AID-GENREG: ANATOMICAL IDENTITY-PRESERVING DEPTH GENERATION FOR ENHANCING PREOPERATIVE -TO-INTRAOPERATIVE LIVER POINT CLOUD REGISTRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate rigid registration between complete preoperative and partial intraoperative liver point clouds is essential for reliable augmented reality visualization of internal hepatic anatomy during laparoscopic liver resection. However, the scarcity of paired clinical data, coupled with the unavailability of ground-truth transformations, hinders the performance of registration models trained via self-supervision under surgical conditions. Diffusion-based generative data construction presents a promising solution to this data scarcity by synthesizing geometrically diverse observations using limited paired data, producing abundant point cloud pairs to advance downstream registration. Nevertheless, adapting this scheme to clinical registration poses a distinct challenge, as existing methods do not explicitly enforce patient-specific anatomical identity between synthetic intraoperative observations and their corresponding preoperative models, yielding unreliable pairs that deteriorate subsequent self-supervised learning. In this work, we present *AID-GenReg*, a novel framework that constructs rich anatomy-consistent synthetic pairs to enhance preoperative-to-intraoperative liver point cloud registration. Specifically, *AID-GenReg* reprojects intraoperative observations and employs a diffusion-based anatomical identity-preserving depth generator to inpaint the resulting missing regions. To preserve anatomical identity represented in the preoperative model, we devise a region-decoupled generation scheme comprising two synergistic mechanisms: Geometry Preservation in observable regions with valid depth and preoperative Anatomy-Guided Completion in complementary missing regions. Such mechanisms ensure that the synthetic pairs offer reliable supervision for registration. Extensive experiments on a public in vivo human liver dataset under three downstream training paradigms demonstrate that *AID-GenReg* outperforms competitive generative alternatives and achieves consistent gains across multiple registration models.

1 INTRODUCTION

During laparoscopic liver resection (LLR), intraoperative augmented reality (AR) visualizes internal hepatic anatomy, helping surgeons localize tumors and major vessels and assess their spatial relationships (Ali et al., 2025). This capability may improve surgical outcomes and lower the risk of postoperative complications (Ramalhinho et al., 2023). The central task in intraoperative AR is to register a patient-specific 3D liver model reconstructed from preoperative CT or MRI to intraoperative 2D laparoscopic images. Hence, accurate preoperative-to-intraoperative (3D-2D) registration is essential for reliable AR guidance during LLR.

Recent learning-based approaches (Zhang et al., 2024; Wu et al., 2025; Yang et al., 2025) reformulate this 3D-2D alignment as rigid 3D-3D point cloud registration between a **complete** preoperative liver model and a **partial** intraoperative surface reconstructed from laparoscopic observations. This formulation enables direct estimation of the global transformation required for AR guidance from their surface geometry. However, most existing methods (Yang et al., 2023; 2025; Wang et al., 2026) are developed and evaluated on in silico, phantom, or porcine data, with limited use of human clinical

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

data. Models trained in these surrogate settings may not reliably generalize to clinical LLR, which involves occlusions from surrounding tissues, blood, and surgical instruments, as well as noise from intraoperative surface reconstruction. These limitations motivate registration learning grounded in real clinical data for robust AR guidance in LLR.

However, collecting sufficient paired preoperative and intraoperative clinical data is challenging. Each case requires a cumbersome acquisition pipeline involving patient-specific preoperative 3D model reconstruction, intraoperative laparoscopic recording, and camera calibration. To the best of our knowledge, P2I-LReg (Zhou et al., 2025) is the only publicly available in vivo clinical dataset for preoperative-to-intraoperative liver point cloud registration. Consequently, building an automatic framework that leverages limited real pairs to construct abundant point cloud training data is highly desirable for developing clinical-domain registration models.

Across natural-scene and surgical registration, prior work has constructed large quantities of point cloud pairs via geometric transformations (Wang & Solomon, 2019; Horache et al., 2021; Yang et al., 2025; Wu et al., 2025), simulation-based rendering (Chen et al., 2023; Zhou et al., 2025), or generative modeling (Chen et al., 2024) to facilitate registration learning. Transformation-based construction poorly reflects real acquisition conditions, while simulation-based rendering suffers from the sim-to-real gap, limiting downstream effectiveness. More recently, diffusion-based pair construction derives additional realistic pairs from real depth observations, improving the generalization of downstream registration models (Chen et al., 2024). Yet, directly applying this paradigm to clinical liver registration does not explicitly enforce anatomical consistency between synthesized intraoperative geometry and its paired patient-specific preoperative model. This limitation is particularly consequential in clinical settings, where GT rigid transformations are generally unavailable and registration learning therefore relies on self-supervision (Zhou et al., 2025), making anatomically mismatched pairs unreliable training signals. This raises a key question: **How can we synthesize intraoperative observations that preserve the *patient-specific anatomical identity* represented by the paired preoperative model for reliable self-supervised registration?**

We propose *AID-GenReg*, a novel framework that constructs diverse anatomy-consistent synthetic pairs to enhance preoperative-to-intraoperative liver point cloud registration. Given a sampled 3D transformation, *AID-GenReg* reprojects real intraoperative depth into the target-view observation, then completes its missing regions using a diffusion-based anatomical identity-preserving depth generator. To preserve the patient-specific anatomical identity encoded in the preoperative model, the generator adopts a region-decoupled design with two synergistic mechanisms: Geometry Preservation anchors valid reprojected measurements in observable regions, while Anatomy-Guided Completion uses the paired preoperative anatomy to guide synthesis in the complementary missing regions without altering the preserved geometry. The resulting synthetic pairs augment limited real pairs for self-supervised liver registration. Our contributions are summarized below:

(1) To our knowledge, *AID-GenReg* is the first to build abundant anatomy-consistent synthetic point cloud pairs from limited clinical data for self-supervised preoperative-to-intraoperative registration.

(2) We develop a diffusion-based depth generator for anatomical identity preservation via a region-decoupled inference scheme. Geometry Preservation maintains reprojected geometry in observable regions, while Anatomy-Guided Completion steers synthesis in the complementary missing regions toward paired preoperative anatomy. Their synergistic operation promotes patient-specific anatomical consistency, yielding more reliable synthetic pairs for downstream registration learning.

(3) We conduct extensive experiments on a public in vivo human liver registration dataset, benchmarking our method against two competitive generative construction approaches and evaluating registration performance across multiple models, each trained under three downstream paradigms. The results show that *AID-GenReg* outperforms both competitors and consistently improves registration performance across models, underscoring its practical potential for registration in LLR.

2 Related Work

2.1 3D Point Cloud Registration

Generic Natural-Scene Registration. 3D point cloud registration estimates the rigid transformation to align two partially overlapping point clouds. Traditional methods like ICP (Besl & McKay,

1992) and Go-ICP (Yang et al., 2016) rely on nearest-neighbor correspondences. However, ICP is sensitive to initialization, while Go-ICP avoids local minima through costly global search. Learning-based methods improve robustness by learning geometric representations and can be broadly divided into direct registration (Aoki et al., 2019; Xu et al., 2021) and correspondence-based approaches (Huang et al., 2021; Yu et al., 2021; Li & Harada, 2022; Qin et al., 2023). The former directly regresses the transformation from global point cloud features, while the latter first establishes corresponding point pairs and then estimates the transformation based on these correspondences. Despite strong performance on natural-scene benchmarks, most target partial-to-partial registration and may not readily generalize to complete-to-partial liver registration.

Preoperative-to-Intraoperative Rigid Liver Registration. In LLR, rigid registration aligns a complete preoperative liver model with a partial intraoperative surface reconstructed from laparoscopic observations. Unlike natural-scene registration, this complete-to-partial setting is particularly challenging because smooth liver surfaces provide few distinctive geometric cues (Wu et al., 2025), while limited visibility, occlusions, and reconstruction noise further hinder reliable correspondence estimation (Yang et al., 2025). Early approaches relied on manual initialization (Teatini et al., 2019), landmark-based alignment (Feuerstein et al., 2008), or handcrafted descriptors (Robu et al., 2018), requiring user intervention or exhibiting limited robustness to occlusion and smooth-surface ambiguity. Recent studies have tailored correspondence-based registration to liver-specific geometry, yielding promising results. EG-Net (Wu et al., 2025) leverages organ-edge cues to improve correspondence estimation in low-overlap settings, while P2P (Yang et al., 2025) mitigates complete-to-partial ambiguity through patches-to-partial matching. Furthermore, Self-P2IR (Zhou et al., 2025) enables landmark-free self-supervised registration with in vivo human data. Nevertheless, the development and empirical validation of most existing methods still rely largely on surrogate datasets, including in silico simulations, physical phantoms, and porcine liver data (Yang et al., 2023; 2025; Zhang et al., 2024; Wu et al., 2025). Models trained on these non-clinical datasets may not generalize reliably to clinical LLR conditions. Appendix A.1 summarizes the datasets used in existing liver registration studies. The scarcity of paired preoperative-intraoperative clinical data motivates us to construct abundant and reliable registration data from the few available clinical pairs.

2.2 Registration Pair Construction

Transformation-Based Construction. This scheme has been employed to construct training pairs in both natural scenes (Wang & Solomon, 2019; Horache et al., 2021) and surgical settings (Yang et al., 2025; Wu et al., 2025), typically by cropping existing point clouds and applying random rigid transformations. However, the resulting target point clouds do not fully reflect real acquisition conditions, limiting model generalization.

Simulation- and Generation-Based Construction. In surgical settings, simulation-based pipelines render anatomical models from virtual laparoscopic viewpoints to construct partial intraoperative observations (Zhou et al., 2025), but the resulting pairs suffer from a sim-to-real gap. For natural scenes, SIRA-PCR (Chen et al., 2023) constructs large-scale pairs from rendered 3D scenes and applies GAN-based adaptation toward real-data distributions, yet its reliance on simulation still limits its generalization. Drawing on the reproject-then-inpaint paradigm of novel-view synthesis (Liu et al., 2023), PointRegGPT (Chen et al., 2024) constructs registration pairs from real depth observations by reprojecting them to new viewpoints and completing missing regions with a diffusion model, thereby mitigating the sim-to-real gap. It employs DDNM-based range-null space decomposition (Wang et al., 2022) to enforce 3D consistency in overlapping regions and maintain reliable transformation supervision. Nevertheless, it is not directly suitable for our clinical setting, which requires anatomical consistency between the constructed intraoperative observation as a whole and its paired preoperative model. Accordingly, our region-decoupled design integrates two synergistic mechanisms for Anatomical Identity Preservation.

3 Method

3.1 Overall Pipeline

Fig. 1 illustrates the pipeline of *AID-GenReg* for enhancing preoperative-to-intraoperative point cloud registration. Our framework comprises two crucial stages: point cloud pair construction and

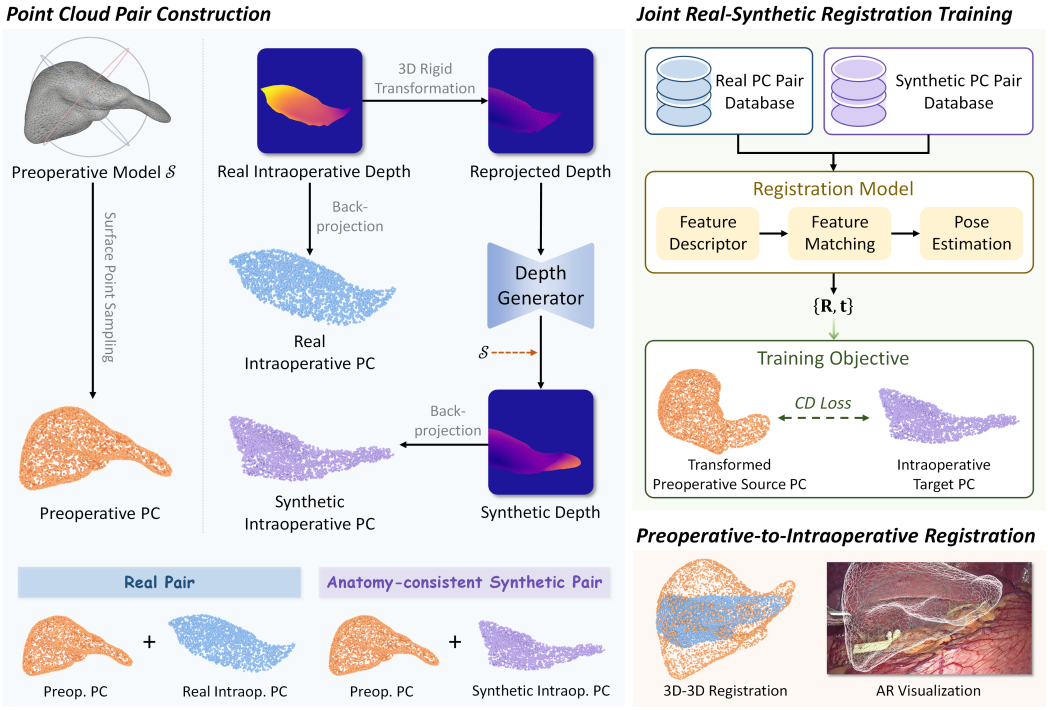


Figure 1: Overview of using our *AID-GenReg* to facilitate preoperative-to-intraoperative point cloud registration in liver surgery.

joint registration training on a combined database of real and synthetic pairs. In the **first** stage, *AID-GenReg* constructs real and synthetic registration pairs from patient-specific preoperative anatomy $\mathcal{S}$ and intraoperative depth observations. To construct a real pair, we sample a preoperative (source) point cloud from $\mathcal{S}$ and back-project a real intraoperative depth map into an intraoperative (target) point cloud. For synthetic pair construction, we reproject the real intraoperative depth map under a randomly sampled 3D rigid transformation and complete the resulting target-view observation using our anatomical identity-preserving depth generator guided by $\mathcal{S}$. The generated depth is then back-projected into 3D and paired with the corresponding preoperative point cloud to form an anatomy-consistent synthetic pair. In the **second** stage, the assembled real and synthetic pair databases are jointly used for self-supervised training of registration models, improving 3D-3D preoperative-to-intraoperative registration for AR guidance.

3.2 Anatomical Identity-Preserving Depth Generator

Overview. To generate reliable intraoperative data for self-supervised point cloud registration, synthetic depth should preserve the patient-specific anatomical identity of its paired preoperative model. Otherwise, the resulting pairs may provide misleading geometric supervision and deteriorate registration learning. To this end, we propose a diffusion-based anatomical identity-preserving depth generator for novel-view intraoperative depth synthesis. As shown in Fig. 2, the generator models the distribution of intraoperative depth observations during training. During inference, it conditions on the reprojected intraoperative observation and camera intrinsics while leveraging the patient-specific preoperative model for anatomical guidance to synthesize target-view depth maps. The generation process introduces two synergistic mechanisms, **Geometry Preservation** and **Anatomy-Guided Completion**, to promote anatomical consistency between generated depth and preoperative model.

3.2.1 Depth Distribution Learning

During training, the generator learns the intraoperative depth distribution from multi-view observations by denoising their corrupted counterparts with a diffusion method (Ho et al., 2020). Given a clean intraoperative depth map $x_0 \sim q(x)$, the T-step forward process progressively adds Gaussian

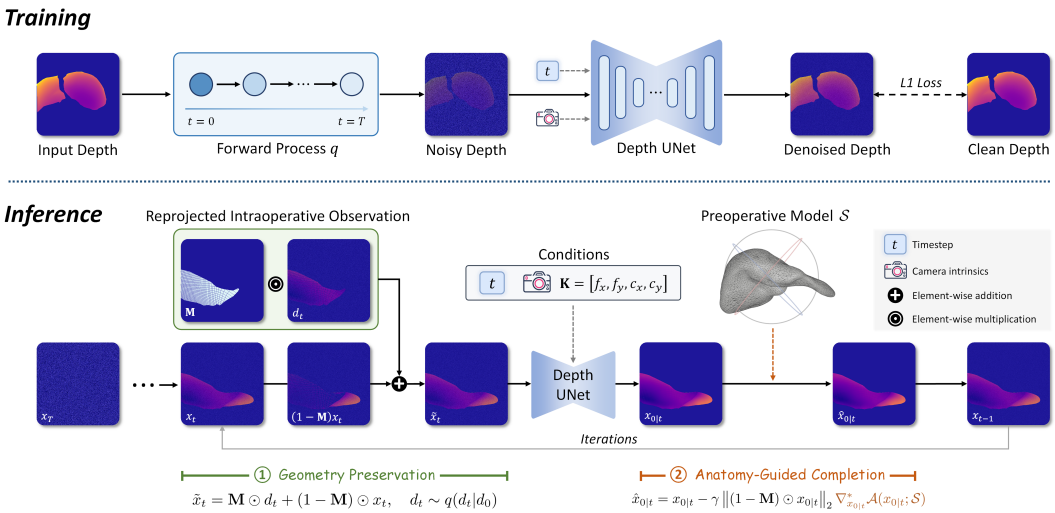


Figure 2: Pipeline of our anatomical identity-preserving depth generator. f_x and f_y denote focal lengths, while c_x and c_y represent principal-point coordinates. ∇^* is the normalized gradient in the missing region, where $\mathcal{A}$ is our proposed anatomical energy.

noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, producing a sequence of noisy depth maps $\{x_t\}_{t=1}^{T}$, where $x_t \sim q(x_t|x_0)$. The denoising process is parameterized by a depth UNet (Ronneberger et al., 2015). Inspired by (Chen et al., 2024), we incorporate camera intrinsics $\mathbf{K}$ into the depth UNet as a control signal, enabling it to generate depth maps that conform to the specified imaging geometry. In our implementation, the conditioned UNet directly predicts the clean depth $x_{0|t}$ from x_t and is optimized with an L_1 reconstruction loss.

3.2.2 Intraoperative Observation Reprojection

To synthesize realistic novel-view depth, we devise an intraoperative observation reprojection strategy that simulates changes in the visible surgical field caused by laparoscope movement. Given a real intraoperative depth map $x_0 \in \mathbb{R}^{H \times W}$, we first back-project it into a 3D point cloud using camera intrinsics $\mathbf{K}$ and apply a randomly sampled rigid transformation $\mathbf{T} \in SE(3)$ parameterized by a rotation $\mathbf{R} \in SO(3)$ and a translation $\mathbf{t} \in \mathbb{R}^3$. The transformed point cloud is then projected onto the target image plane to obtain the reprojected depth map $d_0 \in \mathbb{R}^{H \times W}$. After the viewpoint transformation, d_0 is inherently incomplete since some points fall outside the laparoscopic field of view, and the sparsity of the point cloud leaves several pixels without valid depth. Here, we define a binary mask $\mathbf{M} \in \{0, 1\}^{H \times W}$ over d_0, where $\mathbf{M}$ marks the **observable region** containing valid reprojected depths, while its complement identifies the **missing region** without valid depth. The resulting d_0 retains anatomical cues from the real observation and provides geometric conditioning for realistic target-view depth synthesis.

3.2.3 Geometry Preservation in Observable Regions

To preserve the patient-specific anatomical identity carried by the reprojected depths, we introduce Geometry Preservation, which anchors valid measurements in observable regions throughout denoising. Inspired by diffusion-based image inpainting (Lugmayr et al., 2022; Zhang et al., 2025), we decouple geometry preservation in observable regions from depth completion in missing regions. At each denoising timestep t, we first perturb the reprojected depth map d_0 using the forward diffusion process q, yielding $d_t \sim q(d_t|d_0)$. Then, we form the denoising input by assigning the forward-diffused measurements to observable regions while retaining the current noisy values in missing regions:

$$\tilde{x}_t = \mathbf{M} \odot d_t + (1 - \mathbf{M}) \odot x_t, \tag{1}$$

where $\odot$ denotes element-wise multiplication. The clean depth $x_{0|t}$ is subsequently predicted by the depth UNet. At the end of inference, we restore d_0 in observable regions to ensure exact geometric

consistency. This design accurately preserves the observed geometry and provides reliable anchors for subsequent completion and downstream registration training.

3.2.4 Anatomy-Guided Completion in Missing Regions

Only equipping the generator with Geometry Preservation may produce missing-region geometry that appears plausible yet deviates from the paired preoperative anatomy, yielding unreliable synthetic pairs for downstream registration. This occurs because missing-region completion relies solely on the learned intraoperative depth distribution and reprojected observation, without explicit guidance from the preoperative model. To solve this, we introduce Anatomy-Guided Completion, which incorporates the paired preoperative anatomy as guidance for missing-region depth synthesis.

Anatomical Energy for Guidance. Let x denote a clean intraoperative depth map and $\mathcal{S}$ the corresponding preoperative model. We define a differentiable anatomical energy $\mathcal{A}(x;\mathcal{S})$ that quantifies their anatomical discrepancy via a 3D point cloud distance. Specifically, we sample a preoperative point cloud $\mathcal{P}_{\mathcal{S}}$ from $\mathcal{S}$ and back-project x into an intraoperative point cloud $\mathcal{P}(x;\mathbf{K})$ using camera intrinsics $\mathbf{K}$. To avoid unreliable geometric measurements under large pose discrepancies, we coarsely align $\mathcal{P}_{\mathcal{S}}$ to $\mathcal{P}(x;\mathbf{K})$ using a registration model pretrained on simulation data, yielding $\widetilde{\mathcal{P}}_{\mathcal{S}}$. Since the intraoperative point cloud partially observes the complete preoperative anatomy, we base $\mathcal{A}$ on the unidirectional Chamfer distance (UCD) (Gong et al., 2022), using its squared-distance formulation from $\mathcal{P}(x;\mathbf{K})$ to $\widetilde{\mathcal{P}}_{\mathcal{S}}$:

$$\mathcal{A}(x;\mathcal{S}) = \frac{1}{|\mathcal{P}(x;\mathbf{K})|} \sum_{\mathbf{p}\in\mathcal{P}(x;\mathbf{K})} \min_{\mathbf{q}\in\widetilde{\mathcal{P}}_{\mathcal{S}}} \|\mathbf{p}-\mathbf{q}\|_2^2 . \tag{2}$$

Anatomical Guidance from the Preoperative Model. Inspired by MPGD (He et al., 2024), at each denoising step, we evaluate the anatomical energy at the predicted clean depth $x_{0|t}$ and update $x_{0|t}$ in the direction of $-\nabla_{x_{0|t}}\mathcal{A}(x_{0|t};\mathcal{S})$. The clean-estimate guidance update is given by:

$$\hat{x}_{0|t} = x_{0|t} - c_t \nabla_{x_{0|t}} \mathcal{A}(x_{0|t};\mathcal{S}), \tag{3}$$

where $c_t > 0$ is a time-dependent guidance step size. Eq. (3) performs a local gradient-descent step on $x_{0|t}$, encouraging the clean estimate to better match the preoperative anatomy. Unlike methods such as FreeDoM (Yu et al., 2023), this update avoids backpropagation through the denoising network, thereby reducing computational and memory overhead.

We use the complementary mask $1-\mathbf{M}$ to restrict the guidance update to missing regions, leaving observable regions unchanged. In Eq. (3), the effective guidance depends on both the raw anatomical gradient norm and the predicted depth scale, complicating the choice of c_t. We therefore normalize the missing-region gradient to unit ℓ_2 norm and scale it by $\|(1-\mathbf{M})\odot x_{0|t}\|_2$, simplifying guidance coefficient tuning through direct control of the relative guidance update magnitude. Moreover, we apply anatomical guidance only during the early denoising phase ($t > t_g$) to strike a trade-off between identity faithfulness and inference efficiency. Accordingly, our Anatomy-Guided Completion updates $x_{0|t}$ as follows:

$$\hat{x}_{0|t} = \begin{cases} x_{0|t} - \gamma \left\|(1-\mathbf{M})\odot x_{0|t}\right\|_2 \bar{\mathbf{g}}_t, & t > t_g, \\ x_{0|t}, & t \le t_g, \end{cases} \tag{4}$$

wherein,

$$\bar{\mathbf{g}}_t = \frac{(1-\mathbf{M})\odot \nabla_{x_{0|t}}\mathcal{A}(x_{0|t};\mathcal{S})}{\left\|(1-\mathbf{M})\odot \nabla_{x_{0|t}}\mathcal{A}(x_{0|t};\mathcal{S})\right\|_2}. \tag{5}$$

Here, t_g denotes the guidance cutoff, and $\gamma > 0$ is a predefined anatomical guidance coefficient that controls the guidance strength relative to the missing-region depth norm. The overall procedure for anatomical identity-preserving depth generation is illustrated in Fig. 2.

4 Experiment

4.1 Experimental Setup

Dataset. We evaluate *AID-GenReg* on P2I-LReg (Zhou et al., 2025), a publicly available dataset for preoperative-to-intraoperative liver point cloud registration comprising two subsets: (1) **In vivo**

Table 1: Performance comparison of training paradigms and generative data construction methods across registration models. A, B, and C denote Sim-Pretrain, Real-FT, and RealSyn-FT, respectively. Chamfer accuracy and Dice are reported in %, and Chamfer error in mm. Best results within each model are bolded.

Registration Model	Training Paradigm	Chamfer Accuracy ↑			Chamfer Error ↓		Dice ↑	
		8	10	15	Mean	Median	Mean	Median
CoFiNet (Yu et al., 2021)	A	6.52	13.04	46.74	18.39	16.08	65.78	67.04
	B	8.70	18.48	53.26	16.07	14.30	68.67	70.42
	$C_{\text{SIRA-PCR}}$	5.43	15.22	51.09	17.64	14.79	69.85	72.31
	$C_{\text{PointRegGPT}}$	7.61	16.30	52.17	15.91	14.37	69.63	71.72
	C_{Ours}	**11.96**	**27.17**	**71.74**	**14.07**	**12.17**	**73.17**	**76.01**
GeoTransformer (Qin et al., 2023)	A	16.30	**31.52**	52.17	17.56	14.80	**73.16**	**73.86**
	B	11.96	23.91	40.22	17.96	16.57	70.97	71.63
	$C_{\text{SIRA-PCR}}$	10.87	21.74	47.83	16.45	15.99	69.62	71.05
	$C_{\text{PointRegGPT}}$	15.22	26.09	47.83	16.17	15.42	70.63	71.26
	C_{Ours}	**17.39**	**31.52**	**61.96**	**14.52**	**13.82**	71.36	72.16
P2P (Yang et al., 2025)	A	43.48	68.48	**97.83**	8.86	8.58	78.88	79.44
	B	43.48	70.65	96.74	8.76	8.39	79.21	79.51
	$C_{\text{SIRA-PCR}}$	42.39	69.57	**97.83**	8.78	8.45	79.24	79.93
	$C_{\text{PointRegGPT}}$	39.13	64.13	96.74	9.01	8.63	78.70	79.51
	C_{Ours}	**48.91**	**71.74**	**97.83**	**8.56**	**8.18**	**79.67**	**80.10**
Self-P2IR (Zhou et al., 2025)	A	41.30	72.83	96.74	8.85	8.35	79.13	79.98
	B	45.65	73.91	96.74	8.70	8.16	79.31	80.15
	$C_{\text{SIRA-PCR}}$	43.48	73.91	96.74	8.74	8.25	79.24	80.01
	$C_{\text{PointRegGPT}}$	42.39	70.65	**97.83**	8.73	8.23	79.00	79.27
	C_{Ours}	**51.09**	**76.09**	**97.83**	**8.48**	**7.99**	**79.61**	**80.56**

clinical data. The released in vivo subset contains 344 keyframes extracted from laparoscopic liver resection videos of 21 patients with hepatocellular carcinoma. Each keyframe is accompanied by a liver segmentation mask, while each surgical case provides a CT-derived preoperative liver model and calibrated camera intrinsics. Following the official protocol, we use 16 patient cases for training and the remaining 5 for testing, yielding 252 training and 92 test point cloud pairs, respectively. (2) **Physics-based simulation data.** The simulation subset provides over 2,000 rendered RGB-D images per patient-specific liver model, spanning diverse laparoscopic viewpoints and scene compositions with surgical instrument models and simulated surrounding anatomy, resulting in 49,000 point cloud pairs in total.

Implementation Details. We implement *AID-GenReg* in PyTorch. For each laparoscopic keyframe, we estimate a relative depth map using Depth Anything 3 (Lin et al., 2025) and apply the annotated binary mask to zero out non-liver pixels. All depth maps are resized and zero-padded to 512×512. To reconstruct each in vivo intraoperative point cloud for registration, we back-project the valid liver depth pixels into 3D using the adjusted camera intrinsics. Following (Zhou et al., 2025), we rescale the back-projected point cloud via principal component analysis (Wold et al., 1987) to ensure scale consistency with the corresponding preoperative model. The depth generator is trained on the in vivo depth maps for 400 epochs. By default, we use 200 inference steps and set $\gamma = 0.2$ and $t_g = 100$ for anatomy-guided completion. Unless otherwise specified, we synthesize three novel-view depth maps from each intraoperative training observation using fixed random seeds $\{66, 67, 68\}$, yielding $3\times$ as many synthetic point cloud pairs as real ones. We adopt three training paradigms for downstream registration. **Sim-Pretrain**: trained solely on simulation data; **Real-FT**: initialized from Sim-Pretrain and finetuned on real pairs; **RealSyn-FT**: likewise initialized from Sim-Pretrain and finetuned on mixed real-synthetic pairs. The latter two paradigms employ the same UCD-based self-supervised objective. More implementation details can be found in Appendix A.2.

Evaluation Metrics. We evaluate performance in both 3D and 2D. For 3D evaluation, we use UCD (Gong et al., 2022) to measure the geometric discrepancy between the partial intraoperative point cloud and the registered complete preoperative model. We report Chamfer Accuracy (Acc) as the percentage of test pairs with UCD below 8, 10, and 15 mm, and Chamfer Error as the mean and

median UCD. Following (Zhou et al., 2025), we further assess 2D alignment using the Dice score between the projection of the registered preoperative model and the annotated liver mask, reporting its mean and median.

4.2 Comparison with Generative Methods

We first compare *AID-GenReg* with two representative generative data construction approaches, SIRA-PCR (Chen et al., 2023) and PointRegGPT (Chen et al., 2024). For fairness, all methods construct augmented training sets from the same real pairs at a fixed real-to-synthetic ratio of 1:3 and are evaluated under RealSyn-FT across four registration backbones. As shown in Table 1, *AID-GenReg* consistently achieves the best performance across all registration models and metrics. Averaged across the four backbones, it surpasses SIRA-PCR by 6.80/6.52/8.97 percentage points (pp) in Acc at 8/10/15 mm, while reducing mean error by 1.50 mm. Compared with PointRegGPT, our method delivers Acc@10 gains of 10.87/5.43 pp on CoFiNet/GeoTransformer and 7.61/5.44 pp on P2P/Self-P2IR. Notably, at the stringent 8-mm threshold, neither competitor surpasses the better of Sim-Pretrain and Real-FT on any backbone, whereas *AID-GenReg* consistently achieves the highest Acc@8. These results demonstrate the effectiveness of *AID-GenReg* in building anatomy-consistent synthetic pairs for downstream point cloud registration.

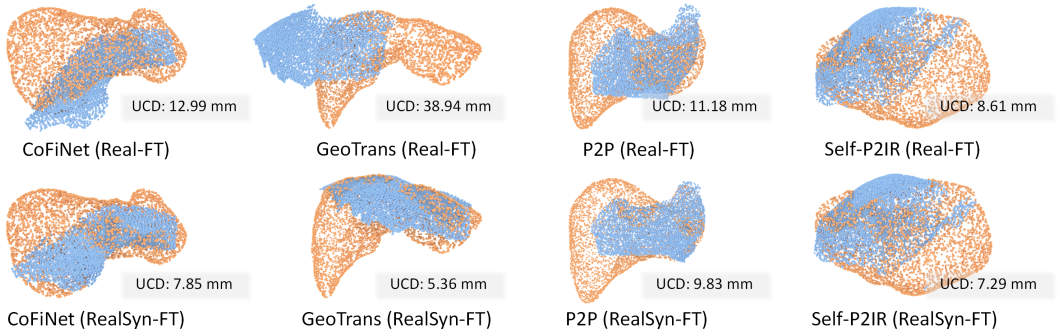


Figure 3: Qualitative comparison between the Real-FT and RealSyn-FT training paradigms across four registration models.

4.3 Evaluation across Multiple Registration Models

To further assess the generalizability of *AID-GenReg* across registration architectures, we evaluate CoFiNet, GeoTransformer, P2P with Lepard (Li & Harada, 2022), and Self-P2IR under all three training paradigms. As detailed in Table 1, the synthetic pairs constructed by *AID-GenReg* improve each registration model over its Real-FT counterpart. CoFiNet benefits markedly, gaining 3.26/8.69/18.48 pp in Acc@8/10/15 and reducing mean/median error by 2.00/2.13 mm. GeoTransformer achieves the largest Acc@15 gain of 21.74 pp and a 3.44-mm mean-error reduction, despite a slight Dice decrease from Sim-Pretrain. Despite the strong Real-FT baselines of P2P and Self-P2IR, *AID-GenReg* further enhances both models. Fig. 3 qualitatively demonstrates the robustness of *AID-GenReg* in challenging liver registration scenarios involving partial intraoperative observations with low overlap. The coverage ratio analysis in Fig. 4 offers a plausible explanation for these findings: the synthetic pairs enrich the real training distribution with diverse coverage patterns and may facilitate robust learning in downstream registration models.

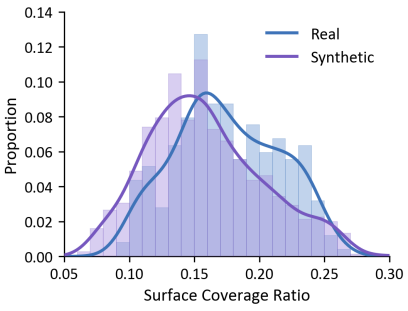


Figure 4: Surface coverage ratio distributions of real and synthetic registration pairs. Coverage is the proportion of preoperative points within 8 mm of the intraoperative surface.

4.4 Model Analysis

Module Ablation. To analyze the effects of key modules, we conduct ablation studies on CoFiNet and Self-P2IR using mixed real and constructed pairs, with Real-FT as the baseline (Table 2). The vanilla variant, relying solely on the learned intraoperative depth distribution, performs comparably to the baseline on both backbones. The reprojection (RP)-only variant performs transformation-based data construction without the depth generator. Compared with the vanilla variant, RP markedly improves CoFiNet, increasing Acc@15 by 11.96 pp and reducing mean error by 2.35 mm, but slightly decreases Acc@8 on Self-P2IR by 1.09 pp. Adding Geometry Preservation (GP) during generation boosts both backbones, yielding 63.04%/97.83% Acc@15 and 14.45/8.47 mm mean error on CoFiNet/Self-P2IR. Further incorporating Anatomy-Guided Completion (AGC) yields the best accuracy across both backbones. These results underscore the value of anatomical identity preservation in constructing clinical registration-oriented synthetic pairs.

Table 2: Ablation studies of key modules. Chamfer accuracy is shown in % and mean error in mm. Best results are bolded.

Model	Module			Chamfer Acc. / Err.		
	RP	GP	AGC	@8 ↑	@15 ↑	Err. ↓
CoFiNet (Baseline)	-	-	-	8.70	53.26	16.07
CoFiNet	×	×	×	6.52	45.65	17.53
	✓	×	×	7.61	57.61	15.18
	✓	✓	×	10.87	63.04	14.45
	✓	✓	✓	**11.96**	**71.74**	**14.07**
Self-P2IR (Baseline)	-	-	-	45.65	96.74	8.70
Self-P2IR	×	×	×	50.00	96.74	8.56
	✓	×	×	48.91	96.74	8.56
	✓	✓	×	48.91	**97.83**	**8.47**
	✓	✓	✓	**51.09**	**97.83**	8.48

Synthetic Data Scale and Anatomical Guidance Coefficient γ. Table 3 reports CoFiNet's performance under varying real-to-synthetic point cloud pair ratios and anatomical guidance coefficients. Performance improves steadily as synthetic data increases, with 1:3 yielding the best results among the evaluated ratios, showing that *AID-GenReg* can produce abundant registration point cloud pairs from limited paired clinical data, thus alleviating the scarcity of training pairs. For anatomical guidance, $\gamma = 0.2$ achieves the best performance in registration, indicating a favorable balance between anatomical identity preservation and generative diversity.

Table 3: Effects of the real-to-synthetic pair ratio and anatomical guidance coefficient γ. Accuracies are in % and errors in mm. Best results are bolded.

Pair Ratio (Real:Syn)	γ	Chamfer Acc. / Err.		
		@8 ↑	@15 ↑	Mean Err. ↓
1:0	-	8.70	53.26	16.07
1:1	0.2	8.70	60.87	15.17
1:2	0.2	10.87	65.22	14.60
1:3	0.2	**11.96**	**71.74**	**14.07**
1:3	0.1	9.78	63.04	15.17
1:3	0.2	**11.96**	**71.74**	**14.07**
1:3	0.3	8.70	63.04	15.74

5 Conclusion

We present *AID-GenReg*, the first framework to construct anatomy-consistent synthetic pairs from limited paired clinical data for self-supervised preoperative-to-intraoperative liver point cloud registration. Its diffusion-based depth generator combines Geometry Preservation and Anatomy-Guided Completion through a region-decoupled inference scheme to preserve patient-specific anatomical identity in synthetic intraoperative observations. Extensive experiments on public in vivo human data demonstrate the effectiveness and robustness of *AID-GenReg*. In the future, we will further extend *AID-GenReg* to kidney registration to validate its cross-organ generalizability.

AI use statement

The core research idea was developed independently by the authors through literature study, summary, and experimental observations. Generative AI tools were not used to originate the central idea of patient-specific anatomical identity preservation. After the research direction was established, these AI tools were used to organize relevant literature, provide suggestions on experimental design,

assist with programming and scripting, and edit the language of parts of the manuscript. Synthetic depth maps were generated using the proposed method to construct training pairs, as described in the Method and Experiment sections. AI-assisted suggestions were critically reviewed and treated only as advisory input. All methods, experimental results, and scientific claims were independently verified by the authors. AI-assisted code was reviewed and tested, and references were manually verified against the original publications. The authors take full responsibility for the final content, results, and conclusions of this study.

REFERENCES

Sharib Ali, Yamid Espinel, Yueming Jin, Peng Liu, Bianca Güttner, Xukun Zhang, Lihua Zhang, Tom Dowrick, Matthew J Clarkson, Shiting Xiao, et al. An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion from the miccai2022 challenge. *Medical image analysis*, 99:103371, 2025.

Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7163–7172, 2019.

P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

E Lee Brewer, Logan W Clements, Jarrod A Collins, Derek J Doss, Jon S Heiselman, Michael I Miga, Chris D Pavas, and Edward H Wisdom III. The image-to-physical liver registration sparse data challenge. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10951, pp. 364–370. SPIE, 2019.

Suyi Chen, Hao Xu, Ru Li, Guanghui Liu, Chi-Wing Fu, and Shuaicheng Liu. Sira-pcr: Sim-to-real adaptation for 3d point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 14394–14405, 2023.

Suyi Chen, Hao Xu, Haipeng Li, Kunming Luo, Guanghui Liu, Chi-Wing Fu, Ping Tan, and Shuaicheng Liu. Pointreggpt: Boosting 3d point cloud registration using generative point-cloud pairs for training. In *European Conference on Computer Vision*, pp. 272–289. Springer, 2024.

Marco Feuerstein, Thomas Mussack, Sandro M Heining, and Nassir Navab. Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Transactions on Medical Imaging*, 27(3):355–369, 2008.

Jingyu Gong, Fengqi Liu, Jiachen Xu, Min Wang, Xin Tan, Zhizhong Zhang, Ran Yi, Haichuan Song, Yuan Xie, and Lizhuang Ma. Optimization over disentangled encoding: Unsupervised cross-domain point cloud completion via occlusion factor manipulation. In *European Conference on Computer Vision*, pp. 517–533. Springer, 2022.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, WeiHsiang Liao, Yuki Mitsufuji, Zico Kolter, Ruslan Salakhutdinov, et al. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, volume 2024, pp. 44819–44850, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Sofiane Horache, Jean-Emmanuel Deschaud, and François Goulette. 3d point cloud registration with multi-scale architecture and unsupervised transfer learning. In *2021 international conference on 3D vision (3DV)*, pp. 1351–1361. IEEE, 2021.

Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. 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, 2021.

Yang Li and Tatsuya Harada. 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, 2022.

Haotong Lin, Sili Chen, Junhao Liew, Donny Y Chen, Zhenyu Li, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth anything 3: Recovering the visual space from any views. *arXiv preprint arXiv:2511.10647*, 2025.

Peng Liu, Bianca Güttner, Yutong Su, Chenyang Li, Jinjing Xu, Mingyang Liu, Zhe Min, Andrey Zhylka, Jasper Smit, Karin Olthof, et al. Pivots: Aligning unseen structures using preoperative to intraoperative volume-to-surface registration for liver navigation. *Medical Image Analysis*, pp. 104139, 2026.

Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9298–9309, 2023.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11461–11471, 2022.

Richard Modrzejewski, Toby Collins, Barbara Seeliger, Adrien Bartoli, Alexandre Hostettler, and Jacques Marescaux. An in vivo porcine dataset and evaluation methodology to measure soft-body laparoscopic liver registration accuracy with an extended algorithm that handles collisions. *International journal of computer assisted radiology and surgery*, 14(7):1237–1245, 2019.

Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, Slobodan Ilic, Dewen Hu, and Kai Xu. Geotransformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9806–9821, 2023.

Navid Rabbani, Lilian Calvet, Yamid Espinel, Bertrand Le Roy, Mathieu Ribeiro, Emmanuel Buc, and Adrien Bartoli. A methodology and clinical dataset with ground-truth to evaluate registration accuracy quantitatively in computer-assisted laparoscopic liver resection. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 10(4):441–450, 2022.

João Ramalhinho, Soojeong Yoo, Thomas Dowrick, Bongjin Koo, Murali Somasundaram, Kurinchi Gurusamy, David J Hawkes, Brian Davidson, Ann Blandford, and Matthew J Clarkson. The value of augmented reality in surgery—a usability study on laparoscopic liver surgery. *Medical Image Analysis*, 90:102943, 2023.

Maria R Robu, João Ramalhinho, Stephen Thompson, Kurinchi Gurusamy, Brian Davidson, David Hawkes, Danail Stoyanov, and Matthew J Clarkson. Global rigid registration of ct to video in laparoscopic liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 13(6):947–956, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Stefan Suwelack, Sebastian Röhl, Sebastian Bodenstedt, Daniel Reichard, Rüdiger Dillmann, Thiago dos Santos, Lena Maier-Hein, Martin Wagner, Josephine Wünscher, Hannes Kenngott, et al. Physics-based shape matching for intraoperative image guidance. *Medical physics*, 41(11): 111901, 2014.

Andrea Teatini, Egidijus Pelanis, Davit L Aghayan, Rahul Prasanna Kumar, Rafael Palomar, Åsmund Avdem Fretland, Bjørn Edwin, and Ole Jakob Elle. The effect of intraoperative imaging on surgical navigation for laparoscopic liver resection surgery. *Scientific Reports*, 9(1):18687, 2019.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022.

Yixuan Wang, Yue Dai, Xiangyue Yang, Binyan Huang, Ronghui Xie, Guohui Mei, and Fucang Jia. Laparoscopic liver point cloud registration via hyperbolic-topology interaction. *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2026.

Yue Wang and Justin M Solomon. Prnet: Self-supervised learning for partial-to-partial registration. *Advances in neural information processing systems*, 32, 2019.

Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

Wenbin Wu, Yifan Gao, Yixiu Wang, Jiayi Zhang, Yiming Zhao, and Xin Gao. Eg-net: An edge-guided network for rigid registration of laparoscopic low-overlap point clouds. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 159–168. Springer, 2025.

Hao Xu, Shuaicheng Liu, Guangfu Wang, Guanghui Liu, and Bing Zeng. Omnet: Learning overlapping mask for partial-to-partial point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3132–3141, 2021.

Jiaolong Yang, Hongdong Li, Dylan Campbell, and Yunde Jia. Go-icp: A globally optimal solution to 3d icp point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2241–2254, 2016.

Zixin Yang, Richard Simon, and Cristian A Linte. Learning feature descriptors for pre-and intra-operative point cloud matching for laparoscopic liver registration. *International journal of computer assisted radiology and surgery*, 18(6):1025–1032, 2023.

Zixin Yang, Jon S Heiselman, Cheng Han, Kelly Merrell, Richard Simon, and Cristian A Linte. Resolving the ambiguity of complete-to-partial point cloud registration for image-guided liver surgery with patches-to-partial matching. *IEEE journal of biomedical and health informatics*, 30 (1):459–472, 2025.

Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884, 2021.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23174–23184, 2023.

Hantao Zhang, Yuhe Liu, Jiancheng Yang, Shouhong Wan, Xinyuan Wang, Wei Peng, and Pascal Fua. Lefusion: Controllable pathology synthesis via lesion-focused diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 13232–13253, 2025.

Yirui Zhang, Yanni Zou, and Peter X Liu. Point cloud registration in laparoscopic liver surgery using keypoint correspondence registration network. *IEEE Transactions on Medical Imaging*, 44 (2):749–760, 2024.

Jun Zhou, Bingchen Gao, Kai Wang, Jialun Pei, Pheng-Ann Heng, and Jing Qin. Landmark-free preoperative-to-intraoperative registration in laparoscopic liver resection. *IEEE Transactions on Medical Imaging*, 44(11):4350–4362, 2025.

A Appendix

A.1 Summary of Liver Registration Datasets

Table 4 summarizes datasets used in existing preoperative-to-intraoperative liver registration studies. The surveyed datasets span four data types: in silico simulations, physical phantoms, porcine data, and in vivo human data. Among the few publicly available in vivo human datasets, LLR-LUS (Rabbani et al., 2022) and P2ILF (Ali et al., 2025) do not permit recovery of intraoperative liver point clouds because they lack the required liver masks or reconstructed liver surfaces. In contrast, P2I-LReg (Zhou et al., 2025) provides a liver segmentation mask for each intraoperative keyframe, thereby supporting monocular reconstruction of the liver surface. Overall, surrogate datasets remain prevalent, whereas publicly accessible paired preoperative and intraoperative data from in vivo human cases are scarce. P2I-LReg (Zhou et al., 2025) is the only publicly available in vivo human dataset that supports preoperative-to-intraoperative liver point cloud registration. Consequently, an automated framework that leverages limited clinical pairs to construct abundant point cloud training data for clinical-domain registration models is highly desirable.

Table 4: Summary of datasets used in existing preoperative-to-intraoperative liver registration studies. Cases refer to individual liver instances. Data types include in silico simulation (D1), physical phantom (D2), porcine (D3), and in vivo human (D4). The last column indicates whether paired preoperative and intraoperative data are publicly accessible. Gray rows denote public D4 datasets.

Dataset	Data Type	# Cases	Public
Open-CAS/PBSM (in silico) (Suwelack et al., 2014)	D1	3	✓
Open-CAS/PBSM (phantom) (Suwelack et al., 2014)	D2	1	✓
DePoLL (Modrzejewski et al., 2019)	D3	1	✓
Sparse Data Challenge (Brewer et al., 2019)	D2	1	✓
LLR-LUS (Rabbani et al., 2022)	D4	4	✓
LiverMatch (Yang et al., 2023)	D1	16	×
P2ILF (Ali et al., 2025)	D4	11	✓
P2P (in silico) (Yang et al., 2025)	D1	131	✓
P2P (in vitro) (Yang et al., 2025)	D2	2	✓
Suzhou (Wu et al., 2025)	D3	30	×
Fudan (Wu et al., 2025)	D4	10	×
P2I-LReg (simulation) (Zhou et al., 2025)	D1	21	✓
P2I-LReg (in vivo) (Zhou et al., 2025)	D4	21	✓
PIVOTS (AMOS) (Liu et al., 2026)	D1	360	×
PIVOTS (phantom) (Liu et al., 2026)	D2	1	×
PIVOTS (HHLBM) (Liu et al., 2026)	D4	7	×

A.2 More Implementation Details

We implement *AID-GenReg* using PyTorch 2.10.0, PyTorch3D 0.7.9, Python 3.10.18, and CUDA 12.8. Table 5 summarizes the training configurations of the proposed depth generator. For intraoperative observation reprojection, roll, pitch, and yaw are uniformly sampled within $\pm 7.5^\circ$, $\pm 18^\circ$, and $\pm 5^\circ$, respectively, and translations along the x-, y-, and z-axes within ± 15, ± 15, and ± 25 mm. To obtain reliable anatomical energy in anatomy-guided completion, we use the Sim-Pretrain version of the downstream registration backbone under evaluation for coarse alignment by default. For transformation estimator adopted by the downstream registration models, Self-P2IR Zhou et al. (2025) and GeoTransformer Qin et al. (2023) use weighted Procrustes, while P2P Yang et al. (2025) and CoFiNet Yu et al. (2021) employ Soft Procrustes. Under Sim-Pretrain, all registration models are trained solely on simulation data for 80 epochs. Self-P2IR and P2P are optimized using

Table 5: Training settings for our depth generator.

Training Settings	Value
GPU	2 × NVIDIA RTX 5880 GPUs
Optimizer	AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Initial Learning Rate	1e-4
Weight Decay	0.01
Scheduler	CosineWithWarmup
Total Epochs	400
Warmup Iterations	500
Batch Size (per GPU)	4
Random Seed	3407

SGD, and GeoTransformer and CoFiNet using Adam, all with a learning rate of 1e-3. Under Real-FT and RealSyn-FT, all models are initialized from their corresponding Sim-Pretrain checkpoints and finetuned for 30 epochs on real-only and mixed real-synthetic pairs, respectively. Across both paradigms, Self-P2IR and P2P retain a learning rate of 1e-3, while GeoTransformer and CoFiNet are finetuned at 1e-5. We use the final-epoch checkpoints of both the depth generator and the downstream registration models in all experiments.