

# End2Reg: Learning Task-Specific Segmentation for Markerless Registration in Spine Surgery

Lorenzo Pettinari<sup>1,2</sup>[0009–0009–1722–9024], Sidaty El Hadramy<sup>1</sup>[0009–0000–2917–0706], Michael Wehrli<sup>1</sup>[0009–0005–8740–3295], Philippe C. Cattin<sup>1</sup>[0000–0001–8785–2713], Daniel Studer<sup>2</sup>[0000–0003–1893–4859], Carol C. Hasler<sup>2</sup>, and Maria Licci<sup>2</sup>[0000–0002–2625–4151]

<sup>1</sup> Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland  
[lorenzo.pettinari@unibas.ch](mailto:lorenzo.pettinari@unibas.ch)

<sup>2</sup> Department of Orthopedics, University Children’s Hospital, Basel, Switzerland

**Abstract.** Intraoperative navigation in spine surgery demands millimeter-level accuracy. Currently, this is achieved through radiation-intensive intraoperative imaging and bone-anchored markers that are invasive and disrupt surgical workflow. Markerless RGB-D registration methods offer a promising alternative. However, existing approaches rely on weak segmentation labels to isolate relevant anatomical structures, potentially propagating errors through the registration process. We present End2Reg, an end-to-end deep learning framework that jointly optimizes segmentation and registration, eliminating the need for segmentation labels and manual steps. The network learns task-specific segmentation masks optimized for registration, guided solely by the registration objective without explicit segmentation supervision. End2Reg achieves state-of-the-art performance on ex- and in-vivo benchmarks, reducing median Target Registration Error by 32% and mean Root Mean Square Error by 61%, while maintaining robust performance under partial occlusions. Ablation results confirm that end-to-end optimization significantly improves registration accuracy. Overall, End2Reg advances towards fully automatic, markerless intraoperative navigation. Code and interactive visualizations are available at: <https://lorenzopettinari.github.io/end-2-reg/>.

**Keywords:** Markerless registration · RGB-D · Point clouds · End-to-end training

## 1 Introduction

Intraoperative navigation systems are essential in orthopedic surgery, where millimeter-level accuracy is crucial for safety and clinical success [16]. This is particularly evident in pedicle screw insertion, where misplacement can cause neural or vascular injury [6,12]. Effective navigation relies on robust registration to align image-derived anatomy with the surgical coordinate system, enabling instrument tracking [5,12]. Preoperative imaging such as CT or MRI allows detailed surgical planning, including implant selection and trajectory optimization

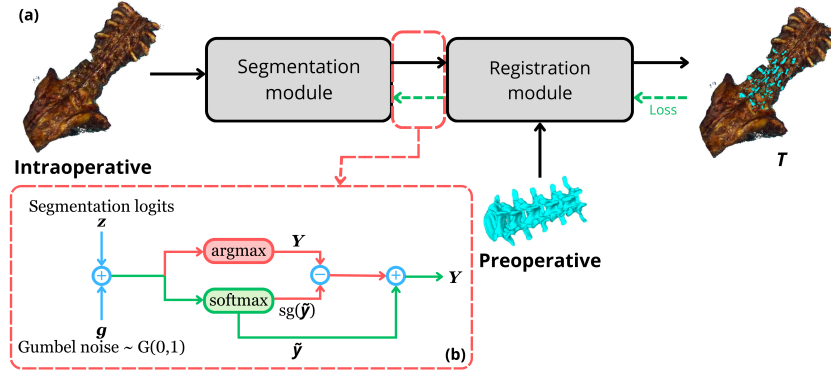


Fig. 1: Overview of the **End2Reg** framework. (a) Network architecture integrates segmentation and registration to estimate the rigid transformation  $T$  aligning intraoperative RGB-D-derived point clouds and preoperative point clouds. (b) Straight-Through Gumbel-Softmax (ST-GS) estimator enabling **end-to-end optimization**, with discrete sampling in the forward pass and soft gradients in the backward pass (Figure inspired by Wei *et al.* [27]).

[5]. For integration into navigation systems, accurate alignment between preoperative images and intraoperative anatomy is essential. Current approaches typically achieve this alignment via intraoperative 3D radiographic imaging and bone-anchored fiducial markers [7,12]. While effective, these strategies have several drawbacks: ionizing radiation exposure for patients and staff, particularly when repeated scans are required in multi-level procedures [22]; high costs and a large operating room footprint; prolonged workflow [24]; and invasive markers, which can loosen during surgery and require repeated registrations [7]. To overcome these limitations, recent work has investigated radiation-free and markerless approaches, often using RGB-D cameras to capture the exposed surgical surface. RGB-D sensors are compact, low-cost and easy to integrate into the operating room. In these methods, intraoperative video frames are converted into 3D point clouds, while preoperative scans are segmented and sampled to obtain point clouds of the anatomy (e.g., the spine). Registration is then formulated as a point cloud alignment problem between pre- and intraoperative data [3,8,13,15].

**Problem formulation:** In this context, intraoperative registration can be formulated as a **heterogeneous point cloud registration** problem performed independently for each RGB-D video frame. Let  $\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3\}_{i=1}^M$  be the *preoperative source point cloud*, and  $\mathcal{Q} = \{\mathbf{q}_j \in \mathbb{R}^3\}_{j=1}^N$  the *intraoperative target point cloud*. Our objective is to estimate a rigid transformation  $T = (\mathbf{R}, \mathbf{t}) \in SE(3)$  that aligns  $\mathcal{P}$  to  $\mathcal{Q}$ , thereby minimizing geometric registration errors such as Target Registration Error (TRE) or Root Mean Square Error (RMSE). This task is challenging due to the **heterogeneous nature** of the data: preoperative and intraoperative point clouds are extracted from different modalities, and intraoperative clouds can include soft tissue **partially obscuring bone**, have **low**

**overlap** due to limited surgical exposure, and be **corrupted by occlusions** from instruments or the surgeon [3,8,13,29].

**Related work:** Point cloud registration has been extensively studied [17]. Traditional methods include ICP [2], which is sensitive to initialization and degrades with low overlap, and RANSAC applied to hand-crafted descriptors such as FPFH [4,21], which improves robustness to outliers and partial overlap but depends on reliable feature extraction. Recently, deep learning-based methods have improved robustness under noise and partial overlap [9,17,18,19,20], nevertheless, their use on heterogeneous surgical point clouds remains limited [26]. In orthopedic surgery, RGB-D-based registration pipelines typically rely on ICP combined with segmentation to isolate the bony anatomy from the surgically exposed surface [3,8,11,13,15,25,29]. When manual annotations are unavailable, segmentation is commonly supervised using automatically generated weak bone labels obtained from the ground-truth-aligned preoperative anatomy [3,8,13,15,25]. Although effective for large-scale data generation, these labels are noisy and can propagate errors into the registration pipeline [25]. Additionally, these methods often still require a manual step to select the expected region of overlap on the preoperative anatomy [8,11,13,29].

**Contributions:** We, therefore, propose **End2Reg** (Figure 1), an end-to-end deep learning framework for vertebral registration. Our main contributions are:

- **Joint learning of segmentation and registration:** segmentation and registration models are trained end-to-end without segmentation labels, allowing End2Reg to learn task-specific segmentations optimized for registration.
- **State-of-the-art registration performance:** validated on the publicly available SpineDepth [14] (*ex vivo*) and SpineAlign [3] (*in vivo*) datasets.
- **Robustness to occlusions:** End2Reg demonstrates improved robustness under partial occlusions, compared to existing methods.

## 2 Method

To address limitations of current RGB-D registration pipelines in orthopedic surgery, we train segmentation and registration end-to-end. By optimizing segmentation through the registration loss, **End2Reg** learns task-specific segmentations tailored for downstream registration without any segmentation label. An overview of the architecture is shown in Figure 1 and the following subsections describe its main components.

### 2.1 Segmentation Module

The segmentation module processes target point clouds  $Q$ , in which each point is defined by its 3D coordinates and normalized RGB color values. Its task is to predict a binary mask  $\mathbf{Y}$  identifying regions relevant for registration (e.g., bony anatomy). We implement it using KPConv [23] in a U-Net-like encoder-decoder architecture. We chose KPConv due to its ability to extract features from irregular and sparse point clouds [23]. Given an input point cloud  $Q$  with

$N$  points, the network outputs logits  $\mathbf{z}_j = [z_j^0, z_j^1]$  for each point  $j = 1, \dots, N$ , corresponding to the irrelevant and relevant classes, respectively. Applying a softmax yields class probabilities  $\boldsymbol{\pi}_j = \text{softmax}(\mathbf{z}_j)$ . A binary segmentation mask  $\mathbf{Y} = [Y_1, \dots, Y_N]$  is obtained by assigning each point to the class with the highest probability, where  $Y_j \in \{0, 1\}$  denotes the predicted label of point  $j$ . This mask is then used as the per-point feature input to the registration module.

## 2.2 Registration Module

The registration module takes the source and target point clouds  $P$  and  $Q$  and predicts the rigid transformation  $T$  that aligns them. We base it upon GeoTransformer as this demonstrated superior registration accuracy compared to other tested methods, is robust to low-overlap, and incorporates outlier filtering [20]. GeoTransformer builds upon KPConv [23] for feature extraction. In the standard KPConv formulation, when no additional features are available, each point is initialized with a constant scalar feature equal to one, forcing the network to rely solely on geometric relationships between neighboring points. Following this convention, points in the source cloud  $P$  are assigned a constant feature of 1. For the target point cloud  $Q$ , we instead leverage the mask  $\mathbf{Y}$  produced by the segmentation module. Specifically, each point  $\mathbf{q}_j \in Q$  is assigned a binary feature  $F(\mathbf{q}_j) = Y_j \in \{0, 1\}$ . The KPConv operation at a query point  $\mathbf{q}$  can be expressed as

$$(F * g)(\mathbf{q}) = \sum_{\mathbf{q}_j \in \mathcal{N}_q} g(\mathbf{q}_j - \mathbf{q}) F(\mathbf{q}_j), \quad (1)$$

where  $\mathbf{q}_j$  denotes the position of a neighboring point,  $\mathcal{N}_q$  is the neighborhood of  $\mathbf{q}$ , and  $g(\cdot)$  is the KPConv kernel function. In this formulation, all points of  $Q$  remain part of the geometry, but only those with  $Y_j = 1$  contribute to the convolution output. This design ensures that the registration module focuses on regions identified by  $\mathbf{Y}$  as relevant.

## 2.3 End-to-End Training

In our formulation, the segmentation module outputs a per-point binary mask  $\mathbf{Y}$  that conditions the registration module. Hard mask generation via  $\arg \max$  yields a deterministic, non-differentiable assignment. As a result, **gradients from the registration loss cannot propagate back to the segmentation module**. To enable end-to-end training, we adopt the **Gumbel-Softmax estimator** [10], which provides a differentiable approximation of categorical sampling via the Gumbel-Max trick. Specifically, by adding i.i.d. noise  $g_j^i \sim \text{Gumbel}(0, 1)$  to each logit  $z_j^i$ , we obtain  $Y_j = \arg \max_i (z_j^i + g_j^i)$ , which is a single sample drawn from the categorical distribution defined by  $\boldsymbol{\pi}_j$ . We then introduce a differentiable relaxation of the categorical sample using the Gumbel-Softmax distribution,

$$\tilde{y}_j^i = \frac{\exp((z_j^i + g_j^i)/\tau)}{\sum_{c=0}^1 \exp((z_j^c + g_j^c)/\tau)}, \quad i \in \{0, 1\}, \quad j = 1, \dots, N \quad (2)$$

where  $\tau > 0$ . Unlike  $\mathbf{Y}$ ,  $\tilde{\mathbf{y}}$  is continuous and differentiable with respect to  $\mathbf{z}$ . As  $\tau \rightarrow 0$ ,  $\tilde{\mathbf{y}}$  converges in distribution to a one-hot vector, recovering categorical samples; for larger  $\tau$ , it produces smoother values. To combine the discrete behavior of  $\mathbf{Y}$  with the differentiability of  $\tilde{\mathbf{y}}$ , we employ the **Straight-Through Gumbel-Softmax Estimator (ST-GS)** [1,10], as illustrated in Figure 1, following the standard **Straight-Through Estimator** [1] formulation:

$$\mathbf{Y}_{\text{STE}} = \mathbf{Y} - \text{sg}(\tilde{\mathbf{y}}) + \tilde{\mathbf{y}}. \quad (3)$$

where  $\text{sg}(\cdot)$  denotes the stop-gradient operator, which treats its input as a constant during backpropagation. In the **forward pass**,  $\mathbf{Y}_{\text{STE}} = \mathbf{Y}$ , because  $\tilde{\mathbf{y}} - \text{sg}(\tilde{\mathbf{y}}) = 0$ . During **backpropagation**, the gradient of  $\mathbf{Y}_{\text{STE}}$  w.r.t.  $\mathbf{z}$  is approximated as:

$$\frac{\partial \mathbf{Y}_{\text{STE}}}{\partial \mathbf{z}} = \frac{\partial (\mathbf{Y} - \text{sg}(\tilde{\mathbf{y}}) + \tilde{\mathbf{y}})}{\partial \mathbf{z}} = \underbrace{\frac{\partial \mathbf{Y}}{\partial \mathbf{z}}}_{=0} - \underbrace{\frac{\partial \text{sg}(\tilde{\mathbf{y}})}{\partial \mathbf{z}}}_{=0} + \frac{\partial \tilde{\mathbf{y}}}{\partial \mathbf{z}} \approx \frac{\partial \tilde{\mathbf{y}}}{\partial \mathbf{z}}. \quad (4)$$

The end-to-end framework is trained with the dual-phase coarse-to-fine loss of GeoTransformer [20]. Further details can be found in Qin *et al.* [20].

### 3 Experiments and Results

#### 3.1 Datasets and implementation details

**End2Reg** was evaluated on two benchmark datasets for vertebral registration: SpineDepth [14] (*ex vivo*) and SpineAlign [3] (*in vivo* and limited surgical exposure), covering both controlled and realistic settings. In the following, we provide details about these datasets as well as their preprocessing:

- **SpineDepth (*ex vivo*)** [14] consists of RGB-D recordings of pedicle screw placement on ten cadaver specimens, with preoperative 3D vertebral models extracted from CT. Ground-truth vertebral poses relative to the RGB-D sensor are provided for each frame via optical tracking. In contrast to Liebmann *et al.* [13], we retain frames with up to 50% occlusion and evaluate the full dataset without viewpoint filtering to reflect realistic intraoperative scenarios. An expert surgeon manually annotated anatomical landmarks (spinous, left and right transverse processes) for each vertebra following Liebmann *et al.* [13]; these annotations were used to compute the TRE. Evaluation was performed using 8-fold cross-validation.
- **SpineAlign (*in vivo*)** [3] contains 24 lumbar open spine surgeries, with preoperative 3D vertebral models extracted from CT or MRI scans. Coarse alignment between pre- and intraoperative point clouds is provided using anatomical landmarks identified by the surgeon. Following Daly *et al.* [3], 20 patients were used for training and 4 for testing, using the same data split.

In both datasets, point clouds were normalized to a unit sphere, and random rigid transformations were applied to each frame to simulate anatomical misalignment. Similar to Pan *et al.* [19] and Qin *et al.* [20], translations were limited to 10% of the point cloud size ( $\approx 51$  mm on SpineDepth,  $\approx 13$  mm on SpineAlign) and rotations to a maximum of  $45^\circ$ , as larger misalignments are usually corrected by generic initialization (e.g., aligning center-of-mass or PCA). The training was performed on an NVIDIA RTX 5090 GPU (32 GB) with a batch size of 1, an initial learning rate of  $10^{-4}$ , a 10 000-iteration warm-up, and cosine annealing.

### 3.2 Results

**SpineDepth registration results:** Table 1 presents the TRE along with translation (RTE) and rotation (RRE) errors for our method, compared to the state-of-the-art results reported by Liebmann *et al.* (AutoReg) [13] using the same dataset split. The qualitative results are illustrated in Figure 2. In addition, to further prove the robustness of our approach to occlusion, we report the TRE as a function of the occlusion ratio in Figure 3.



Fig. 2: **End2Reg registration results** of the preoperative to the intraoperative point cloud on three SpineDepth specimens. Colors indicate the per-point displacement magnitude relative to the ground-truth pose.

**SpineAlign registration results:** Table 1 reports the registration results, compared to the state-of-the-art reported by Daly *et al.* (CorrNet) [3]. The dataset provides only a coarse pre-to-intraoperative alignment, which Daly *et al.* use as initialization, whereas our approach operates without relying on this prior. As no reliable ground truth is available, the results should be interpreted as a preliminary benchmark. Following Daly *et al.* [3], we report RMSE and fitness.

**Further comparing methods:** We additionally compared **End2Reg** against a set of segmentation-registration baselines. All the baselines use the same KPConv-based segmentation network as End2Reg while differing in the registration module. In contrast to End2Reg, segmentation and registration networks were trained independently. The segmentation module was trained using a cross-entropy loss with weak bone labels, generated by assigning points in  $Q$  within a threshold distance of the surface  $P$  aligned to the ground-truth pose [3,25]. Thresholds of 3 mm for SpineDepth and 10 mm for SpineAlign were used, yielding segmentation

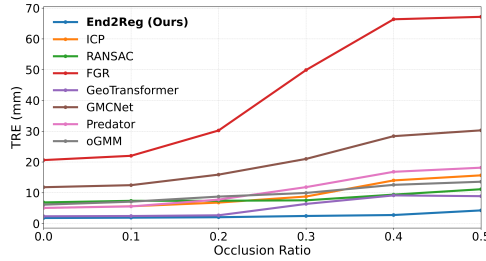


Fig. 3: **TRE under increasing occlusion.** Comparison of different methods as the occlusion ratio increases from 0 to 0.5. End2Reg demonstrates the highest robustness, showing the smallest TRE across all occlusion levels.

performance (Table 2) consistent with prior work [3,13]. For registration, we evaluated both classical [2,4,28] and learning-based methods [9,18,19,20]. Consistent with our problem formulation, we selected learning-based rigid registration methods designed for low-overlap scenarios. To aid classical baselines on SpineDepth, we follow prior work [13] and include an additional manual step consisting of selecting the expected region of overlap on the preoperative anatomy. Quantitative analysis of the inference time for comparing methods is provided in Table 3.

**Task-specific segmentation:** Qualitative results (Figure 4) show that our model highlights regions informative for registration, corresponding primarily to bony anatomy while suppressing surrounding soft tissue. Importantly, these task-specific segmentations are not intended to reproduce anatomical ground truth or weak bone labels but are optimized for downstream registration. Accordingly, overlap-based metrics are not used as primary evaluation criteria. Instead (Table 2), we report the one-sided Chamfer distance and the 95th percentile Hausdorff distance from the weak-label point cloud to the predicted segmentation, measuring the containment of weak bone labels.

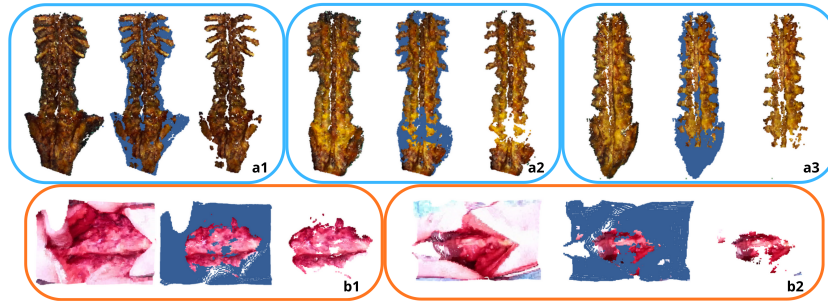


Fig. 4: Task-specific segmentations on three specimens for SpineDepth (a) and two cases on SpineAlign (b). For each, three point clouds are shown: intraoperative input point cloud, segmentation (background in blue), and registration-relevant points, which mainly correspond to bone structures.

Table 1: Registration results. SpineDepth: RRE (deg), RTE (mm), TRE (mm) reported as median [Q1, Q3], following Liebmann *et al.* [13]. SpineAlign: Fitness, RMSE (mm) reported as mean ( $\pm$  st. dev.), following Daly *et al.* [3].

| Method                             | SpineDepth           |                      |                      | SpineAlign                      |                                   |
|------------------------------------|----------------------|----------------------|----------------------|---------------------------------|-----------------------------------|
|                                    | RRE $\downarrow$     | RTE $\downarrow$     | TRE $\downarrow$     | Fitness $\uparrow$              | RMSE $\downarrow$                 |
| <i>Classical methods</i>           |                      |                      |                      |                                 |                                   |
| ICP(1)                             | 2.3[1.5, 3.4]        | 5.0[3.4, 6.7]        | 5.4[3.8, 7.2]        | .49( $\pm$ .25)                 | 12.26( $\pm$ 4.51)                |
| RANSAC+ICP(1)                      | 1.9[1.2, 4.0]        | 3.7[2.3, 5.7]        | 4.2[2.7, 6.7]        | .51( $\pm$ .27)                 | 12.70( $\pm$ 6.42)                |
| FGR+ICP(1)                         | 2.3[1.3, 11.5]       | 15.9[10.3, 30.6]     | 17.2[10.8, 37.2]     | .21( $\pm$ .18)                 | 17.89( $\pm$ 5.13)                |
| <i>Deep learning-based methods</i> |                      |                      |                      |                                 |                                   |
| GMCNet [19]                        | 11.1[7.1, 17.1]      | 10.0[7.2, 13.9]      | 12.7[9.5, 17.1]      | .22( $\pm$ .20)                 | 9.31( $\pm$ 2.91)                 |
| GeoTransformer [20]                | 1.4[1.0, 2.0]        | 1.9[1.2, 3.1]        | 2.2[1.5, 3.4]        | .70( $\pm$ .12)                 | 3.04( $\pm$ 1.76)                 |
| OverlapPredator [9]                | 2.8[1.9, 4.1]        | 4.8[3.5, 6.6]        | 5.1[3.8, 7.0]        | .34( $\pm$ .18)                 | 9.93( $\pm$ 4.58)                 |
| oGMM [18]                          | 3.1[1.8, 5.5]        | 6.4[4.0, 9.2]        | 6.7[4.3, 9.7]        | .14( $\pm$ .09)                 | 9.73( $\pm$ 2.48)                 |
| AutoReg [13](1, 2)                 | –                    | –                    | 2.7[1.7, 3.6]        | –                               | –                                 |
| CorrNet [3](3)                     | –                    | –                    | –                    | .58( $\pm$ .11)                 | 7.14( $\pm$ 0.47)                 |
| <b>End2Reg (Ours)</b>              | <b>1.3[0.8, 1.8]</b> | <b>1.9[1.2, 2.7]</b> | <b>1.8[1.2, 2.7]</b> | <b>.71(<math>\pm</math>.18)</b> | <b>2.80(<math>\pm</math>1.31)</b> |

- (1) Requires manual selection of expected overlapping regions on SpineDepth data.  
(2) Results obtained on a filtered dataset with specific viewpoints and no occlusions.  
(3) Requires manual initialization of the registration.

**Ablation Study:** We evaluated the effect of end-to-end training by comparing **End2Reg** with the same architecture trained in a two-step manner, where segmentation and registration are optimized sequentially rather than jointly (reported as *GeoTransformer* in Table 1). Wilcoxon signed-rank tests on SpineDepth results show that the improvement of End2Reg is significant ( $p \ll 0.05$ ) with a moderate effect size ( $r = 0.47$ ) and reduced outlier ratio (3.8% vs 7.0%).

## 4 Conclusion

We presented **End2Reg**, an **end-to-end framework** for pre-to-intraoperative vertebral point cloud registration achieving state-of-the-art results on SpineDepth

Table 2: Task-specific segmentation quantitative evaluation: Dice, IoU, one-sided Chamfer (mm) and HD95 (mm). Results as median [Q1, Q3] for SpineDepth, as Liebmann *et al.* [13], and mean ( $\pm$  st. dev.) for SpineAlign, as Daly *et al.* [3].

| Method         | SpineDepth           |                      |                      | SpineAlign                      |                                 |                                 |
|----------------|----------------------|----------------------|----------------------|---------------------------------|---------------------------------|---------------------------------|
|                | Dice $\uparrow$      | Chamfer $\downarrow$ | HD95 $\downarrow$    | IoU $\uparrow$                  | Chamfer $\downarrow$            | HD95 $\downarrow$               |
| AutoReg [13]   | <b>.74[.72, .74]</b> | –                    | –                    | –                               | –                               | –                               |
| CorrNet [3]    | –                    | –                    | –                    | .78( $\pm$ .00)                 | –                               | –                               |
| KPConv         | .73[.68, .77]        | 1.5[1.0, 2.4]        | 10.5[7.2, 15.3]      | <b>.84(<math>\pm</math>.05)</b> | <b>0.5(<math>\pm</math>1.3)</b> | <b>3.3(<math>\pm</math>3.2)</b> |
| <b>End2Reg</b> | .50[.42, .54]        | <b>0.9[0.6, 1.3]</b> | <b>5.6[4.3, 7.3]</b> | .70( $\pm$ .06)                 | 0.8( $\pm$ 0.3)                 | 4.2( $\pm$ 0.9)                 |

Table 3: Inference times (ms) for automatic registration methods. For multi-stage approaches, the time includes both segmentation and registration.

| Dataset    | GMCNet | GeoTransformer | Overlap      | oGMM | End2Reg |
|------------|--------|----------------|--------------|------|---------|
|            | [19]   | [20]           | Predator [9] | [18] | (Ours)  |
| SpineDepth | 400    | 638            | 960          | 1150 | 620     |
| SpineAlign | 280    | 455            | 485          | 738  | 450     |

(*ex vivo*) [14] and SpineAlign (*in vivo*) [3]. Our method **removes the need for segmentation labels** by learning task-specific segmentations optimized for registration. It enables accurate alignment of the full preoperative anatomy while remaining robust to occlusions. Evaluation was limited by available datasets: SpineDepth lacks preoperative deformation, and SpineAlign provides only coarse manual alignment. Future work will address intervertebral deformation and acquire an *in vivo* dataset with CT-based ground-truth alignment. Overall, End2Reg shows strong potential to improve accuracy and automation of markerless intraoperative registration, supporting its integration into clinical workflows.

**Disclosure of Interests.** The authors have no competing interests to declare relevant to this article’s content.

## References

1. Bengio, Y., Léonard, N., Courville, A.: Estimating or propagating gradients through stochastic neurons for conditional computation. arXiv preprint arXiv:1308.3432 (2013)
2. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures. vol. 1611, pp. 586–606. Spie (1992)
3. Daly, C., Marconi, E., Riva, M., Ekanayake, J., Elson, D.S., Rodriguez Y Baena, F.: Towards markerless intraoperative tracking of deformable spine tissue. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 627–637. Springer (2025)
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
5. Floyd, E., Cohn, P., Ferguson, J., Kalantar, S.B.: A review of preoperative planning technologies for spinal deformity correction. In: Seminars in Spine Surgery. vol. 32, p. 100787. Elsevier (2020)
6. Gelalis, I.D., Paschos, N.K., Pakos, E.E., Politis, A.N., Arnaoutoglou, C.M., Karageorgos, A.C., Ploumis, A., Xenakis, T.A.: Accuracy of pedicle screw placement: a systematic review of prospective in vivo studies comparing free hand, fluoroscopy guidance and navigation techniques. European spine journal **21**(2), 247–255 (2012)
7. Holly, L.T., Foley, K.T.: Intraoperative spinal navigation. Spine **28**(15S), S54–S61 (2003)
8. Hu, X., Nguyen, A., Rodriguez Y Baena, F.: Occlusion-robust visual markerless bone tracking for computer-assisted orthopedic surgery. IEEE Transactions on Instrumentation and Measurement **71**, 1–11 (2021)

9. Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K.: Predator: Registration of 3d point clouds with low overlap. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. pp. 4267–4276 (2021)
10. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)
11. Ji, S., Fan, X., Paulsen, K.D., Roberts, D.W., Mirza, S.K., Lollis, S.S.: Patient registration using intraoperative stereovision in image-guided open spinal surgery. *IEEE Transactions on Biomedical Engineering* **62**(9), 2177–2186 (2015)
12. Karkenny, A.J., Mendelis, J.R., Geller, D.S., Gomez, J.A.: The role of intraoperative navigation in orthopaedic surgery. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* **27**(19), e849–e858 (2019)
13. Liebmann, F., von Atzigen, M., Stütz, D., Wolf, J., Zingg, L., Suter, D., Cavalcanti, N.A., Leoty, L., Esfandiari, H., Snedeker, J.G., et al.: Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery. *Medical Image Analysis* **91**, 103027 (2024)
14. Liebmann, F., Stütz, D., Suter, D., Jecklin, S., Snedeker, J.G., Farshad, M., Fürnstahl, P., Esfandiari, H.: 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), 164 (2021)
15. Liu, H., Rodriguez Y Baena, F.: Automatic markerless registration and tracking of the bone for computer-assisted orthopaedic surgery. *IEEE Access* **8**, 42010–42020 (2020)
16. Luther, N., Iorgulescu, J.B., Geannette, C., Gebhard, H., Saleh, T., Tsiouris, A.J., Härtl, R.: Comparison of navigated versus non-navigated pedicle screw placement in 260 patients and 1434 screws: screw accuracy, screw size, and the complexity of surgery. *Clinical Spine Surgery* **28**(5), E298–E303 (2015)
17. Lyu, M., Yang, J., Qi, Z., Xu, R., Liu, J.: Rigid pairwise 3d point cloud registration: A survey. *Pattern Recognition* **151**, 110408 (2024)
18. Mei, G., Poesi, F., Saltori, C., Zhang, J., Ricci, E., Sebe, N.: Overlap-guided gaussian mixture models for point cloud registration. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 4511–4520 (2023)
19. Pan, L., Cai, Z., Liu, Z.: Robust partial-to-partial point cloud registration in a full range. *IEEE Robotics and Automation Letters* **9**(3), 2861–2868 (2024)
20. Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Ilic, S., Hu, D., Xu, K.: Geotransformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(8), 9806–9821 (2023)
21. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: 2009 IEEE international conference on robotics and automation. pp. 3212–3217. IEEE (2009)
22. Striano, B.M., Crawford, A.M., Verhofste, B.P., Hresko, A.M., Hedequist, D.J., Schoenfeld, A.J., Simpson, A.K.: Intraoperative navigation increases the projected lifetime cancer risk in patients undergoing surgery for adolescent idiopathic scoliosis. *The Spine Journal* **24**(6), 1087–1094 (2024)
23. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6411–6420 (2019)
24. Tonetti, J., Boudissa, M., Kerschbaumer, G., Seurat, O.: Role of 3d intraoperative imaging in orthopedic and trauma surgery. *Orthopaedics & Traumatology: Surgery & Research* **106**(1), S19–S25 (2020)

25. Warner, W.R., Bhattacharya, I., Evans, L.T., Mirza, S.K., Paulsen, K.D., Fan, X.: Sparse-xm: Spine pose adjustment with rgb-d bone segmentation via cross-modality label transfer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 532–541. Springer (2025)
26. Weber, M., Wild, D., Kleesiek, J., Egger, J., Gsaxner, C.: Deep learning-based point cloud registration for augmented reality-guided surgery. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
27. Wei, T., Patel, Y., Shekhovtsov, A., Matas, J., Barath, D.: Generalized differentiable ransac. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 17649–17660 (2023)
28. Zhou, Q.Y., Park, J., Koltun, V.: Fast global registration. In: European conference on computer vision. pp. 766–782. Springer (2016)
29. Zhu, S., Zhao, Z., Pan, Y., Zheng, G.: Markerless robotic pedicle screw placement based on structured light tracking. International Journal of Computer Assisted Radiology and Surgery **15**(8), 1347–1358 (2020)