2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Towards Global Sparse and Partial Point Set Registration with Pose-Robust Completion for Computer-Assisted Orthopedic Surgery

Xinzhe Du, Yuxin Zhai, Shixing Ma, Mingyang Liu, Yi Liu, Qingfeng Yin,
Rui Song, Yibin Li, Max Q.-H. Meng, *IEEE Fellow*, Zhe Min*

***Abstract*— In computer-assisted orthopedic surgery (CAOS), accurately registering sparse and partial intraoperative point sets with a complete preoperative model remains highly challenging due to limited overlap, extreme sparsity, and point localization noise. In this paper, we propose a novel end-to-end completion then registration framework to accurately register partial and sparse point sets in CAOS. First, we develop a three-branch network that separately encodes intraoperative pose and geometry, while extracting rotation-invariant geometric priors from the preoperative model in a canonical space. This structure-aware design provides strong and beneficial cues for completing missing regions using sparse and partial data. Second, to address the sensitivity of the completion to random input poses, the completion is specifically conducted in a canonical frame and a learned $SE(3)$ transform maps the output back to the observed intraoperative space. Third, we introduce a probabilistic registration module based on a bidirectional hybrid mixture model that aligns the completed intraoperative and preoperative point sets in distribution space by jointly optimizing the source-to-target and target-to-source objectives, addressing density mismatch and geometric inconsistencies that may arise from completion. Finally, we present the individual loss formulations for both supervised and unsupervised learning paradigms, enabling robust end-to-end optimization of the entire pipeline. We systematically validate our approach on $1,757$ femur, $1,301$ hip, and 397 tibia models, as well as real-world phantom experiments. Our method achieves state-of-the-art performance under low overlap (15–30%), sparse observations (64–128 points), and large initial misalignments (up to $[-180, 180]^\circ$ rotation and $[-100, 100]mm$ translation), demonstrating strong robustness and generalization.**

I. INTRODUCTION

Registration is a core step in robotic systems, especially in surgical robotics [1]–[5]. The goal is to estimate the best rigid transformation that optimally aligns a complete preoperative bone model with possibly partial and sparse intraoperative point sets [2]. In CAOS, preoperative point sets are extracted from volumetric CT or MRI and provide a complete model of the bone [1], [2]. Intraoperative point sets are typically acquired with optically tracked pointers, which sample only accessible surface regions and yield few measurements. As a result, the intraoperative data are sparse and exhibit low overlap with the preoperative model [2]. Both sets are noisy due to image resolution limits and tracking/localization errors. Aligning such sparse, partial observations to a full model remains challenging.

These characteristics expose systematic weaknesses in prevailing methods. (i) Overlap-guided matching predicts per-point overlap to suppress non-overlapping regions and works well at moderate overlap (e.g., $\gtrsim 30\%$) [6]–[8]. At low overlap the estimates become unstable and can down-weight genuine inliers, leading to registration failures. (ii) Keypoint-free (superpoint/patch-level) matching builds superpoint or patch correspondences with coarse-to-fine refinement [9]–[11]. This strategy relies on the prediction of superpoints by stable local neighborhoods. However, when dealing with only tens of points and fragmented surfaces, representative superpoints become unreliable, and the performance of multi-level matching deteriorates. (iii) Local-descriptor pipelines design rotation-tolerant features [12], [13], but sparse neighborhoods and anisotropic depth noise reduce descriptor discriminability and weaken graph consistency, degrading rigid alignment. (iv) Probabilistic registration models each set as a mixture and aligns distributions via soft correspondences [8], [14], [15], which alleviates density mismatch and noise but degrades under extreme sparsity because components receive too few observations to estimate reliable parameters. Existing studies [16] have shown that completing partial observations before registration can improve registration robustness. However, such frameworks typically do not explicitly model pose variation in the observations, and the completion stage may therefore remain sensitive to arbitrary input poses. This limitation motivates a pose-robust completion-aware formulation that restores geometric support before alignment and remains robust under low overlap, extreme sparsity, and arbitrary pose changes.

To address these challenges, we propose an end-to-end *Completion then Registration* framework. First, to compensate for severe geometric incompleteness, we design a three-branch feature extractor: two branches operate on the intraoperative point set to encode pose and geometry, and a third branch processes the preoperative model in a canonical frame to learn rotation-invariant geometric priors, providing strong structural cues for completion. Second, to reduce sensitivity to input pose, we complete geometry in

This work was supported in part by the National Natural Science Foundation of China under Grant 62303275, National Natural Science Fund for Excellent Young Scientists Fund Program (Overseas) under Grant 22IAA01849, and Jinan Science and Technology Bureau under Grant 202333011. (*Corresponding author: Zhe Min (minzhe@sdu.edu.cn)).
Xinzhe Du, Yuxin Zhai, Shixing Ma, Mingyang Liu, Rui Song, Yibin Li, and Zhe Min are with the School of Control Science and Engineering, Shandong University, Jinan, China. Xinzhe Du and Zhe Min are also with the Shenzhen Loop Area Institute (SLAI), Shenzhen, China. Zhe Min is also with the UCL Hawkes Institute and the Department of Medical Physics and Biomedical Engineering, University College London, London, U.K. Yi Liu and Qingfeng Yin are with the Department of Orthopedic Surgery and Sports Medicine, the Second Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China. Max Q.-H. Meng is with the Shenzhen Key Laboratory of Robotics Perception and Intelligence, Southern University of Science and Technology, Shenzhen, China.

the canonical frame and then map the completed points back to the intraoperative frame using a learned $SE(3)$ transform. This improves completion stability and benefits downstream registration. Third, we introduce a probabilistic registration module based on a bidirectional Hybrid Mixture Model (HMM) that aligns the completed and preoperative sets in distribution space using source-to-target and target-to-source terms, which handles density mismatch and geometric inconsistencies. Finally, we design supervised and unsupervised losses for the joint completion–registration pipeline, yielding an end-to-end trainable framework. Our approach achieves robust registration under extremely sparse inputs, low overlap, and large initial misalignment.

Our main contributions are as follows.

1) **Completion then Registration Framework.** We propose an end-to-end Completion then Registration framework, trainable in both supervised and unsupervised modes, that delivers strong robustness to extreme sparsity and low overlap and supports initialization-free global registration.
2) **Pose-robust completion and bidirectional registration.** We propose a pose-robust completion then registration strategy for intraoperative point sets under arbitrary $SE(3)$ poses, low overlap, and extreme sparsity. Completion is performed in a canonical frame with a learned $SE(3)$ back-mapping, followed by bidirectional HMM-based probabilistic registration that improves robustness to density mismatch and completion-induced geometric inconsistencies.
3) **Comprehensive validation.** We validate the proposed framework on large-scale bone datasets (3,455 shapes) and real phantom experiments, demonstrating state-of-the-art accuracy and strong robustness under extreme sparsity, low overlap, and global-registration settings.

II. Related Work

A. Correspondence-based Registration Methods

The most widely used correspondence-based registration approaches fall into three families: (i) descriptor matching, (ii) keypoint-free (superpoint/patch-level) matching, and (iii) overlap-guided matching. Local descriptor pipelines detect salient points and compute rotation-tolerant descriptors to establish correspondences. MAC [13] constructs a compatibility graph over putative correspondences and extracts maximal cliques to generate pose hypotheses that satisfy local geometric consistency. TurboReg [17] replaces exponential maximal-clique enumeration with fixed 3-cliques and a pivot-guided search, improving both efficiency and accuracy. These methods require sufficiently dense and reliable neighborhoods to form distinctive descriptors or large cliques. Keypoint-free methods form superpoint or patch-level correspondences with coarse-to-fine refinement. RegTR [9] employs Transformers to model long-range context for superpoint correspondences. GeoTransformer [10] encodes geometric relations to improve matching under low overlap. These methods depend on the availability of stable local patches and sufficient sampling to build reliable superpoint features. Overlap-guided matching methods such as Predator [6] and RorNet [7] predict overlap confidence to suppress non-overlapping regions and improve robustness. When the intraoperative set is extremely sparse or the overlap is very small, stable superpoints and reliable overlap masks are difficult to obtain, which often reduces accuracy.

B. Correspondence-free Registration Methods

Direct regression methods estimate the rigid transform from global features without explicit correspondences. FMR [18] aligns global descriptors efficiently but is less robust in highly partial settings. EquivAlign [19] learns SE(3)-equivariant features in an RKHS and recovers pose by feature-space distance minimization. Probabilistic alignment represents point sets as mixtures and uses soft assignments to mitigate density mismatch and noise. DeepGMR learns point–to–component probabilities and then sequentially estimates mixture parameters and the rigid pose [14]. UGMM removes supervision by maximizing a likelihood-style objective that jointly refines mixture parameters and soft responsibilities along with the pose [15]. OGMM augments this paradigm with an overlap predictor that down-weights nonoverlapping regions during responsibility assignment and pose recovery, improving partial-to-full registration [8]. When overlap is very low or the intraoperative set is extremely sparse, the estimated mixtures and responsibilities can become unreliable, which motivates first completing geometry and then aligning it with normal-aware probabilistic models.

III. Method

A. Problem Formulation

Let the preoperative (complete) set be $\mathcal{Y} = \{\mathbf{y}_n \in \mathbb{R}^3\}_{n=1}^{N}$ and the sparse intraoperative set be $\mathcal{X}^s = \{\mathbf{x}_k^s \in \mathbb{R}^3\}_{k=1}^{K}$ with $K \ll N$. When available, each point has a unit normal $\widehat{\mathbf{y}}_n, \widehat{\mathbf{x}}_k^s \in \mathbb{R}^3$ (with $\|\widehat{\mathbf{y}}_n\|_2 = \|\widehat{\mathbf{x}}_k^s\|_2 = 1$). We define 6D tokens $\mathbf{d}_n^y = [\mathbf{y}_n^\top, \widehat{\mathbf{y}}_n^\top]^\top$, and the sets $\mathcal{D}_Y = \{\mathbf{d}_n^y\}$. Our network $f_\theta(\mathcal{X}^s, \mathcal{Y})$ predicts a dense completion $\mathcal{X} = \{\mathbf{x}_m\}_{m=1}^{M}$ (M=N)in the observation frame. When needed, per-point normals $\widehat{\mathbf{x}}_m$ are estimated via local PCA, yielding $\mathcal{D}_X = \{\mathbf{d}_m^x\}$ with $\mathbf{d}_m^x = [\mathbf{x}_m^\top, \widehat{\mathbf{x}}_m^\top]^\top$. We then estimate a rigid transform $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in SO(3) \times \mathbb{R}^3$ that aligns $\mathcal{D}_X$ to $\mathcal{D}_Y$ by minimizing a generic discrepancy $\mathrm{dis}(\mathbf{T}(\mathcal{D}_X), \mathcal{D}_Y), \mathbf{T}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}, \mathbf{T}(\widehat{\mathbf{x}}) = \mathbf{R}\widehat{\mathbf{x}}$. A bi-directional objective can be used as $\mathrm{dis}_{\mathrm{bi}} = \mathrm{dis}(\mathbf{T}(\mathcal{D}_X), \mathcal{D}_Y) + \mathrm{dis}(\mathbf{T}^{-1}(\mathcal{D}_Y), \mathcal{D}_X)$, where $\mathbf{T}^{-1}$ is the inverse transform. To clarify, we reuse the same symbol for a point set and its matrix representation (rows as points).

B. Completion Tri-Encoder

1) Pose Equivariant Encoder (PEE): Given a sparse intraoperative set $\mathcal{X}^s = \{\mathbf{x}_k^s \in \mathbb{R}^3\}_{k=1}^{K}$, PEE builds on Vector Neurons (VN) [20] to learn (i) an $SO(3)$-invariant global code for shape decoding and (ii) an explicit rigid pose $\mathbf{T}_{\mathrm{v}} = (\mathbf{R}_{\mathrm{v}}, \mathbf{t}_{\mathrm{v}})$ that maps a canonical completion back to the observation frame.

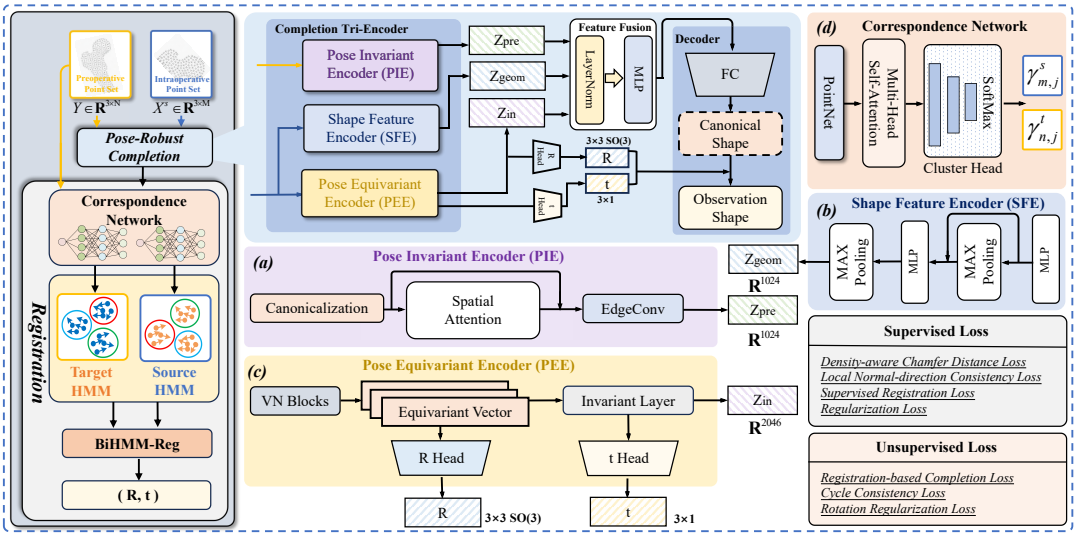

Fig. 1. Overview of the proposed pipeline. The left panel shows the end to end flow from completion to registration. Panels (a) to (d) present the key modules including the tri-encoder (PIE/SFE/PEE) and the correspondence network

a) Vector Neurons Backbone: To handle inputs under arbitrary poses, it is essential to extract *equivariant* features. We therefore adopt the Vector Neurons (VN) paradigm [20], which models each channel as a 3D vector and applies equivariant linear maps with shared scalar nonlinearities. Let $\mathbf{v}_{m,c}^{(\ell)} \in \mathbb{R}^3$ denote the c-th vector channel at point m in the ℓ-th VN block. Following the *Invariant* layer [20], we design a VN-style invariant module that learns a right-handed local frame $\mathbf{F}_m \in \mathbb{R}^{3\times3}$ for each point, where two axes are predicted and orthogonalized via Gram–Schmidt and the third is obtained by cross product. We then express the vector features in this frame as $\mathbf{y}_{m,c}^{(\ell)} = \mathbf{F}_m^\top \mathbf{v}_{m,c}^{(\ell)}$. Since $\mathbf{F}_m$ co-rotates with the input, the resulting representation is rotation-invariant. We build the global invariant code from the last-layer invariant feature by concatenating it with a point-wise mean branch, flattening the vector dimension, and applying symmetric max pooling over the point dimension. This yields the global invariant descriptor $\mathbf{z}_{\text{inv}} \in \mathbb{R}^{2046}$, which is used for completion and to condition the translation head.

b) Rotation Head ($\mathbf{R}$): From each level we take a global average pooling of the vector field, $\mathbf{G}^{(\ell)} = \text{Pool}_m(\mathbf{v}_{m,:}^{(\ell)}) \in \mathbb{R}^{C_\ell\times3}$, concatenate across levels to $\mathbf{G} \in \mathbb{R}^{C\times3}$ with $C = \sum_\ell C_\ell$, and regress a raw 3×3 matrix which is projected to the nearest rotation:

$$\mathbf{E} = \text{mat}\big(\mathbf{W}\,\text{vec}(\mathbf{G}) + \mathbf{b}\big) \in \mathbb{R}^{3\times3}, \mathbf{E} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \quad (1)$$

$$\mathbf{R}_{\text{v}} = \mathbf{U}\,\text{diag}\big(1, 1, \text{signdet}(\mathbf{U}\mathbf{V}^\top)\big)\mathbf{V}^\top. \quad (2)$$

Here $\text{Pool}_m(\cdot)$ is a symmetric pooling over points (default: global average pooling), $\text{vec}(\cdot)$ flattens $\mathbb{R}^{C\times3}$ to $\mathbb{R}^{3C}$, and $\text{mat}(\cdot)$ reshapes to $\mathbb{R}^{3\times3}$. The SVD projection and right-handed correction ensure $\mathbf{R} \in SO(3)$ and stabilize training. Optionally, we regularize $\mathbf{E}$ by $\lambda\|\mathbf{E}^\top\mathbf{E} - \mathbf{I}\|_F^2$ as a soft orthogonality prior.

c) Translation Head ($\mathbf{t}$): Translation is anchored at the centroid and refined by an invariant residual from $\mathbf{z}_{\text{inv}}$:

$$\bar{\mathbf{x}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k^s, \quad \Delta\mathbf{t}_{\text{v}} = \phi(\mathbf{z}_{\text{inv}}) \in \mathbb{R}^3, \quad \mathbf{t}_{\text{v}} = \bar{\mathbf{x}} + \Delta\mathbf{t}_{\text{v}}, \quad (3)$$

where ϕ is a small MLP. By construction, $\bar{\mathbf{x}}$ handles global translation while $\Delta\mathbf{t}_{\text{v}}$ captures object-specific shifts that are invariant to rotations.

d) Outputs: PEE returns $(\mathbf{z}_{\text{inv}}, \mathbf{R}_{\text{v}}, \mathbf{t}_{\text{v}})$. The invariant descriptor $\mathbf{z}_{\text{inv}}$ drives the decoder in a canonical frame, and the predicted pose maps the canonical completion (i.e., $\mathbf{P}_{\text{canon}}$) to the observation frame as $\mathcal{X} = \mathbf{P}_{\text{canon}}\mathbf{R}_{\text{v}}^\top + \mathbf{t}_{\text{v}}$.

2) Shape Feature Encoder (SFE): Given a partial cloud $\mathcal{X}^s = \{\mathbf{x}_k^s \in \mathbb{R}^3\}_{k=1}^K$, SFE is a PointNet-style shared MLP with a global context broadcast:

$$\mathbf{F} = \phi_1(\mathcal{X}^s) \in \mathbb{R}^{K\times256}, \mathbf{g}_1 = \max_{k=1..K}\mathbf{F} \in \mathbb{R}^{256},$$
$$\mathbf{g}_{\text{geom}} = \max_{k=1..K}\phi_2\big(\text{cat}[\mathbf{F}, \mathbf{g}_1^\top]\big) \in \mathbb{R}^{1024}, \quad (4)$$

where $\phi_1 : \mathbb{R}^3 \to \mathbb{R}^{256}$ and $\phi_2 : \mathbb{R}^{512} \to \mathbb{R}^{1024}$ are shared 1×1 Conv–BN–ReLU blocks, $\text{cat}[\cdot,\cdot]$ is channel concatenation. The resulting observation–frame global feature is $\mathbf{g}_{\text{geom}} \in \mathbb{R}^{1024}$, which is later concatenated with the invariant codes from the other branches and fed to the decoder.

3) Pose Invariant Encoder (PIE): Given a preoperative point set $\mathcal{Y} = \{\mathbf{y}_n \in \mathbb{R}^3\}_{n=1}^N$ (in a frame different from the intraoperative one), PIE produces an $SE(3)$-invariant global prior $\mathbf{z}_{\text{pre}} \in \mathbb{R}^{1024}$ that summarizes object shape.

a) Canonicalization: We select S anchors via farthest–point sampling (FPS), form a mean-centered matrix $\mathbf{P} \in \mathbb{R}^{S\times3}$, and compute its SVD $\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{R}_c^\top$. We enforce a right-handed basis by flipping one column of $\mathbf{R}_c$ if $\det(\mathbf{R}_c) < 0$. All points are then expressed in this canonical frame $\mathbf{Y} = \mathbf{Y}\mathbf{R}_c \in \mathbb{R}^{N\times3}$, which removes the global pose while preserving intrinsic geometry.

b) Spatial alignment attention: Spatial alignment attention highlights salient structures in the canonical frame. Let $\mathbf{F} = \mathrm{MLP}(\tilde{\mathbf{Y}}) \in \mathbb{R}^{N\times C}$, $\mathbf{Q},\mathbf{K} = \mathrm{Linear}_1(\mathbf{F}) \in \mathbb{R}^{N\times d}$, $\mathbf{V} = \mathrm{Linear}_2(\mathbf{F}) \in \mathbb{R}^{N\times 3}$. We compute

$$\mathbf{A}_{\text{att}} = \mathrm{softmax}\big(\mathbf{Q}\mathbf{K}^\top\big) \in \mathbb{R}^{N\times N}, \tag{5}$$

$$\tilde{\mathbf{Y}}' = \tilde{\mathbf{Y}} + \alpha\, \mathbf{A}_{\text{att}}\mathbf{V} \in \ \mathbb{R}^{N\times 3}, \tag{6}$$

with learnable α. Because attention operates on canonical coordinates and uses permutation-symmetric weighting, $\tilde{\mathbf{Y}}'$ is invariant to global rigid motions of the input.

c) EdgeConv feature extraction: We then apply EdgeConv [21] on the canonical coordinates $\tilde{\mathbf{Y}}'$ to encode local–global geometry and obtain a preoperative invariant prior via symmetric global pooling:

$$\mathbf{z}_{\text{pre}} = \mathrm{Proj}\Big(f_{\text{pool}}\big(\mathrm{EdgeConv}(\tilde{\mathbf{Y}}')\big)\Big) \in \mathbb{R}^{1024}, \tag{7}$$

Here, $f_{\text{pool}}(\cdot)$ is global max pooling over the point dimension. $\mathrm{Proj}(\cdot)$ is a linear projection to the target dimension.

C. Completion Equivariant Decoder

1) Feature Fusion: We concatenate the three global descriptors to form a single feature vector $\mathbf{f} = [\,\mathbf{z}_{\text{inv}}, \mathbf{g}_{\text{geom}}, \mathbf{z}_{\text{pre}}\,] \in \mathbb{R}^D$, where $D = 2046 + 1024 + 1024 = 4094$ in our setup. We then apply LayerNorm over the feature dimension, $\tilde{\mathbf{f}} = \mathrm{LN}(\mathbf{f})$, to stabilize optimization and balance the scales among branches. A lightweight MLP $\phi_{\text{fuse}} : \mathbb{R}^D \to \mathbb{R}^d$ produces the fused latent $\mathbf{z}_{\text{f}} = \phi_{\text{fuse}}(\tilde{\mathbf{f}}) \in \mathbb{R}^d$ (we use $d = 1024$ in all experiments). This fused latent conditions the canonical-frame decoder.

2) Equivariant Decoder: The decoder is a fully connected head that predicts a set of N canonical 3D points. Concretely, $\mathrm{vec}(\mathbf{P}_{\text{canon}}) = \phi_{\text{dec}}(\mathbf{z}_{\text{f}}) \in \mathbb{R}^{3N}$, which is reshaped to $\mathbf{P}_{\text{canon}} \in \mathbb{R}^{N\times 3}$. Using the pose $\mathbf{T} = (\mathbf{R},\mathbf{t})$ estimated by PEE (i.e., Sect. III-B.1), the observation-frame completion (i.e., intraoperative frame) is $\mathcal{X} = \mathbf{P}_{\text{canon}}\,\mathbf{R}^\top + \mathbf{t}$, where $\mathbf{t}$ is broadcast to all points.

D. Registration

We register the completed intraoperative point set tokens $\mathbf{d}_m^x = [(\mathbf{x}_m)^\top, (\widehat{\mathbf{x}}_m)^\top]^\top$ and the preoperative point set tokens $\mathbf{d}_n^y = [\mathbf{y}_n^\top, \widehat{\mathbf{y}}_n^\top]^\top$, with unit normals $\widehat{\mathbf{x}}_m$ and $\widehat{\mathbf{y}}_n$ estimated online via local PCA on k-NN neighborhoods.

1) Hybrid Mixture Model (HMM)[22]: We model each 6D geometric datum $d_n = (\mathbf{p}_n, \widehat{\mathbf{p}}_n)$, where $\mathbf{p}_n \in \mathbb{R}^3$ is a 3D position and $\widehat{\mathbf{p}}_n \in \mathbb{S}^2$ is a unit normal, using a J-component hybrid mixture that couples an isotropic Gaussian for positions and a von Mises–Fisher (vMF) distribution for normals:

$$p(d_n \mid \Theta) = \sum_{j=1}^{J} w_j \underbrace{\mathcal{N}\big(\mathbf{p}_n \,\big|\, \boldsymbol{\mu}_j,\, \sigma_j^2\mathbf{I}_3\big)}_{\text{position}} \underbrace{f_{\text{vMF}}\big(\widehat{\mathbf{p}}_n \,\big|\, \widehat{\boldsymbol{\mu}}_j,\, \kappa_j\big)}_{\text{normal}},$$

with parameters $\Theta = \{w_j, \boldsymbol{\mu}_j, \sigma_j^2, \widehat{\boldsymbol{\mu}}_j, \kappa_j\}_{j=1}^J$, $w_j \geq 0$, $\sum_j w_j = 1$, $\|\widehat{\boldsymbol{\mu}}_j\|_2 = 1$. For $\mathbb{S}^2$, the vMF density is $f_{\text{vMF}}(\widehat{\mathbf{p}} \mid \widehat{\boldsymbol{\mu}}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp\big(\kappa\, \widehat{\boldsymbol{\mu}}^\top \widehat{\mathbf{p}}\big)$.

a) EM estimation: Given responsibilities $\gamma_{n,j} = \Pr(z_n{=}j \mid d_n, \Theta)$ (E-step) with $N_j = \sum_{n=1}^N \gamma_{n,j}$ and $w_j = N_j/N$, the M-step updates are (isotropic positional covariance and vMF on $\mathbb{S}^2$):

$$\boldsymbol{\mu}_j = \frac{1}{N_j}\sum_{n=1}^{N} \gamma_{n,j}\, \mathbf{p}_n, \quad \widehat{\boldsymbol{\mu}}_j = \frac{\sum_{n=1}^N \gamma_{n,j}\, \widehat{\mathbf{p}}_n}{\|\sum_{n=1}^N \gamma_{n,j}\, \widehat{\mathbf{p}}_n\|_2}, \tag{8}$$

$$\sigma_j^2 = \frac{1}{3N_j}\sum_{n=1}^{N} \gamma_{n,j}\, \|\mathbf{p}_n - \boldsymbol{\mu}_j\|_2^2, \tag{9}$$

$$r_j = \frac{1}{N_j}\Big\|\sum_{n=1}^{N} \gamma_{n,j}\, \widehat{\mathbf{p}}_n\Big\|_2, \quad \kappa_j = \frac{r_j(3 - r_j^2)}{1 - r_j^2}. \tag{10}$$

2) Point–to–distribution Correspondence Network: We learn soft correspondences for each point set and then compute mixture parameters in a single M-step. Inputs are 6D tokens $\mathbf{d}_m^x = [\mathbf{x}_m^\top, \widehat{\mathbf{x}}_m^\top]^\top$ and $\mathbf{d}_n^y = [\mathbf{y}_n^\top, \widehat{\mathbf{y}}_n^\top]^\top$. A PointNet encoder with a T-Net maps them to per-point descriptors $\mathbf{h}_m = \phi(\mathbf{d}_m^x) \in \mathbb{R}^D$ and $\mathbf{g}_n = \phi(\mathbf{d}_n^y) \in \mathbb{R}^D$. We refine intra-set context with multi-head self-attention, $\tilde{\mathbf{h}}_m = \mathrm{MSA}(\{\mathbf{h}_{m'}\})$ and $\tilde{\mathbf{g}}_n = \mathrm{MSA}(\{\mathbf{g}_{n'}\})$. Local–global fusion concatenates each token with the set-wise global maximum, $\mathbf{h}_m^{\text{lg}} = [\tilde{\mathbf{h}}_m,\ \max_{m'} \tilde{\mathbf{h}}_{m'}]$, $\mathbf{g}_n^{\text{lg}} = [\tilde{\mathbf{g}}_n,\ \max_{n'} \tilde{\mathbf{g}}_{n'}]$. A clustering head (a linear layer with J outputs) produces J logits for each token, and a softmax over j yields $\gamma_{m,j}^s = \mathrm{softmax}_j\big(\mathrm{Cluster}(\mathbf{h}_m^{\text{lg}})\big)$ and $\gamma_{n,j}^t = \mathrm{softmax}_j\big(\mathrm{Cluster}(\mathbf{g}_n^{\text{lg}})\big)$. With $\gamma_{m,j}^s$ and $\gamma_{n,j}^t$ fixed, we compute mixture parameters by the closed-form updates in Eqs. (8)–(10) and then align the two HMMs using the closed-form $\mathrm{SE}(3)$ solver in Sec. III-E.

E. Bidirectional HMM Registration (BiHMM-Reg)

We estimate a rigid transform $(\mathbf{R},\mathbf{t}) \in \mathrm{SO}(3)\times\mathbb{R}^3$ between the completed intraoperative source $\mathcal{X} = \{(\mathbf{x}_m, \widehat{\mathbf{x}}_m)\}_{m=1}^M$ and the preoperative target $\mathcal{Y} = \{(\mathbf{y}_n, \widehat{\mathbf{y}}_n)\}_{n=1}^N$. Given fixed soft responsibilities $\boldsymbol{\Gamma}^s = [\gamma_{m,j}^s] \in \mathbb{R}^{M\times J}$ and $\boldsymbol{\Gamma}^t = [\gamma_{n,j}^t] \in \mathbb{R}^{N\times J}$, and mixture statistics $\{(\boldsymbol{\mu}_j^s, \widehat{\boldsymbol{\mu}}_j^s)\}_{j=1}^J$ and $\{(\boldsymbol{\mu}_j^t, \widehat{\boldsymbol{\mu}}_j^t)\}_{j=1}^J$, we solve for $(\mathbf{R},\mathbf{t})$ as follows.

a) Objective: We adopt a *bidirectional optimization strategy* that jointly minimizes the source→target and target→source KL divergences. Following the per-direction derivation in [14], the objective is $\mathcal{L}_{\text{bi}}(\mathbf{R},\mathbf{t}) = \mathcal{L}_{\text{fwd}} + \mathcal{L}_{\text{bwd}}$, with

$$\mathcal{L}_{\text{fwd}} = \sum_{m,j} \gamma_{m,j}^s \Big[\frac{1}{2\sigma_j^2}\, \|\mathbf{R}\mathbf{x}_m + \mathbf{t} - \boldsymbol{\mu}_j^t\|_2^2 - \kappa_j\, \widehat{\mathbf{x}}_m\, \mathbf{R}\, \widehat{\boldsymbol{\mu}}_j^t\Big],$$

$$\mathcal{L}_{\text{bwd}} = \sum_{n,j} \gamma_{n,j}^t \Big[\frac{1}{2\sigma_j^2}\, \|\mathbf{R}\boldsymbol{\mu}_j^s + \mathbf{t} - \mathbf{y}_n\|_2^2 - \kappa_j\, \big(\widehat{\boldsymbol{\mu}}_j^s\big)^\top\, \mathbf{R}\, \widehat{\mathbf{y}}_n\Big].$$

b) Closed form.: Eliminating $\mathbf{t}$ by the first order condition gives $\mathbf{t}^*(\mathbf{R}) = \bar{\boldsymbol{\mu}}_t - \mathbf{R}\bar{\mathbf{x}}$. With $\gamma_m^s = \sum_j \gamma_{m,j}^s$, $\gamma_n^t = \sum_j \gamma_{n,j}^t$, $\gamma_j^s = \sum_m \gamma_{m,j}^s$, $\gamma_j^t = \sum_n \gamma_{n,j}^t$ and $Z = \sum_m \gamma_m^s + \sum_n \gamma_n^t$, the fused centroids are

$$\bar{\mathbf{x}} = \frac{\sum_m \gamma_m^s \mathbf{x}_m + \sum_j \gamma_j^t \boldsymbol{\mu}_j^s}{Z}, \quad \bar{\boldsymbol{\mu}}_t = \frac{\sum_j \gamma_j^s \boldsymbol{\mu}_j^t + \sum_n \gamma_n^t \mathbf{y}_n}{Z}.$$

Define centered positions $\mathbf{X}_c$, $\mathbf{Y}_c$ and centered mixture means $\mathbf{M}_s$, $\mathbf{M}_t$, and their stacked normals $\mathbf{N}_x$, $\mathbf{N}_y$, $\mathbf{N}_{\mu_s}$, $\mathbf{N}_{\mu_t}$. This yields an alignment matrix independent of $\mathbf{t}$, $\mathbf{H} = \mathbf{H}_{\text{pos}} + \mathbf{H}_{\text{nor}}$, with

$$\mathbf{H}_{\text{pos}} = \tfrac{1}{\sigma^2}\left(\mathbf{X}_c^\top \mathbf{\Gamma}^s \mathbf{M}_t + \mathbf{M}_s^\top (\mathbf{\Gamma}^t)^\top \mathbf{Y}_c\right),$$

$$\mathbf{H}_{\text{nor}} = \kappa\left(\mathbf{N}_x^\top \mathbf{\Gamma}^s \mathbf{N}_{\mu_t} + \mathbf{N}_{\mu_s}^\top (\mathbf{\Gamma}^t)^\top \mathbf{N}_y\right).$$

Maximizing $\operatorname{tr}(\mathbf{R}^\top \mathbf{H})$ over $\mathbf{R} \in \mathrm{SO}(3)$ gives the Kabsch SVD solution. Let $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the singular value decomposition. The optimal rotation is

$$\mathbf{R}^* = \mathbf{V}\mathbf{S}\mathbf{U}^\top, \quad \mathbf{S} = \operatorname{diag}\big(1,\, 1,\, \operatorname{sign}(\det(\mathbf{V}\mathbf{U}^\top))\big).$$

The optimal translation is $\mathbf{t}^* = \bar{\boldsymbol{\mu}}_t - \mathbf{R}^* \bar{\mathbf{x}}$. The homogeneous transform is $\mathbf{T} = \begin{bmatrix} \mathbf{R}^* & \mathbf{t}^* \\ \mathbf{0} & 1 \end{bmatrix}$.

F. Loss Function

1) Supervised Loss Function: The supervised objective combines geometric fidelity, normal consistency, and pose regularization.

a) Density-aware Chamfer Distance (DCD): Given two point sets $\mathcal{A} = \{\mathbf{a}_i\}_{i=1}^{|\mathcal{A}|}$ and $\mathcal{B} = \{\mathbf{b}_j\}_{j=1}^{|\mathcal{B}|}$, define

$$\begin{aligned}\mathcal{L}_{\text{DCD}}(\mathcal{A}, \mathcal{B}) = \tfrac{1}{2}\Big[\frac{1}{|\mathcal{A}|}\sum_{\mathbf{a}\in\mathcal{A}}\big(1 - \tfrac{e^{-\alpha\|\mathbf{a}-\widehat{\mathbf{b}}\|^2}}{n_{\mathcal{B}}(\widehat{\mathbf{b}})^\lambda}\big) \\ + \frac{1}{|\mathcal{B}|}\sum_{\mathbf{b}\in\mathcal{B}}\big(1 - \tfrac{e^{-\alpha\|\mathbf{b}-\widehat{\mathbf{a}}\|^2}}{n_{\mathcal{A}}(\widehat{\mathbf{a}})^\lambda}\big)\Big],\end{aligned} \tag{11}$$

where $\widehat{\mathbf{b}} = \mathrm{NN}_{\mathcal{B}}(\mathbf{a})$ and $\widehat{\mathbf{a}} = \mathrm{NN}_{\mathcal{A}}(\mathbf{b})$ denote nearest neighbors, and $n_{\mathcal{B}}(\widehat{\mathbf{b}})$ (resp. $n_{\mathcal{A}}(\widehat{\mathbf{a}})$) is the number of points in $\mathcal{A}$ (resp. $\mathcal{B}$) whose nearest neighbor equals $\widehat{\mathbf{b}}$ (resp. $\widehat{\mathbf{a}}$). Hyperparameters $\alpha > 0$ and $\lambda \in [0, 1]$ control distance sharpness and density re-weighting, respectively (we set α=100, λ=0.5) [23]. For supervised completion we use $\mathcal{A} = \mathcal{X}$ and $\mathcal{B} = \tilde{\mathcal{Y}}$, where $\tilde{\mathcal{Y}} = \mathbf{T}_{\text{gt}}(\mathcal{Y})$ is the preoperative set mapped to the observation frame by the ground-truth transform.

b) Local Normal-direction Consistency (LNC): While DCD mitigates density artifacts, it may still yield isolated outliers. We therefore encourage local agreement of predicted normals on the completed set $\mathcal{X}$. For each completed point $\mathbf{x}_m$ with normal $\widehat{\mathbf{x}}_m$ and its K_{nn} nearest neighbors $\{\widehat{\mathbf{x}}_{mj}\}_{j=1}^{K_{\text{nn}}}$, define the directional discrepancy $\mathrm{D}_{mj} = 1 - \widehat{\mathbf{x}}_m^\top \widehat{\mathbf{x}}_{mj}$ and its local mean $\bar{\mathrm{D}}_m = \frac{1}{K_{\text{nn}}}\sum_{j=1}^{K_{\text{nn}}} \mathrm{D}_{mj}$. We penalize the local variance:

$$\sigma_m = \sqrt{\frac{1}{K_{\text{nn}}}\sum_{j=1}^{K_{\text{nn}}}(\mathrm{D}_{mj} - \bar{\mathrm{D}}_m)^2}, \mathcal{L}_{\text{LNC}} = \frac{1}{|\mathcal{X}|}\sum_{m=1}^{|\mathcal{X}|}\sigma_m. \tag{12}$$

c) Regularization Loss: Let $\mathbf{E} \in \mathbb{R}^{3\times 3}$ be the regressed rotation proxy and $\widehat{\mathbf{R}} = \Pi_{\mathrm{SO}(3)}(\mathbf{E})$ its projection onto $\mathrm{SO}(3)$ (via SVD with $\det = +1$ correction). We use

$$\mathcal{L}_{\text{orth}} = \left\|\mathbf{E}^\top\mathbf{E} - \mathbf{I}_3\right\|_F^2, \quad \mathcal{L}_{\text{svd}} = \left\|\mathbf{E} - \widehat{\mathbf{R}}\right\|_F^2. \tag{13}$$

d) Supervised Registration Loss: Given the ground-truth rigid motion $\mathbf{T}_{\text{gt}} \in \mathrm{SE}(3)$ that maps $\mathcal{Y}$ to the observation frame, and an estimate $\mathbf{T} \in \mathrm{SE}(3)$, we minimize the symmetric Frobenius discrepancy

$$\mathcal{L}_{\text{Reg}} = \left\|\mathbf{T}\,\mathbf{T}_{\text{gt}}^{-1} - \mathbf{I}_4\right\|_F^2 + \left\|\mathbf{T}_{\text{gt}}\,\mathbf{T}^{-1} - \mathbf{I}_4\right\|_F^2. \tag{14}$$

e) Total supervised objective: The overall supervised loss combines geometric fidelity, normal consistency, and rotation/pose regularization: $\mathcal{L}_{\text{sup}} = w_{\text{dcd}}\mathcal{L}_{\text{DCD}}(\mathcal{X}, \tilde{\mathcal{Y}}) + w_{\text{reg}}\mathcal{L}_{\text{Reg}} + w_{\text{lnc}}\mathcal{L}_{\text{LNC}} + w_{\text{orth}}\mathcal{L}_{\text{orth}} + w_{\text{svd}}\mathcal{L}_{\text{svd}}$, with unit weights ($w_{\text{dcd}}$=1, w_{reg}=0.05, w_{lnc}=0.05, w_{orth}=5e-4, w_{svd}=10e-3) by default unless otherwise stated.

2) Unsupervised Loss Function: Thanks to the bidirectional design and a closed-form recovery of the rigid motion, we directly obtain two transforms between the completed and preoperative sets: $\widehat{\mathbf{T}}_{XY} = (\widehat{\mathbf{R}}_{XY}, \widehat{\mathbf{t}}_{XY})$ and $\widehat{\mathbf{T}}_{YX} = (\widehat{\mathbf{R}}_{YX}, \widehat{\mathbf{t}}_{YX})$.

a) Registration-based completion Loss: We align each set to the other and measure the density-aware discrepancy using Eq. (11):

$$\mathcal{L}_{\text{comp}} = \mathcal{L}_{\text{DCD}}(\widehat{\mathbf{T}}_{YX}(\mathcal{Y}),\, \mathcal{X}) + \mathcal{L}_{\text{DCD}}(\widehat{\mathbf{T}}_{XY}(\mathcal{X}),\, \mathcal{Y}), \tag{15}$$

with α=100 and λ=0.5 as in the supervised case.

b) Cycle Consistency Loss: We promote inverse consistency by composing the two predictions on the same stream and penalizing the deviation from identity in point space:

$$\mathcal{L}_{\text{cycle}} = \mathcal{L}_{\text{DCD}}\big(\widehat{\mathbf{T}}_{XY}(\widehat{\mathbf{T}}_{YX}(\mathcal{Y})), \mathcal{Y}\big). \tag{16}$$

A symmetric variant on $\mathcal{X}$ can be added if desired.

c) Rotation regularization: We encourage the predicted rotation blocks to be orthonormal:

$$\mathcal{L}_{\text{regul}} = \left\|\widehat{\mathbf{R}}_{YX}^\top\widehat{\mathbf{R}}_{YX} - \mathbf{I}_3\right\|_F^2 + \left\|\widehat{\mathbf{R}}_{XY}^\top\widehat{\mathbf{R}}_{XY} - \mathbf{I}_3\right\|_F^2. \tag{17}$$

d) Total unsupervised objective: The overall loss is the weighted sum $\mathcal{L}_{\text{unsup}} = \mathcal{L}_{\text{comp}} + \beta\mathcal{L}_{\text{cycle}} + \gamma\mathcal{L}_{\text{regul}}$, with unit weights by default (β=0.5, γ=0.05) unless specified otherwise.

IV. Experimental Results

To demonstrate the effectiveness of our method, we conducted extensive simulation experiments on human femur, hip and tibia models from the MedShapeNet dataset [24] and the bonePC dataset [25]. Our method is compared against the following categories of methods: *Traditional Methods* ($\star$): ICP [26] and BCPD [27]; *Probabilistic Registration Methods* ($\triangle$): DeepGMR [14], UGMM [15], and OGMM [8]; *Overlap-guided Methods* ($\odot$): Predator [6] and RorNet [7]; *Superpoint-based Methods* ($\bullet$): RegTr [9], GeoTransformer [10], and RoITr [11]; *Descriptor Matching Methods* ($\blacktriangle$): MAC [13] and TurboReg [17] both leverage handcrafted descriptors (i.e., FPFH [12]); *Unsupervised Methods* ($\circ$): FMR [18] and EquivAlign [19]

Datasets. We use 1,301 hip and 1,399 femur models from MedShapeNet [24] and 358 femur and 397 tibia models from the BonePC dataset [25], covering anatomical regions commonly involved in computer-assisted orthopedic

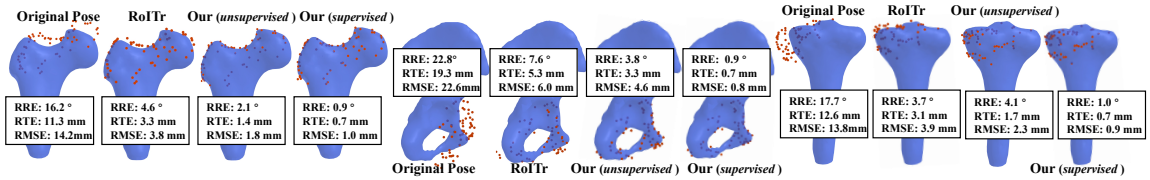

Fig. 2. Registration results on femur (left four columns), hip (middle four columns), and tibia (right four columns) at a 30% overlap ratio with 64 intraoperative points. The preoperative full mesh is rendered in blue and the sparse intraoperative points are shown in orange.

procedures such as total hip arthroplasty (THA) and total knee arthroplasty (TKA). All models are derived from clinical CT or MRI scans of real patients, as reported in the respective datasets [24], [25]. We adopt an 80/20 stratified split per category, yielding $N_{\text{train}}^{\text{hip}}$=1,041, $N_{\text{test}}^{\text{hip}}$=260; $N_{\text{train}}^{\text{femur}}$=1,406, $N_{\text{test}}^{\text{femur}}$=351; and $N_{\text{train}}^{\text{tibia}}$=318, $N_{\text{test}}^{\text{tibia}}$=79, for totals N_{train}=2,765 and N_{test}=690. All point sets are uniformly sampled to 1,024 points and normalized to $[0,1]^3$ during training, while original metric scales are preserved for error evaluation. Because the models come from different patients and clinical sites, their global poses are heterogeneous and do not follow a shared canonical orientation. This motivates our pose-robust completion in a learned canonical frame, followed by a mapping back to the observation frame.

Evaluation Metrics. To assess registration accuracy, we use the Relative Rotation Error (RRE) calculated as $\text{RRE} = \arccos[\frac{\text{tr}(\mathbf{R}_{\text{gt}}\mathbf{R}_{\text{pred}}^{\top})-1}{2}] \times \frac{180^{\circ}}{\pi}$, measured in degrees (°), and the Relative Translation Error (RTE) given by $\text{RTE} = \|\mathbf{t}_{\text{pred}} - \mathbf{t}_{\text{gt}}\|_2$, measured in millimeters (mm). We also compute the correspondence RMSE as $\text{RMSE} = \frac{1}{K}\sqrt{\sum_{k=1}^{K} \|\mathbf{T}_{\text{gt}}(\mathbf{x}_k) - \mathbf{T}_{\text{pred}}(\mathbf{x}_k)\|^2}$. Registration Recall (RR) is determined using a 10 mm RMSE threshold. **Implementation Details.** Unless otherwise stated, the number of HMM components is J=16. We train for 500 epochs using Adam (batch size 32). For supervised training, the initial learning rate is η_0=10^{-4} with a step decay at epoch 50 by a factor of γ=0.7 (i.e., $\eta \leftarrow 0.7\eta$). For unsupervised training, the initial learning rate is η_0=10^{-3} with a step decay at epoch 20 by a factor of γ=0.5. All learning-based registration models are implemented in PyTorch and trained on a single NVIDIA GeForce RTX 4090 GPU.

A. Low-overlap robustness.

We evaluate robustness under low overlap by simulating intraoperative point sets $\mathcal{X}^s$ as *partial* point sets with overlap ratios of 15% and 30%. We then apply a random rigid motion to $\mathcal{X}^s$, where the rotation is sampled in axis–angle form with the angle uniformly drawn from $[-45, 45]^{\circ}$ about a random axis, and the translation components are drawn independently from $[-50, 50]$ mm along each axis. To mimic CAOS noise characteristics (i.e., the z-axis noise is 2–5 times that of the x and y axes) [2], we inject zero-mean anisotropic Gaussian noise with covariance $\mathbf{\Sigma}_{\text{aniso}} = \text{diag}(0.25, 0.25, 2.25)\,\text{mm}^2$ (per-axis standard deviations $[0.5, 0.5, 1.5]$ mm).

We report rotation error (RRE, degrees), translation error (RTE, mm), and RMSE (mm), and benchmark against six families of baselines, namely *Traditional* (⋆), *Probabilistic Registration* (△), *Overlap-guided based* (⊙), *Superpoint–based* (●), *Descriptor–matching* (▲), and *Unsupervised* (○), as listed in Table I. As summarized in Table I, our method achieves the lowest errors across both overlap levels under *supervised* training, and delivers competitive performance in the *unsupervised* setting as well.

TABLE I

PERFORMANCE UNDER LOW-OVERLAP REGIMES. BEST RESULTS HIGHLIGHTED IN LIGHT GREEN, SECOND-BEST IN LIGHT GRAY.

Method	15%			30%		
	RRE	RTE	RMSE	RRE	RTE	RMSE
ICP [26](⋆)	20.20	17.72	20.14	22.56	19.33	24.85
BCPD [27](⋆)	24.89	32.59	43.91	20.23	28.68	31.24
DeepGMR [14](△)	18.83	21.03	25.84	17.32	20.21	23.22
UGMM [15](△/○)	31.35	26.87	30.08	25.73	23.12	27.50
OGMM [8](△)	11.84	10.35	12.76	9.56	7.19	9.63
FMR [18](○)	16.53	28.07	31.41	12.74	24.04	26.08
Predator [6](⊙)	17.92	13.46	15.65	12.94	11.93	13.12
RorNet [7](⊙)	19.23	14.52	15.09	19.13	12.67	13.37
RegTr [9](●)	9.23	8.79	8.92	7.53	9.73	10.81
GeoTrans [10](●)	**3.23**	5.61	6.28	3.21	4.29	4.70
RoITr [11](●)	3.82	6.78	6.92	**2.91**	5.02	5.33
EquivAlign [19](○)	19.42	14.79	15.38	17.21	12.83	13.57
MAC [13](▲)	8.14	8.73	9.17	8.01	8.13	8.42
TurboReg [17](▲)	8.81	8.41	9.08	7.22	8.31	8.77
Ours (*supervised*)	**1.17**	**0.98**	**1.18**	**0.83**	**0.90**	**0.97**
Ours (*unsupervised*)	5.12	**1.82**	**3.37**	4.79	**1.48**	**2.89**

TABLE II

PERFORMANCE UNDER SPARSE POINT COUNTS. BEST RESULTS HIGHLIGHTED IN LIGHT GREEN, SECOND-BEST IN LIGHT GRAY.

Method	30% 64 pts			30% 128 pts		
	RRE	RTE	RMSE	RRE	RTE	RMSE
GeoTrans [10](●)	**4.68**	6.01	6.92	4.20	5.12	6.88
RoITr [11](●)	5.61	5.37	6.81	**4.02**	5.87	6.25
EquivAlign [19](○)	22.54	16.23	16.49	21.73	14.48	14.91
MAC [13](▲)	14.31	10.28	11.65	12.23	9.84	10.12
TurboReg [17](▲)	8.82	11.03	11.54	8.84	10.61	11.09
Ours (*supervised*)	**1.27**	**1.01**	**1.26**	**1.07**	**0.97**	**1.11**
Ours (*unsupervised*)	6.79	**2.88**	**4.58**	5.57	**2.48**	**4.89**

TABLE III

PERFORMANCE UNDER LARGE POSE CHANGES. BEST RESULTS HIGHLIGHTED IN LIGHT GREEN, SECOND-BEST IN LIGHT GRAY.

Method	$[-180, 180]^{\circ}$		$[-100, 100]_{\text{mm}}$	
	RRE	RTE	RMSE	RR
GeoTrans [10](●)	**7.82**	**11.96**	**12.68**	**71.17**
RoITr [11](●)	9.79	12.37	13.95	66.09
MAC [13](▲)	14.69	18.92	20.10	56.23
TurboReg [17](▲)	17.63	17.93	19.66	57.39
Ours (*supervised*)	**6.23**	**1.17**	**5.43**	**99.57**

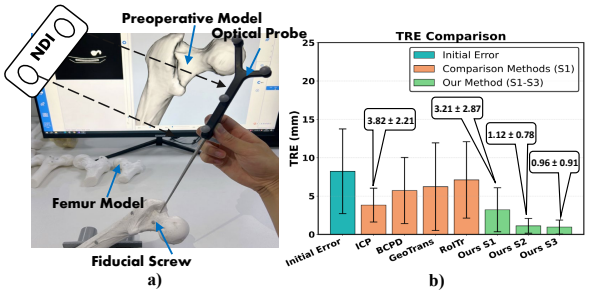

Fig. 3. Real femur phantom experiments. (a) Experimental system and acquisition workflow. (b) Quantitative registration results (TRE, mm) under adaptation protocols S1–S3 and competing baselines.

B. Sparsity robustness at fixed overlap.

To further assess robustness under extreme sparsity, we perform comparisons at a constant overlap. We adopt the protocol of Sec. IV-A and fix the overlap of the intraoperative point sets $\mathcal{X}^s$ at 30%. Each $\mathcal{X}^s$ is randomly downsampled to 64 or 128 points, while the preoperative counterpart remains dense (i.e., 1024 points). The ranges of the random rigid transform and the anisotropic noise model are identical to those in Sec. IV-A. For this setting we evaluate the top performers from Sec. IV-A, including GeoTransformer [10], RoITr [11], EquivAlign [19], MAC [13], and TurboReg [17]. As summarized in Table II, our method achieves the lowest errors under *supervised* training and remains competitive in the *unsupervised* setting. Qualitative examples under this protocol are shown in Fig. 2.

C. Global registration under large pose changes.

To further validate robustness and generalization under substantial pose variation, we perform an initialization-free global registration study. We follow the 128-point setting of Sec. IV-B and use the same anisotropic noise model. A large random rigid motion is applied to the intraoperative point set $\mathcal{X}^s$, with rotation angles uniformly sampled in $[-180, 180]^\circ$ about random axes and translation components independently sampled in $[-100, 100]$ mm along each axis. We report RRE (°), RTE (mm), RMSE (mm), and Registration Recall (RR), where RR counts cases with $\mathrm{RMSE} < 10\,\mathrm{mm}$.

Table III summarizes the results. Our method achieves the lowest RRE, RTE, and RMSE, and the highest RR (99.57%), substantially outperforming *all* compared methods. While the competing baselines claim robustness to large-pose *global* registration, we observed their performance degraded significantly in this challenging CAOS scenario. Hence, the proposed method is resilient to large initial misalignments even at 128-point sparsity with anisotropic noise, confirming its pose robustness.

D. Real phantom experiments and adaptation protocols

We evaluate our method on 3D-printed femur phantoms. Preoperative point sets (1024 pts) were extracted from CT volumes acquired on a Revolution CT (GE Healthcare, USA). Intraoperative point sets (50 pts) were collected with an optically tracked pointer using a Polaris Vega XT (Northern Digital Inc., Canada). As shown in Fig. 3b, fiducial screws provide paired target points used both to solve the ground-truth rigid transform $\mathbf{T}_{\mathrm{gt}}$ and to compute TRE. Specifically, TRE is defined as the mean Euclidean distance between the intraoperative target points transformed by the estimated rigid transform and those transformed by $\mathbf{T}_{\mathrm{gt}}$: $\mathrm{TRE} = \frac{1}{N}\sum_{n=1}^{N} \left\| \mathbf{R}_{\mathrm{pred}}\mathbf{p}_{\mathrm{tar},n}^{\mathrm{intra}} + \mathbf{t}_{\mathrm{pred}} - \left(\mathbf{R}_{\mathrm{gt}}\mathbf{p}_{\mathrm{tar},n}^{\mathrm{intra}} + \mathbf{t}_{\mathrm{gt}}\right) \right\| \mathrm{mm}$. We used six phantoms and acquired five intraoperative trials per phantom, yielding 30 paired datasets.

a) Adaptation protocols: Let $\mathcal{D}_{\mathrm{syn}}^{\mathrm{femur}}$ denote the synthetic femur training set used in prior sections and $\mathcal{D}_{\mathrm{phant}} = \{(\mathcal{P}_i, \mathcal{Q}_i)\}_{i=1}^{30}$ the phantom cohort. We split $\mathcal{D}_{\mathrm{phant}}$ into a fine-tuning set with two phantoms (10 pairs) and a held-out test set with four phantoms (20 pairs). We also define a small set of CT-derived preoperative shapes $\mathcal{S}_{\mathrm{pre}} = \{S_j\}_{j=1}^{5}$. We consider three clinically feasible training/adaptation settings:

S1 (pre-train on 'synthetic' only). Train on $\mathcal{D}_{\mathrm{syn}}^{\mathrm{femur}}$ with pose sampling $\theta \sim \mathcal{U}(-15^\circ, 15^\circ)$ (axis–angle) and $\mathbf{t} \sim \mathcal{U}(-10, 10)^3$ mm, then evaluate on $\mathcal{D}_{\mathrm{phant}}^{\mathrm{test}}$.
S2 (pre-train with five preoperative shapes). Train on $\mathcal{D}_{\mathrm{syn}}^{\mathrm{femur}} \cup \mathcal{S}_{\mathrm{pre}}$ under the same pose ranges $\theta \sim \mathcal{U}(-15^\circ, 15^\circ)$ and $\mathbf{t} \sim \mathcal{U}(-10, 10)^3$ mm, then evaluate on $\mathcal{D}_{\mathrm{phant}}^{\mathrm{test}}$. This setting examines whether a small amount of preoperative anatomy improves the learned prior.
S3 (fine-tune on 10 paired acquisitions). Start from S1 weights and fine-tune on $\mathcal{D}_{\mathrm{phant}}^{\mathrm{FT}}$ (10 pairs) with the learning rate reduced by a factor of 10 for 20 epochs, then evaluate on $\mathcal{D}_{\mathrm{phant}}^{\mathrm{test}}$.

These protocols mirror the data availability in practice and allow us to measure how lightweight adaptation improves real-world performance.

b) Experimental procedure: We follow the workflow of a standard computer-assisted THA procedure [28]. First, a coarse alignment is obtained by selecting four anatomical landmark pairs per case. After this step the mean TRE across the 20 held-out trials is 8.23 ± 5.52,mm. We then perform fine registration using our method and the baselines. Classical ICP [26] and BCPD [27] are included as widely used fine registration methods. In addition, we evaluate the strongest learning-based baselines identified in Secs. IV-A and IV-B, namely GeoTransformer [10] and RoITr [11]. Our method is tested under three settings S1–S3, while learning-based baselines use their S1-style pretrained models. Quantitative TREs are summarized in Fig. 3b. Relative to S1 (3.21 ± 2.87,mm), adding five preoperative shapes (S2) reduces TRE to 1.12 ± 0.78,mm (≈65% reduction), and fine-tuning on 10 phantom pairs (S3) further reduces it to 0.96 ± 0.91,mm (≈70% reduction) on the same 20 trials.

E. Ablation study.

We conduct ablations under the *supervised* 64-point protocol of Sec. IV-B to quantify the contribution of five components: the *Pose Equivariant Encoder* (PEE), *Shape Feature Encoder* (SFE), *Pose Invariant Encoder* (PIE), *Bidirectional optimization* (BiO) strategy, and the *Local Normal-direction Consistency* (LNC) loss. The first row in Table IV

TABLE IV
ABLATION STUDY RESULTS

Modules					Metrics			
PEE	**SFE**	**PIE**	**BiO**	**LNC**	**RRE**	**RTE**	**RMSE**	**DCD**
✓	✓	✓	✓	✓	**1.27**	**1.01**	**1.26**	**0.24**
	✓	✓	✓	✓	1.54	1.22	1.39	0.28
✓		✓	✓	✓	2.23	1.42	1.74	0.30
✓	✓		✓	✓	1.33	1.12	1.27	0.25
✓	✓	✓		✓	1.43	1.12	1.38	0.27
✓	✓	✓	✓		1.31	1.08	1.29	0.28

(blue shading) is our full model evaluated on the default setting of Sec. IV-B. We report registration metrics (RRE, RTE and RMSE) and a completion metric (DCD[23], lower is better). As summarized in Table IV, each module contributes to the final accuracy.

V. CONCLUSIONS AND FUTURE WORK

We presented an end-to-end *Completion then Registration* pipeline for CAOS that explicitly targets the hard regime of sparse, low-overlap, and arbitrarily posed intraoperative observations. The method combines a three-branch encoder that separates pose and shape cues while injecting rotation-invariant priors from the preoperative model, a pose-robust completion strategy performed in a canonical frame, and a probabilistic registration module based on a bidirectional hybrid mixture model. We designed supervised and unsupervised objectives for joint optimization. Experiments on large bone cohorts and real femur phantoms demonstrated state-of-the-art accuracy under severe sparsity, low overlap, anisotropic noise, and large pose changes.

REFERENCES

[1] Z. Min, J. Lai, and H. Ren, "Innovating robot-assisted surgery through large vision models," *Nature Reviews Electrical Engineering*, pp. 1–14, 2025.

[2] Z. Min, A. Zhang, Z. Zhang, J. Wang, S. Song, H. Ren, and M. Q.-H. Meng, "3-d rigid point set registration for computer-assisted orthopedic surgery (caos): A review from the algorithmic perspective," *IEEE Transactions on Medical Robotics and Bionics*, vol. 5, no. 2, pp. 156–169, 2023.

[3] Y. Ma, X. An, Q. Yang, M. Cai, Z. Tang, J. Chang, V. Iacovacci, T. Xu, L. Zhang, and Q. Wang, "Magnetic continuum robot for intelligent manipulation in medical applications," *SmartBot*, vol. 1, no. 2, p. e12011, 2025.

[4] Z. Feng, J. Liu, and H. Wang, "Optimizing scene flow with neural rigidity prior," *Robot Learning*, vol. 1, no. 1, pp. 1–15, 2024.

[5] Z. Min, Z. M. Baum, S. U. Saeed, S. Ma, X. Du, M. Emberton, D. C. Barratt, Z. A. Taylor, and Y. Hu, "Biomechanics-informed non-rigid medical image registration with elasticity theories," *IEEE Transactions on Medical Imaging*, 2026.

[6] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler, "Predator: Registration of 3d point clouds with low overlap," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2021, pp. 4267–4276.

[7] Y. Wu, Y. Zhang, W. Ma, M. Gong, X. Fan, M. Zhang, A. K. Qin, and Q. Miao, "Rornet: Partial-to-partial registration network with reliable overlapping representations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 11, pp. 15 453–15 466, 2023.

[8] G. Mei, F. Poiesi, C. Saltori, J. Zhang, E. Ricci, and N. Sebe, "Overlap-guided gaussian mixture models for point cloud registration," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4511–4520.

[9] Z. J. Yew and G. H. Lee, "Regtr: End-to-end point cloud correspondences with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 6677–6686.

[10] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, S. Ilic, D. Hu, and K. Xu, "Geotransformer: Fast and robust point cloud registration with geometric transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9806–9821, 2023.

[11] H. Yu, Z. Qin, J. Hou, M. Saleh, D. Li, B. Busam, and S. Ilic, "Rotation-invariant transformer for point cloud matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 5384–5393.

[12] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 3212–3217.

[13] J. Yang, X. Zhang, P. Wang, Y. Guo, K. Sun, Q. Wu, S. Zhang, and Y. Zhang, "Mac: Maximal cliques for 3d registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[14] W. Yuan, B. Eckart, K. Kim, V. Jampani, D. Fox, and J. Kautz, "Deepgmr: Learning latent gaussian mixture models for registration," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 2020, pp. 733–750.

[15] X. Huang, S. Li, Y. Zuo, Y. Fang, J. Zhang, and X. Zhao, "Unsupervised point cloud registration by learning unified gaussian mixture models," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7028–7035, 2022.

[16] X. Du, S. Ma, Z. Zhang, R. Song, Y. Li, M. Q.-H. Meng, and Z. Min, "Registration after completion: Towards sparse and partial point set registration for computer-assisted orthopedic surgery," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 7710–7717.

[17] S. Yan, P. Shi, Z. Zhao, K. Wang, K. Cao, J. Wu, and J. Li, "Turboreg: Turboclique for robust and efficient point cloud registration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 26 371–26 381.

[18] X. Huang, G. Mei, and J. Zhang, "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*, 2020, pp. 11 366–11 374.

[19] R. Zhang, Z. Zhou, M. Sun, O. Ghasemalizadeh, C.-H. Kuo, R. M. Eustice, M. Ghaffari, and A. Sen, "Correspondence-free se(3) point cloud registration in rkhs via unsupervised equivariant learning," in *Computer Vision – ECCV 2024*. Cham: Springer Nature Switzerland, 2025, pp. 68–86.

[20] C. Deng, O. Litany, Y. Duan, A. Poulenard, A. Tagliasacchi, and L. J. Guibas, "Vector neurons: A general framework for so (3)-equivariant networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 200–12 209.

[21] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *ACM Transactions on Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[22] Z. Min, J. Wang, and M. Q.-H. Meng, "Robust generalized point cloud registration using hybrid mixture model," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4812–4818.

[23] T. Wu, L. Pan, J. Zhang, T. Wang, Z. Liu, and D. Lin, "Balanced chamfer distance as a comprehensive metric for point cloud completion," *Advances in Neural Information Processing Systems*, vol. 34, pp. 29 088–29 100, 2021.

[24] J. Li, A. Pepe, C. Gsaxner, G. Luijten, Y. Jin, N. Ambigapathy, E. Nasca, N. Solak, G. M. Melito, A. R. Memon *et al.*, "Medshapenet–a large-scale dataset of 3d medical shapes for computer vision," *arXiv preprint arXiv:2308.16139*, 2023.

[25] F. Chen, Q. Du, J. Zhao, Z. Zhao, D. Zhang, and H. Liao, "A generalized full-to-partial registration framework of 3d point sets for computer-aided orthopedic surgery," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 3, pp. 1010–1021, 2023.

[26] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[27] O. Hirose, "A bayesian formulation of coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 7, pp. 2269–2286, 2020.

[28] S. Ma, Q. Wang, X. Du, A. Zhang, Q. Jin, Y. Liu, Q. Yin, W. Liu, R. Song, Y. Li *et al.*, "A comparative study of augmented reality-assisted orthopedic surgical navigation systems," in *2025 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE, 2025, pp. 959–964.