 2015. Go-ICP: A globally optimal solution to 3D ICP point-set registration. IEEE Trans. Pattern Anal. Mach. Intell. 38 (11), 2241–2254.

Yang, Z., Simon, R., Linte, C.A., 2023. Learning feature descriptors for pre-and intra-operative point cloud matching for laparoscopic liver registration. Int. J. Comput. Assist. Radiol. Surg. 18 (6), 1025–1032.

Yang, Z., Simon, R.A., Merrell, K., Linte, C.A., 2024. Boundary constraint-free biomechanical model-based surface matching for intraoperative liver deformation correction. IEEE Trans. Med. Imaging 44, 1723–1734, URL https://api.semanticscholar.org/CorpusID:268510028.

Yang, J., Zhang, X., Fan, S., Ren, C., Zhang, Y., 2023. Mutual voting for ranking 3D correspondences. IEEE Trans. Pattern Anal. Mach. Intell..

Yew, Z.J., Lee, G.H., 2022. Regtr: End-to-end point cloud correspondences with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6677–6686.

Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T., 2017. 3Dmatch: Learning local geometric descriptors from rgb-d reconstructions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1802–1811.

Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V., 2021. Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 16259–16268.

Zhao, X., Wu, X., Miao, J., Chen, W., Chen, P.C., Li, Z., 2022. Alike: Accurate and lightweight keypoint detection and descriptor extraction. IEEE Trans. Multimed. 25, 3101–3112.

Zhou, J., Gao, B., Wang, K., Pei, J., Heng, P.-A., Qin, J., 2025. Landmark-free preoperative-to-intraoperative registration in laparoscopic liver resection. IEEE Trans. Med. Imaging.

Zhu, L., Guan, H., Lin, C., Han, R., 2022. Neighborhood-aware geometric encoding network for point cloud registration. 4. arXiv preprint arXiv:2201.12094.

Zhylka, A., Smit, J., Olthof, K., ter Beek, L., Ruers, T., Fusaglia, M., 2024. Simulated liver deformations with intraoperative ultrasound: steps towards a multi-modal dataset. "Int. J. Comput. Assist. Radiol. Surgery" 19, 135–136. http://dx.doi.org/10.1007/s11548-024-03128-9, Computer Assisted Radiology and Surgery 2024, CARS ; Conference date: 18-06-2024 Through 21-06-2024.