This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Shape-Aware Registration Without Pretraining: CT-Ultrasound Bone Surface Alignment via Instance-Level Shape Completion

Luohong Wu[1], Elise Cho[1,2], Yunke Ao[1,3], Lennard Marx[1], Nicola A. Cavalcanti[1], Yiru Yang[4], Matthias Seibold[1], and Philipp Fürnstahl[1]

[1] Research in Orthopedic Computer Science, Balgrist University Hospital, University of Zurich, Zurich, Switzerland `luohong.wu@balgrist.ch`
[2] Computer Science Department, ETH Zurich, Zurich, Switzerland
[3] AI Center, ETH Zurich, Zurich, Switzerland,
[4] University of Zurich, Zurich, Switzerland

Abstract. Accurate preoperative–intraoperative registration is crucial for transferring patient-specific surgical plans to the intraoperative setting in computer- and robot-assisted orthopedic surgery. Ultrasound (US) has gained increasing attention as an intraoperative modality due to its radiation-free and real-time capabilities. However, the inherent imaging constraints, such as bone shadowing and the limited anatomical coverage during surgery, often result in incomplete US-derived bone surfaces, compromising registration accuracy and robustness. Moreover, the scarcity of paired CT–US datasets restricts the applicability of powerful learning-based approaches to different anatomies. To address these challenges, we propose ComReg, an end-to-end framework for CT–US bone surface registration via instance-specific shape completion, jointly optimizing surface completion and cross-modal registration within a unified network. In addition, we introduce a self-supervised registration strategy based on patient-specific US bone surface simulation to alleviate data scarcity, enabling effective learning of shape priors without requiring cross-instance datasets. Experiments on the publicly-available UltraBones100k dataset demonstrate the effectiveness of our approach, achieving performance comparable to state-of-the-art fully supervised methods trained on cross-instance datasets. Overall, our work provides a promising solution for cross-modality bone surface registration in limited-data settings. The code is available at: https://comreg-page.github.io.

Keywords: Bone Surface Registration · Ultrasound · Data Synthesis · Bone Shape Completion.

1 Introduction

In computer- and robot-assisted orthopedic surgery (CAOS), patient-specific surgical plans are derived from preoperative medical imaging [1]. To guide accurate execution, these plans must be accurately transferred to the intraoperative setting. This transfer relies on registration between preoperative images (e.g., CT, MRI) and intraoperative imaging modalities such as fluoroscopy and ultrasound (US) [2]. Recently, especially US has gained increasing attention as an intraoperative modality owing to its radiation-free nature and portability [3].

US-based registration has been explored using both image-based and surface-based approaches. Image-based methods align raw US volumes with preoperative imaging by maximizing a cross-modality similarity metric [4]. Surface-based methods first extract bone surface point clouds from both modalities and perform registration using either instance-specific non-learning-based methods (e.g., RANSAC followed by iterative closest point (ICP)) or cross-instance learning-based methods [5,6]. Despite these advances, even best-in-class methods remain fundamentally limited by the characteristics of B-mode US [6,9]. US inherently suffers from incomplete bone visualization due to acoustic shadowing effects [7], further compounded by intraoperative scanning constraints [8]. This partial observation degrades the robustness and accuracy of registration, particularly for long and symmetric bones [9].

Building on recent progress in shape-aware representation learning in computer vision [10], we hypothesize that patient-specific completion of intraoperative US-derived bone surfaces using preoperative priors can enhance registration performance in limited data setting. Yu et al. proposed AdaPoinTr, which leverages a transformer to model short- and long-range dependencies between observed and missing structures of natural objects [11]. Beyond standalone completion methods, recent studies have explored networks that jointly perform point cloud completion and registration, enabling more coherent structural reasoning across partial observations [12,13]. These models learn shape priors from cross-instance datasets [10].

Recent efforts have extended shape completion to bone surface reconstruction in CAOS [14,15,16]. Massalimova et al. adapted AdaPoinTr for spinal vertebra shape completion from partial RGB-D data, demonstrating promising performance [14]. Gafencu et al. used ray casting to simulate US imaging for synthetic data simulation from CT-derived lumbar spine models and trained an autoencoder-based network for shape completion [15]. Li et al. adapted PointAttN for spinal vertebra completion, followed by a PointNet-based network for anatomical landmark detection [16]. While these approaches demonstrate promising reconstruction performance, they fundamentally rely on cross-instance training datasets to learn shape priors. As a result, completion is primarily optimized for geometric fidelity rather than downstream registration robustness. Consequently, joint modeling of completion and registration under realistic CAOS data constraints remains largely unexplored [3,17].

To address these limitations, we propose ComReg, a shape-aware CT-US bone registration framework based on instance-specific shape completion. Our contributions are twofold: (1) unlike prior work, our method jointly optimizes completion and registration within a unified end-to-end architecture explicitly designed for registration under US-induced partial observations; (2) to alleviate data scarcity, we introduce a self-supervised training strategy based on patient-specific synthetic US simulation, enabling effective learning without requiring cross-instance training data. Experimental results demonstrate its effectiveness in limited-data settings. The code is available at: https://comreg-page.github.io.

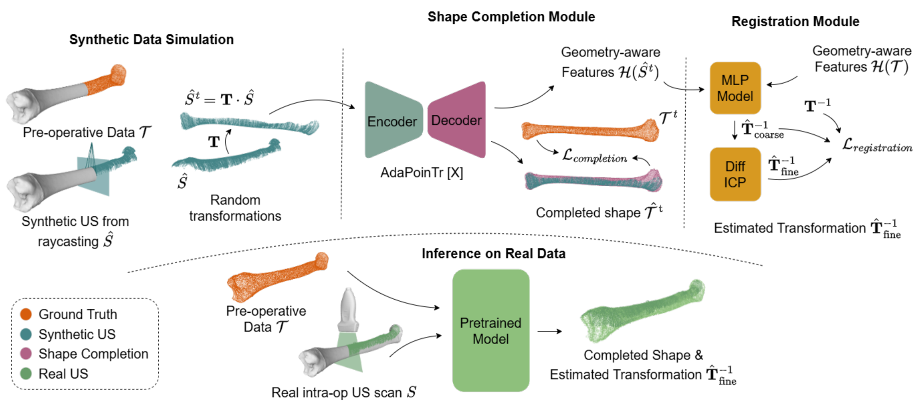

Fig. 1. Overview of ComReg. In the preoperative stage, synthetic US point clouds are generated from the patient-specific preoperative bone model to train the shape completion and registration modules. In the intraoperative stage, real US data are fed into the pretrained modules for inference.

2 Methods

Given a complete CT-derived point cloud $\mathcal{T} = \{\mathbf{t}_i \in \mathbb{R}^3\}$ and a partial US-derived point cloud $\mathcal{S} = \{\mathbf{s}_i \in \mathbb{R}^3\}$, we formulate CT–US registration under partial observations as a joint completion-and-alignment problem. A unified end-to-end network infers both the completed representation of $\mathcal{S}$ and the rigid transformation aligning $\mathcal{S}$ to $\mathcal{T}$. Figure 1 illustrates an overview of the proposed framework, which comprises a synthetic data simulation module, a shape completion module, and a registration module. To enable registration limited data conditions, the network is trained using patient-specific simulated data and deployed on real intraoperative US data at inference time. To mitigate the sim-to-real gap, we introduce a tailored synthetic US generation strategy derived directly from preoperative bone models.

Patient-specific Synthetic US Bone Surface Simulation. Given only the bone model segmented from preoperative CT, we sample a synthetic 3D point cloud $\hat{\mathcal{S}} = \{\hat{\mathbf{s}}_i \in \mathbb{R}^3\}$ from a stochastic generative process such that $\hat{\mathcal{S}} \sim p_{\hat{\mathcal{S}}}$ approximates the distribution of real US point clouds $\mathcal{S} \sim p_{\mathcal{S}}$. These synthetic observations are used to train the shape completion and registration modules. In clinical practice, 3D US volumes are reconstructed by sweeping a tracked 2D probe over the bone surface, acquiring sequential 2D slices that are compounded into a 3D volume. This acquisition process introduces spatial inconsistencies due to calibration inaccuracies, tracking errors, and patient respiration [7]. Inspired by prior work [15,16], we emulate this process by casting rays from a plane along anatomically plausible trajectories. To model acoustic shadowing at the cortical bone interface, we retain only the first intersection along each ray as the observed bone surface. To further enhance realism and downstream generalizability, we inject stochastic perturbations into the simulation process including variations in global sweep direction, local probe tilts, non-uniform sweep speed, and randomly dropped slices. To simulate segmentation noise, each ray produces a false intersection with probability 0.003, affecting both the ray itself and neighboring rays. The corresponding depths are uniformly sampled from $[0.1, 0.9]$ of the image height. Smooth inter-frame motion is enforced by generating per-slice probe tilts as a bounded random walk followed by moving-average smoothing (tilt range $|\theta_x|_{\max} = 20°$, $|\theta_y|_{\max} = 5°$; smoothing window $w_\theta = 25$). Non-uniform sweep speed is implemented by sampling slice positions $\{s_i\}_{i=0}^{N-1}$ along the sweep direction using lognormal step noise, with $\Delta = \frac{s_{\max}-s_{\min}}{N-1}$ and $\epsilon_i \sim \mathcal{N}(0,1)$. We smooth the noise using a moving-average window w_s to obtain $\tilde{\epsilon}_i = \mathrm{MA}_{w_s}(\epsilon_i)$, and define lognormal step multipliers as $m_i = \mathrm{clip}(\exp(\sigma_s\,\tilde{\epsilon}_i),\ f_{\min},\ f_{\max})$. Step sizes are computed as $\Delta s_i = \Delta\, m_i$, yielding a monotone sequence $s_{i+1} = s_i + \Delta s_i$ with $s_0 = s_{\min}$, which is finally linearly rescaled to satisfy $s_{N-1} = s_{\max}$. In our implementation, we empirically set $\sigma_s = 1.8$, $w_s = 61$, and $(f_{\min}, f_{\max}) = (0.03, 10)$.

In clinical practice, preoperative and intraoperative data are not pre-aligned, resulting in unknown initial pose discrepancies. To simulate this scenario during training, each iteration samples a random rigid transformation by first randomly selecting an axis, then applying a per-axis rotation uniformly sampled from $[-90°, 90°]$ and a translation uniformly sampled from $[-100, 100]$ mm along the selected axis. We denote $\mathbf{T} = [\mathbf{R} \mid \mathbf{t}]$ and $\mathbf{T}^{-1} = [\mathbf{R}^\top \mid -\mathbf{R}^\top\mathbf{t}]$. The transformed synthetic US point cloud is denoted as $\hat{\mathcal{S}}^{\mathrm{t}} = \mathbf{T}\cdot\hat{\mathcal{S}} = \mathbf{R}\mathbf{s}_i + \mathbf{t}$ and, similarly, the transformed CT point cloud as $\mathcal{T}^{\mathrm{t}} = \mathbf{T}\cdot\mathcal{T}$.

Shape Completion Module. Given the sampled partial point cloud $\hat{\mathcal{S}}^{\mathrm{t}}$, the shape completion module $f(\cdot)$ predicts the complete shape as $f(\hat{\mathcal{S}}^{\mathrm{t}}) = \hat{\mathcal{T}}^{\mathrm{t}} \approx \mathcal{T}^{\mathrm{t}}$. Following prior work, we adopt AdaPoinTr as the backbone, which formulates shape completion as a set-to-set translation problem [11]. $\hat{\mathcal{S}}^{\mathrm{t}}$ is first converted into a set of point proxies by sampling representative centers using farthest point sampling, followed by extracting local geometric features $\mathcal{F}$ using a DGCNN backbone [18]. A geometry-aware Transformer encoder M_E maps these proxies with positional embeddings to latent features $\mathcal{V} = M_E(\mathcal{F})$, capturing both local surface characteristics and long-range spatial relationships. Conditioned on $\mathcal{V}$, a query generator predicts representative centers of the missing regions and constructs dynamic query embeddings $\mathcal{Q}$ representing both observed and inferred regions. The Transformer decoder M_D refines these queries and outputs geometry-aware features $\mathcal{H} = M_D(\mathcal{Q}, \mathcal{V})$ representing the complete shape. Finally, a FoldingNet-based reconstruction head generates dense point sets centered at the predicted representative centers $\hat{\mathcal{C}}^{\mathrm{t}}$, yielding the completed point cloud $\hat{\mathcal{T}}^{\mathrm{t}}$ [19].

We supervise both the center prediction and point cloud reconstruction using the Chamfer Distance (CD). For a given sample $\hat{\mathcal{S}}^{\mathrm{t}}$, the center-level loss and point-level loss are defined as

$$\mathcal{L}_{\text{sparse}}(\hat{\mathcal{S}}^{\mathrm{t}}) = \frac{1}{|\hat{\mathcal{C}}^{\mathrm{t}}|} \sum_{c \in \hat{\mathcal{C}}^{\mathrm{t}}} \min_{t \in \mathcal{T}^{\mathrm{t}}} \|c - t\| + \frac{1}{|\mathcal{T}^{\mathrm{t}}|} \sum_{t \in \mathcal{T}^{\mathrm{t}}} \min_{c \in \hat{\mathcal{C}}^{\mathrm{t}}} \|t - c\|,$$

$$\mathcal{L}_{\text{dense}}(\hat{\mathcal{S}}^{\mathrm{t}}) = \frac{1}{|\hat{\mathcal{T}}^{\mathrm{t}}|} \sum_{p \in \hat{\mathcal{T}}^{\mathrm{t}}} \min_{t \in \mathcal{T}^{\mathrm{t}}} \|p - t\| + \frac{1}{|\mathcal{T}^{\mathrm{t}}|} \sum_{t \in \mathcal{T}^{\mathrm{t}}} \min_{p \in \hat{\mathcal{T}}^{\mathrm{t}}} \|t - p\|.$$

The final training objective minimizes the expected completion loss over the distribution of simulated partial inputs,

$$\mathcal{L}_{\text{completion}} = \mathbb{E}_{\hat{\mathcal{S}}^{\mathrm{t}} \sim p_{\hat{\mathcal{S}}^{\mathrm{t}}}} \left[\mathcal{L}_{\text{sparse}}(\hat{\mathcal{S}}^{\mathrm{t}}) + \mathcal{L}_{\text{dense}}(\hat{\mathcal{S}}^{\mathrm{t}}) \right].$$

Registration Module. The registration module $g(\cdot)$ aligns the sampled synthetic intraoperative point cloud $\hat{\mathcal{S}}^{\mathrm{t}}$ with the untransformed preoperative point cloud $\mathcal{T}$ by regressing the inverse rigid transformation $\mathbf{T}^{-1}$ in a coarse-to-fine manner. Exploiting the fact that decoded completion features encode complete bone geometry, we concatenate the decoded features of $\hat{\mathcal{S}}^{\mathrm{t}}$ and $\mathcal{T}$ and predict a coarse inverse transformation using $\mathrm{MLP}(\mathcal{H}(\hat{\mathcal{S}}^{\mathrm{t}}) \oplus \mathcal{H}(\mathcal{T})) \to \hat{\mathbf{T}}^{-1}_{\text{coarse}}$. The MLP comprises five blocks, each with a fully connected layer, group normalization, and GELU activation, and a hidden dimension of 512. It outputs a normalized quaternion and a translation vector.

Rather than refining $\hat{\mathbf{T}}^{-1}_{\text{coarse}}$ in a separate ICP post-processing stage as in prior work [6,20], we embed a differentiable ICP module with soft-correspondence estimation into the network [21], enabling joint end-to-end optimization. Given $\hat{\mathcal{S}}^{\mathrm{t}} = \{\hat{\mathbf{s}}_i^{\mathrm{t}}\}$, we apply $\hat{\mathbf{T}}^{-1}_{\text{coarse}}$ to obtain the coarse-aligned point cloud $\hat{\mathcal{S}}_{\text{coarse}} = \{\hat{\mathbf{s}}_i^{c}\}$ and iteratively minimize the soft point-to-point objective

$$\min_{\Delta \mathbf{T}} \sum_i \|\Delta \mathbf{T} \cdot \hat{\mathbf{s}}_i^{c} - \tilde{\mathbf{t}}_i\|^2, \qquad \tilde{\mathbf{t}}_i = \sum_j w_{ij} \mathbf{t}_j,$$

with correspondence weights $w_{ij} = \frac{\exp(-\|\hat{\mathbf{s}}_i^{c} - \mathbf{t}_j\|^2 / 2\sigma^2)}{\sum_k \exp(-\|\hat{\mathbf{s}}_i^{c} - \mathbf{t}_k\|^2 / 2\sigma^2)}$. Here, the constant σ controls the softness of the correspondence assignment, with smaller values enforcing harder, more selective correspondences. Each iteration estimates $\Delta \mathbf{T}$ via a differentiable SVD solver, yielding the refined transformation $\hat{\mathbf{T}}^{-1}_{\text{fine}}$.

Registration is supervised by a pose loss defined on rigid transformations. Given an estimated transformation $\hat{\mathbf{T}} = [\hat{\mathbf{R}} | \hat{\mathbf{t}}]$ and the ground-truth transformation $\mathbf{T} = [\mathbf{R} | \mathbf{t}]$, the pose loss is defined as $\mathcal{L}_{\text{reg}}(\hat{\mathbf{T}}, \mathbf{T}) = \|\hat{\mathbf{t}} - \mathbf{t}\|_2 + \arccos\left(\frac{\mathrm{tr}(\hat{\mathbf{R}}^\top \mathbf{R}) - 1}{2}\right)$, where the first term corresponds to the relative translation error and the second term measures the rotation error. The final registration loss is defined as

$$\mathcal{L}_{\text{registration}} = \mathbb{E}_{\hat{\mathcal{S}}^{\mathrm{t}} \sim p_{\hat{\mathcal{S}}^{\mathrm{t}}}} \left[\lambda_{\text{coarse}} \mathcal{L}_{\text{reg}}(\hat{\mathbf{T}}^{-1}_{\text{coarse}}, \mathbf{T}^{-1}) + \lambda_{\text{fine}} \mathcal{L}_{\text{reg}}(\hat{\mathbf{T}}^{-1}_{\text{fine}}, \mathbf{T}^{-1}) \right].$$

The overall objective jointly optimizes completion and registration, $\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{completion}} + \lambda_{\text{reg}}\mathcal{L}_{\text{registration}}$. Rather than a fixed weight, λ_{reg} is set adaptively at each step so that the registration term contributes a constant fraction ρ of the completion-loss magnitude, $\lambda_{\text{reg}} = \rho\,\mathcal{L}_{\text{completion}}/\mathcal{L}_{\text{registration}}$, keeping completion dominant while letting the registration gradient scale with it.

Inference on Real Data. Following patient-specific pre-training on synthetic data, the model is directly applied to real intraoperative US data $\mathcal{S}$ and the preoperative point cloud $\mathcal{T}$ without additional fine-tuning.

3 Experiments and Results

Dataset and Evaluation Metrics. We evaluate ComReg on UltraBones100k [7], the only published and publicly-available dataset that has real-world ground-truth data for US shape completion. UltraBones100k contains registered CT–US pairs of the fibula and tibia from 14 human cadaveric specimens, together with bone surface segmentation models. To evaluate alignment accuracy, we use relative rotation error (RRE) and relative translation error (RTE) between the predicted transformation matrix and the ground-truth transformation. For distance-based evaluation, we use Chamfer Distance (CD) and the 95% Hausdorff Distance (HD95) between the registered source point clouds and the target point clouds. For the completion task, we compute the CD and HD95 between the completed point cloud and the ground-truth.

Baselines. Given that ICP is often regarded as a de-facto standard when initialized close to the true alignment, we simulate this gold-standard scenario by running ICP with near-perfect initialization, denoted as pseudo ground-truth. As an instance-specific baseline, we evaluate the classical RANSAC+ICP pipeline using the Open3D implementation with default parameters[5]. For cross-instance learning-based baselines, we include the SOTA methods CASTv2 [25], GeoTransformer [24], and Predator [23], using specimen-wise leave-one-out cross-validation. Since our approach is training-free at the cross-instance level, we analyze the dependence of the baseline methods on training data by varying the training–validation splits and evaluate their performance under progressively limited training data regimes.

Results. On the registration task, the main quantitative results are summarized in Table 1. Compared with RANSAC, ComReg achieves superior accuracy with lower runtime, yielding a CD of 1.27 ± 1.54 mm, HD95 of 3.27 ± 3.95 mm, RTE of 3.19 ± 4.63 mm, and RRE of $2.93 \pm 19.98\ ^\circ$. Compared with SOTA cross-instance methods, ComReg achieves comparable performance. Table 2 further reports the performance of cross-instance methods under limited data scenarios. As the amount of training data decreases, their performance degrades consistently, whereas ComReg remains unaffected due to its training-free formulation. On the completion task, ComReg achieves a CD of 1.23 ± 0.61 mm and HD95 of 3.16 ± 2.23 mm between the completed point cloud and the ground truth. For

[5] https://www.open3d.org/docs/release/tutorial/pipelines/global_registration.html

Table 1. Quantitative results on UltraBones100k.

Method	CD (mm) ↓	HD95 (mm) ↓	RTE (mm) ↓	RRE (degree) ↓	Runtime (s) ↓
Instance-specific methods					
RANSAC100M [6]	1.76 ± 2.33	5.15 ± 7.26	7.02 ± 22.16	11.90 ± 37.95	121.86
RANSAC250M [6]	1.50 ± 1.94	4.61 ± 5.45	3.76 ± 13.14	5.65 ± 22.71	234.61
ComReg (ours)	1.27 ± 1.54	3.27 ± 3.95	3.19 ± 4.63	2.93 ± 19.98	0.28
Cross-instance methods					
CASTv2 [25]	0.95 ± 0.62	2.68 ± 0.92	2.71 ± 2.29	1.74 ± 1.53	0.19
GeoTransformer [24]	0.97 ± 0.41	2.83 ± 0.73	3.34 ± 8.55	2.13 ± 10.73	0.25
Predator [23]	0.91 ± 0.33	2.72 ± 0.67	2.36 ± 1.80	1.43 ± 1.12	0.34
Pseudo ground-truth	0.90 ± 0.13	2.12 ± 0.29	1.09 ± 0.43	1.39 ± 1.03	-

Table 2. Performance of cross-instance methods trained on a limited dataset and a single instance.

Scenario	Method	CD (mm) ↓	HD95 (mm) ↓	RTE (mm) ↓	RRE (degree) ↓
Limited ($n = 5$)	CASTv2 [25]	2.11 ± 1.57	4.73 ± 4.16	5.02 ± 5.75	11.21 ± 27.05
	GeoTransformer [24]	3.25 ± 6.13	7.17 ± 15.44	19.91 ± 51.67	21.80 ± 52.04
	Predator [23]	2.38 ± 2.19	5.15 ± 5.72	14.51 ± 43.16	16.49 ± 34.83
Instance ($n = 1$)	CASTv2 [25]	2.89 ± 3.38	8.30 ± 7.74	16.62 ± 40.33	47.05 ± 56.71
	GeoTransformer [24]	6.12 ± 9.43	13.59 ± 34.47	49.27 ± 104.13	101.21 ± 91.84
	Predator [23]	3.40 ± 3.14	9.85 ± 10.83	38.67 ± 86.60	97.45 ± 81.20
	ComReg (ours)	1.27 ± 1.54	3.27 ± 3.95	3.19 ± 4.63	2.93 ± 19.98

a comparative evaluation of different models on shape completion, we refer to previous work [14,15,16].

Ablation Studies. To isolate the effect of the differentiable refinement term, we evaluate a variant with $\lambda_{\text{fine}} = 0$ on UltraBones100k, followed by a conventional non-differentiable ICP refinement. To assess whether the decoded complete shape feature $\mathcal{H} = M_D(\mathcal{Q}, \mathcal{V})$ improves registration, we feed the encoded latent feature $\mathcal{V} = M_E(\mathcal{F})$, representing only observed regions, into the registration module. Finally, to assess whether US-based synthetic data simulation improves performance, we adopt direct random and uniform point sampling from $\mathcal{T}$, following the original AdaPoinTr implementation [11]. The results of the ablation studies are shown in Table 3.

Implementation Details. We set the point cloud size to 20,000 (including $\mathcal{T}$, $\mathcal{S}$, and $\hat{\mathcal{S}}$) and sample 2048 $\hat{\mathcal{S}} \sim p_{\hat{\mathcal{S}}}$ per $\mathcal{T}$ with a batch size of 4 in each epoch. The network is trained for 500 epochs with joint completion and registration optimization using the adaptive weight $\rho = 0.1$. We set $\lambda_{\text{coarse}} = 1$, $\lambda_{\text{fine}} = 0.5$, and $\sigma = 5 \times 10^{-3}$ in normalized coordinates. The differentiable ICP runs $K = 25$ iterations on $N_c = 4096$ subsampled points per cloud. The completion backbone uses AdamW (lr 1×10^{-4}, weight decay 5×10^{-4}), and the registration head uses Adam (lr 1×10^{-4}). Both employ a LambdaLR scheduler with a 0.9 decay factor every 20 epochs. Experiments are conducted on a workstation with an NVIDIA GeForce RTX 4090 GPU, an Intel Core i9-14900K CPU, and 64 GB RAM. All

Table 3. Quantitative results of the ablation studies.

Method	CD	HD95	RTE ($\hat{\mathbf{T}}_{\text{coarse}}^{-1}$)	RRE ($\hat{\mathbf{T}}_{\text{coarse}}^{-1}$)	RTE ($\hat{\mathbf{T}}_{\text{fine}}^{-1}$)	RRE ($\hat{\mathbf{T}}_{\text{fine}}^{-1}$)
Variant(w/o completion)	1.61 ± 1.69	4.53 ± 5.10	14.31 ± 11.12	18.92 ± 24.48	4.30 ± 7.66	5.02 ± 21.04
Variant($\lambda_{\text{fine}} = 0$)	1.67 ± 1.44	4.96 ± 5.28	13.22 ± 9.29	16.69 ± 25.14	4.01 ± 5.97	4.42 ± 18.40
Variant(uniform sampling)	2.04 ± 1.93	5.47 ± 6.28	18.71 ± 27.04	25.31 ± 33.62	8.11 ± 19.74	9.50 ± 27.59
Full model	1.27 ± 1.54	3.27 ± 3.95	10.81 ± 9.08	12.90 ± 18.74	3.19 ± 4.63	2.93 ± 19.98

hyperparameters are empirically selected. For details of the training process, we refer readers to our code[6].

4 Discussion

Cross-instance vs Instance-specific. As shown in Table 2, reducing the training data substantially degrades registration accuracy for dataset-driven methods (e.g., CD increases to 230%). This reduction underscores their sensitivity to training set size. Moreover, to the best of our knowledge, publicly-available CT–US datasets exist only for the lower leg, making it infeasible to apply these methods to other anatomies. In contrast, our approach simulates intraoperative US from patient-specific data using US imaging principles, eliminating the need for large paired datasets. This makes it suitable for new specimens and anatomies, especially in rare-disease or low-incidence cases such as bone fractures where paired MRI/CT–US data are scarce. Comparable performance highlights its feasibility.

Runtime and Clinical Feasibility. Conventional instance-specific methods such as RANSAC require online optimization during surgery, taking ≈ 230 s for RANSAC250M. This limits their applicability in surgery where real-time registration is critical, particularly for tracking bone motion. In contrast, our model is fully optimized preoperatively, and registration is completed via a single forward pass within 1 second, enabling development towards real-time use.

Ablation Studies. We hypothesized that the completed global shape features in our shape completion module facilitate coarse initialization and therefore contribute to a more robust registration performance. The higher RTE and RRE of the variant (without shape completion) in Table 3 support our hypothesis. In addition, variant ($\lambda_{\text{fine}} = 0$) underperforms the full model, suggesting that differential ICP encourages $\hat{\mathbf{T}}_{\text{coarse}}^{-1}$ to fall within the ICP convergence basin, thereby enhancing stability and lowering failure rate. When multiple coarse solutions exist (e.g., near-symmetric anatomy), supervising $\hat{\mathbf{T}}_{\text{fine}}^{-1}$ promotes ICP-resolvable initializations, whereas coarse-only supervision may converge to suboptimal local minima. Finally, the significant performance degradation of the uniform sampling variant highlights the importance of realistic synthetic data generation. Applying the original uniform point sampling strategy from AdaPoinTr increases this discrepancy, particularly for slice-wise structures and bone shadowing patterns. This domain gap ultimately degrades final inference performance.

[6] https://comreg-page.github.io.

Robustness and Limitations. ComReg assumes that the synthetic data distribution $\hat{\mathcal{S}}$ approximates the real ultrasound distribution $\mathcal{S}$. However, the current ray-casting-based simulation has several limitations. It does not model realistic segmentation noise distributions. It neglects ultrasound-specific imaging artifacts, such as angle-dependent reflectivity and refraction. These limitations increase the domain gap and degrade performance. The remaining discrepancy to the pseudo ground-truth suggests substantial room for improvement. Therefore, more advanced simulation strategies should be explored in future work [26,27]. In addition, rare flipped completion results ($\approx 1.46\%$) produce flipped registration outcomes ($\approx 180°$), resulting in a substantially larger RRE standard deviation than that of other methods. This suggests that mechanisms for detecting and mitigating flipped failures should be considered in future work. Finally, although ComReg is anatomically agnostic in principle, its generalization across different anatomies remains to be validated.

5 Conclusion

We proposed a framework for CT–US bone surface registration that does not rely on paired training datasets. To address the data scarcity in CAOS, ComReg enables self-supervision by generating patient-specific synthetic data based on ultrasound imaging principles. The results demonstrate competitive performance comparable to cross-instance baselines. Overall, this work provides a practical approach for robust, real-time, end-to-end bone surface registration under limited-data conditions.

Acknowledgments. This research has been funded by the Innosuisse Flagship project PROFICIENCY No. PFFS-21-19. This work has also been supported by the OR-X, a Swiss national research infrastructure for translational surgery, and associated funding by the University of Zurich, Switzerland and University Hospital Balgrist . This work is based on experiments performed at the Swiss Center for Musculoskeletal Imaging, SCMI, Balgrist Campus AG, Zürich.

Disclosure of Interests. None

References

1. Cleary, K., Peters, T.M.: Image-guided interventions: technology review and clinical applications. Annu. Rev. Biomed. Eng. **12**(1), 119–142 (2010)
2. Picard, F., Deakin, A.H., Riches, P.E., Deep, K., Baines, J.: Computer assisted orthopaedic surgery: past, present and future. Med. Eng. Phys. **72**, 55–65 (2019)
3. Hohlmann, B., Broessner, P., Radermacher, K.: Ultrasound-based 3D bone modelling in computer assisted orthopedic surgery–a review and future challenges. Comput. Assist. Surg. **29**(1), 2276055 (2024)
4. Pandey, P., Guy, P., Hodgson, A.J., Abugharbieh, R.: Fast and automatic bone segmentation and registration of 3D ultrasound to CT for the full pelvic anatomy: a comparative study. Int. J. Comput. Assist. Radiol. Surg. **13**(10), 1515–1524 (2018)

5. Ma, L., Zhao, Z., Chen, F., Zhang, B., Fu, L., Liao, H.: Augmented reality surgical navigation with ultrasound-assisted registration for pedicle screw placement: a pilot study. Int. J. Comput. Assist. Radiol. Surg. **12**(12), 2205–2215 (2017)
6. Ao, Y., Esfandiari, H., Carrillo, F., Laux, C.J., As, Y., Li, R., Van Assche, K., Davoodi, A., Cavalcanti, N.A., Farshad, M., et al.: SafeRPlan: Safe deep reinforcement learning for intraoperative planning of pedicle screw placement. Med. Image Anal. **99**, 103345 (2025)
7. Wu, L., Cavalcanti, N.A., Seibold, M., Loggia, G., Reissner, L., Hein, J., Beeler, S., Viehoefer, A., Wirth, S., Calvet, L., Fuernstahl, P.: UltraBones100k: A reliable automated labeling method and large-scale dataset for ultrasound-based bone surface extraction. Comput. Biol. Med. **194**, 110435 (2025). https://doi.org/10.1016/j.compbiomed.2025.110435
8. Wu, L., Seibold, M., Cavalcanti, N.A., Loggia, G., Reissner, L., Sigrist, B., Hein, J., Calvet, L., Viehöfer, A., Fürnstahl, P.: UltraBoneUDF: Self-supervised bone surface reconstruction from ultrasound based on neural unsigned distance functions. Comput. Med. Imaging Graph. 102690 (2025)
9. Ciganovic, M., Ozdemir, F., Pean, F., Fuernstahl, P., Tanner, C., Goksel, O.: Registration of 3D freehand ultrasound to a bone model for orthopedic procedures of the forearm. Int. J. Comput. Assist. Radiol. Surg. **13**(6), 827–836 (2018)
10. Fei, B., Yang, W., Chen, W.-M., Li, Z., Li, Y., Ma, T., Hu, X., Ma, L.: Comprehensive review of deep learning-based 3D point cloud completion processing and analysis. IEEE Trans. Intell. Transp. Syst. **23**(12), 22862–22883 (2022)
11. Yu, X., Rao, Y., Wang, Z., Lu, J., Zhou, J.: AdaPoinTr: Diverse point cloud completion with adaptive geometry-aware transformers. IEEE Trans. Pattern Anal. Mach. Intell. **45**(12), 14114–14130 (2023). https://doi.org/10.1109/TPAMI.2023.3309253
12. Liu, Y., Liu, Z.: Low overlapping point cloud registration using mutual prior based completion network. IEEE Trans. Image Process. **33**, 4781–4795 (2024). https://doi.org/10.1109/TIP.2024.3437234
13. Chen, G., Wang, M., Yuan, L., Yang, Y., Yue, Y.: Rethinking point cloud registration as masking and reconstruction. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 17671–17681 (2023). https://doi.org/10.1109/ICCV51070.2023.01624
14. Massalimova, A., Liebmann, F., Jecklin, S., Carrillo, F., Farshad, M., Fuernstahl, P.: SurgPointTransformer: transformer-based vertebra shape completion using RGB-D imaging. Comput. Assist. Surg. **30**(1), 2511126 (2025)
15. Gafencu, M.-A., Velikova, Y., Saleh, M., Ungi, T., Navab, N., Wendler, T., Azampour, M.F.: Shape completion in the dark: completing vertebrae morphology from 3D ultrasound. Int. J. Comput. Assist. Radiol. Surg. **19**(7), 1339–1347 (2024)
16. Li, R., Cai, Y., Davoodi, A., Borghesan, G., Vander Poorten, E.: 3D ultrasound shape completion and anatomical feature detection for minimally invasive spine surgery. Med. Biol. Eng. Comput., 1–14 (2025)
17. Pandey, P.U., Quader, N., Guy, P., Garbi, R., Hodgson, A.J.: Ultrasound bone segmentation: A scoping review of techniques and validation practices. Ultrasound Med. Biol. **46**(4), 921–935 (2020)
18. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. **38**(5), 1–12 (2019)
19. Yang, Y., Feng, C., Shen, Y., Tian, D.: FoldingNet: Point cloud auto-encoder via deep grid deformation. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 206–215 (2018)

20. Liebmann, F., von Atzigen, M., Stuetz, D., Wolf, J., Zingg, L., Suter, D., Cavalcanti, N.A., Leoty, L., Esfandiari, H., Snedeker, J.G., et al.: Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery. Med. Image Anal. **91**, 103027 (2024)
21. Wei, T., Patel, Y., Shekhovtsov, A., Matas, J., Barath, D.: Generalized differentiable RANSAC. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 17649–17660 (2023)
22. Wu, L., Seibold, M., Cavalcanti, N.A., Ao, Y., Flepp, R., Massalimova, A., Calvet, L., Fuernstahl, P.: NeuralBoneReg: A novel self-supervised method for robust and accurate multi-modal bone surface registration. arXiv preprint arXiv:2511.14286 (2025)
23. Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K.: PREDATOR: Registration of 3D Point Clouds with Low Overlap. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4265–4274 (2021). https://doi.org/10.1109/CVPR46437.2021.00425
24. Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Ilic, S., Hu, D., Xu, K.: GeoTransformer: Fast and Robust Point Cloud Registration with Geometric Transformer. IEEE Trans. Pattern Anal. Mach. Intell. **45**(8), 9806–9821 (2023). https://doi.org/10.1109/TPAMI.2023.3259038
25. Huang, R., Chai, L., Tang, Y., Li, Z., Chen, J., Li, L.: Consistency-Aware Spot-Guided Transformer for Accurate and Versatile Point Cloud Registration. IEEE Trans. Pattern Anal. Mach. Intell., 1–17 (2026). https://doi.org/10.1109/TPAMI.2026.3653989
26. Spencer, L., Cardoza, R.A., Dubey, V.K., Haese, C.E., Kreidel, F., Moussa, I., Rausch, M.K., Fuhg, J.N.: Fully differentiable ultrasound simulation utilizing raytracing. arXiv preprint arXiv:2604.15017 (2026)
27. Bertramo, N., Duguey, G., Gopalakrishnan, V.: DiffUS: Differentiable ultrasound rendering from volumetric imaging. In: International Workshop on Advances in Simplifying Medical Ultrasound, pp. 23–32. Springer (2025)