Registration Teaches Registration: Transform-Derived Crown Guidance for Semi-Supervised CBCT–IOS Alignment

Yi Zhu[1][0009−0001−1159−6853], Razmig Kéchichian[1][0000−0001−7974−8705], Raphaël Richert[2][0000−0002−9298−1293], and Sébastien Valette[1][0000−0001−7549−4808]

[1] CREATIS, CNRS UMR 5220, Inserm U1294, INSA Lyon, Université Claude Bernard Lyon 1, Villeurbanne, France
[2] Hospices Civils de Lyon, PAM Odontologie, Lyon, France
yi.zhu@creatis.insa-lyon.fr

Abstract. Accurate registration of cone-beam computed tomography (CBCT) and intraoral scans (IOS) is difficult because CBCT contains roots, bone, and metal artifacts, whereas IOS observes only crown and gingival surfaces. STSR 2026 Task 2 makes this problem semi-supervised, providing transforms for 30 cases alongside 300 unlabeled cases. We present Registration Teaches Registration (RTR). Instead of treating each labeled transform only as an output to predict, RTR uses it as a teacher: the transform places an IOS crown in CBCT and thereby reveals where that crown is supported. These aligned surfaces form weak upper/lower targets for a semi-supervised 3D U-Net. At inference, the predicted support narrows the search over multiple IOS crown hypotheses; parity-aware multiscale ICP generates valid candidate transforms, and learned geometric ranking with jaw consistency selects the final pair. Learning identifies where the modalities overlap, geometry determines how they align, and ranking resolves which solution to trust. Fusing supervised and self-trained support ensembles reduces grouped target Chamfer distance from 1.118 to 1.082 mm. RTR achieves 5.7848 mm mean translation error and 2.8637° mean rotation error on 50 validation cases and ranks first on the official hidden test.

Keywords: Dental registration · CBCT · Intraoral Scanner · Weak supervision · Self-training

1 Introduction

CBCT and IOS describe complementary dental anatomy. CBCT reveals roots and surrounding bone, but crown boundaries can be degraded by partial volume effects and metal artifacts. IOS resolves crown surfaces accurately, yet contains neither roots nor a direct CBCT coordinate frame. Registering both modalities is therefore a prerequisite for implant planning, orthodontic analysis, and high-fidelity crown–root fusion [18]. This is not conventional full-overlap registration: IOS is a partial surface, CBCT contains many non-dental structures, and tooth arrangements can be geometrically symmetric.

The STSR challenge formalizes the problem as semi-supervised rigid registration [17,16]. Each case contains a CBCT volume and paired upper/lower IOS meshes; the output is one physical-space 4×4 transform per jaw. The 2026 training set contains 30 cases with transforms and 300 cases without transforms, but no tooth-segmentation labels. We additionally identified an acquisition/export protocol whose provided orthogonal transforms have $\det(R) = -1$. We call the determinant sign the transform *parity*: parity $+1$ is an ordinary rotation, whereas parity -1 includes a reflection introduced by the coordinate convention. A conventional SVD head that projects every estimate to $SO(3)$ cannot express these parity--1 labels.

Related work. Classical CBCT–IOS pipelines segment dental structures and then use RANSAC or centroid alignment followed by ICP [8,7]. ICP is effective when initialization and target extraction are reliable [2,14], but partial overlap and repeated tooth geometry produce local minima. The winning STSR 2025 solution combined PointNetLK and ICP [4,1]; the runner-up used dual PointNet++ encoders, feature matching, and an SVD head [6,13]. Three-dimensional U-Nets are standard volumetric predictors [3], while ToothSeg demonstrates the value of explicit tooth segmentation [12]. Task 2 does not provide such labels, and RTR uses no external segmentation model. Pseudo-labeling extends sparse supervision [9], and ExtraTrees can rank heterogeneous geometric measurements without feature normalization [5].

Motivation and contributions. A global HU threshold retains skull and cortical bone, while direct SVD regression can remove reflections. We instead turn Task 2 transforms into cross-modal weak supervision: aligned IOS surfaces teach a compact CBCT crown-support predictor that guides, rather than replaces, geometry. RTR (i) learns upper/lower support from transformed IOS surfaces and confidence-filtered self-training, (ii) preserves NIfTI physical coordinates and both transform parities during coarse-to-fine search, and (iii) ranks physically valid candidates before joint upper/lower selection. Fixed correspondence banks use only organizer-provided Task 2 training data. Grouped support evaluation, public validation, and the official hidden test evaluate the submitted system; the scope of each result is stated explicitly in Sec. 4.

2 Method

2.1 Registration as Support Discovery

Let $X_j = \{x_i\}_{i=1}^{N_j}$ be the IOS point set of jaw $j \in \{u, l\}$ and let I be the CBCT volume with affine A. The task is to estimate

$$T_j = \begin{bmatrix} R_j & t_j \\ 0 & 1 \end{bmatrix}, \qquad R_j^\top R_j = I_3, \quad \det(R_j) \in \{-1, +1\}, \tag{1}$$

which maps IOS points into CBCT physical space.

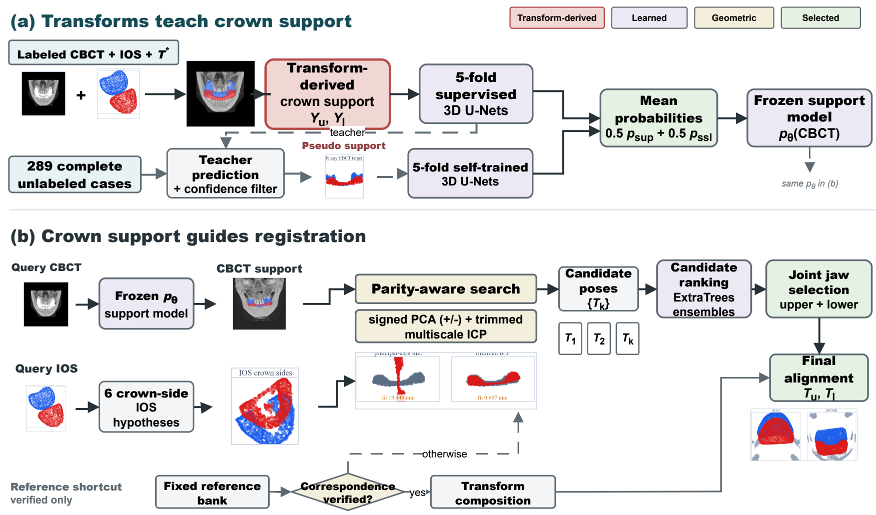

Fig. 1. Overview of RTR. During training, supplied transforms convert IOS crown-side surfaces into weak upper/lower CBCT support; supervised teachers and confidence-filtered self-training form the fused support model. At inference, verified reference correspondence is used when available; otherwise, predicted support and six IOS hypotheses drive parity-aware PCA/ICP, candidate ranking, and joint jaw selection. Blue/red denote upper/lower jaws.

A labeled transform is more than a six-degree-of-freedom target: after applying it to an IOS crown surface, it becomes a spatial annotation of where that surface is supported in CBCT. RTR first learns *where* an IOS crown can plausibly align, then asks geometry to generate rigid explanations of that support and learns only *which* explanation to trust. This converts 30 labeled transforms into dense voxel supervision while preserving rigid geometry. Figure 1 summarizes both the learned geometric route and the verified-reference shortcut described in Sec. 2.5.

2.2 Transforms as Weak Supervision

We obtain an affine-preserving dental region from multithreshold high-HU connected components, expand it by 20 mm, and resample it to a 128^3 grid at 1.25-mm isotropic spacing. For each labeled jaw, the lower and upper 35% tails of the smallest-variance IOS principal axis define two possible crown-side surfaces. We map both with the supplied transform T_j^* and retain the side with stronger local CBCT density. If $C_j = T_j^* X_j^{crown}$ is the selected surface in CBCT coordinates and A_g is the grid affine, the weak support is

$$Y_j(v) = \mathbb{1}[d(A_g v, C_j) \leq 0.7 \text{ mm}]\,. \tag{2}$$

Overlapping upper/lower supports are assigned to the nearer surface. A five-level 3D U-Net maps the normalized CBCT region to background, upper-support, and lower-support probabilities. Encoder widths are 24, 48, 96, 192, and 384; each level uses two 3^3 convolutions, GroupNorm, and SiLU, with max-pooling and transposed-convolution decoding. Each member has 12.7 million parameters. With class weights $(0.10, 1, 1)$, the loss is weighted cross-entropy plus foreground Dice,

$$\mathcal{L}_{support} = \mathcal{L}_{CE}^{w} + 1 - \frac{1}{2} \sum_{c \in \{u,l\}} \frac{2 \sum_v p_c(v) y_c(v) + 1}{\sum_v p_c(v) + \sum_v y_c(v) + 1}. \tag{3}$$

Five CBCT-identity-grouped folds provide supervised teachers. Each teacher predicts the 289 complete unlabeled cases; pseudo targets are retained when each jaw contains 500–6000 voxels, foreground confidence is at least 0.45, entropy is at most 1.0, and no identical-CBCT group is repeated. The top 80 pseudo targets per teacher train five additional members at effective weight 0.2. The supervised and self-training ensemble means are fused as

$$p_{fused} = 0.5 p_{sup} + 0.5 p_{ssl}. \tag{4}$$

Thresholding at 0.25 and removing components smaller than four voxels yields the two CBCT targets.

2.3 Support-Guided Hypothesis Search

Crown-side extraction is uncertain, so RTR does not commit to one cut: low/high tails of the smallest-variance IOS principal axis at 25%, 35%, and 45% form six source hypotheses per jaw. Each is matched to the corresponding predicted CBCT support. The observed 266×266×200 volumes at 0.3-mm spacing use parity −1; the remaining protocols use parity +1. We enumerate signed principal-axis frames with the required parity and add a jaw-specific mean-rotation initialization fitted on labeled cases.

Trimmed ICP refines each initialization at correspondence radii of 8, 5, 3, 2, and 1.25 mm while retaining the best 70% of matches. We keep up to eight source- and target-diverse basins, then explore $(12°, 8$ mm$)$, $(6°, 4$ mm$)$, and $(3°, 2$ mm$)$ perturbation neighborhoods. The resulting candidates express uncertainty in crown-side choice, parity, and initialization.

2.4 Learning to Select, Not to Regress

Each valid candidate is represented by 97 measurements of bidirectional distance, trimmed fit, millimetric coverage, normal agreement, full-arch plausibility, target geometry, parity, source hypothesis, and within-case rank. Seven ExtraTrees regressors estimate candidate error, while seven pairwise ExtraTrees classifiers compare candidates. Their fractional ranks are fused as

$$b_k = \alpha \, \mathrm{rank}(c_k^{reg}) + (1-\alpha) \, \mathrm{rank}(c_k^{pair}), \qquad \alpha = 0.575. \tag{5}$$

At most 30 candidates are generated per jaw and the best 20 enter the rankers. The top eight per jaw enter joint selection. If $\Delta\theta$ and Δt denote deviation from the mean labeled upper-to-lower rotation and translation, the selected pair minimizes

$$J(u,l) = b_u + b_l + 0.01\,\Delta\theta(u,l) + 0.075\,\Delta t(u,l). \tag{6}$$

Pairs with inconsistent parity are excluded. This weak coupling resolves mirrored or swapped arch solutions without forcing both jaws to share a transform.

2.5 Evidence First, Search Second

Some supplied cases repeat a CBCT acquisition with an identical or remeshed IOS. RTR therefore checks a fixed training-derived reference bank before global search. The banks begin with 60 labeled-jaw references and are expanded, before validation/test inference, with quality-gated pseudo-transforms obtained only from organizer-provided unlabeled Task 2 cases. Their final paired-vertex and surface banks contain 94 and 97 entries, respectively; no validation/test ground truth or external data enter them.

Preserved IOS vertex ordering permits transform composition only when fitted correspondence has RMS ≤ 0.02 mm and P95 ≤ 0.05 mm. For a remeshed IOS on the same CBCT, reuse additionally requires predicted teacher TRE $\leq$ 1.5 mm, bidirectional score ≤ 0.8 mm, median ≤ 1 mm, P90 ≤ 2 mm, and at least 95% coverage within 2 mm in both directions. These gates depend only on input geometry and frozen training-derived assets. If they fail, the learned geometric route runs. A predicted support with fewer than 32 voxels triggers a multithreshold target; implausible full-arch fit triggers one jaw-specific ROI retry.

3 Experiments

The data comprise 30 labeled cases (60 jaws), 300 unlabeled cases (289 with complete upper/lower IOS), and 50 public-validation cases (100 jaws). Sixteen labeled jaws use parity -1. Development folds are grouped by CBCT identity. The official metrics are mean translation error e_t in millimeters and mean rotation error e_R in degrees [15]:

$$e_t = \|\hat{t} - t^*\|_2, \qquad e_R = \frac{180}{\pi}\arccos\left(\frac{\mathrm{tr}(\hat{R}R^{*\top}) - 1}{2}\right), \tag{7}$$

with equal transform parity enforced and the arccosine argument clipped to $[-1, 1]$. Support quality is measured against held-out transform-derived targets using symmetric Chamfer distance in physical millimeters and foreground Dice.

Preprocessing preserves the NIfTI affine, extracts the dental ROI, resamples it to 128^3 at 1.25-mm spacing, and samples each IOS into six hypotheses. Training uses random flips along the first two grid axes, in-plane 90° rotations,

Table 1. Development environments and requirements.

System	Windows 11, 64-bit
CPU	Intel Core i5-12600K, 10 cores/16 threads
RAM	4×8 GB DDR4, 2667 MT/s
GPU	NVIDIA GeForce RTX 5070 Ti, 16 GB
Environment	CUDA 12.8, Python 3.10, PyTorch 2.8.0

Table 2. Held-out agreement with transform-derived crown support. Lower Chamfer and higher Dice are better.

Target model	Symmetric Chamfer (mm)	Dice
Supervised ensemble (five grouped folds)	1.1181	0.5238
Self-training ensemble (five grouped folds)	1.0909	0.5321
Probability fusion (0.5/0.5)	**1.0817**	**0.5339**

intensity scaling in $[0.9, 1.1]$, offsets in $[-0.1, 0.1]$, and Gaussian noise with standard deviation 0.025. The ten U-Nets are trained from scratch with batch size 1 and AdamW (learning rate 3×10^{-4}, weight decay 10^{-4}); the rankers use 4,929 candidates from the 60 labeled jaws. Development used the system in Table 1; the submitted Linux/AMD64 container uses Python 3.11, PyTorch 2.5.1, CUDA 12.1, and an 8-GB host-memory limit. U-Nets run sequentially. Correspondence routes take tens of seconds, whereas difficult randomized searches may require several minutes. Source code and reproduction instructions are available at `https://github.com/avalanchezy/RegistrationTeachesRegistration`.

4 Results and Discussion

4.1 Learning the Registration Target

Table 2 evaluates the support branch under CBCT-identity-grouped validation.

Self-training lowers Chamfer by 0.0272 mm and raises Dice by 0.0083; fusion contributes a further 0.0092 mm and 0.0018, respectively. Thus Table 2 directly supports improved localization of the learned registration target. It does *not* establish how much this modest support gain changes final registration error. Moderate Dice is expected because the target is a thin crown-surface neighborhood rather than a full tooth mask.

4.2 End-to-End Registration and Scope of Evidence

The frozen container obtains 5.7848 mm mean translation error and 2.8637° mean rotation error on public validation, produces valid matrices for all 100 jaws, and ranks first on the official hidden test. These results support the complete deployed cascade, but do not isolate the causal contribution of each component.

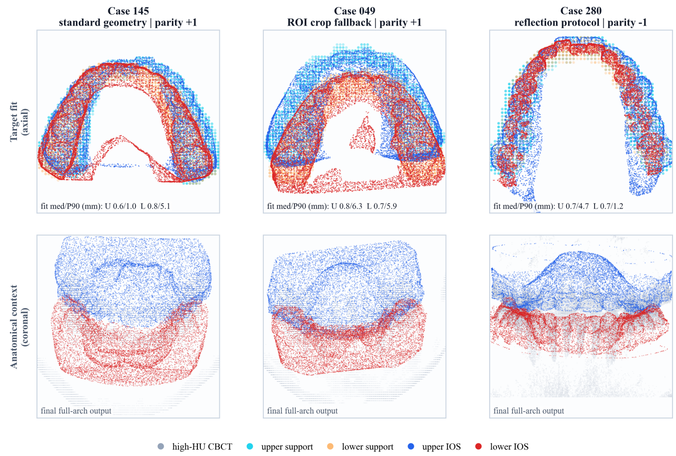

Fig. 2. Crown-guided public-validation outputs in axial and coronal projection. Gray: high-HU CBCT context; blue/red: transformed upper/lower IOS. The cases were selected before obtaining the public server score.

The expected roles are nevertheless distinct: support prediction suppresses non-dental CBCT structures; parity-aware PCA/ICP ensures candidate coverage in both coordinate conventions; tree ranking resolves local geometric ambiguity; and jaw coupling discourages anatomically inconsistent upper/lower pairs. Of these claims, Table 2 directly tests support prediction, while the server scores evaluate only their combination. We therefore make no component-level end-to-end accuracy claim without a matched registration ablation.

On public validation, the frozen correspondence policy accepts 26 preserved-vertex jaws and 10 same-CBCT surface jaws; the remaining 64 jaws use the general learned geometric route, including two ROI retries. Correspondence routing is a transductive use of organizer-provided Task 2 training and unlabeled data: banks and gates were fixed before evaluation, and acceptance uses only the query CBCT/IOS geometry. It was part of the submitted Docker and therefore of the official evaluation. Because public reference transforms are hidden, route-specific translation/rotation errors and the routes' contribution to the aggregate score cannot be measured; the route counts should not be interpreted as an ablation.

Three general-route registrations are shown in Fig. 2. The overlays inspect anatomical overlap and failure handling; public transforms are unavailable for case-level accuracy.

4.3 Limitations and Future Work

Weak targets inherit errors from principal-axis side selection and supplied transforms. Confidence filtering uses only the most reliable subset of the 289 complete unlabeled cases. Correspondence routing is useful for repeated challenge acquisitions but may rarely apply in routine clinical data, where the general route carries the workload. Parity--1 protocols and symmetric arches enlarge the search space, and difficult searches can require minutes. A parity-conditioned canonical frame, uncertainty-aware pseudo-label use, and a distilled coarse-pose network could reduce ambiguity and runtime.

5 Conclusion

RTR reuses registration labels to learn where CBCT and IOS overlap, then combines that support with parity-aware geometry and learned candidate selection. The frozen system produces all required outputs and ranks first on the hidden STSR 2026 test. The results directly demonstrate improved support prediction and strong performance of the complete cascade; component-level registration effects remain an important target for future controlled evaluation.

Acknowledgements. This work was developed and evaluated within the ODIN 2026 challenge cluster, including the STS, ToothFairy, and Bite2Text challenges. We acknowledge the STS challenge series [17,16] and the related ODIN contributions [10,11]. We thank the organizers and data owners for providing the challenge data and evaluation platform.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aoki, Y., Goforth, H., Srivatsan, R.A., Lucey, S.: PointNetLK: Robust & efficient point cloud registration using PointNet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7163–7172 (2019). https://doi.org/10.1109/CVPR.2019.00733 2
2. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2), 239–256 (1992). https://doi.org/10.1109/34.121791 2
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science, vol. 9901, pp. 424–432. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49 2
4. George, A.B., Gadha Lekshmi, P., Bathini, S., Govind, A.: Semi-supervised CBCT–IOS registration using PointNetLK. In: Oral and Dental Image Analysis. Lecture Notes in Computer Science, vol. 16473, pp. 211–220. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-20711-1_20 2
5. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Machine Learning **63**(1), 3–42 (2006). https://doi.org/10.1007/s10994-006-6226-1 2

6. Ji, C., Liu, Y., Jiang, Y., Wang, L.: Learning-based CBCT–IOS registration with PointNet++ and SVD. In: Oral and Dental Image Analysis. Lecture Notes in Computer Science, vol. 16473, pp. 181–190. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-20711-1_17 2
7. Kim, S., Choi, Y., Na, J., Song, I.S., Lee, Y.S., Hwang, B.Y., Lim, H.K., Baek, S.J.: Best of both modalities: Fusing CBCT and intraoral scan data into a single tooth image. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Lecture Notes in Computer Science, vol. 15002, pp. 553–563. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72069-7_52 2
8. Kim, Y.J., Ahn, J.H., Lim, H.K., Nguyen, T.P., Jha, N., Kim, A., Yoon, J.: Novel procedure for automatic registration between cone-beam computed tomography and intraoral scan data supported with 3D segmentation. Bioengineering **10**(11), 1326 (2023). https://doi.org/10.3390/bioengineering10111326 2
9. Lee, D.H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: ICML Workshop on Challenges in Representation Learning (2013) 2
10. Lumetti, L., Di Bartolomeo, M., Pellacani, A., Anesi, A., Grana, C., Bolelli, F.: Ontology-grounded structured prediction for dental CBCT reporting. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2026 (2026), in press 8
11. Lumetti, L., Rizzo, F., Cremonini, F., Candeloro, E., Luca, L., Grana, C., Bolelli, F.: Do multimodal LLMs understand intraoral dental data? dataset, platform, and baselines. In: Computer Vision – ECCV 2026 (2026), in press 8
12. van Nistelrooij, N., Krämer, L., Kempers, S., Beyer, M., Bolelli, F., Xi, T., Bergé, S., Heiland, M., Maier-Hein, K.H., Vinayahalingam, S., Isensee, F.: ToothSeg: Robust tooth instance segmentation and numbering in CBCT using deep learning and self-correction. IEEE Journal of Biomedical and Health Informatics **30**(8), 6801–6812 (2026). https://doi.org/10.1109/JBHI.2025.3650444 2
13. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems. vol. 30, pp. 5099–5108 (2017) 2
14. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling. pp. 145–152. IEEE Computer Society (2001). https://doi.org/10.1109/IM.2001.924423 2
15. STSR 2026 Organizing Committee: MICCAI STSR 2026 Task 2: Semi-supervised teeth crown and root registration in 3D CBCT and IOS scans. Codabench Competition 16042 (2026), https://www.codabench.org/competitions/16042/, accessed 24 August 2026 5
16. Wang, Y., Li, Z., Wu, C., Liu, J., Zhang, Y., Ni, J., Luo, Q., Chen, J., Zhang, H., Liu, J., Han, C., Fu, K., Ji, C., Cai, X., Hao, J., Zheng, Z., Xu, S., Chen, J., Yu, X., Zhang, Q., Qian, D., Wang, S., Zhou, H.: MICCAI STS 2024 Challenge: Semi-supervised instance-level tooth segmentation in panoramic X-Ray and CBCT images. Medical Image Analysis **110**, 103986 (2026). https://doi.org/10.1016/j.media.2026.103986 2, 8
17. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Dan, R., Zhang, Q., Huang, X., Huang, Z., Liu, J., Zheng, Z., Wu, C., Li, Y., Li, Z., Ma, Z., Cui, W., Luo, S., Wang, C., Chen, Y., Li, T., Liu, Y., Feng, X., Ni, J., Liu, D., Wang, Q., Lin, Z., Song, W., Li, Y., Wang, B., Wang, C., Chen, Q.,

Li, M., Zhou, H., Jin, Q.: MICCAI 2023 STS Challenge: A retrospective study of semi-supervised approaches for teeth segmentation. Pattern Recognition **170**, 112049 (2026). https://doi.org/10.1016/j.patcog.2025.112049 2, 8

18. Zhu, Y., Kéchichian, R., Richert, R., Ikehata, S., Valette, S.: High-fidelity 3D tooth reconstruction by fusing intraoral scans and CBCT data via deep implicit representations. In: 2026 IEEE 23rd International Symposium on Biomedical Imaging (ISBI). pp. 1–4. IEEE (2026). https://doi.org/10.1109/ISBI61048.2026.11515997 1