

NeuralBoneReg: A Novel Self-Supervised Method for Robust and Accurate Multi-Modal Bone Surface Registration

Luohong Wu^{a,*}, Matthias Seibold^a, Nicola A. Cavalcanti^a, Yunke Ao^{a,b}, Roman Flepp^a, Aidana Massalimova^a, Lilian Calvet^a, Philipp Frnstahl^{a,b}

^a*Research in Orthopedic Computer Science, Balgrist University Hospital, University of Zurich, Lengghalde 5, Zurich, 8008, Zurich, Switzerland*

^b*AI Center, ETH Zurich, Ramistrasse 101, Zurich, 8092, Zurich, Switzerland*

Abstract

Background: In computer- and robot-assisted orthopedic surgery (CAOS), patient-specific surgical plans are generated from preoperative medical imaging data to define target locations and implant trajectories. During surgery, these plans must be precisely transferred to the intraoperative setting to guide accurate execution. The accuracy and success of this transfer relies on cross-registration between preoperative and intraoperative data. However, the substantial heterogeneity across imaging modalities and devices renders this registration process challenging and error-prone, leading to inaccuracies. Consequently, more robust and accurate methods for automatic, modality-agnostic multimodal registration of bone surfaces would have a substantial clinical impact.

Methods: We propose NeuralBoneReg, a self-supervised, surface-based framework for bone surface registration using 3D point clouds as a modality-agnostic representation. NeuralBoneReg comprises two key components: an implicit neural unsigned distance field (UDF) module and a multilayer perceptron (MLP)-based registration module. The UDF module learns a neural representation of the preoperative bone model. The registration module solves both global initialization and local refinement by generating a set of transformation hypotheses to register the intraoperative point cloud with the

*Corresponding author: Lengghalde 5, Zurich, 8008, Zurich, Switzerland
Email address: luohong.wu@balgrist.ch (Luohong Wu)

preoperative neural UDF. Compared to state-of-the-art (SOTA) supervised registration, NeuralBoneReg operates in a self-supervised manner, without requiring inter-subject training data with ground truth transformations. We evaluated NeuralBoneReg against baseline methods on two publicly available multi-modal datasets: a CT–ultrasound dataset of the fibula and tibia (UltraBones100k) and a CT–RGB-D dataset of spinal vertebrae (SpineDepth). The evaluation also includes a newly introduced CT–ultrasound dataset of cadaveric subjects containing femur and pelvis (UltraBones-Hip), which will be made publicly available.

Results: Quantitative and qualitative results demonstrate the effectiveness of NeuralBoneReg, matching or surpassing existing methods across all datasets. It achieves a mean relative rotation error (RRE) of 1.68° and a mean relative translation error (RTE) of 1.86 mm on the UltraBones100k dataset; a mean RRE of 1.88° and a mean RTE of 1.89 mm on the UltraBones-Hip dataset; and a mean RRE of 3.79° and a mean RTE of 2.45 mm on the SpineDepth dataset. These results highlight the method’s strong generalizability across different anatomies and imaging modalities.

Conclusion: NeuralBoneReg achieves robust, accurate, and modality-agnostic registration of bone surfaces, offering a promising solution for reliable cross-modal alignment in computer- and robot-assisted orthopedic surgery.

Keywords: Multimodal Bone Surface Registration, Rigid CT-ultrasound Registration, Implicit Neural Representation, Computer Assisted Orthopedic Surgery

1. Introduction

Computer- and robot-assisted orthopedic surgery (CAOS) is a widely adopted discipline that has significantly enhanced accuracy and reproducibility in surgical interventions[1], such as bone tumor resection [2], total knee arthroplasty (TKA) [3], and bone deformity correction [4]. A central prerequisite in many CAOS applications is the accurate registration between preoperative imaging data and surgery plans, including computed tomography (CT) and magnetic resonance imaging (MRI), with intraoperative data acquired from modalities such as optical tracking, ultrasound, X-ray and RGB-D imaging [1, 5, 6, 7]. The accuracy of this registration process determines how precisely the surgery plan is realized intraoperatively, directly affecting navigation accuracy, instrument guidance, and surgical outcomes

[8, 9, 10]. Achieving high accuracy across heterogeneous imaging modalities remains a key technical challenge and field of active research.

Existing registration approaches can be broadly categorized into image-based and surface-based methods, many designed for specific modality pairs[11, 12]. Image-based methods operate directly on volumetric or projection data without requiring explicit bone surface segmentation. Typical modality combinations include CT–CT [13, 14] and MRI–CT [15, 16] for 3D/3D registration, and CT–X-ray [17, 18, 19] and MRI–X-ray [17, 20] for 3D/2D registration. More recently, registration of CT/MRI with ultrasound has gained considerable attention due to ultrasound’s radiation-free and cost-effective nature [21, 22, 23, 24, 25]. Image-based methods generally optimize information-theoretic or intensity-based similarity measures such as mutual information in 3D/3D registration [22, 14, 24, 25], or match digitally reconstructed radiographs (DRRs) to intraoperative X-rays in 3D/2D registration [17]. Despite promising results in specific use cases, image-based methods remain limited in robustness and generalization due to modality-specific artifacts, resolution disparities, and imaging device heterogeneity, constraining their clinical adoption[17, 12].

Surface-based registration mitigates these limitations by segmenting the anatomy of interest and representing its geometry as a point cloud, providing a compact and modality-agnostic representation. In open surgeries such as TKA, a common strategy involves manually sampling intraoperative point clouds of the target bone using a tracked pointer, which are then registered to point clouds derived from segmented CT or MRI data[26, 27, 28]. To automate sampling, RGB-D sensors have been employed as alternatives to the tracked pointer [29, 30, 31, 32, 33]. Direct surface sampling is effective in open surgeries with sufficient bone exposure but infeasible in minimally invasive procedures. Here, intraoperative imaging is utilized for surface reconstruction, such as intraoperative CT, X-rays [34, 35, 36] and ultrasound [37, 38, 39, 40, 41, 42, 43]. Although surface-based registration facilitates cross-modality alignment, it remains challenging due to inherent differences in imaging physics that lead to noisy point clouds, varying spatial resolution and incomplete surface correspondences [44, 45, 46, 47, 48, 29, 30]. Registration is further complicated by geometric ambiguities often present in bone anatomy, leading to multiple equivalent solutions in the registration space [41, 46]. To mitigate these challenges, several strategies have been explored. Manual initialization followed by local refinement can achieve accurate results but prolongs procedures and requires expert intervention [49, 42, 43].

General-purpose point cloud alignment methods such as RANSAC [50], fast global registration (FAST) [51, 52], or principal component analysis (PCA)-based alignment can automate initialization but remain error-prone and sensitive to noise and partial data [53, 54, 31, 32, 33, 55, 56, 48, 54]. Prior-driven methods, which exploit prior knowledge such as bone shape priors or estimated inter-modality transformations, improve robustness within specific settings but lack generalization across anatomies [21, 30, 57, 41]. State-of-the-art (SOTA) deep learning-based methods have also shown promise for surface registration, but their application to multimodal bone registration remains limited due to the lack of large-scale paired training datasets [58, 59]. Consequently, achieving robust and accurate automatic multimodal bone surface registration, especially for emerging intraoperative modalities such as ultrasound and RGB-D, remains an open research challenge. Our proposed method is designed to address these challenges.

In this work, we introduce NeuralBoneReg, a modality-agnostic and self-supervised framework for robust and accurate multimodal bone surface registration. NeuralBoneReg models bone geometry using a continuous implicit neural representation (INR) that enables accurate and differentiable point-to-surface distance evaluation across modalities. The framework comprises two main modules: a preoperative bone surface representation module (NeuralUDF) and an intraoperative registration module (NeuralReg). The NeuralUDF module employs a multi-layer perceptron (MLP) to learn an implicit unsigned distance field (UDF) from the preoperative point cloud. During the intraoperative stage, the NeuralReg module aligns the intraoperative point clouds within the learned UDF through parallel hypothesis optimization and a dedicated registration loss.

The main contributions of this study include:

- A self-supervised approach for multimodal bone surface registration. We introduce the first self-supervised, INR-based registration method, eliminating the need for large paired datasets or ground-truth transformations. This instance-wise training strategy enhances adaptability across anatomies and facilitates clinical translation.
- INR-based bone surface modeling. The proposed NeuralUDF module learns a continuous UDF from preoperative point clouds, enabling efficient and differentiable distance computation without nearest-neighbor search [41, 60] or voxel interpolation [38, 37, 60]. This design shifts

computational load from the intraoperative to the preoperative stage, aligning with real-time requirements in CAOS.

- A parallel neural solver addressing surface ambiguities. To avoid convergence to incorrect local optima[41], our NeuralReg module explores the $SE(3)$ transformation space through multiple lightweight hypothesis generation heads sharing a common backbone. This shared-parameter, cooperative search improves performance over independent parallel optimization and accelerates global convergence.
- Comprehensive evaluation across three datasets. NeuralBoneReg and SOTA methods are evaluated on (i) an open-source CT–ultrasound dataset of the fibula and tibia (UltraBones100k) [47], (ii) an open-source CT–RGB-D dataset of spinal vertebrae (SpineDepth) [61], and (iii) an in-house CT–ultrasound dataset of the femur and pelvis (UltraBones-Hip). Ablation studies are conducted to quantify the contribution of each proposed component.
- Release of the UltraBones-Hip dataset. To support future research, we publicly release the UltraBones-Hip dataset, comprising paired CT and ultrasound volumes of human ex-vivo specimens with annotated bone segmentations and CT–ultrasound ground-truth transformations.

To ensure reproducibility of the results, the code and all data have been made publicly available¹.

2. Related Work

In this section, we present prior work on surface-based multimodal bone surface registration as well as the SOTA in deep learning-based point cloud registration.

Surface-Based Multimodal Bone Surface Registration. Surface-based methods represent anatomical structures as point clouds extracted from preoperative and intraoperative imaging data. A typical registration workflow consists of coarse global initialization followed by local refinement, most commonly via iterative closest point (ICP)[62]. Early studies achieved

¹<https://neuralbonereg.github.io/>

global initialization through manually selected correspondences. For instance, Ma et al. [49], Fanti et al. [42], and Li et al. [43] registered point clouds derived from CT and ultrasound data of the spine and lower extremities using manual landmark selection and subsequent ICP refinement, achieving mean target registration errors of < 2 mm. While accurate, such approaches are labor-intensive, require expert knowledge, and erroneous point sampling requires re-registration, prolonging the procedure by up to 20 minutes [63, 26]. To reduce manual effort, fully automatic global initialization approaches have been proposed that typically rely on feature-based or probabilistic methods such as RANSAC [50], FAST [51, 52], or PCA-based alignment. Ao et al. applied fast point feature histograms (FPFH)-based RANSAC and FAST followed by ICP for CT-ultrasound spine registration in a robotic surgery simulator [54]. Similar RANSAC-ICP pipelines have been employed for CT/MRI to RGB-D registration in TKA [31, 33], and for femoral registration between optical scanner and RGB-D scans [32]. FPFH-based RANSAC with ICP refinement has also been used for CT to intraoral scanner registration of alveolar bone [55] and crown-root-bone alignment [56]. Liu et al. reformed CT to ultrasound registration of lumbar vertebrae by PCA-based alignment of centroids and orientation, followed by ICP refinement [53]. Although these approaches achieve reliable alignment under controlled conditions and moderate overlap (error $\approx 2^\circ/1$ mm [54]), they remain sensitive to noise, outliers, and incomplete surfaces in real-world datasets (error $\approx 5^\circ/6$ mm [31, 33]). To improve robustness, prior knowledge and geometric constraints have been incorporated into automated pipelines. Hacıhaliloglu et al. aligned ultrasound-derived lumbar spine reconstructions with a preoperative statistical shape-and-pose model, incorporating ultrasound probe position as a prior for the initialization [21]. Hu et al. registered intraoperative RGB-D-derived point clouds of the femur to preoperative MRI-derived point clouds using a prior-correspondence-based global stage followed by ICP refinement [30]. Ciganovic et al. [41] and Chan et al. [57] exploited PCA-based initialization that embeds geometric shape and pose priors for MRI-ultrasound and CT-ultrasound registration, followed by ICP refinement. A similar prior-driven strategy has been employed for CT-ultrasound registration of the pelvis based on known geometry symmetry and main bone axis [64]. Although prior-driven approaches achieve strong performance on real-world datasets (error $\approx 1^\circ/\approx 1$ mm [41, 57]), their strong dependence on anatomy-specific priors and initialization hinders generalization across different bone anatomies.

Deep Learning-based Point Cloud Registration. The growing availability of large-scale, annotated datasets has enabled deep learning-based approaches for point cloud registration. These methods have emerged as powerful alternatives to traditional approaches, offering greater robustness, scalability, and the ability to learn data-driven priors [59, 65]. They can be categorized into three main categories: neural feature extractors for correspondence matching, initialization models for global alignment, and end-to-end registration models for regressing the final transformation parameters[59]. Unlike handcrafted descriptors such as FPFH, neural feature extractors provide robust correspondence estimation under noise, clutter, or partial overlap by learning discriminative, rotation-invariant representations directly from raw point sets, including FCGF[66], 3DSmoothNet [67], SpineNet [68], 3DMatch[69], YOHO[70]. Directly relevant to our partial-to-complete registration task, Huang et al. proposed Predator to improve correspondence extraction by predicting overlapping regions and keypoint saliency under low-overlap conditions[71]. Based on the deep features, a traditional coarse-to-fine registration workflow can then be applied to estimate the final transformation parameters. Neural feature extractors have shown to outperform handcrafted feature descriptors such as FPFH, achieving mean rotation/translation error of $1.74^\circ/19\text{ mm}$ on ModelNet and $2.03^\circ/60\text{ mm}$ on 3DMatch [71]. Although effective, the optimal transformation estimator requires dataset-specific parameter tuning. To overcome this limitation, deep networks such as 3DRegNet [72] and DeepGMR [73] have been proposed to predict global initialization followed by a conventional local refinement method. For instance, Liebmann et al. proposed a supervised network where the global alignment was learned using an MLP that learns a PCA-based initialization for aligning RGB-D-derived lumbar spine point clouds (L1–L5 posterior surface) with pre-operative CT bone models [29], achieving an average mean target registration error of 2.72 mm. Recent studies further streamlined registration by jointly solving global and local alignment in an end-to-end network, such as DCP [74], DeepVCP [75], and FINet [76]. Qin et al. introduced Geo-Transformer [77], which encodes pairwise distances and triplet-wise angles between superpoints to learn transformation-invariant features for robust correspondence matching. The correspondences are then processed by a differentiable singular value decomposition (SVD) layer for transformation regression, achieving rotation/translation errors of $0.27^\circ/68\text{ mm}$ on KITTI and $1.74^\circ/19\text{ mm}$ on ModelNet[77]. Although end-to-end methods achieve strong performance, they rely on ground-truth transformations for supervision. Self-

supervised methods, such as PPF-FoldNet [78] and UDPReg [79], optimize geometric consistency objectives (e.g., Earth Mover’s Distance or cycle consistency) without requiring explicit labels, achieving comparable accuracy as supervised networks (3.578 °/67 mm on ModelNet and 2.951 °/86 mm on 3DMatch). However, in contrast to our approach, these methods still require pretraining on large datasets of paired point clouds, which are rarely available for multimodal bone surfaces [58].

We hypothesize that limited dataset size in our task is the primary bottleneck for the performance of deep learning-based methods. To confirm this, our experiments will benchmark Predator [71] and GeoTransformer [77] on all datasets.

3. Methods

NeuralBoneReg is a surface-based approach designed to register preoperative imaging data (e.g. CT or MRI) with intraoperative imaging data (e.g. ultrasound or RGB-D). We begin by presenting the problem formulation in Section 3.1. Our proposed method is detailed in Section 3.2, comprising the NeuralUDF module in Section 3.2.1 and the NeuralReg module in Section 3.2.2.

3.1. Problem Formulation

We assume that a segmented bone surface from preoperative imaging modality is available, and denote it as $\mathcal{S}$. A point cloud $\mathcal{C}^{\text{raw}}$ of size M^{raw} is uniformly and randomly sampled from $\mathcal{S}$, represented as

$$\mathcal{C}^{\text{raw}} = \left\{ \mathbf{c}_i^{\text{raw}} = [x_i, y_i, z_i]^T \in \mathbb{R}^3 \mid i \in [1, \dots, M^{\text{raw}}] \right\} \quad (1)$$

The spatial scale of $\mathcal{C}^{\text{raw}}$ is defined as the diagonal length of the 3-dimensional bounding box:

$$\ell_{\mathcal{C}}^{\text{raw}} = \sqrt{\left(\max_i(x_i) - \min_i(x_i)\right)^2 + \left(\max_i(y_i) - \min_i(y_i)\right)^2 + \left(\max_i(z_i) - \min_i(z_i)\right)^2} \quad (2)$$

The centroid of $\mathcal{C}^{\text{raw}}$ is computed as:

$$\overline{\mathcal{C}^{\text{raw}}} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \frac{1}{M^{\text{raw}}} \sum_{i=1}^{M^{\text{raw}}} \mathbf{c}_i^{\text{raw}} \quad (3)$$

Following prior work [80, 81, 82], we apply spatial normalization to map $\mathcal{C}^{\text{raw}}$ into the normalized cube $(-1, 1)^3$ using the transformation:

$$\frac{\mathbf{c}_i^{\text{raw}} - \overline{\mathcal{C}^{\text{raw}}}}{\ell_{\mathcal{C}}^{\text{raw}}} \quad (4)$$

We then discretize the space into a voxel grid and perform downsampling to achieve uniform point density. The resulting preoperative point cloud is denoted as:

$$\mathcal{C} = \{\mathbf{c}_i \in (-1, 1)^3 \mid i \in [1, \dots, M], M \leq M^{\text{raw}}\} \quad (5)$$

The intraoperative imaging data provide a partial and noisy observation of the target bone surface $\mathcal{S}'$, acquired through modalities such as point-sampling, ultrasound, or RGB-D imaging. For point-sampling with a tracked pointer, the intraoperative point cloud $\mathcal{U}^{\text{raw}}$ is directly constructed from the sensor measurements [26]. For ultrasound and RGB-D imaging, $\mathcal{U}^{\text{raw}}$ is generated via 2D segmentation followed by 3D reconstruction using recorded pose information [47, 29]. The resulting intraoperative point cloud $\mathcal{U}^{\text{raw}}$ is denoted as:

$$\mathcal{U}^{\text{raw}} = \{\mathbf{u}_i^{\text{raw}} = [x_i, y_i, z_i]^T \in \mathbb{R}^3 \mid i \in [1, \dots, N^{\text{raw}}]\} \quad (6)$$

Analogous to $\mathcal{C}^{\text{raw}}$, the spatial scale $\ell_{\mathcal{U}}^{\text{raw}}$ and the shape centroid $\overline{\mathcal{U}^{\text{raw}}}$ are computed. As intraoperative data typically covers only a partial surface, $\ell_{\mathcal{U}}^{\text{raw}} < \ell_{\mathcal{C}}^{\text{raw}}$ generally holds. $\mathcal{U}^{\text{raw}}$ is normalized using $\ell_{\mathcal{C}}^{\text{raw}}$ to ensure scale consistency with $\mathcal{C}$, followed by voxel discretization and downsampling. The resulting intraoperative point cloud $\mathcal{U}$ is denoted as:

$$\mathcal{U} = \{\mathbf{u}_i \in (-1, 1)^3 \mid i \in [1, \dots, N], N \leq N^{\text{raw}}\} \quad (7)$$

In summary, the registration task is defined as estimating the 6-Degrees-of-Freedom (6-DoF) rigid transformation $\mathbf{T}_{\mathcal{C} \leftarrow \mathcal{U}} = [\mathbf{R} \mid \mathbf{t}] \in \text{SE}(3)$ that aligns the intraoperative point cloud $\mathcal{U}$ with the preoperative point cloud $\mathcal{C}$ by minimizing the following objective function:

$$\min_{\mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3} \frac{1}{N} \sum_{i=1}^N \|\mathbf{c}_{\pi(i)} - (\mathbf{R}\mathbf{u}_i + \mathbf{t})\|_2 \quad \text{s.t. } \mathbf{c}_{\pi(i)} \in \mathcal{C}, \mathbf{u}_i \in \mathcal{U}. \quad (8)$$

where $\pi(i) = \arg \min_{j=1, \dots, M} \|\mathbf{c}_j - (\mathbf{R}\mathbf{u}_i + \mathbf{t})\|$ denotes the index of the nearest neighbor of the transformed point $(\mathbf{R}\mathbf{u}_i + \mathbf{t})$ in $\mathcal{C}$.

3.2. Proposed Method

Aligned with the CAOS workflow, comprising preoperative and intraoperative stages, NeuralBoneReg includes two corresponding modules: preoperative NeuralUDF and intraoperative NeuralReg, as illustrated in Figure 1. In the preoperative stage, a neural UDF is learned from the preoperative point cloud $\mathcal{C}$. During the intraoperative stage, NeuralReg generates multiple registration transformation hypotheses in parallel to align the intraoperative point cloud $\mathcal{U}$ with the learned neural UDF, solving both global initialization and local refinement efficiently. Each hypothesis is evaluated based on the mean distance within the learned UDF, and the hypothesis yielding the lowest mean distance is selected as the final result.

3.2.1. NeuralUDF

Equation 8 corresponds to the objective optimized in conventional point-cloud-based registration, such as in a RANSAC-ICP pipeline. Minimizing this objective requires nearest-neighbor search, which is non-differentiable and computationally expensive, particularly in iterative optimization pipelines. To overcome this challenge, Wein et al. [37] fused ultrasound data with a fixed-resolution signed distance field (SDF) derived from preoperative CT. The fixed-resolution SDFs represent the soft tissue layer between bone and skin, providing gradients that enable simulated probe-induced compression during intensity-based local refinement. Building on this idea, we propose to learn a neural UDF from the preoperative point cloud $\mathcal{C}$, which offers several advantages. The learned UDF supports direct distance queries from arbitrary 3D points, eliminating the need for explicit, non-differential nearest-neighbor search. In addition, unlike fixed-resolution distance volumes, a neural UDF provides effectively arbitrary resolution and smooth, differentiable objectives that enable efficient gradient-based optimization. As we will show in Section 5, fixed-resolution representations degrade when the target anatomy spans larger scales (e.g., pelvis vs. vertebra).

Several methods have been proposed for learning a neural implicit representation from a point cloud [80, 83, 82], all of which can be applied to our NeuralUDF module. Among these methods, UltraBoneUDF has shown effectiveness in learning neural UDF for both partial and complete bone anatomies[80]. The network learns to approximate the unsigned distance of any spatial location to the underlying surface by minimizing a tangent-plane consistency loss and a coarse global-shape regularizer. This yields a smooth,

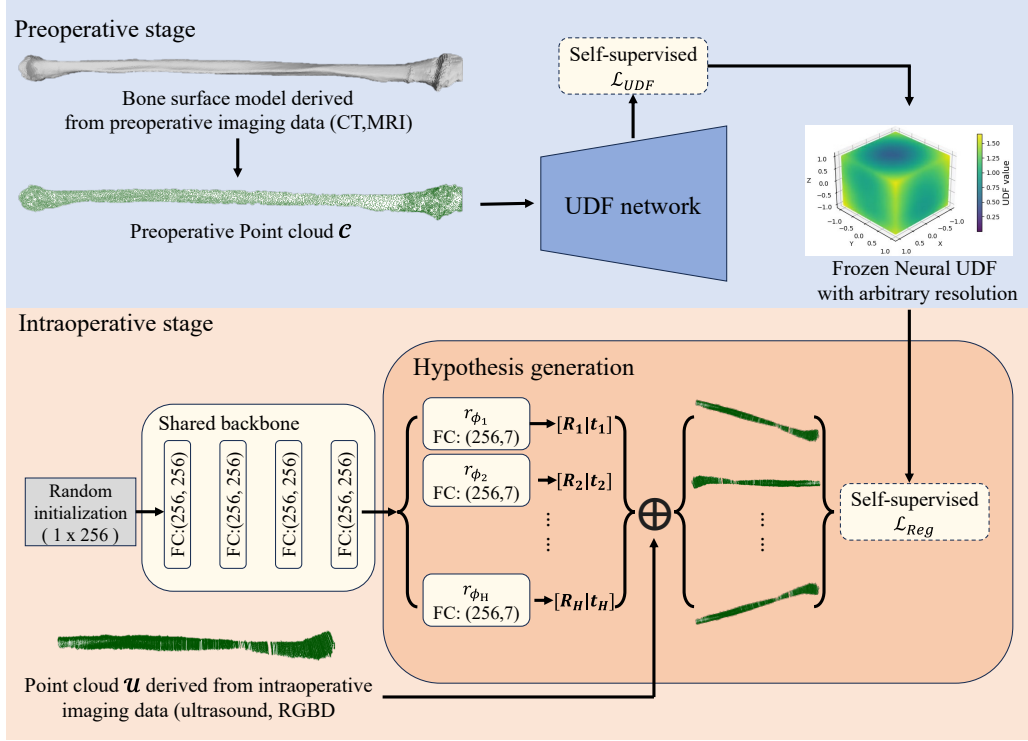


Figure 1: Network architecture of the two-staged approach of NeuralBoneReg. In the preoperative stage, the NeuralUDF module learns an implicit neural representation of the target bone surface from the preoperative data $\mathcal{C}$. During the intraoperative stage, the NeuralReg module leverages multiple lightweight neural heads built on a randomly initialized shared backbone to efficiently generate registration hypotheses aligning the intraoperative data $\mathcal{U}$ with the preoperative data $\mathcal{C}$. FC(dim_{in}, dim_{out}) denotes a fully connected layer with input dimension dim_{in} and output dimension dim_{out}, followed by a Tanh activation.

differentiable distance field that implicitly encodes the bone surface and supports direct geometric reasoning. In our work, we apply UltraBoneUDF with default parameters. For details of the training procedure, including the architecture and loss function $\mathcal{L}_{UDF}$, we refer to [80].

Though more time-consuming to train, the NeuralUDF module is computed preoperatively, offloading nearest-neighbor computations and allowing faster intraoperative optimization. After training, the UDF defines a mapping $d(\mathbf{q}_i) : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$, assigning each query point $\mathbf{q}_i \in \mathbb{R}^3$ a non-negative value representing its predicted distance to $\mathcal{S}$. The frozen NeuralUDF is used for distance queries in the NeuralReg module.

3.2.2. *NeuralReg*

Leveraging the learned UDF, the registration task is reformulated as minimizing the mean UDF value over the transformed intraoperative point cloud:

$$\min_{[\mathbf{R}|\mathbf{t}] \in \text{SE}(3)} \frac{1}{N} \sum_{i=1}^N d(\mathbf{R}\mathbf{u}_i + \mathbf{t}) \quad (9)$$

In contrast to Equation 8, the proposed formulation is continuous and differentiable with respect to $[\mathbf{R} \mid \mathbf{t}]$, enabling direct gradient-based optimization.

However, geometric symmetries in $\mathcal{S}$ induce numerous local minima in the objective, making direct optimization of the 6-DoF transformation, whether deterministic or stochastic, unreliable. As shown later in our ablation studies, this naive approach often converges to suboptimal solutions, highlighting the need for a more robust optimization strategy.

A strategy to mitigate such ambiguities is to explore multiple candidate solutions in parallel. For instance, Bishop’s Mixture Density Networks predict the parameters of a mixture probability distribution, allowing the modeling of multiple plausible outputs for a shared input [84]. Inspired by this concept, our registration network is designed to generate and evaluate multiple transformation hypotheses in parallel. Given the input point cloud $\mathcal{U}$ and the pretrained UDF, the key idea is to leverage multiple hypothesis generation heads to explore different regions of $\text{SE}(3)$ simultaneously. This parallel hypothesis generation increases the likelihood of finding the correct alignment without relying on sequential local refinements, while the shared backbone allows information exchange between hypotheses to improve overall convergence stability. The depth of the shared backbone determines how effectively information is shared across , while the number of heads controls the diversity of explored transformations. Section 4.5 presents ablation studies analyzing the influence of both factors on registration performance.

At the architectural level, *NeuralReg* comprises a shared MLP backbone f_θ and H hypothesis generation heads $\{r_{\phi_h}\}_{h=1}^H$, as illustrated in Figure 1. Given a randomly sampled initialization vector $\bar{\mathbf{x}} \in (-1, 1)^{256}$, the shared backbone maps a latent feature $f_\theta(\bar{\mathbf{x}}) \in (-1, 1)^{256}$, that serves as a common conditional signal for all heads. During training, $\bar{\mathbf{x}}$ is held fixed over the overall training for a single input point cloud while the backbone parameters θ are optimized jointly. Conditioned on d , each head implements a mapping $r_\phi(\mathcal{U}, f_\theta(\bar{\mathbf{x}}) \mid d) \rightarrow \mathbb{R}^3 \times \mathbb{R}^4$ that outputs a translation vector $\mathbf{t} \in \mathbb{R}^3$ and a quaternion $\eta \in \mathbb{R}^4$ ($\|\eta\|_2 = 1$). The quaternion η is converted into a

rotation matrix $\mathbf{R} \in \text{SO}(3)$, which is then combined with $\mathbf{t}$ to define a rigid transformation hypothesis $[\mathbf{R} \mid \mathbf{t}] \in \text{SE}(3)$. The alignment quality of each transformation $[\mathbf{R} \mid \mathbf{t}]$ is evaluated through Equation 9, which serves as the training objective for the network.

To efficiently exploit GPU parallelism, each hypothesis generation head is implemented as a lightweight fully connected layer followed by a tanh activation ($\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$), chosen for its normalized output range $\tanh(x) \in (-1, 1)$. To ensure valid unit-norm representation, each output quaternion is subsequently normalized such that $\|\eta\|_2 = 1$. While all heads share the backbone parameters θ , they maintain distinct parameter sets $\{\phi_h\}_{h=1}^H$.

The registration network is trained in a self-supervised manner using the following loss function $\mathcal{L}_{\text{Reg}}$:

$$\mathcal{L}_{\text{Reg}} = \frac{1}{H} \sum_{h \in [1, H]} \left[\frac{1}{B} \sum_{i \in [1, B]} d(\mathbf{R}_h \cdot \mathbf{u}_i + \mathbf{t}_h) \right] \quad (10)$$

where B denotes the batch size.

During inference, the hypothesis yielding the lowest mean UDF value is selected as the final solution.

4. Experiments

To evaluate NeuralBoneReg, we perform a comprehensive comparison against existing SOTA registration methods. Section 4.1 introduces the evaluation datasets, followed by implementation details in Section 4.2. The SOTA approaches are described in Section 4.3, and the evaluation metrics are summarized in Section 4.4. Section 4.5 describes the ablation studies.

4.1. Datasets

Given the growing body of research in CT-ultrasound registration, we conduct experiments on two CT-ultrasound datasets. To further assess the generalizability of NeuralBoneReg beyond ultrasound, an additional CT-RGB-D dataset is included. Together, these datasets cover a range of anatomies such as the fibula, tibia, femur, pelvis, and lumbar spine.

4.1.1. UltraBones100k

UltraBones100k is an open-source dataset comprising CT-derived bone surfaces and tracked 3D ultrasound recordings from 14 human ex-vivo lower

Specimen No.	Height (cm)	Weight (kg)	BMI	Age	Gender	Arthritis
1	188	98	27.86	66	Male	History of falls
2	180	93	28.59	66	Male	History of falls
3	168	75	26.63	72	Female	History of falls
4	165	76	27.95	71	Female	None specified
5	178	77	24.39	49	Male	None specified

Table 1: Demographic characteristics of the subjects in UltraBones-Hip. Each subject includes pelvis and proximal femur.

extremity specimens including the tibia and fibula [47]. For each subject, multiple ultrasound scans are available per bone. A pretrained segmentation model is provided to generate bone segmentation masks from the ultrasound data. Following the preprocessing steps described in Section 3.1, we generate $\mathcal{C}$ and $\mathcal{U}$ for each bone of each subject, resulting in 28 paired point clouds (14 subjects $\times$ 2 bones). All preprocessed point clouds ($\mathcal{C}$, $\mathcal{U}$) are available for downloading via our repository ².

4.1.2. UltraBones-Hip

To expand the evaluation to cover more anatomies, we collected an additional CT-ultrasound dataset of the hip region, including the femur and pelvis. Data acquisition followed the UltraBones100k protocol [47] and involved five human ex-vivo specimens (see Table 1). Each specimen was scanned along the coronal and axial directions. Ethical approval for conducting this study was obtained from the Zurich Cantonal Ethical Committee (BASEC Nr. 2023-01652). We directly applied the pretrained segmentation model from UltraBones100k without any fine-tuning[47]. Due to anatomical differences, the segmentation masks were noisier than those in UltraBones100k (Figure 4). The processing steps described in Section 3.1 were applied to generate $\mathcal{C}$ and $\mathcal{U}$, resulting in 5 subjects $\times$ 3 anatomical structures (left femur, right femur, and pelvis) = 15 pairs of point clouds. The raw data, ground-truth transformations, and preprocessed point clouds are publicly available via our repository.

4.1.3. SpineDepth

The SpineDepth dataset comprises CT and RGB-D data of lumbar spines from 10 human ex-vivo subjects [61]. For each subject, CT-derived bone

²<https://neuralbonereg.github.io/>

surfaces are available for each lumbar vertebra (L1 to L5). In addition, RGB-D recordings of the exposed spine were acquired during the ex-vivo surgery using a ZED Mini camera (Stereolabs Inc., San Francisco, CA, USA) from two viewpoints, with corresponding ground-truth transformations between modalities. Following prior work [29], we exclude subjects #1 and #10 due to limited anatomical exposure. For each of the remaining 8 subjects, RGB-D scans of the same vertebral level and viewpoint were combined, yielding $8 \times 5 \times 2 = 80$ $\mathcal{C}$ - $\mathcal{U}$ pairs. We make preprocessed point clouds publicly available via our repository.

4.2. Implementation Details

In our experiments, we set $|\mathcal{C}| = M = 40,000$ and $|\mathcal{U}| = N = 40,000$. For the NeuralUDF module, we apply UltraBoneUDF with default parameters [80]. For NeuralReg, the number of hypothesis generation heads was set to $H = 1,000$, and training was performed using the Adam optimizer (learning rate: 0.001, momentum: 0.9) for 1,000 iterations.

The entire framework was implemented in PyTorch v2.7.0 with CUDA 12.8 support. Differential operations, including quaternion-rotation matrix conversions, were implemented using PyTorch3D v0.7.8 [85]. The code is publicly available via our repository.

4.3. Baseline Methods

To simulate realistic unregistered point cloud scenarios, random initialization transformations $\mathbf{T}_{\text{per}} = [\mathbf{R} \in \text{SO}(3) \mid \mathbf{t} \in [-1, 1]^3]$ were applied to each intraoperative point cloud $\mathcal{U}$. Each registration method was tasked with recovering the inverse transformation $\mathbf{T}_{\text{per}}^{-1}$. To ensure robustness, each experiment was repeated 20 times with independent random initializations per $\mathcal{C}$ - $\mathcal{U}$ pair.

In the experiments, we compare the performance of NeuralBoneReg against the following baseline methods:

- **Traditional methods.** We evaluate three commonly used two-stage approaches combining coarse global initialization with local refinement: PCA+ICP [41, 53, 64], RANSAC+ICP [31, 32, 33, 55, 54, 56], and FAST+ICP [52, 54, 86]. Following prior work, FPFH features are used for RANSAC and FAST. In PCA+ICP, we improve global initialization robustness by sampling multiple rotations around the first three principal axes [41]. For ICP, both point-to-point and point-to-plane variants

are evaluated, and the better result is reported for each dataset. This policy is applied to all ICP-based experiments. Implementations are based on Open3D ³.

- **Deep learned methods.** To benchmark against learning-based approaches for point cloud registration [59], we include two SOTA networks: Predator [71] and GeoTransformer [77]. We train the official implementations ^{4 5} on each evaluation dataset using subject-wise leave-one-out cross-validation. Reported results represent the mean performance across all validation splits. Additionally, to simulate the scenario where training data are unavailable in CAOS, we evaluate their cross-dataset generalizability.
- **Pseudo ground-truth.** To estimate the upper performance bound, we apply ICP initialized with the ground-truth transformation for each $\mathcal{C}\text{-}\mathcal{U}$ pair, and define the results as the pseudo ground-truth.

The runtime of all methods was measured on a system equipped with an Intel Xeon Platinum 8450H CPU and an NVIDIA H100 GPU.

4.4. Evaluation Metrics

For each experiment, given the ground truth transformation $\mathbf{T}_{\text{GT}} = [\mathbf{R}_{\text{GT}} \mid \mathbf{t}_{\text{GT}}]$ and the estimated transformation $\mathbf{T}_{\text{est}} = [\mathbf{R}_{\text{est}} \mid \mathbf{t}_{\text{est}}]$, performance is quantified using the evaluation metrics established in prior work [59, 71]:

- Relative rotation error (RRE) in degrees: RRE measures the rotation angle between $\mathbf{R}_{\text{GT}}$ and $\mathbf{R}_{\text{est}}$ by:

$$\text{RRE}(\mathbf{R}_{\text{est}}) = \arccos \left(\frac{\text{Tr}(\mathbf{R}_{\text{GT}}^\top \cdot \mathbf{R}_{\text{est}}) - 1}{2} \right) \quad (11)$$

- Relative translation error (RTE) in mm, which is defined as:

$$\text{RTE}(\mathbf{t}_{\text{est}}) = \|\mathbf{t}_{\text{est}} - \mathbf{t}_{\text{GT}}\|_2 \quad (12)$$

³https://www.open3d.org/docs/release/tutorial/pipelines/global_registration.html

⁴<https://github.com/prs-eth/OverlapPredator>

⁵<https://github.com/qinzheng93/GeoTransformer>

For a series of O experiments conducted on a dataset, the mean RRE and RTE are reported, and the recall rate (RR) is defined as:

$$\text{RR}(x) = \frac{1}{O} \sum_{i=1}^O \begin{cases} 1, & \text{if RRE}_i < x \text{ and RTE}_i < x, \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where x denotes the tolerance threshold.

4.5. Ablation Studies

To assess the contribution of individual design choices in NeuralBoneReg, we conducted ablation studies focusing on four key aspects.

The optimization approach (neural vs traditional). NeuralBoneReg integrates a neural distance field with a neural optimization module. To systematically assess the contribution of these components, we compare neural, hybrid, and traditional variants. Moreover, we analyze combinations of implicit and grid-based distance field representations with either neural or traditional optimizers across all datasets:

- **NeuralUDF+BFGS:** To quantify the performance gain over a traditional gradient-based optimizer, we employ Broyden–Fletcher–Goldfarb–Shanno (BFGS) to optimize Equation 9. BFGS, a quasi-Newton method that approximates the Hessian for efficient optimization of smooth functions, has been used in CT–ultrasound bone surface registration [87, 88]. The implementation provided in the SciPy library is used ⁶.
- **NeuralUDF+DE:** The BFGS variant optimizes Equation 9 by exploiting gradient information from the initialization. However, the pronounced symmetry of most human bones results in an optimization landscape with numerous local minima, increasing the likelihood of incorrect convergence. As an alternative to gradient-based methods, the differential evolution (DE) algorithm has been widely employed for such highly non-convex optimization problems[88, 89, 90]. The DE algorithm mitigates local minima by stochastically exploring the search

⁶<https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>

space in parallel through a population of candidate solutions, and converges when the population stabilizes. The implementation provided in the SciPy library is used ⁷.

- **GridVolume+NeuralReg:** To quantify the performance gain achieved by an implicit neural representation, a discrete fixed-resolution distance volume is constructed from the preoperative point cloud $\mathcal{C}$, where each voxel stores the distance to its nearest neighbor in $\mathcal{C}$ [38, 37, 60]. Distance at arbitrary query points are obtained by linear interpolation. To balance resolution and runtime, we set the grid resolution to 512^3 across all datasets and use NeuralReg for optimization.
- **GridVolume+BFGS:** This variant combines conventional grid-based point queries and a traditional gradient-based optimizer.

Grid resolutions. Grid-based methods employ fixed-resolution distance volumes. While this representation may adequately capture smaller anatomies, larger bones require substantially higher grid resolutions to achieve comparable surface detail. To investigate the effect of grid resolution, we evaluate the GridVolume variant using volumes of $[128^3, 256^3, 512^3]$ on the UltraBones-Hip dataset (largest anatomy) and the SpineDepth dataset (smallest anatomy).

Shared backbone for joint learning. Sharing backbone parameters allows gradient updates to propagate across heads, potentially accelerating convergence over independent optimization. The number of shared parameters is therefore expected to influence the effectiveness of this propagation mechanism. To quantify this effect, we vary the backbone depth within the range $[0, 1, 2, 3, 4, 5, 6]$ (0 indicating no parameter sharing) and evaluate the variants on the SpineDepth dataset, which contains the largest number of $\mathcal{C}$ - $\mathcal{U}$ pairs among the three datasets.

Head counts. To assess the impact of the number of heads, we evaluate NeuralBoneReg on the SpineDepth dataset using head counts ranging from 100 to 1000.

Method	UltraBones100k		UltraBones-Hip		SpineDepth		Mean runtime [s] ↓
	RRE [°] ↓	RTE [mm] ↓	RRE [°] ↓	RTE [mm] ↓	RRE [°] ↓	RTE [mm] ↓	
RANSAC50M+ICP	15.55 ± 42.31	9.20 ± 22.52	38.24 ± 66.09	17.94 ± 35.92	7.89 ± 28.04	5.77 ± 14.38	62.17
RANSAC100M+ICP	12.74 ± 39.07	7.21 ± 19.24	26.86 ± 55.78	16.68 ± 37.08	5.54 ± 23.43	4.53 ± 14.07	117.64
RANSAC250M+ICP	5.36 ± 20.39	3.67 ± 12.24	15.15 ± 44.84	11.18 ± 33.20	4.25 ± 20.68	3.69 ± 11.51	226.58
FAST+ICP	19.09 ± 46.58	4.17 ± 5.74	26.04 ± 32.52	20.06 ± 15.30	31.25 ± 25.71	15.10 ± 8.33	0.17
PCA+ICP	2.18 ± 10.86	3.07 ± 7.58	63.54 ± 73.58	31.36 ± 34.88	63.19 ± 60.96	13.00 ± 11.05	24.30
Predator	1.43 ± 1.12	2.36 ± 1.81	15.94 ± 45.96	13.82 ± 30.26	3.43 ± 16.39	1.95 ± 1.50	0.34
GeoTransformer	1.93 ± 8.39	3.29 ± 12.81	8.12 ± 18.59	10.74 ± 27.47	3.77 ± 10.47	2.89 ± 4.75	0.27
Ours	1.68 ± 1.18	1.86 ± 1.09	1.88 ± 1.75	1.89 ± 0.73	3.79 ± 13.74	2.45 ± 2.41	41.72 (+119.32)
Pseudo ground-truth	1.40 ± 1.02	1.09 ± 0.43	1.00 ± 0.64	1.46 ± 0.72	1.53 ± 1.14	1.76 ± 0.75	-

Table 2: Quantitative results on three datasets: UltraBones100k, UltraBones-Hip, and SpineDepth. RRE: relative rotation error (degrees). RTE: relative translation error (mm). RANSAC250M+ICP denotes 250 million iterations in the global initialization stage. NeuralBoneReg method involves a training phase using preoperative data, with an average duration of 119.32 s across all datasets.

Method	UltraBones100k		SpineDepth	
	RRE [°] ↓	RTE [mm] ↓	RRE [°] ↓	RTE [mm] ↓
Predator (trained on SpineDepth)	38.69 ± 67.34	35.29 ± 66.14	-	-
GeoTransformer (trained on SpineDepth)	14.82 ± 34.44	20.85 ± 49.82	-	-
Predator (trained on UltraBones100k)	-	-	106.04 ± 58.39	31.51 ± 21.61
GeoTransformer (trained on UltraBones100k)	-	-	60.57 ± 35.57	26.98 ± 19.74

Table 3: Quantitative results of the cross-dataset generalization experiments. RRE: relative rotation error (degrees). RTE: relative translation error (mm).

5. Results

5.1. Main Results

Table 2 summarizes the quantitative results across all evaluation datasets. NeuralBoneReg achieves RREs of $1.68 \pm 1.18^\circ$, $1.88 \pm 1.75^\circ$, and $3.79 \pm 13.74^\circ$ on UltraBones100k, UltraBones-Hip, and SpineDepth, respectively, and RTEs of 1.86 ± 1.09 mm, 1.89 ± 0.73 mm, and 2.45 ± 2.41 mm. Across datasets, NeuralBoneReg often matches or surpasses the accuracy of SOTA supervised methods such as Predator and GeoTransformer, while operating fully self-supervised. In all cases, the registration accuracy remained within approximately $1^\circ/1$ mm of the pseudo ground truth, demonstrating robustness across different modalities and anatomical regions. As shown in Figure 2, NeuralBoneReg attains a high recall (close to 1) before $x = 5$ across all datasets. Representative qualitative results are presented in Figure 3 for UltraBones100k, Figure 4 for UltraBones-Hip, and Figure 5 for SpineDepth.

⁷https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html

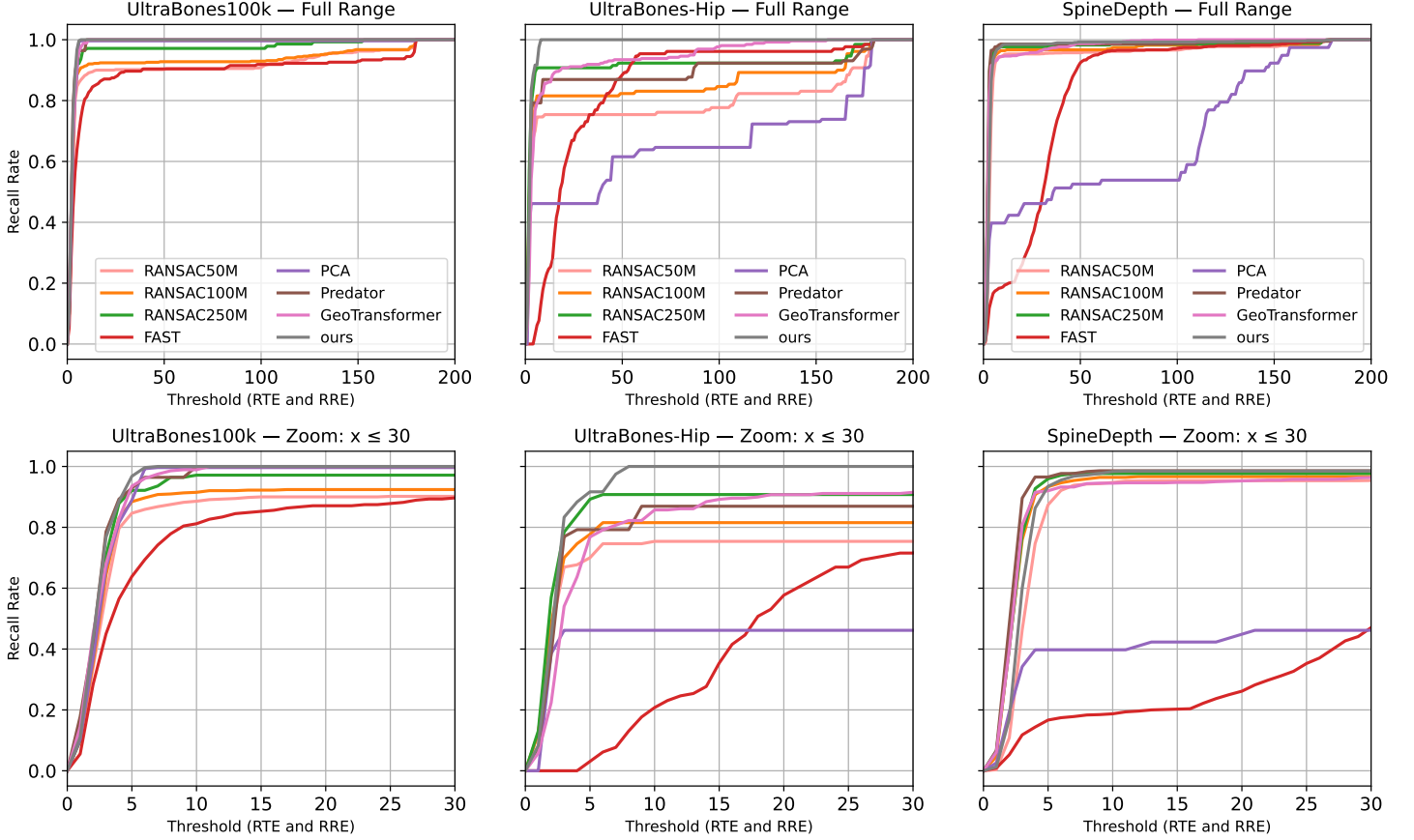


Figure 2: Recall-rate curves for UltraBones100k (first column), UltraBones-Hip (second column), and SpineDepth (last column). x denotes the threshold and y denotes $RR(x)$. The top row shows the full range, while the bottom row focuses on $0 \leq x \leq 30$.

3D visualization of registration results are provided via our repository.

Table 3 summarizes the cross-dataset generalization results of Predator and GeoTransformer. Without annotated training data, both methods experience substantial performance degradation. On UltraBones100k, the mean RRE/RTE of Predator increase from $1.43^\circ/2.36$ mm to $38.69^\circ/35.29$ mm, while those of GeoTransformer rise from $1.93^\circ/3.29$ mm to $14.82^\circ/20.85$ mm.

5.2. Results of Ablation Studies

Within the ablations shown in Table 4, GridVolume+NeuralReg performs best, with RREs of $4.51 \pm 3.52^\circ$ (UltraBones100k), $6.67 \pm 9.50^\circ$ (UltraBones-

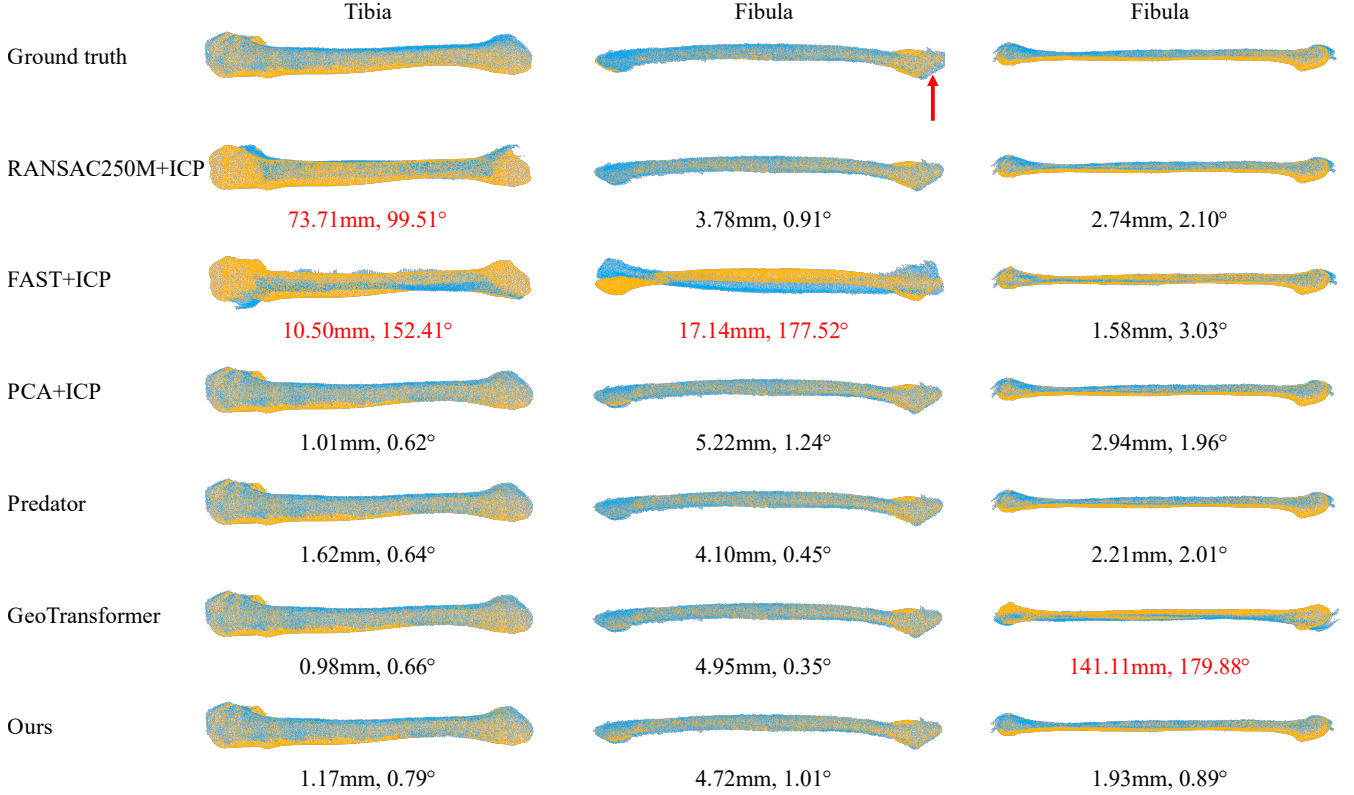


Figure 3: Exemplary qualitative results on UltraBones100k. The preoperative point cloud $\mathcal{C}$ is visualized in yellow, while the intraoperative point cloud $\mathcal{U}$ is visualized in blue. The mean relative translation error (mm) and rotation error (°) are reported below each figure. Sever errors with $RRE > 10^\circ$ or $RTE > 10$ mm are highlighted in red.

Hip), and $4.75 \pm 16.86^\circ$ (SpineDepth), and RTEs of 2.44 ± 0.93 mm, 3.18 ± 5.48 mm, and 2.55 ± 2.68 mm, respectively.

Across the grid-based method, we evaluated grid resolutions 128^3 , 256^3 , and 512^3 on the UltraBones-Hip and SpineDepth datasets. Increasing the grid resolution markedly improved registration accuracy for UltraBones-Hip, reducing mean RRE/RTE from $31.87^\circ/11.08$ mm (128^3) to $11.19^\circ/4.73$ mm (256^3) and $6.67^\circ/3.18$ mm (512^3). In contrast, the improvement on SpineDepth was modest, decreasing from $5.70^\circ/2.79$ mm (128^3) to $4.88^\circ/2.64$ mm (256^3) and $4.75^\circ/2.55$ mm (512^3).

The effect of varying the depth of the shared backbone in the NeuralReg

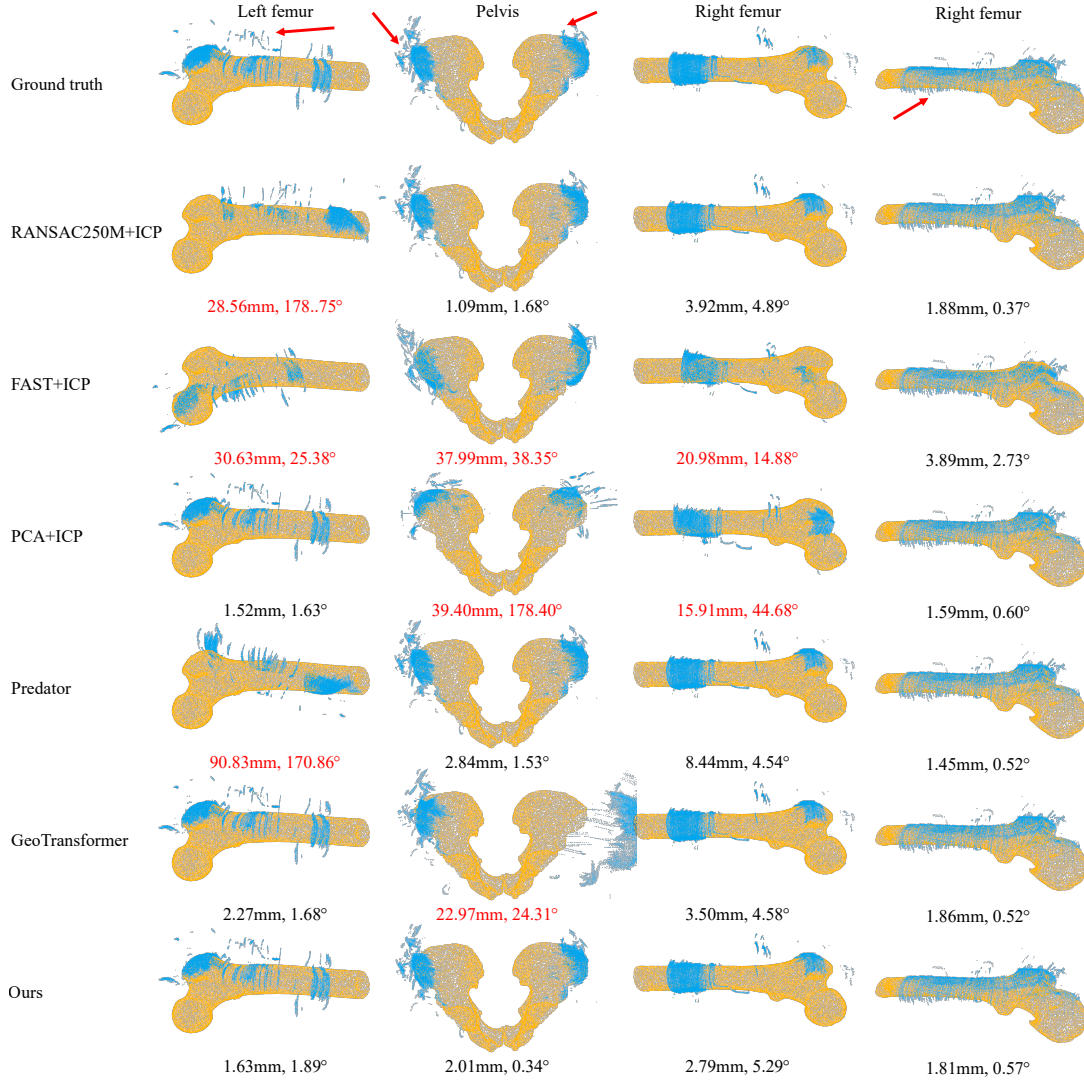


Figure 4: Exemplary qualitative results on UltraBones-Hip. The preoperative point cloud $\mathcal{C}$ is visualized in yellow, while the intraoperative point cloud $\mathcal{U}$ is visualized in blue. The mean relative translation error (mm) and rotation error (°) are reported below each figure. Sever errors with $RRE > 10^\circ$ or $RTE > 10$ mm are highlighted in red.

module is visualized in Figure 6. Notably, when no parameters are shared between different hypothesis generation heads (depth = 0), the mean RRE and RTE are 9.43° and 4.17 mm, respectively, while the performance generally improves with increasing depth.

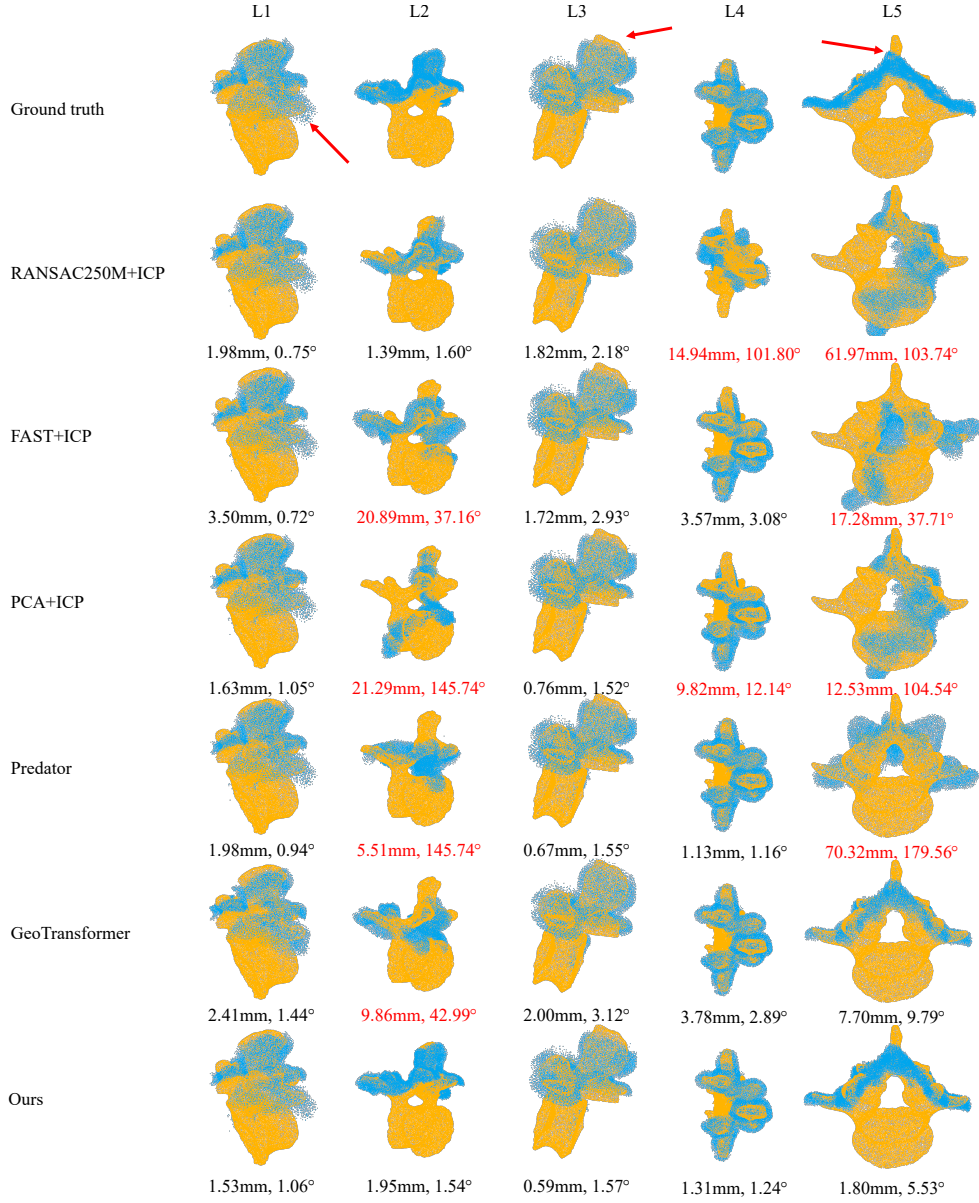


Figure 5: Exemplary qualitative results on SpineDepth. The preoperative point cloud $\mathcal{C}$ is visualized in yellow, while the intraoperative point cloud $\mathcal{U}$ is visualized in blue. The mean relative translation error (mm) and rotation error (°) are reported below each figure. Sever errors with $RRE > 10^\circ$ or $RTE > 10$ mm are highlighted in red.

Method	UltraBones100k		UltraBones-Hip		SpineDepth	
	RRE [°] ↓	RTE [mm] ↓	RRE [°] ↓	RTE [mm] ↓	RRE [°] ↓	RTE [mm] ↓
NeuralUDF+BFGS	117.72 ± 77.52	14.06 ± 18.18	133.75 ± 55.54	44.48 ± 28.57	92.82 ± 63.30	15.46 ± 10.45
NeuralUDF+DE	17.47 ± 41.33	3.46 ± 4.33	47.98 ± 34.18	19.09 ± 22.01	67.94 ± 51.16	12.34 ± 9.65
GridVolume+NeuralReg	4.51 ± 3.52	2.44 ± 0.93	6.67 ± 9.50	3.18 ± 5.48	4.75 ± 16.86	2.55 ± 2.68
GridVolume+BFGS	115.88 ± 76.92	33.44 ± 58.81	130.80 ± 59.74	46.17 ± 28.29	98.79 ± 60.43	16.38 ± 9.08
NeuralUDF+NeuralReg (ours)	1.68 ± 1.18	1.86 ± 1.09	1.88 ± 1.75	1.89 ± 0.73	3.79 ± 13.74	2.45 ± 2.41

Table 4: Quantitative results of the ablation studies on three datasets: UltraBones100k, UltraBones-Hip, and SpineDepth. RRE: relative rotation error (degrees). RTE: relative translation error (mm).

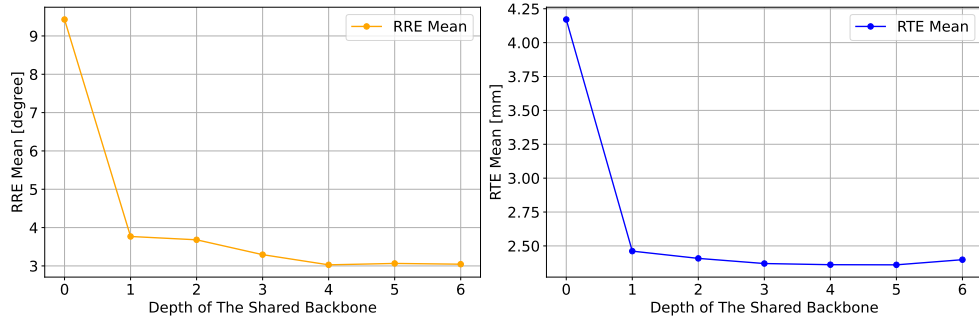


Figure 6: Plots of mean RTE and RRE versus depth of the shared backbone within the NeuralReg module on the SpineDepth dataset.

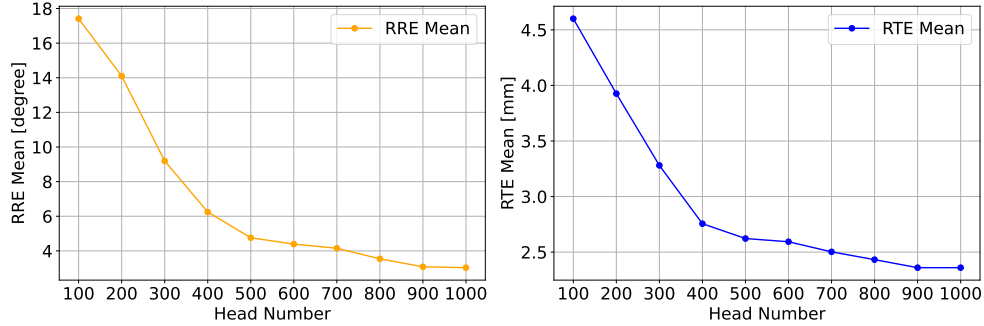


Figure 7: Plots of mean RTE and RRE versus head counts on the SpineDepth dataset.

Head counts. Figure 7 illustrates the impact of varying the number of heads in the NeuralReg module. To further characterize head behavior, Figure 8 depicts representative traces of UDF values across training iterations for the three datasets.

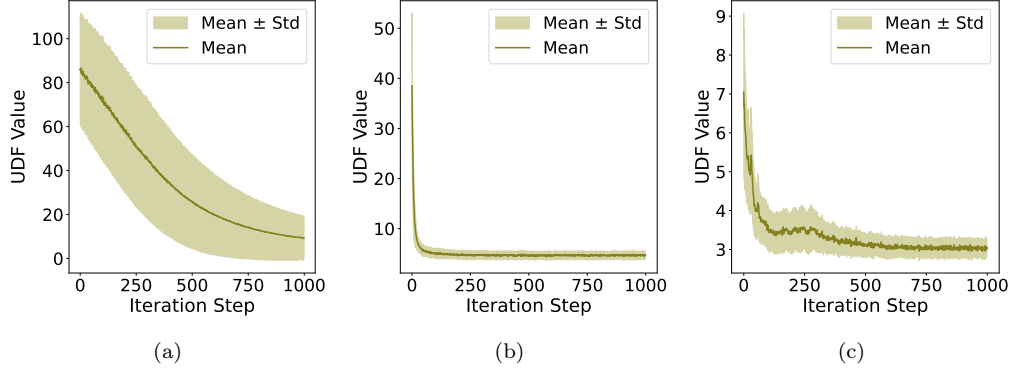


Figure 8: Example UDF traces of all heads during the training process for three datasets: (a) UltraBones100k, (b) UltraBones-Hip, and (c) SpineDepth.

6. Discussion

This study presents NeuralBoneReg, an instance-level, self-supervised framework for modality-agnostic bone surface registration. This method integrates implicit neural surface representations with parallel hypothesis generation to improve registration robustness in challenging CAOS scenarios. Through experiments on three multimodal datasets covering different anatomies, NeuralBoneReg demonstrated SOTA accuracy and robustness, matching or surpassing both traditional optimization-based pipelines and supervised deep learning methods. In the following, we discuss the observed performance in terms of accuracy, runtime, methodology and limitations.

Accuracy. The pronounced symmetry of bone anatomy leads to an optimization landscape with numerous local minima, particularly under poor initialization. This effect is evident in the RR plots (Figure 2), which increase sharply as the threshold values approach 180. SOTA registration pipelines handle these local minima differently. RANSAC+ICP attempts to mitigate them by randomized correspondence sampling, but even with 250 million iterations (≈ 226 s runtime), its results remain unstable, exhibiting high RRE/RTE variability ($5.36 \pm 20.39^\circ$ and 3.67 ± 12.24 mm on UltraBones100k). PCA+ICP performs well when the anatomical structures exhibit distinct and stable principal axes (UltraBones100k), but degrades when PCA-derived axes become ambiguous or noisy, as in UltraBones-Hip and SpineDepth. Supervised networks such as Predator and GeoTransformer learn robust data-driven features and perform well with large datasets (UltraBones100k, SpineDepth), but their accuracy drops when data are scarce

(UltraBones-Hip) or when anatomical variability is high. The cross-dataset generalization results in Table 3 further emphasize their limitation when training data for unseen anatomies are unavailable. In contrast, NeuralBoneReg operates self-supervised at an instance level and does not require large training data sets. It achieves low mean RRE/RTE with small standard deviations across all datasets, outperforming all traditional and most supervised methods, particularly in limited-data settings.

Runtime. Fast execution is critical in CAOS, as prolonged intraoperative time directly impacts surgical costs and patient outcomes. Many SOTA methods perform computations during surgery and become increasingly expensive as the number of sampling iterations grows. NeuralBoneReg exploits the two-stage CAOS workflow by shifting the computationally intensive nearest-neighbor queries to the preoperative phase. Nevertheless, its intraoperative runtime (≈ 42 s) remains longer than that of fully supervised networks, which complete a single forward pass within one second. However, NeuralBoneReg operates entirely self-supervised without requiring prior training datasets, offering a practical balance between computational efficiency, data independence, and generalization capability.

Methodology. NeuralBoneReg combines a neural UDF with a neural optimizer to achieve smooth and efficient registration, while other variants are evaluated in the ablation studies (Table 4). BFGS-based variants exhibit large performance variability across datasets, reflecting the highly non-convex optimization landscape of the registration problem. NeuralUDF+DE addresses this issue via population-based parallel search. Its higher accuracy over BFGS-based variants confirms the advantage of exploring multiple transformation hypotheses (Table 4). However, the achievable population size is constrained by computational resources: our experiments with 50 candidates exceeded the runtime of NeuralBoneReg. In contrast, our lightweight neural hypothesis-generation heads enables highly parallel GPU exploration. As shown in Figure 7, increasing the number of heads systematically improves performance, validating the effectiveness of this design. In addition, performance benefits from parameter sharing within the backbone: As shown in Figure 6 and Figure 8, the largest gain occurs from depth 0 (no sharing) to 1, supporting our hypothesis that parameter sharing accelerates global convergence. In addition, as summarized in Table 4, NeuralBoneReg achieves higher accuracy than its fixed-resolution variants, indicating the advantages of implicit neural surface representations. Fixed-resolution distance volumes approximate distances via interpolation, which can introduce discretization

artifacts and degrades gradient smoothness. As shown in Section 5.2, increasing the grid size improves performance on larger anatomies (UltraBones-Hip), but provides only minor gains on smaller ones (SpineDepth), illustrating the limited scalability of fixed-resolution volumes whose memory and runtime grow cubically with resolution. In contrast, NeuralBoneReg learns a smooth, continuous representation that allows distance queries at arbitrary resolution and enables stable, efficient gradient-based optimization via backpropagation.

Limitations. NeuralBoneReg has several limitations. As described in Equation 10, the method addresses a partial-to-complete registration problem, assuming that the preoperative surface $\mathcal{S}$ is complete, and the intraoperative surface $\mathcal{S}'$ is a subset ($\mathcal{S}' \subset \mathcal{S}$). If $\mathcal{S}'$ contains significant artifacts from other bones, optimization may fail. This limits the applicability of NeuralBoneReg to problems, where the set difference $[\mathcal{S}' - \mathcal{S} \cap \mathcal{S}']$ is large. In addition, registration relies on accurate bone surface segmentation. Future work could address this by integrating a classification or weighting module to down-weight outliers during loss computation. Moreover, NeuralBoneReg remains more computationally expensive than some other baseline approaches. Future work should focus on reducing runtime, through multi-GPU execution or a C++ implementation of the inference pipeline. Finally, the current study evaluated bone structures only in ex-vivo datasets. Future extensions should address larger in-vivo patient populations to confirm clinical applicability.

7. Conclusion

In this work, we introduced NeuralBoneReg, a self-supervised, modality-agnostic framework for bone surface registration that leverages implicit neural representations and parallel hypothesis generation. The method addresses fundamental challenges in CAOS: (i) symmetry induced ambiguities that create unstable optimization landscapes, (ii) limited generalizability in supervised registration networks to unseen anatomies and modalities, and (iii) scalability constraints in fixed-resolution distance volumes that hinder accurate registration of large anatomical structures. By shifting computation to the preoperative stage, NeuralBoneReg enables efficient intraoperative optimization. Our findings underline the potential of self-supervised approaches to advance CAOS by providing a more reliable foundation for surgical guidance and navigation. Future work will focus on translating our approach into

clinical application.

CRedit authorship contribution statement

Luohong Wu: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. **Matthias Seibold:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing. **Nicola A. Cavalcant:** Conceptualization, Data curation, Methodology, Resources, Writing – review and editing. **Yunke Ao:** Conceptualization, Resources, Software, Writing – review and editing. **Roman Flepp:** Software, Validation, Writing – original draft. **Aidana Massalimova:** Data curation, Resources. **Lilian Calvet:** Supervision, Methodology, Writing – review and editing. **Philipp Fürnstah:** Conceptualization, Funding acquisition, Methodology, Resources, Project administration, Supervision, Writing – review and editing.

Acknowledgement

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 and University Hospital Balgrist. This work is based on experiments performed at the Swiss Center for Musculoskeletal Imaging, SCMI, Balgrist Campus AG, Zürich.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used ChatGPT in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- [1] K. Cleary, T. M. Peters, Image-guided interventions: technology review and clinical applications, *Annual review of biomedical engineering* 12 (1) (2010) 119–142.

- [2] R. Evrard, T. Schubert, L. Paul, P.-L. Docquier, Quality of resection margin with patient specific instrument for bone tumor resection, *Journal of bone oncology* 34 (2022) 100434.
- [3] N. Sugano, Computer-assisted orthopedic surgery, *Journal of Orthopaedic Science* 8 (2003) 442–448.
- [4] N. Baraza, C. Chapman, S. Zakani, K. Mulpuri, 3d-printed patient specific instrumentation in corrective osteotomy of the femur and pelvis: a review of the literature, *3D Printing in Medicine* 6 (2020) 1–7.
- [5] K.-M. Scheufler, J. Franke, A. Eckardt, H. Dohmen, Accuracy of image-guided pedicle screw placement using intraoperative computed tomography-based navigation with automated referencing, part i: cervicothoracic spine, *Neurosurgery* 69 (4) (2011) 782–795.
- [6] E. M. Kanlić, F. DeLaRosa, M. Pirela-Cruz, Computer assisted orthopaedic surgery–caos, *Bosnian journal of basic medical sciences* 6 (1) (2006) 7.
- [7] F. Picard, A. H. Deakin, P. E. Riches, K. Deep, J. Baines, Computer assisted orthopaedic surgery: past, present and future, *Medical engineering & physics* 72 (2019) 55–65.
- [8] D. Gautam, S. Vivekanandan, M. D. Mazur, Robotic spine surgery: Systematic review of common error types and best practices, *Operative Neurosurgery* 28 (3) (2025) 295–302.
- [9] C. Wandvik, M. E. Greil, S. Colby, D. Gautam, M. D. Mazur, Limitations of current robot-assisted pedicle screw insertion systems, *Neurosurgical Focus* 57 (6) (2024) E14.
- [10] N. J. Lee, I. A. Buchanan, V. Boddapati, J. Mathew, G. Marciano, P. J. Park, E. Leung, A. L. Buchholz, J. Pollina, E. Jazini, et al., Do robot-related complications influence 1 year reoperations and other clinical outcomes after robot-assisted lumbar arthrodesis? a multicenter assessment of 320 patients, *Journal of Orthopaedic Surgery and Research* 16 (1) (2021) 308.
- [11] J. Kubicek, F. Tomanec, M. Cerny, D. Vilimek, M. Kalova, D. Oczka, Recent trends, technical concepts and components of computer-assisted

- orthopedic surgery systems: a comprehensive review, *Sensors* 19 (23) (2019) 5199.
- [12] G. Zheng, L. P. Nolte, Computer-assisted orthopedic surgery: current state and future perspective, *Frontiers in surgery* 2 (2015) 66.
 - [13] J. Tonetti, M. Boudissa, G. Kerschbaumer, O. Seurat, Role of 3d intra-operative imaging in orthopedic and trauma surgery, *Orthopaedics & Traumatology: Surgery & Research* 106 (1) (2020) S19–S25.
 - [14] S. Park, W. Plishker, H. Quon, J. Wong, R. Shekhar, J. Lee, Deformable registration of ct and cone-beam ct with local intensity matching, *Physics in Medicine & Biology* 62 (3) (2017) 927.
 - [15] P.-L. Docquier, L. Paul, O. Cartiaux, C. Delloye, X. Banse, Computer-assisted resection and reconstruction of pelvic tumor sarcoma, *Sarcoma* 2010 (1) (2010) 125162.
 - [16] M. Pojskić, M. Bopp, B. Saß, A. Kirschbaum, C. Nimsky, B. Carl, Intraoperative computed tomography-based navigation with augmented reality for lateral approaches to the spine, *Brain Sciences* 11 (5) (2021) 646.
 - [17] P. Markelj, D. Tomaževič, B. Likar, F. Pernuš, A review of 3d/2d registration methods for image-guided interventions, *Medical image analysis* 16 (3) (2012) 642–661.
 - [18] Y. Otake, M. Armand, R. S. Armiger, M. D. Kutzer, E. Basafa, P. Kazanzides, R. H. Taylor, Intraoperative image-based multiview 2d/3d registration for image-guided orthopaedic surgery: incorporation of fiducial-based c-arm tracking and gpu-acceleration, *IEEE transactions on medical imaging* 31 (4) (2011) 948–962.
 - [19] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill, et al., A comparison of similarity measures for use in 2-d-3-d medical image registration, *IEEE transactions on medical imaging* 17 (4) (1998) 586–595.
 - [20] P.-C. Ku, A. Martin-Gomez, C. Gao, R. Grupp, S. C. Mears, M. Armand, Towards 2d/3d registration of the preoperative mri to intraoperative fluoroscopic images for visualisation of bone defects, *Computer*

Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 11 (4) (2023) 1096–1105.

- [21] I. Hacihaliloglu, A. Rasouljan, R. N. Rohling, P. Abolmaesumi, Local phase tensor features for 3-d ultrasound to statistical shape+ pose spine model registration, *IEEE transactions on Medical Imaging* 33 (11) (2014) 2167–2179.
- [22] A. Roche, X. Pennec, G. Malandain, N. Ayache, Rigid registration of 3-d ultrasound with mr images: a new approach combining intensity and gradient information, *IEEE transactions on medical imaging* 20 (10) (2002) 1038–1049.
- [23] G. P. Penney, D. C. Barratt, C. S. Chan, M. Slomczykowski, T. J. Carter, P. J. Edwards, D. J. Hawkes, Cadaver validation of intensity-based ultrasound to ct registration, *Medical image analysis* 10 (3) (2006) 385–395.
- [24] P. Pandey, P. Guy, A. J. Hodgson, R. Abugharbieh, Fast and automatic bone segmentation and registration of 3d ultrasound to ct for the full pelvic anatomy: a comparative study, *International Journal of Computer Assisted Radiology and Surgery* 13 (10) (2018) 1515–1524.
URL <http://dx.doi.org/10.1007/s11548-018-1788-5>
- [25] W. Wein, A. Khamene, D.-A. Clevert, O. Kutter, N. Navab, Simulation and fully automatic multimodal registration of medical ultrasound, in: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29–November 2, 2007, Proceedings, Part I* 10, Springer, 2007, pp. 136–143.
- [26] D. K. Bae, S. J. Song, Computer assisted navigation in knee arthroplasty, *Clinics in orthopedic surgery* 3 (4) (2011) 259–267.
- [27] K. C. Anderson, K. C. Buehler, D. C. Markel, Computer assisted navigation in total knee arthroplasty: comparison with conventional methods, *The Journal of arthroplasty* 20 (2005) 132–138.
- [28] D. A. Wilson, C. Anglin, F. Ambellan, C. M. Grewe, A. Tack, H. Lamecker, M. Dunbar, S. Zachow, Validation of three-dimensional

models of the distal femur created from surgical navigation point cloud data for intraoperative and postoperative analysis of total knee arthroplasty, *International Journal of Computer Assisted Radiology and Surgery* 12 (12) (2017) 2097–2105.

- [29] F. Liebmman, M. von Atzigen, D. Stütz, J. Wolf, L. Zingg, D. Suter, N. A. Cavalcanti, L. Leoty, H. Esfandiari, J. G. Snedeker, M. R. Oswald, M. Pollefeys, M. Farshad, P. Fürnstahl, Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery, *Medical Image Analysis* 91 (2024) 103027.
- [30] X. Hu, A. Nguyen, F. R. y Baena, Occlusion-robust visual markerless bone tracking for computer-assisted orthopedic surgery, *IEEE Transactions on Instrumentation and Measurement* 71 (2021) 1–11.
- [31] I. Félix, C. Raposo, M. Antunes, P. Rodrigues, J. P. Barreto, Towards markerless computer-aided surgery combining deep segmentation and geometric pose estimation: application in total knee arthroplasty, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 9 (3) (2021) 271–278.
- [32] X. Hu, F. R. y. Baena, Automatic bone surface restoration for markerless computer-assisted orthopaedic surgery, *Chinese Journal of Mechanical Engineering* 35 (1) (2022) 18.
- [33] P. Rodrigues, M. Antunes, C. Raposo, P. Marques, F. Fonseca, J. P. Barreto, Deep segmentation leverages geometric pose estimation in computer-aided total knee arthroplasty, *Healthcare technology letters* 6 (6) (2019) 226–230.
- [34] R. B. Grupp, M. Unberath, C. Gao, R. A. Hegeman, R. J. Murphy, C. P. Alexander, Y. Otake, B. A. McArthur, M. Armand, R. H. Taylor, Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2d/3d registration, *International journal of computer assisted radiology and surgery* 15 (5) (2020) 759–769.
- [35] P.-C. Ku, M. Liu, R. Grupp, A. Harris, J. K. Oni, S. C. Mears, A. Martin-Gomez, M. Armand, End-to-end 2d/3d registration from pre-operative mri to intra-operative fluoroscopy for orthopedic procedures,

International Journal of Computer Assisted Radiology and Surgery (2025) 1–12.

- [36] K. Leung, N. Tang, L. Cheung, E. Ng, Image-guided navigation in orthopaedic trauma, *The Journal of Bone & Joint Surgery British Volume* 92 (10) (2010) 1332–1337.
- [37] W. Wein, A. Karamalis, A. Baumgartner, N. Navab, Automatic bone detection and soft tissue aware ultrasound–ct registration for computer-aided orthopedic surgery, *International journal of computer assisted radiology and surgery* 10 (2015) 971–979.
- [38] M. Salehi, R. Prevost, J.-L. Moctezuma, N. Navab, W. Wein, Precise ultrasound bone registration with learning-based segmentation and speed of sound calibration, paper presented at Medical Image Computing and Computer-Assisted Intervention – MICCAI (2017).
- [39] K. A. Eppenhof, M. W. Lafarge, M. Veta, J. P. Pluim, Progressively trained convolutional neural networks for deformable image registration, *IEEE transactions on medical imaging* 39 (5) (2019) 1594–1604.
- [40] H. Guo, X. Xu, S. Xu, B. J. Wood, P. Yan, End-to-end ultrasound frame to volume registration, in: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV* 24, Springer, 2021, pp. 56–65.
- [41] M. Ciganovic, F. Ozdemir, F. Pean, P. Fuernstahl, C. Tanner, O. Goksel, Registration of 3d freehand ultrasound to a bone model for orthopedic procedures of the forearm, *International Journal of Computer Assisted Radiology and Surgery* 13 (6) (2018) 827–836.
URL <http://dx.doi.org/10.1007/s11548-018-1756-0>
- [42] Z. Fanti, F. Torres, E. Hazan-Lasri, A. Gastelum-Strozzi, L. Ruiz-Huerta, A. Caballero-Ruiz, F. A. Cosío, Improved surface-based registration of ct and intraoperative 3d ultrasound of bones, *Journal of healthcare engineering* 2018 (1) (2018) 2365178.
- [43] R. Li, A. Davoodi, Y. Cai, K. Niu, G. Borghesan, N. Cavalcanti, A. Masalimova, F. Carrillo, C. J. Laux, M. Farshad, P. Fürnstahl, E. V.

- Poorten, Robot-assisted ultrasound reconstruction for spine surgery: from bench-top to pre-clinical study, *International Journal of Computer Assisted Radiology and Surgery* 18 (9) (2023) 1613–1623.
URL <http://dx.doi.org/10.1007/s11548-023-02932-z>
- [44] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: A survey, *IEEE transactions on medical imaging* 32 (7) (2013) 1153–1190.
 - [45] J. Woo, M. Stone, J. L. Prince, Multimodal registration via mutual information incorporating geometric and spatial context, *IEEE Transactions on Image Processing* 24 (2) (2014) 757–769.
 - [46] R. Li, A. Davoodi, Y. Cai, G. Borghesan, N. Cavalcanti, C. J. Laux, M. Farshad, F. Carrillo, P. Frnstahl, E. Vander Poorten, Development and evaluation of robot-assisted ultrasound navigation system for pedicle screw placement: An ex-vivo animal validation, *The International Journal of Medical Robotics and Computer Assisted Surgery* 20 (1) (10 2023).
URL <http://dx.doi.org/10.1002/rcs.2590>
 - [47] L. Wu, N. A. Cavalcanti, M. Seibold, G. Loggia, L. Reissner, J. Hein, S. Beeler, A. Viehfer, S. Wirth, L. Calvet, P. Frnstahl, Ultrabones100k: A reliable automated labeling method and large-scale dataset for ultrasound-based bone surface extraction, *Computers in Biology and Medicine* 194 (2025) 110435.
 - [48] J. N. Zhang, Y. Fan, D. J. Hao, Risk factors for robot-assisted spinal pedicle screw malposition, *Scientific reports* 9 (1) (2019) 3025.
 - [49] L. Ma, Z. Zhao, F. Chen, B. Zhang, L. Fu, H. Liao, Augmented reality surgical navigation with ultrasound-assisted registration for pedicle screw placement: a pilot study, *International journal of computer assisted radiology and surgery* 12 (12) (2017) 2205–2215.
 - [50] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24 (6) (1981) 381–395.
 - [51] Q.-Y. Zhou, J. Park, V. Koltun, Fast global registration, in: *European conference on computer vision*, Springer, 2016, pp. 766–782.

- [52] C. Gsaxner, A. Pepe, J. Wallner, D. Schmalstieg, J. Egger, Markerless image-to-face registration for untethered augmented reality in head and neck surgery, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 236–244.
- [53] L. Liu, Y. Zhao, A. Li, X. Yu, X. Xiao, S. Liu, M. Q.-H. Meng, A photoacoustics-enhanced drilling probe for radiation-free pedicle screw implantation in spinal surgery, *Frontiers in Bioengineering and Biotechnology* 10 (2022) 1000950.
- [54] Y. Ao, H. Esfandiari, F. Carrillo, C. J. Laux, Y. As, R. Li, K. Van Assche, A. Davoodi, N. A. Cavalcanti, M. Farshad, et al., Saferplan: Safe deep reinforcement learning for intraoperative planning of pedicle screw placement, *Medical Image Analysis* 99 (2025) 103345.
- [55] J. Liu, J. Hao, H. Lin, W. Pan, J. Yang, Y. Feng, G. Wang, J. Li, Z. Jin, Z. Zhao, et al., Deep learning-enabled 3d multimodal fusion of cone-beam ct and intraoral mesh scans for clinically applicable tooth-bone reconstruction, *Patterns* 4 (9) (2023).
- [56] J. Hao, J. Liu, J. Li, W. Pan, R. Chen, H. Xiong, K. Sun, H. Lin, W. Liu, W. Ding, et al., Ai-enabled automatic multimodal fusion of cone-beam ct and intraoral scans for intelligent 3d tooth-bone reconstruction and clinical applications, *arXiv preprint arXiv:2203.05784* (2022).
- [57] A. Chan, B. Coutts, E. Parent, E. Lou, Development and evaluation of ct-to-3d ultrasound image registration algorithm in vertebral phantoms for spine surgery, *Annals of Biomedical Engineering* 49 (1) (2021) 310–321.
- [58] M. Unberath, C. Gao, Y. Hu, M. Judish, R. H. Taylor, M. Armand, R. Grupp, The impact of machine learning on 2d/3d registration for image-guided interventions: A systematic review and perspective, *Frontiers in Robotics and AI* 8 (2021) 716007.
- [59] Y.-X. Zhang, J. Gui, X. Cong, X. Gong, W. Tao, A comprehensive survey and taxonomy on point cloud registration based on deep learning, *arXiv preprint arXiv:2404.13830* (2024).
- [60] U. Castellani, A. Bartoli, 3d shape registration, in: *3D Imaging, Analysis and Applications*, Springer, 2020, pp. 353–411.

- [61] F. Liebmam, D. Stütz, D. Suter, S. Jecklin, J. G. Snedeker, M. Farshad, P. Fürnstahl, H. Esfandiari, Spinedepth: A multi-modal data collection approach for automatic labelling and intraoperative spinal shape reconstruction based on rgb-d data, *Journal of Imaging* 7 (9) (2021) 164.
URL <http://dx.doi.org/10.3390/jimaging7090164>
- [62] P. J. Besl, 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.
- [63] E. W. Nottmeier, T. L. Crosby, Timing of paired points and surface matching registration in three-dimensional (3d) image-guided spinal surgery, *Clinical Spine Surgery* 20 (4) (2007) 268–270.
- [64] P. Foroughi, R. H. Taylor, G. Fichtinger, Automatic initialization for 3d bone registration, in: *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, Vol. 6918, SPIE, 2008, pp. 895–902.
- [65] Z. Zhang, Y. Dai, J. Sun, Deep learning based point cloud registration: an overview, *Virtual Reality & Intelligent Hardware* 2 (3) (2020) 222–246.
- [66] C. Choy, J. Park, V. Koltun, Fully convolutional geometric features, in: *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8958–8966.
- [67] Z. Gojcic, C. Zhou, J. D. Wegner, A. Wieser, The perfect match: 3d point cloud matching with smoothed densities, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5545–5554.
- [68] S. Ao, Q. Hu, B. Yang, A. Markham, Y. Guo, Spinnet: Learning a general surface descriptor for 3d point cloud registration, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 11753–11762.
- [69] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, T. Funkhouser, 3dmatch: Learning local geometric descriptors from rgb-d reconstructions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1802–1811.

- [70] H. Wang, Y. Liu, Z. Dong, W. Wang, You only hypothesize once: Point cloud registration with rotation-equivariant descriptors, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 1630–1641.
- [71] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, 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.
- [72] G. D. Pais, S. Ramalingam, V. M. Govindu, J. C. Nascimento, R. Chellappa, P. Miraldo, 3dregnet: A deep neural network for 3d point registration, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 7193–7203.
- [73] W. Yuan, B. Eckart, K. Kim, V. Jampani, D. Fox, J. Kautz, Deepgmr: Learning latent gaussian mixture models for registration, in: European conference on computer vision, Springer, 2020, pp. 733–750.
- [74] Y. Wang, J. M. Solomon, Deep closest point: Learning representations for point cloud registration, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 3523–3532.
- [75] W. Lu, G. Wan, Y. Zhou, X. Fu, P. Yuan, S. Song, Deepvcv: An end-to-end deep neural network for point cloud registration, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 12–21.
- [76] H. Xu, N. Ye, G. Liu, B. Zeng, S. Liu, Finet: Dual branches feature interaction for partial-to-partial point cloud registration, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 2848–2856.
- [77] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, S. Ilic, D. Hu, K. Xu, Geotransformer: Fast and robust point cloud registration with geometric transformer, IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (8) (2023) 9806–9821.
- [78] H. Deng, T. Birdal, S. Ilic, Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 602–618.

- [79] G. Mei, H. Tang, X. Huang, W. Wang, J. Liu, J. Zhang, L. Van Gool, Q. Wu, Unsupervised deep probabilistic approach for partial point cloud registration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 13611–13620.
- [80] L. Wu, M. Seibold, N. A. Cavalcanti, G. Loggia, L. Reissner, B. Sigrist, J. Hein, L. Calvet, A. Viehöfer, P. Fürnstahl, Ultraboneudf: Self-supervised bone surface reconstruction from ultrasound based on neural unsigned distance functions, arXiv preprint arXiv:2505.17912 (2025).
- [81] H. Chen, L. Kumaralingam, S. Zhang, S. Song, F. Zhang, H. Zhang, T.-T. Pham, K. Punithakumar, E. H. Lou, Y. Zhang, L. H. Le, R. Zheng, Neural implicit surface reconstruction of freehand 3d ultrasound volume with geometric constraints, Medical Image Analysis 98 (2024) 103305.
URL <https://www.sciencedirect.com/science/article/pii/S1361841524002305>
- [82] B. Ma, Z. Han, Y.-S. Liu, M. Zwicker, Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces (2020).
URL <https://arxiv.org/abs/2011.13495>
- [83] H. Chen, L. Kumaralingam, J. Li, K. Punithakumar, L. H. Le, R. Zheng, Neural implicit representation for three-dimensional ultrasound carotid surface reconstruction using unsigned distance function, in: 2023 IEEE International Ultrasonics Symposium (IUS), IEEE, 2023, pp. 1–3.
- [84] C. M. Bishop, Mixture density networks (1994).
- [85] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson, G. Gkioxari, Accelerating 3d deep learning with pytorch3d, arXiv:2007.08501 (2020).
- [86] E. Kerkhof, A. Thabit, M. Benmahdjoub, P. Ambrosini, T. van Ginhoven, E. B. Wolvius, T. van Walsum, Depth-based registration of 3d preoperative models to intraoperative patient anatomy using the hololens 2, International Journal of Computer Assisted Radiology and Surgery (2025) 1–12.
- [87] R. Fletcher, Practical methods of optimization, John Wiley & Sons, 2000.

- [88] S. Winter, B. Brendel, I. Pechlivanis, K. Schmieder, C. Igel, Registration of ct and intraoperative 3-d ultrasound images of the spine using evolutionary and gradient-based methods, *IEEE Transactions on Evolutionary Computation* 12 (3) (2008) 284–296.
- [89] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization* 11 (1997) 341–359.
- [90] M. Pant, H. Zaheer, L. Garcia-Hernandez, A. Abraham, et al., Differential evolution: A review of more than two decades of research, *Engineering Applications of Artificial Intelligence* 90 (2020) 103479.