ROOTFLOW: 3D TOOTH ROOT SYNTHESIS FROM INTRAORAL SCANS AND PANORAMIC RADIOGRAPHS VIA LATENT FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Three-dimensional tooth surface morphology and spatial arrangement are important for dental diagnosis and treatment planning. Intraoral scans (IOS) capture detailed crown surfaces but cannot reveal tooth roots, whose three-dimensional morphology and spatial arrangement typically require cone-beam computed tomography (CBCT) to assess, at the cost of additional radiation exposure. Panoramic X-ray (PX) images provide root-related morphological and spatial cues at a lower radiation dose, but their inherent depth ambiguity and geometric distortion limit accurate three-dimensional root reconstruction. We propose ROOTFlow (Root-Oriented Oral Tomography via Latent Flow Matching), a framework that combines IOS-derived crown geometry with radiographic context to generate quadrant-level CBCT volumes and corresponding tooth segmentations, focusing on root surface morphology and spatial arrangement. The framework comprises four components: (i) quadrant-level data construction, in which unsigned distance fields (UDFs) encode IOS meshes as volumetric geometric priors; (ii) latent representation learning with two 3D variational autoencoders that compress CBCT and UDF inputs; (iii) cross-modal contrastive alignment that maps CBCT and PX features into a shared representation space; and (iv) conditional 3D latent flow matching, in which a velocity network conditioned on UDF latent features and PX features guides flow integration to generate a CBCT latent representation. The generated latent representation is jointly decoded into a quadrant-level CBCT volume and corresponding tooth segmentation. PX features provide complementary morphological and contextual cues rather than metrically reliable three-dimensional measurements. We construct a quadrant-level dataset comprising matched CBCT–IOS–PX trios derived from 5,682 patients. On a common cohort of 161 evaluation cases, ROOTFlow achieves the highest whole-tooth Dice and lowest average surface distance among the evaluated configurations, while improving volume fidelity over the real-PX volumetric baselines.

1 INTRODUCTION

Intraoral scans (IOS) capture the visible crowns and their spatial arrangement, but leave the underlying root anatomy unobserved. Cone-beam computed tomography (CBCT) provides 3D access to these hidden structures, but typically entails a higher ionizing radiation dose than standard 2D dental radiographs, with particular caution advised for pediatric and adolescent patients (U.S. Food and Drug Administration, n.d.). Panoramic X-ray (PX) images provide two-dimensional cues to root morphology, complementing the crown geometry captured by IOS. We therefore seek to reconstruct three-dimensional tooth-root geometry from IOS-derived crown surfaces and panoramic radiographs.

This recovery problem is underdetermined: multiple distinct root geometries can be consistent with the same crown and panoramic observations. Existing methods combine crown observations with templates or learned implicit representations (Barone et al., 2015; Yang et al., 2022), or reconstruct teeth and volumes from panoramic inputs (Song et al., 2021; Ma et al., 2024; Park et al., 2024). Paired clinical CBCT and PX observations enable learning cross-modal features that link radio-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

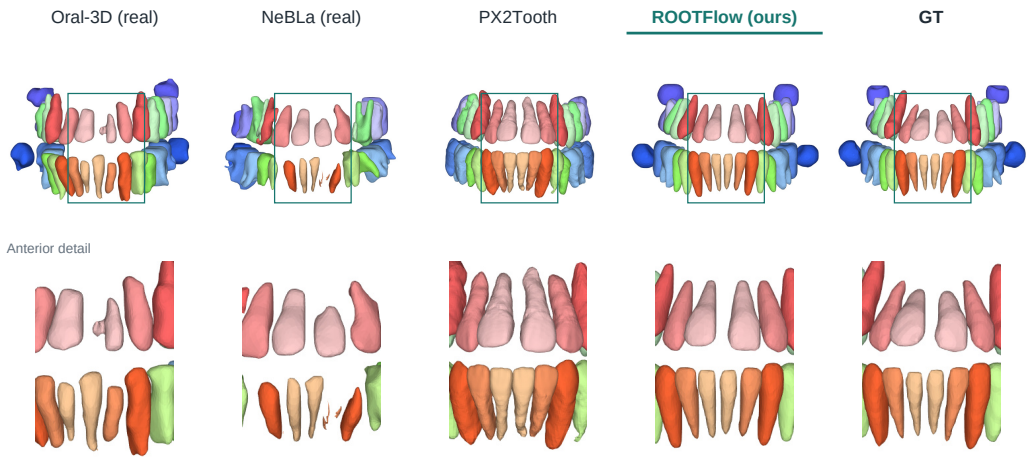
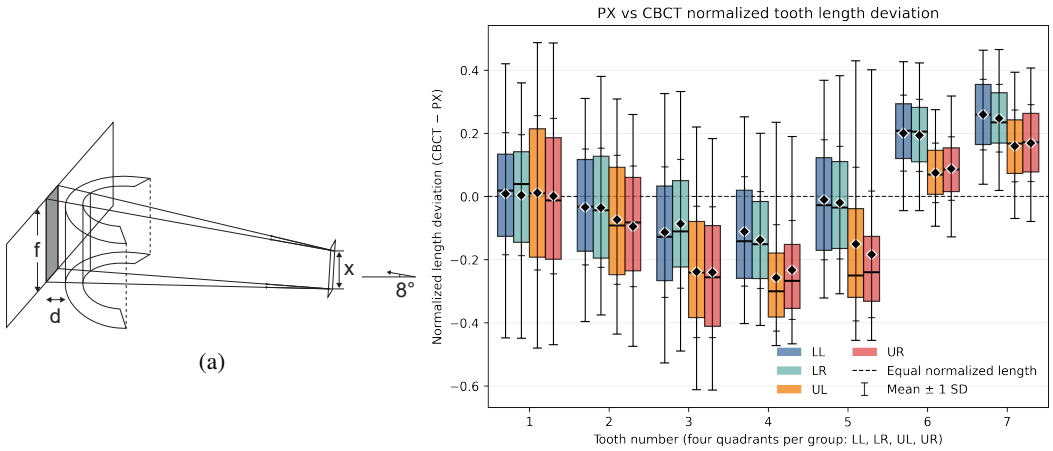


Figure 1: Tooth-surface reconstruction for Case A. The upper row compares real-PX baseline configurations, ROOTFlow, and the reference (GT); the lower row enlarges the anterior regions outlined above. ROOTFlow more closely preserves the reference anterior tooth shapes and arrangement in this example. All detail panels use the same image-coordinate crop.



(a) (b)

Figure 2: 2a is reprinted from (Whaites, 2002) illustrating the height magnification from panoramic imaging focal trough's angulation. 2b are tooth-wise normalized tooth length differences between PX and CBCT using data from our evaluation cases. The distribution reflects the non-uniform magnification of panoramic imaging, with anterior/mesial teeth experiencing greater length magnification than posterior molars. Incisors show smaller net deviations because their inclination shortens their projected height in PX, partially compensating for the magnification effect.

graphic appearance to three-dimensional tooth morphology. These features can serve as morphological priors for inferring unobserved roots, while IOS provides patient-specific crown geometry to guide the reconstruction. However, panoramic magnification depends on object position and imaging geometry (Devlin & Yuan, 2013; Yeom et al., 2018), while patient positioning can affect apparent tooth lengths and angulations (Stramotas et al., 2002), as illustrated in Figure 2. These unavoidable distortions compound the depth ambiguity inherent in two-dimensional projections. We thus hypothesize that directly using measurements taken from PX images as explicit root reconstruction guidance can lead to tooth size deviations.

We therefore propose ROOTFlow (Root-Oriented Oral Tomography via Latent Flow Matching), a framework for synthesizing quadrant-level CBCT volumes and corresponding tooth segmentations

from IOS and PX, with a focus on root surface morphology and spatial arrangement. IOS-derived unsigned distance fields (UDFs) provide the geometric starting point in latent space, while cross-modal contrastive alignment learns morphological priors from paired clinical CBCT–PX observations. A conditional velocity network transports UDF latents toward CBCT latents through flow integration, guided by crown geometry and features extracted from the patient's PX image. Figure 1 illustrates the resulting tooth geometry alongside baseline reconstructions and the reference.

Our contributions are: (1) IOS-anchored quadrant-level synthesis. We formulate tooth-root reconstruction as joint CBCT and tooth-segmentation synthesis, using IOS-derived crown geometry as both the latent transport source and a persistent condition. (2) Cross-modal morphological priors for latent flow matching. We combine contrastive CBCT–PX alignment with conditional UDF-to-CBCT latent transport, using patient-specific PX features to guide reconstruction of unobserved tooth-root structure. (3) A large paired clinical cohort and evaluation. We curate matched IOS–PX–CBCT observations from 5,682 patients and evaluate volume fidelity and whole-tooth geometry, including anterior orientation, posterior morphological peaks, and position-wise comparisons with explicit coverage analysis.

2 Related Work

Tooth-root completion and panoramic reconstruction. Parametric methods construct tooth models from panoramic measurements (Mazzotta et al., 2013), while template-based approaches combine optically scanned crowns with root templates adapted to panoramic observations (Barone et al., 2015). ToothInpaintor learns an implicit complete-tooth representation and recovers roots from partial dental models and panoramic images (Yang et al., 2022); its input setting is the closest to ours. Panoramic reconstruction methods use different geometric priors and outputs: Oral-3D uses a dental-arch prior (Song et al., 2021); X2Teeth separates localization and shape estimation (Liang et al., 2021); Occudent learns implicit tooth representations from actual radiographs (Park et al., 2023); and PX2Tooth generates tooth point clouds with a prior-fusion module (Ma et al., 2024). ROOTFlow combines IOS crown geometry and panoramic observations to jointly synthesize CBCT volumes and per-tooth segmentations within each dental quadrant, with a focus on tooth-root reconstruction.

Volumetric reconstruction and acquisition modeling. DTR-Net combines intensity reconstruction, segmentation, and implicit tooth modeling (Mei et al., 2024). NeBLa uses acquisition-aware simulation and real-to-synthetic image translation without requiring matched clinical PX–CBCT training pairs or patient-specific arch priors (Park et al., 2024). Other approaches include progressive reconstruction in 3DPX (Li et al., 2024) and acquisition-guided Gaussian representations in X-Splat (Szczepański et al., 2026). Patient-tailored panoramic synthesis further addresses discrepancies between simulated and clinical acquisition (Sunilkumar et al., 2025). ROOTFlow learns cross-modal morphological priors directly from paired clinical CBCT–PX observations to guide IOS-conditioned reconstruction; it does not rely on acquisition simulation as the primary link between modalities.

Latent generation and cross-modal learning. Variational autoencoders learn compact representations (Kingma & Welling, 2014), and latent diffusion demonstrates their utility for iterative generation (Rombach et al., 2022). Flow matching and rectified flow learn velocity fields for transport along specified paths (Lipman et al., 2023; Liu et al., 2023). CLIP and SigLIP provide precedents for learning representations from paired modalities (Radford et al., 2021; Zhai et al., 2023); our PX branch uses a SigLIP 2 visual backbone (Tschannen et al., 2025). ROOTFlow combines these directions by using aligned PX features to condition flow matching from IOS-derived UDF latents to CBCT latents, using a conditioned 3D UNet model (von Platen et al., 2022).

3 Method

3.1 Problem Formulation and Quadrant Representation

Given IOS crown meshes M and a same-patient quadrant-level radiograph image P, ROOTFlow generates a CBCT-like intensity volume $\hat{V}$ and a per-tooth segmentation $\hat{Y}$. We represent each dental quadrant on a $96 \times 128 \times 96$ grid with 0.5 mm isotropic spacing, with consistent flips to

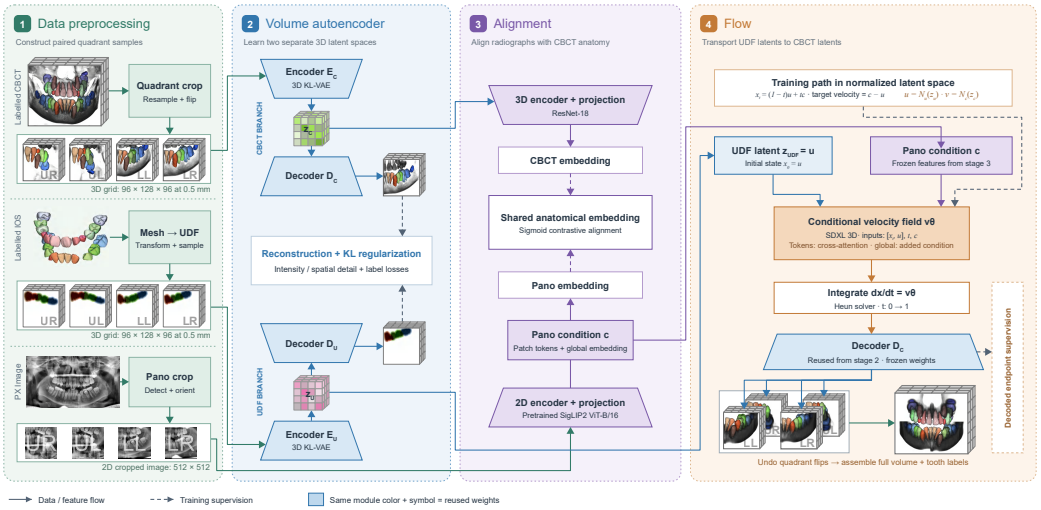


Figure 3: Overview of ROOTFlow. Quadrant-level preprocessing constructs CBCT targets, IOS-derived UDF inputs, and panoramic crops. Two volumetric autoencoders learn latent representations, and CBCT–PX alignment supplies radiographic conditions. A conditional velocity network transports UDF latents toward CBCT latents through flow integration, followed by joint decoding of a quadrant-level CBCT volume and segmentation. Real CBCT provides training supervision; generation uses IOS-derived latents and panoramic features. Solid arrows indicate data or feature flow, and dashed arrows indicate training supervision.

standardize left/right and upper/lower orientations. The target stack X_c contains one normalized intensity channel and 13 tooth-label channels, giving 14 input channels in total.

The IOS stack X_u combines one truncated unsigned surface-distance channel with eight tooth-identity channels, giving nine input channels in total. For crown meshes M_j in voxel coordinates, we use

$$d(\mathbf{x}) = \min_j \inf_{\mathbf{y} \in M_j} \|\mathbf{x} - \mathbf{y}\|_2, \qquad U(\mathbf{x}) = 1 - \frac{\min(d(\mathbf{x}), \tau)}{\tau}, \tag{1}$$

with $\tau = 5$ voxels (2.5 mm). This representation encodes observed crown surfaces with per-tooth segmentation. Storage quantization, label conventions, and coordinate preparation are detailed in Appendix A.

The pipeline comprises volumetric representation learning, CBCT–PX alignment, and conditional latent transport. At inference, only the IOS encoder, PX embedding branch, velocity network, and CBCT decoder are evaluated. Coordinate preparation remains a separate requirement, discussed in Appendix A. Figure 3 summarizes the four components and their training and generation paths.

3.2 Latent Representations and Cross-Modal Conditioning

Two independent 3D variational autoencoders learn representations of X_u and X_c. Their posterior-mean encoders E_u, E_c yield latents with shape $12 \times 24 \times 32 \times 24$. Both decoders are trained to reconstruct their respective continuous and categorical channels. Downstream stages use cached posterior means with fixed representation models; the UDF decoder is unnecessary for CBCT generation. Modality-specific channel statistics estimated from training data define normalized latents $u = N_u(E_u(X_u))$ and $c = N_c(E_c(X_c))$. Normalization controls channel scales but does not assume alignment of the two latent spaces.

For radiographic conditioning, a 3D encoder processes CBCT latents and a SigLIP 2 image encoder processes PX crops. Projection heads map them to a shared embedding space. A symmetric pairwise sigmoid objective aligns matched observations, supplemented by binary supervision of 13 tooth-presence indicators from CBCT labels. Duplicate matches share positive weight, and unmatched observations from the same patient are excluded from negatives. Appendix B gives the objective and architecture settings.

The PX branch supplies a global feature through image-conditioned embeddings and patch tokens through cross-attention in the velocity network. These features provide morphological and contextual evidence, not metric 3D measurements. The shared embedding projection and patch-token representation serve different conditioning roles; neither establishes local geometric correspondence.

3.3 Paired Conditional Latent Flow Matching

For each matched sample, the normalized UDF latent u is the transport source and the normalized CBCT latent c is its target. Following straight-path flow matching (Lipman et al., 2023; Liu et al., 2023), we define

$$x_t = (1-t)u + tc, \qquad v^* = c - u, \qquad t \sim 0.25\delta_0 + 0.75\mathcal{U}(0,1). \tag{2}$$

The mass at zero explicitly exposes training to source states used at inference. A 3D U-Net predicts velocity from the channelwise concatenation of x_t and u, a time embedding, and PX features f_p. Its flow-matching objective is

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{u,c,f_p,t}\left[\mathrm{MSE}\left(v_\theta(x_t,t;u,f_p), c-u\right)\right]. \tag{3}$$

Thus, real CBCT latents define the training path and target velocity, but are not inference inputs. Unlike noise-initialized generation, the source is an encoded observation; UDF geometry also remains an explicit condition throughout integration.

3.4 Training Objectives and Inference

Decoded endpoint supervision. We supplement velocity regression with supervision of a one-step endpoint estimate:

$$\tilde{c} = x_t + (1-t)v_\theta(x_t,t;u,f_p), \qquad (a,L) = D_c\big(N_c^{-1}(\tilde{c})\big). \tag{4}$$

Here $\sigma(a)$ is the predicted intensity and L contains categorical logits. The CBCT decoder is frozen, while gradients through its input train the velocity network. The total objective is

$$\begin{aligned} \mathcal{L}_{\mathrm{flow}} &= \mathcal{L}_{\mathrm{FM}} + 0.1\mathcal{L}_{\mathrm{end}}, \\ \mathcal{L}_{\mathrm{end}} &= \mathcal{L}_{\mathrm{region}} + \mathcal{L}_{\mathrm{grad}} + \mathcal{L}_{\mathrm{CE}} + \mathcal{L}_{\mathrm{Dice}} + \mathcal{L}_{\mathrm{absent}}. \end{aligned} \tag{5}$$

Endpoint loss definitions. The region term $\mathcal{L}_{\mathrm{region}}$ computes mean absolute intensity error over separately normalized tooth and background regions. The gradient term $\mathcal{L}_{\mathrm{grad}}$ sums mean absolute differences between predicted and target signed first-order finite differences along the three spatial axes over the full crop. It supervises local intensity variation.

The label-classification term $\mathcal{L}_{\mathrm{CE}}$ uses classification loss for 13 tooth-label channels. The soft Dice term $\mathcal{L}_{\mathrm{Dice}}$ is averaged over tooth classes present in the target. For each absent tooth slot, $\mathcal{L}_{\mathrm{absent}}$ averages the 64 largest predicted voxel probabilities and then averages over absent slots. Absence is determined from the reference CBCT labels, not from the lack of an IOS crown, so a tooth unobserved by IOS is not automatically treated as absent. Endpoint terms have equal internal weights, with the outer coefficient of 0.1 in Equation 5.

These losses supervise volumetric intensity and tooth labels rather than directly minimizing root-surface distance. Gradients through the frozen decoder update the velocity network without changing decoder parameters. The endpoint estimate in Equation 4 is computed from a ground-truth-interpolated training state, not from source-only integration; validation therefore evaluates integrated predictions.

Flow integration and joint decoding. At inference, we solve

$$\frac{dx_t}{dt} = v_\theta(x_t,t;u,f_p), \quad x_0 = u, \qquad (\hat{a},\hat{L}) = D_c\big(N_c^{-1}(x_1)\big). \tag{6}$$

We use 16 Heun steps (32 velocity evaluations) with exponential-moving-average weights. A sigmoid and inverse intensity scaling yield the reconstructed volume, while the tooth-label outputs provide the corresponding per-tooth segmentation.

4 EXPERIMENTS

4.1 DATASET AND EXPERIMENTAL SETUP

Dataset. Our dataset comprises matched CBCT volumes, intraoral scans, and clinical panoramic radiographs from 5,682 patients. After preprocessing and cross-modal matching, each patient contributed four dental quadrants, yielding 22,728 quadrant-level samples. CBCT volumes provide volumetric reference data, while IOS and clinical PX provide the inputs for reconstruction. The cohort was randomly divided into training (4,420 patients), testing (1,101 patients), and evaluation (161 patients) sets. Of the total cohort, 2,229 patients have mixed dentition. We preprocessed all CBCT volumes by cropping them to contain only the dentition regions, and resampled them to 0.5 mm isotropic spacing. To obtain human-verified segmentations for both CBCT and panoramic (PX) images, we employed a manually assisted pipeline. Additionally, manual guidance was used to compute the registration transformations aligning the IOS mesh space with the CBCT volume space used for UDF generation.

For the 161 cases in the evaluation set, volume evaluation includes three methods with five configurations in total, while geometry evaluation additionally includes PX2Tooth, which does not generate intensity volumes.

Implementation. Autoencoding, alignment, and flow learning are trained separately. During flow learning, the autoencoders and PX representation are fixed, and the velocity network is optimized using Equation 5. The velocity U-Net has widths $(64, 128, 256)$ and uses an EMA decay of 0.999. Generation uses 16 Heun steps. Flow training uses the AdamW optimizer with a learning rate of 1×10^{-5} and a CosineAnnealingLR scheduler. The model is trained with a batch size of 8 for 200 epochs on four NVIDIA A6000 GPUs. Additional architecture and training details are provided in Appendix B.

Evaluation. We evaluate volume fidelity using PSNR, SSIM, and LPIPS, and tooth geometry using presence accuracy, Dice, average surface distance (ASD), whole-tooth PCA length mean absolute error (MAE), anterior PCA-axis angular difference, and posterior morphological-peak count MAE. Anterior and posterior subsets cover positions 1–3 and 4–8 in all four quadrants, respectively, excluding embedded teeth. Geometry is scored only when the predicted and reference tooth are both present. Tooth metrics are averaged within each case and then equally across cases. The results are reported as mean $\pm$ population standard deviation. These descriptors measure whole-tooth geometry and morphological peak counts. Appendix C specifies the metric definitions. For volumes from baseline methods Oral3D and NeBLa, as well as the ground truth volumes, we use a shared external automatic segmentation model to produce their tooth instances.

4.2 COMPARISON WITH EXISTING METHODS

We compare ROOTFlow with Oral-3D, NeBLa, and PX2Tooth (Song et al., 2021; Park et al., 2024; Ma et al., 2024). Oral-3D and NeBLa each include real-PX and synthetic-PX configurations. Synthetic radiographs are obtained through projection using CBCT-derived arch or acquisition trajectories described in respective papers. We therefore distinguish the real-PX comparisons from the synthetic-input references. PX2Tooth participates in geometry evaluation but has no intensity-volume metrics; its generated root presence is purely based on the crowns detected from IOS, hence its tooth presence accuracy is not reported. All baseline methods are retrained using the same dataset and data splits as ROOTFlow. Notably, the volume of data used for these baselines' retrainings is significantly larger than that reported in their original papers.

Tooth presence and geometry. Table 1 reports the geometry comparisons. ROOTFlow achieves the highest whole-tooth Dice (0.8353) and lowest ASD (0.4694 mm). Relative to PX2Tooth, Dice increases by 7.31 percentage points and ASD decreases by 29.69%, which indicating closer overlap and surface agreement. ROOTFlow also has the lowest PCA length MAE (1.3707 mm) among the evaluated configurations. Using each method's valid teeth, ROOTFlow has the highest mean Dice at 28 of the 32 ordinary tooth positions. To control tooth coverage, we also evaluate the same 3,235 tooth instances available in all six configurations across 161 cases. ROOTFlow retains better Dice (0.8746) and ASD (0.3480 mm) than PX2Tooth (0.7678 and 0.6448 mm). Angular results are

Table 1: Tooth geometry on the common cohort. Entries are mean ± population standard deviation; bold denotes the best mean across configurations. Geometry is conditional on both teeth being present. Panels separate overlap/surface metrics from shape descriptors for readability.

Method	Presence acc.↑	Dice↑	ASD (mm)↓
Oral-3D (real)	0.944 ± 0.060	0.305 ± 0.132	2.691 ± 1.272
NeBLa (real)	0.826 ± 0.131	0.120 ± 0.070	5.581 ± 3.350
PX2Tooth	—	0.762 ± 0.048	0.668 ± 0.159
ROOTFlow	$\mathbf{0.969 \pm 0.044}$	$\mathbf{0.835 \pm 0.050}$	$\mathbf{0.469 \pm 0.191}$
Oral-3D (syn)	0.969 ± 0.041	0.722 ± 0.054	0.772 ± 0.206
NeBLa (syn)	0.968 ± 0.043	0.654 ± 0.071	0.953 ± 0.246
Method	Length MAE (mm)↓	Anterior angle (°)↓	Posterior peaks MAE↓
Oral-3D (real)	2.396 ± 0.701	13.292 ± 5.700	0.449 ± 0.145
NeBLa (real)	3.456 ± 1.022	$14.586 \pm 6.714^{*}$	0.632 ± 0.184
PX2Tooth	1.929 ± 1.019	6.310 ± 1.885	0.577 ± 0.225
ROOTFlow	$\mathbf{1.371 \pm 0.491}$	$\mathbf{3.219 \pm 1.268}$	$\mathbf{0.192 \pm 0.128}$
Oral-3D (syn)	1.910 ± 0.565	7.123 ± 3.726	0.385 ± 0.159
NeBLa (syn)	2.089 ± 0.605	7.526 ± 3.946	0.486 ± 0.156

*NeBLa (real) has 159 valid anterior-angle cases; other entries use 161 cases. Anterior/posterior subsets are positions 1–3/4–8. Best means are determined before rounding; effective tooth counts vary across methods.

more sensitive: on each method's common teeth (a tooth is included in the statistics only when it is present across all methods), ROOTFlow and PX2Tooth yield $7.3438°$ and $10.7091°$, respectively. Per-position means, standard deviations, and counts are provided in Appendix D.

Anterior orientation and posterior morphology. ROOTFlow yields an anterior angular difference of $3.2191°$, a 48.98% reduction from PX2Tooth's $6.3098°$. Its posterior peak-count MAE is 0.1918, 50.14% below the closest configuration, Oral-3D (syn). These subsets complement global overlap and surface metrics by assessing anterior orientation and posterior local morphology. ROOTFlow's signed length and posterior peak-count differences are -0.1442 ± 0.8317 mm and -0.0751 ± 0.1651, respectively. We use MAE to characterize error magnitude because positive and negative signed errors can cancel.

Volume fidelity. Table 2 shows that ROOTFlow achieves the best results on all three image metrics among the evaluated real-PX configurations. Compared with Oral-3D (real) and NeBLa (real), PSNR increases by 2.33 and 1.41 dB, while LPIPS decreases by 24.26% and 29.12%, respectively. Across all configurations, predictably, NeBLa (syn) has the highest PSNR and SSIM, and Oral-3D (syn) has the lowest LPIPS. Both have lower tooth Dice and higher ASD than ROOTFlow. Synthetic-PX are generated using deterministic, known mathematical projections from CBCT volumes. Consequently, both geometric and intensity discrepancies between the synthetic-PX and their source CBCT volumes are significantly smaller than those between real-PX and CBCT volumes. This reduced difference makes it conceivably easier for the model to reconstruct CBCT from synthetic panoramic data.

4.3 Qualitative Results and Discussion

Qualitative comparison. Figures 4 and 5 compare Case B (adolescent) and Case C (adult). Mesh views illustrate differences in tooth completeness, surface shape, and arrangement, while volume renderings and slices show tooth boundaries and separation. Real- and synthetic-PX configurations are labeled separately. The mesh left views highlight ROOTFlow's closer agreement with GT in tooth inclination and visible root morphology in these examples. Differences in surrounding-tissue coverage are not interpreted as tooth-reconstruction errors by themselves.

Table 2: Volume fidelity on 161 common cases (mean ± population standard deviation). Synthetic-PX configurations are separated from real-PX comparisons. Bold denotes the best mean overall; underlining denotes the best mean among real-PX configurations. PX2Tooth has no intensity-volume output.

Method	PSNR (dB)↑	SSIM↑	LPIPS↓
Oral-3D (real)	20.527 ± 1.402	0.831 ± 0.021	0.162 ± 0.025
NeBLa (real)	21.444 ± 1.542	0.846 ± 0.021	0.173 ± 0.035
PX2Tooth	—	—	—
ROOTFlow	$\underline{22.858} \pm 2.001$	$\underline{0.872} \pm 0.023$	$\underline{0.122} \pm 0.026$
Oral-3D (syn)	23.525 ± 1.832	0.889 ± 0.020	$\mathbf{0.113 \pm 0.025}$
NeBLa (syn)	$\mathbf{24.659 \pm 1.969}$	$\mathbf{0.895 \pm 0.020}$	0.134 ± 0.035

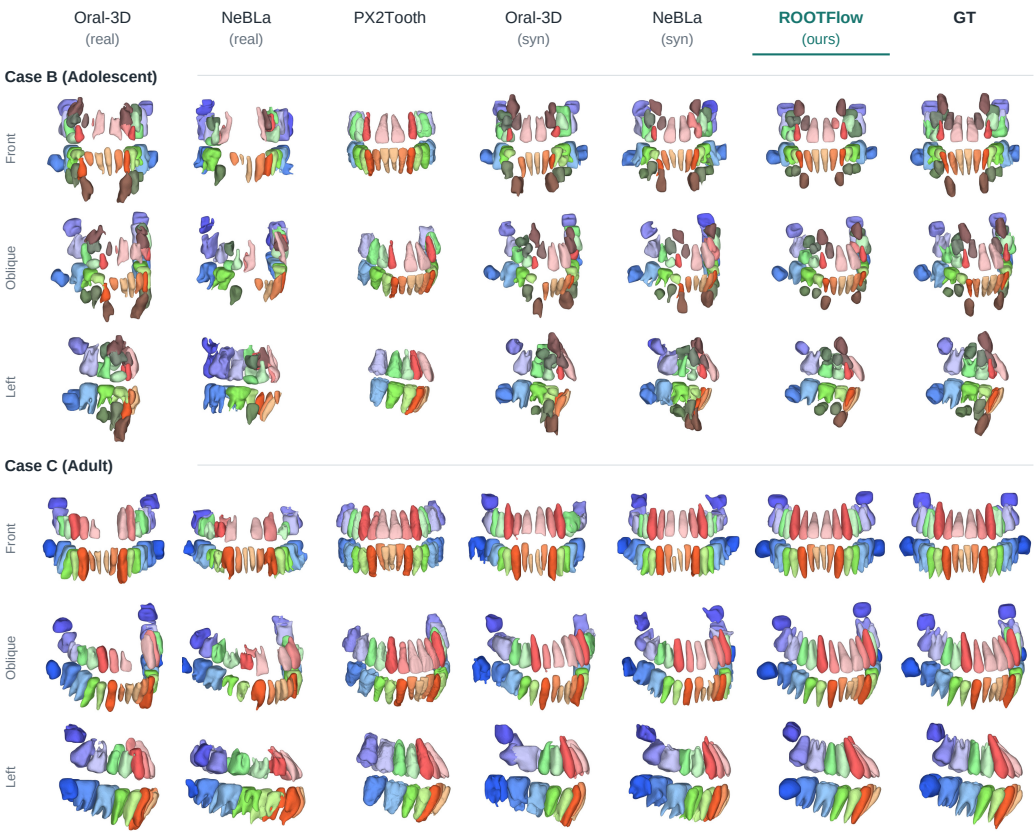


Figure 4: Tooth-surface comparisons for Case B (adolescent) and Case C (adult), shown from frontal, oblique, and left views. Columns compare real-PX baselines, PX2Tooth, synthetic-PX baselines, ROOTFlow, and GT. Colors indicate tooth instances.

Learning from paired clinical observations. ROOTFlow is motivated by the association between projected radiographic appearance and three-dimensional tooth morphology. CBCT–PX alignment uses paired clinical observations to learn a representation of this association before generation, while IOS supplies observed crown geometry. This is a learned statistical prior rather than a geometric inversion of PX. The comparisons assess the complete framework; without component ablations, they do not isolate the contributions of alignment, geometric conditioning, or flow matching.

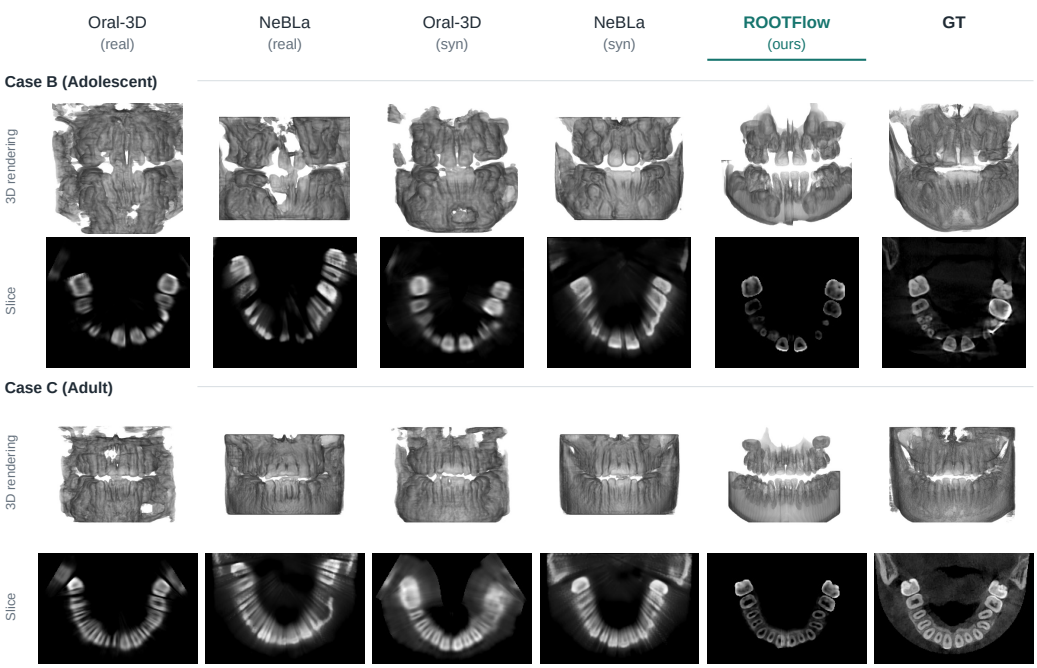


Figure 5: Volumetric comparisons for Case B (adolescent) and Case C (adult). Each case shows a three-dimensional rendering and a slice for the real- and synthetic-PX baselines, ROOTFlow, and GT. PX2Tooth is omitted because it does not generate intensity volumes.

Limitations. IOS and PX do not uniquely determine hidden roots; deterministic generation does not quantify this ambiguity. Current descriptors use whole-tooth masks and do not directly measure isolated root surfaces, apex positions, or clinically annotated root counts.

This study employs quadrant-level volume cropping to increase the number of training samples, simplify geometric modeling, and reduce annotation burden. However, this design does not model inter-quadrant dependencies. For example, the mandibular left and right central incisors are reconstructed independently, despite their adjacency and shared arch-level context. Extending the framework to model the complete dentition, or to incorporate cross-quadrant constraints, may provide a more information-rich representation and improve global anatomical consistency.

In clinical imaging workflows, lateral radiographs or facial photographs may also be available for some patients. These complementary views could provide structural information that is limited in a PX and IOS, particularly regarding mandibular morphology and overall craniofacial geometry. Incorporating such multimodal information is a promising direction for future work and may improve three-dimensional reconstruction beyond tooth-shape fidelity alone.

5 Conclusion

ROOTFlow combines IOS-derived crown geometry with a radiographic representation learned from paired CBCT–PX observations to generate quadrant-level CBCT volumes and tooth segmentations through latent flow matching. On the common evaluation cohort, it improves whole-tooth overlap and surface agreement and outperforms the real-PX volumetric baselines on image fidelity. Position-wise and common-tooth analyses further characterize the effects of evaluation coverage. These results support geometry-conditioned reconstruction, while isolated-root and task-specific clinical validation remain necessary.

AI USE STATEMENT

We disclose the use of generative AI tools in preparing this manuscript. Codex 5.6 was used to draft and polish sections of the text, including the presentation of the method, experiments, and discussion, and to write and debug the experimental code; Perplexity was used for literature retrieval and discovery, including identification and organization of related work. All research ideas, technical contributions, experimental designs, results, and claims are the authors' own work. The authors reviewed and verified all AI-assisted text and code, and verified all references against the original publications, taking full responsibility for the content of this manuscript. No AI tool was used for research ideation, synthetic data generation, or proving mathematical claims.

ETHICS STATEMENT

This study uses paired CBCT volumes, intraoral scans, and clinical panoramic radiographs to develop and evaluate a computational reconstruction method. The generated volumes and tooth geometries have not been validated as replacements for diagnostic CBCT.

REPRODUCIBILITY STATEMENT

Section 3 defines the model and training objectives. Appendices A–C describe data preprocessing, network configurations, numerical integration, and evaluation protocols. Appendix D reports supplementary results on tooth coverage and intensity-threshold sensitivity.

REFERENCES

Sandro Barone, Alessandro Paoli, and Armando Viviano Razionale. Geometrical modeling of complete dental shapes by using panoramic x-ray, digital mouth data and anatomical templates. *Computerized Medical Imaging and Graphics*, 43:112–121, 2015. doi: 10.1016/j.compmedimag.2015.01.005.

Hugh Devlin and J. Yuan. Object position and image magnification in dental panoramic radiography: A theoretical analysis. *Dentomaxillofacial Radiology*, 42(1):29951683, 2013. doi: 10.1259/dmfr/29951683.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

Xiaoshuang Li, Mingyuan Meng, Zimo Huang, Lei Bi, Eduardo Delamare, Dagan Feng, Bin Sheng, and Jinman Kim. 3DPX: Progressive 2D-to-3D oral image reconstruction with hybrid MLP–CNN networks. In *Medical Image Computing and Computer Assisted Intervention*, pp. 25–34, 2024. doi: 10.1007/978-3-031-72104-5_3. URL https://papers.miccai.org/miccai-2024/006-Paper2442.html.

Yuan Liang, Weinan Song, Jiawei Yang, Liang Qiu, Kun Wang, and Lei He. X2Teeth: 3D teeth reconstruction from a single panoramic radiograph. *arXiv preprint arXiv:2108.13004*, 2021. URL https://arxiv.org/abs/2108.13004.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.02747.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.03003.

Wen Ma, Huikai Wu, Zikai Xiao, Yang Feng, Jian Wu, and Zuozhu Liu. PX2Tooth: Reconstructing the 3D point cloud teeth from a single panoramic x-ray. In *Medical Image Computing and Computer Assisted Intervention*, pp. 411–421, 2024. doi: 10.1007/978-3-031-72384-1_39. URL https://papers.miccai.org/miccai-2024/631-Paper1637.html.

Laura Mazzotta, Mauro Cozzani, Armando Razionale, Sabrina Mutinelli, Attilio Castaldo, and Armando Silvestrini-Biavati. From 2D to 3D: Construction of a 3D parametric model for detection of dental roots shape and position from a panoramic radiograph—a preliminary report. *International Journal of Dentistry*, 2013:964631, 2013. doi: 10.1155/2013/964631.

Lanzhuju Mei, Yu Fang, Yue Zhao, Xiang Sean Zhou, Min Zhu, Zhiming Cui, and Dinggang Shen. DTR-Net: Dual-space 3D tooth model reconstruction from panoramic x-ray images. *IEEE Transactions on Medical Imaging*, 43(1):517–528, 2024. doi: 10.1109/TMI.2023.3313795.

Sihwa Park, Seongjun Kim, In-Seok Song, and Seung Jun Baek. 3D teeth reconstruction from panoramic radiographs using neural implicit functions. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 2023. doi: 10.1007/978-3-031-43999-5_36. URL https://arxiv.org/abs/2311.16524.

Sihwa Park, Seongjun Kim, Doeyoung Kwon, Yohan Jang, In-Seok Song, and Seung Jun Baek. NeBLa: Neural Beer–Lambert for 3D reconstruction of oral structures from panoramic radiographs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(5):4433–4441, 2024. doi: 10.1609/aaai.v38i5.28241. URL https://ojs.aaai.org/index.php/AAAI/article/view/28241.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 8748–8763, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022. URL https://arxiv.org/abs/2112.10752.

Weinan Song, Yuan Liang, Jiawei Yang, Kun Wang, and Lei He. Oral-3D: Reconstructing the 3D structure of oral cavity from panoramic x-ray. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):566–573, 2021. doi: 10.1609/aaai.v35i1.16135. URL https://ojs.aaai.org/index.php/AAAI/article/view/16135.

Steve Stramotas, Joseph P. Geenty, Peter Petocz, and M. Ali Darendeliler. Accuracy of linear and angular measurements on panoramic radiographs taken at various positions in vitro. *European Journal of Orthodontics*, 24(1):43–52, 2002. doi: 10.1093/ejo/24.1.43.

Anusree P. Sunilkumar, Bikram Keshari Parida, Seong-Yong Moon, and Wonsang You. Panoramic x-ray synthesis from dental CBCT using patient-tailored simulated geometry with uniform projection sampling. *IEEE Access*, 2025. doi: 10.1109/ACCESS.2025.3573744.

Tomasz Szczepański, Szymon Płotka, Michal K. Grzeszczyk, Tomasz Trzciński, and Arkadiusz Sitek. X-Splat: Gaussian splatting for 3D CBCT generation from single panoramic radiograph. *arXiv preprint arXiv:2607.02099*, 2026. URL https://arxiv.org/abs/2607.02099.

Michael Tschannen, Alexey Gritsenko, Xiao Wang, Muhammad Ferjad Naeem, Ibrahim Alabdulmohsin, Nikhil Parthasarathy, Talfan Evans, Lucas Beyer, Ye Xia, Basil Mustafa, Olivier Hénaff, Jeremiah Harmsen, Andreas Steiner, and Xiaohua Zhai. SigLIP 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features. *arXiv preprint arXiv:2502.14786*, 2025. URL https://arxiv.org/abs/2502.14786.

U.S. Food and Drug Administration. Dental cone-beam computed tomography. https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/dental-cone-beam-computed-tomography, n.d. Accessed September 21, 2026.

Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, William Berman, Yiyi Xu, Steven Liu, and Thomas Wolf. Diffusers: State-of-the-art diffusion models. https://github.com/huggingface/diffusers, 2022.

Eric Whaites. *Essentials of dental radiography and radiology*. Churchill Livingstone; Edinburgh, 2002.

Yuezhi Yang, Zhiming Cui, Changjian Li, and Wenping Wang. ToothInpaintor: Tooth inpainting from partial 3D dental model and 2D panoramic image. *arXiv preprint arXiv:2211.15502*, 2022. URL https://arxiv.org/abs/2211.15502.

Han-Gyeol Yeom, Jo-Eun Kim, Kyung-Hoe Huh, Won-Jin Yi, Min-Suk Heo, Sam-Sun Lee, and Soon-Chul Choi. Development of a new ball-type phantom for evaluation of the image layer of panoramic radiography. *Imaging Science in Dentistry*, 48(4):255–259, 2018.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11975–11986, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Zhai_Sigmoid_Loss_for_Language_Image_Pre-Training_ICCV_2023_paper.html.

A Data Construction and Preprocessing

A.1 Coordinates and Quadrant Crops

CBCT and label maps are resampled to 0.5 mm isotropic spacing. Each quadrant has shape $96 \times 128 \times 96$ in depth–height–width order. Supplied IOS-to-volume transforms establish a common coordinate system. Lower quadrants are flipped along depth and right quadrants along width, with consistent operations on all volumetric channels. Training crops use tooth-label bounds with randomized placement. The standalone prediction script instead anchors crops to crown meshes and reads a reference image header for size, spacing, origin, and direction. These protocols and the provenance of transforms must be distinguished in evaluation.

Matching requires patient and quadrant agreement and selects an eligible radiograph within the configured time window. Panoramic images are cropped into quadrant observations; standalone prediction uses a tooth detector. A quadrant is a correlated sample from a patient, not an independent patient.

A.2 Intensity and Label Channels

The CBCT representation contains one normalized intensity channel and 13 tooth-label channels, giving 14 channels in total. Intensity normalization clips values to $[-1000, 4000]$ HU and rescales them to $[0, 1]$. The tooth-label channels encode the quadrant-level tooth identities. This normalization does not establish cross-scanner physical calibration.

The tooth vocabulary contains eight ordinary quadrant slots and five additional slots corresponding to I-prefixed annotation identifiers at positions 1–5. I-prefixed identifiers at positions 6–8 are mapped to the corresponding ordinary slots. IOS mesh identifiers use the quadrant prefixes UR, UL, LL, and LR followed by positions 1–8.

A.3 UDF Storage and Identity Encoding

Distances in Equation 1 are computed after mesh vertices are transformed into voxel coordinates. The stored unsigned 8-bit field is

$$Q(\mathbf{x}) = \left\lfloor 255 \frac{\min(d(\mathbf{x}), \tau)}{\tau} \right\rfloor, \qquad U_q(\mathbf{x}) = 1 - Q(\mathbf{x})/255. \tag{7}$$

The encoder receives U_q, the quantized version of U. Nearest-mesh identities define eight crown slots within its support. Concatenating this field with eight tooth-identity channels gives the nine-channel input X_u. These channels describe proximity to crown surfaces, not complete-tooth interiors. The truncation is five voxel units, equivalent to 2.5 mm on the stated grid.

Table 3: Representation and flow settings for the retained method configuration.

Setting	Value
Quadrant grid / spacing	$96 \times 128 \times 96$ / 0.5 mm
CBCT / UDF input channels	14 / 9
CBCT / UDF tooth-label channels	13 / 8
VAE widths / latent channels	$(16, 32, 64)$ / 12
Latent spatial grid	$24 \times 32 \times 24$
PX image size	512×512
Velocity U-Net widths	$(64, 128, 256)$
Velocity input / output channels	24 / 12
Latent-statistics subset	Up to 512 training quadrants
Integration / steps	Heun / 16
EMA decay	0.999

A.4 Quadrant Export

Generated quadrants can be unflipped and placed in a common output frame. The supplied export takes voxelwise maxima in overlaps for intensity and integer labels. This convention can introduce seams or resolve label conflicts by identifier; it is not a separately learned full-jaw reconstruction model.

B Network Architectures and Training Details

B.1 Volumetric Representations

Each VAE uses 3D residual blocks with two spatial downsampling operations and corresponding decoder upsampling. The configuration has widths $(16, 32, 64)$ and 12 latent channels. Encoders parameterize diagonal Gaussian posteriors; log variances are clamped to $[-12, 8]$. VAE training uses reparameterized samples, while downstream caches contain posterior means. Both decoders participate in reconstruction training; only the CBCT decoder is used for subsequent volume generation.

B.2 Alignment Objective

A 3D ResNet and a pretrained SigLIP 2 visual tower supply CBCT and PX representations. Projection heads output normalized 256-dimensional embeddings h_c, h_p. The PX tower is fine-tuned at 5% of the learning rate of the other parameters. Let P_i contain candidates with the same matched radiograph path as anchor i, and A_i contain these positives and candidates from other patients. One directional objective is

$$\mathcal{L}_{c\to p} = -\frac{1}{B}\sum_i \sum_{j\in A_i} w_{ij} \log \sigma\left[y_{ij}(s h_{c,i}^\top h_{p,j} + b)\right], \tag{8}$$

where positives have $y_{ij} = 1$ and $w_{ij} = 1/|P_i|$; negatives have $y_{ij} = -1$ and unit weight. The scale s is learned and bias b is fixed by configuration. Both directions are averaged. Two auxiliary heads predict tooth-presence vectors, with the weighted mean of their binary cross-entropy losses added to alignment. Presence targets come from CBCT annotations, including teeth unobserved by IOS.

The global PX projection enters image-conditioned embeddings and last-layer patch tokens enter cross-attention. Feature caches retain checkpoint identity to keep these representations consistent across training and inference.

B.3 Normalization and Numerical Integration

Channelwise means and standard deviations are estimated separately for CBCT and UDF using a fixed hash-selected training subset. For modality m, normalization is $N_m(z) = (z - \mu_m)/\sigma_m$. Flow learning uses the normalized latents; decoding reverses CBCT normalization only. For step

size $h = 1/16$, Heun integration evaluates

$$\begin{aligned} k_1 &= v_\theta(x_n, t_n; u, f_p), \\ k_2 &= v_\theta(x_n + hk_1, t_n + h; u, f_p), \\ x_{n+1} &= x_n + \tfrac{h}{2}(k_1 + k_2). \end{aligned} \tag{9}$$

EMA weights are used for generation. The terminal state is decoded after inverse normalization. No CBCT encoder or UDF decoder is evaluated in this path.

C Evaluation Protocols

C.1 Volume Fidelity and Shared Segmentation

Predicted volumes are cropped or padded onto the reference grid, with missing regions filled at -1000 HU. Intensities are clipped to $[-1000, 4000]$ HU, normalized to $[0, 1]$, and set to zero outside the union of reference bone and tooth masks. PSNR uses data range 1 and mean squared error over the full masked grid. SSIM is computed in three dimensions using a $7 \times 7 \times 7$ uniform window, sample covariance, and data range 1. LPIPS uses AlexNet (v0.1), with each grayscale axial slice repeated across three channels and rescaled from $[0, 1]$ to $[-1, 1]$, without resizing. Scores are averaged equally over all axial slices. These image metrics are not computed for PX2Tooth. Segmentation is preserved from an external tooth instance segmenter. PX2Tooth geometry is evaluated from precomputed per-tooth voxel masks.

C.2 Tooth Presence, Overlap, and Surface Distance

Instances are matched by exact tooth-label names. The candidate vocabulary comprises 32 ordinary positions (UR, UL, LL, LR, each 1–8) and 20 I-prefixed labels (each quadrant, 1–5). Presence requires at least one labeled voxel, with no additional size threshold. Presence accuracy is $(TP + TN)/(TP+TN+FP+FN)$. Here, TP denotes presence in both prediction and reference, without an overlap threshold. All geometry metrics are undefined for FP, FN, and TN labels and excluded from geometric averages. PX2Tooth presence accuracy is not reported.

For nonempty binary tooth masks S and G, Dice is $2|S \cap G|/(|S|+|G|)$. ASD is the arithmetic mean of the two directed mean nearest-surface distances, using six-connected boundary voxel centers in physical coordinates. Units are millimeters. Both metrics use whole-tooth masks. I-prefixed labels are tabulated separately and excluded from the anterior and posterior subsets.

C.3 Whole-Tooth Length, Axis Direction, and Morphological Peaks

PCA uses centered whole-tooth voxel centers scaled by physical spacing. The eigenvector associated with the largest eigenvalue defines the long axis. Length is the maximum minus minimum projection onto that axis, without adding voxel width. Signed length difference is prediction minus reference; MAE takes the absolute difference for each tooth before aggregation. The sign of each PCA axis is fixed in array coordinates: its z component is positive for UR/UL and negative for LL/LR, including I-prefixed labels. For unit axes $\hat{\mathbf{u}}$ and $\mathbf{u}$, angular error is

$$E_{\text{axis}} = \frac{180}{\pi} \arccos\big(\text{clip}(\hat{\mathbf{u}}^\top \mathbf{u}, -1, 1)\big), \qquad \|\hat{\mathbf{u}}\|_2 = \|\mathbf{u}\|_2 = 1. \tag{10}$$

The error is in degrees and ranges from 0 to 180; opposite directions are distinct. The anterior subset comprises UR1–UR3, UL1–UL3, LL1–LL3, and LR1–LR3, excluding I-prefixed labels. No independent crown–root separation is performed.

Posterior morphological peaks are detected on a PCA height map oriented by the same axis-sign convention, retaining the maximum axial projection in each transverse bin. Detection uses 0.5 mm transverse bins, nearest-neighbor hole filling, Gaussian smoothing with a standard deviation of 0.5 mm, 0.5 mm peak prominence, and 2 mm transverse peak suppression. The subset comprises positions 4–8 in all four quadrants, excluding I-prefixed labels. We report signed peak-count difference and MAE. These are morphological descriptors, not validated anatomical root counts; no erupted-only restriction is applied.

C.4 Coordinates, Subgroups, and Aggregation

Segmentation evaluation requires matching voxel spacing and orientation and integer-voxel origin offsets. Reference and prediction masks are placed on a shared physical grid covering the union of their fields of view, retaining predicted foreground outside the reference extent. No registration transform is estimated by the evaluator. For each metric, finite, defined tooth values are averaged within each case, followed by an unweighted mean over cases with a defined value. Standard deviations use denominator N, where N is the number of contributing cases. Per-position tables report means, population standard deviations, and valid instance counts. I-prefixed labels are tabulated separately.

D Additional Results

D.1 Sensitivity to Coverage and Intensity Threshold

All six configurations have defined anterior angular errors for 159 common cases. On this subset, ROOTFlow and PX2Tooth yield $3.2057°$ and $6.2983°$; the ranking agrees with the main comparison. Common-tooth results use 3,235 instances across 161 cases and retain the same within-case then between-case averaging scheme.

Threshold Dice is evaluated at 0–1000 HU in 100 HU increments within the union of reference bone and tooth masks, applying each threshold separately to prediction and reference. Rankings vary with threshold: ROOTFlow is not uniformly best and falls below Oral-3D (real) at some higher thresholds. We therefore report the complete sweep rather than choosing a threshold based on favorable ranking. These intensity-set Dice scores differ from tooth-instance Dice.

Table 4 reports common-tooth means, and Table 5 reports the complete intensity-threshold sweep. Tables 6, 7, and 8 provide the per-position Dice statistics for all 52 labels. These comparisons are descriptive; no statistical significance tests are reported.

Table 4: Common-tooth sensitivity analysis on 3,235 instances from 161 cases. Entries are case-averaged means; ASD and length MAE are in millimeters, angle is in degrees, and peaks denotes count MAE. Angle and peak-count columns here cover all common teeth, unlike the restricted subsets in the main table.

Method	Dice↑	ASD↓	Angle↓	Length↓	Peaks↓
Oral-3D (real)	0.301	2.651	17.159	2.336	0.309
NeBLa (real)	0.124	5.304	22.389	3.235	0.426
PX2Tooth	0.768	0.645	10.709	1.876	0.316
ROOTFlow	0.875	0.348	7.344	1.337	0.123
Oral-3D (syn)	0.730	0.727	12.309	1.874	0.245
NeBLa (syn)	0.661	0.906	13.866	2.003	0.304

Table 5: Intensity-threshold Dice over the common 161-case cohort. All computed thresholds are reported; PX2Tooth has no intensity-volume output.

HU	Oral-3D (real)	NeBLa (real)	ROOTFlow	Oral-3D (syn)	NeBLa (syn)
0	0.772	0.850	0.857	0.879	0.922
100	0.739	0.807	0.815	0.860	0.896
200	0.702	0.752	0.766	0.835	0.861
300	0.664	0.693	0.717	0.806	0.822
400	0.628	0.634	0.665	0.777	0.782
500	0.593	0.573	0.609	0.748	0.741
600	0.559	0.514	0.552	0.719	0.702
700	0.525	0.460	0.494	0.690	0.665
800	0.493	0.412	0.439	0.660	0.632
900	0.461	0.368	0.390	0.628	0.603
1000	0.429	0.327	0.349	0.596	0.575

Table 6: Per-position Dice: upper labels. Each cell gives mean ± population standard deviation and valid instance count. Means use each method's own teeth present in both prediction and reference; they are not strictly paired across methods.

Label	Oral-3D (real)	NeBLa (real)	PX2Tooth	ROOTFlow	Oral-3D (syn)	NeBLa (syn)
UR1	0.278 ± 0.207 $n = 158$	0.135 ± 0.170 $n = 141$	0.795 ± 0.050 $n = 156$	0.894 ± 0.032 $n = 161$	0.712 ± 0.112 $n = 161$	0.670 ± 0.164 $n = 161$
UR2	0.249 ± 0.212 $n = 150$	0.097 ± 0.150 $n = 127$	0.754 ± 0.064 $n = 153$	0.868 ± 0.046 $n = 158$	0.692 ± 0.141 $n = 158$	0.607 ± 0.185 $n = 158$
UR3	0.302 ± 0.214 $n = 135$	0.121 ± 0.165 $n = 115$	0.740 ± 0.112 $n = 140$	0.864 ± 0.071 $n = 150$	0.692 ± 0.188 $n = 150$	0.593 ± 0.212 $n = 148$
UR4	0.313 ± 0.222 $n = 143$	0.142 ± 0.167 $n = 128$	0.765 ± 0.088 $n = 135$	0.875 ± 0.052 $n = 146$	0.726 ± 0.157 $n = 147$	0.558 ± 0.199 $n = 147$
UR5	0.319 ± 0.242 $n = 141$	0.096 ± 0.141 $n = 129$	0.766 ± 0.086 $n = 139$	0.864 ± 0.080 $n = 150$	0.741 ± 0.137 $n = 150$	0.604 ± 0.179 $n = 143$
UR6	0.424 ± 0.223 $n = 156$	0.138 ± 0.187 $n = 127$	0.803 ± 0.034 $n = 151$	0.884 ± 0.079 $n = 161$	0.773 ± 0.106 $n = 161$	0.713 ± 0.112 $n = 160$
UR7	0.377 ± 0.233 $n = 153$	0.115 ± 0.158 $n = 119$	0.805 ± 0.039 $n = 79$	0.822 ± 0.090 $n = 157$	0.749 ± 0.122 $n = 159$	0.719 ± 0.113 $n = 160$
UR8	0.301 ± 0.202 $n = 64$	0.033 ± 0.094 $n = 55$	0.734 ± 0.073 $n = 13$	0.630 ± 0.191 $n = 80$	0.695 ± 0.119 $n = 72$	0.626 ± 0.149 $n = 75$
UL1	0.258 ± 0.207 $n = 154$	0.169 ± 0.183 $n = 149$	0.804 ± 0.044 $n = 156$	0.898 ± 0.032 $n = 160$	0.722 ± 0.111 $n = 160$	0.695 ± 0.145 $n = 160$
UL2	0.226 ± 0.228 $n = 149$	0.116 ± 0.160 $n = 137$	0.749 ± 0.070 $n = 149$	0.873 ± 0.043 $n = 156$	0.713 ± 0.133 $n = 157$	0.655 ± 0.165 $n = 157$
UL3	0.255 ± 0.228 $n = 147$	0.125 ± 0.176 $n = 118$	0.735 ± 0.116 $n = 145$	0.869 ± 0.056 $n = 155$	0.731 ± 0.140 $n = 155$	0.646 ± 0.173 $n = 152$
UL4	0.267 ± 0.230 $n = 148$	0.137 ± 0.161 $n = 141$	0.767 ± 0.090 $n = 140$	0.875 ± 0.070 $n = 154$	0.744 ± 0.128 $n = 152$	0.586 ± 0.176 $n = 153$
UL5	0.281 ± 0.233 $n = 146$	0.097 ± 0.155 $n = 131$	0.778 ± 0.070 $n = 139$	0.880 ± 0.059 $n = 149$	0.733 ± 0.160 $n = 150$	0.638 ± 0.153 $n = 141$
UL6	0.386 ± 0.235 $n = 154$	0.142 ± 0.202 $n = 130$	0.801 ± 0.076 $n = 151$	0.879 ± 0.108 $n = 159$	0.786 ± 0.089 $n = 158$	0.735 ± 0.092 $n = 159$
UL7	0.340 ± 0.239 $n = 153$	0.121 ± 0.180 $n = 127$	0.793 ± 0.122 $n = 85$	0.803 ± 0.135 $n = 159$	0.748 ± 0.111 $n = 160$	0.740 ± 0.112 $n = 159$
UL8	0.300 ± 0.249 $n = 71$	0.032 ± 0.090 $n = 55$	0.767 ± 0.047 $n = 15$	0.614 ± 0.205 $n = 83$	0.693 ± 0.154 $n = 72$	0.640 ± 0.176 $n = 84$

Table 7: Per-position Dice: lower labels. Each cell gives mean ± population standard deviation and valid instance count. Means use each method's own teeth present in both prediction and reference; they are not strictly paired across methods.

Label	Oral-3D (real)	NeBLa (real)	PX2Tooth	ROOTFlow	Oral-3D (syn)	NeBLa (syn)
LL1	0.153 ± 0.194 $n = 158$	0.104 ± 0.174 $n = 144$	0.745 ± 0.059 $n = 153$	0.873 ± 0.057 $n = 160$	0.655 ± 0.161 $n = 160$	0.612 ± 0.191 $n = 160$
LL2	0.170 ± 0.211 $n = 152$	0.105 ± 0.173 $n = 141$	0.754 ± 0.066 $n = 152$	0.874 ± 0.045 $n = 157$	0.680 ± 0.136 $n = 157$	0.634 ± 0.176 $n = 158$
LL3	0.243 ± 0.227 $n = 149$	0.153 ± 0.188 $n = 136$	0.738 ± 0.099 $n = 146$	0.869 ± 0.051 $n = 151$	0.710 ± 0.150 $n = 154$	0.622 ± 0.192 $n = 153$
LL4	0.270 ± 0.224 $n = 149$	0.126 ± 0.156 $n = 140$	0.750 ± 0.091 $n = 149$	0.873 ± 0.067 $n = 152$	0.722 ± 0.132 $n = 153$	0.639 ± 0.164 $n = 149$
LL5	0.320 ± 0.218 $n = 145$	0.137 ± 0.165 $n = 136$	0.762 ± 0.086 $n = 139$	0.849 ± 0.134 $n = 148$	0.728 ± 0.123 $n = 147$	0.643 ± 0.149 $n = 148$
LL6	0.445 ± 0.200 $n = 158$	0.176 ± 0.176 $n = 142$	0.814 ± 0.036 $n = 151$	0.885 ± 0.047 $n = 157$	0.776 ± 0.116 $n = 158$	0.730 ± 0.108 $n = 159$
LL7	0.451 ± 0.210 $n = 157$	0.195 ± 0.187 $n = 136$	0.822 ± 0.045 $n = 100$	0.826 ± 0.100 $n = 160$	0.751 ± 0.129 $n = 159$	0.742 ± 0.120 $n = 160$
LL8	0.448 ± 0.203 $n = 76$	0.175 ± 0.182 $n = 63$	0.761 ± 0.055 $n = 14$	0.622 ± 0.225 $n = 91$	0.714 ± 0.165 $n = 83$	0.607 ± 0.163 $n = 93$
LR1	0.160 ± 0.205 $n = 147$	0.065 ± 0.131 $n = 117$	0.754 ± 0.062 $n = 151$	0.867 ± 0.063 $n = 154$	0.669 ± 0.131 $n = 154$	0.623 ± 0.154 $n = 153$
LR2	0.186 ± 0.208 $n = 150$	0.058 ± 0.129 $n = 126$	0.737 ± 0.080 $n = 151$	0.863 ± 0.084 $n = 154$	0.672 ± 0.143 $n = 155$	0.623 ± 0.159 $n = 155$
LR3	0.258 ± 0.210 $n = 153$	0.102 ± 0.163 $n = 132$	0.733 ± 0.105 $n = 148$	0.859 ± 0.063 $n = 155$	0.694 ± 0.169 $n = 155$	0.614 ± 0.183 $n = 155$
LR4	0.294 ± 0.213 $n = 151$	0.122 ± 0.163 $n = 130$	0.736 ± 0.113 $n = 149$	0.869 ± 0.090 $n = 155$	0.721 ± 0.138 $n = 155$	0.638 ± 0.163 $n = 151$
LR5	0.330 ± 0.216 $n = 150$	0.102 ± 0.157 $n = 135$	0.759 ± 0.084 $n = 146$	0.864 ± 0.071 $n = 151$	0.737 ± 0.110 $n = 152$	0.678 ± 0.126 $n = 150$
LR6	0.471 ± 0.170 $n = 151$	0.178 ± 0.177 $n = 126$	0.808 ± 0.036 $n = 152$	0.884 ± 0.046 $n = 153$	0.792 ± 0.062 $n = 153$	0.749 ± 0.079 $n = 153$
LR7	0.471 ± 0.201 $n = 158$	0.179 ± 0.184 $n = 137$	0.821 ± 0.042 $n = 98$	0.827 ± 0.095 $n = 157$	0.772 ± 0.097 $n = 157$	0.747 ± 0.087 $n = 157$
LR8	0.435 ± 0.219 $n = 79$	0.125 ± 0.145 $n = 55$	0.755 ± 0.072 $n = 11$	0.576 ± 0.231 $n = 89$	0.685 ± 0.164 $n = 82$	0.577 ± 0.195 $n = 81$

Table 8: Per-position Dice: I-prefixed labels. Each cell gives mean ± population standard deviation and valid instance count. Means use each method's own teeth present in both prediction and reference; they are not strictly paired across methods.

Label	Oral-3D (real)	NeBLa (real)	PX2Tooth	ROOTFlow	Oral-3D (syn)	NeBLa (syn)
IUR1	—	—	—	—	—	—
	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$
	0.247 ± 0.216	0.000 ± 0.000	—	0.615 ± 0.133	0.520 ± 0.083	0.625 ± 0.086
IUR2	$n = 3$	$n = 1$	$n = 0$	$n = 3$	$n = 2$	$n = 2$
	0.262 ± 0.206	0.140 ± 0.129	0.314 ± 0.069	0.659 ± 0.098	0.585 ± 0.192	0.545 ± 0.208
IUR3	$n = 31$	$n = 6$	$n = 4$	$n = 32$	$n = 36$	$n = 35$
	0.302 ± 0.231	0.141 ± 0.210	—	0.658 ± 0.114	0.658 ± 0.154	0.489 ± 0.168
IUR4	$n = 23$	$n = 6$	$n = 0$	$n = 22$	$n = 25$	$n = 24$
	0.218 ± 0.209	0.080 ± 0.104	0.000 ± 0.000	0.534 ± 0.240	0.546 ± 0.229	0.428 ± 0.203
IUR5	$n = 30$	$n = 8$	$n = 1$	$n = 39$	$n = 36$	$n = 36$
	—	—	—	—	—	—
IUL1	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$
	—	0.008 ± 0.000	—	0.380 ± 0.000	0.547 ± 0.036	0.606 ± 0.133
IUL2	$n = 0$	$n = 1$	$n = 0$	$n = 1$	$n = 2$	$n = 3$
	0.256 ± 0.198	0.097 ± 0.089	0.269 ± 0.029	0.654 ± 0.141	0.643 ± 0.153	0.558 ± 0.174
IUL3	$n = 32$	$n = 7$	$n = 2$	$n = 33$	$n = 33$	$n = 32$
	0.232 ± 0.209	—	—	0.704 ± 0.089	0.684 ± 0.128	0.542 ± 0.149
IUL4	$n = 24$	$n = 0$	$n = 0$	$n = 22$	$n = 24$	$n = 24$
	0.220 ± 0.204	0.051 ± 0.135	0.092 ± 0.000	0.661 ± 0.148	0.653 ± 0.143	0.473 ± 0.159
IUL5	$n = 37$	$n = 8$	$n = 1$	$n = 37$	$n = 28$	$n = 38$
	—	—	—	—	—	—
ILL1	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$
	—	—	—	—	—	—
ILL2	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$
	0.252 ± 0.182	0.199 ± 0.225	0.238 ± 0.145	0.621 ± 0.111	0.688 ± 0.065	0.570 ± 0.156
ILL3	$n = 27$	$n = 7$	$n = 4$	$n = 26$	$n = 25$	$n = 25$
	0.270 ± 0.202	0.077 ± 0.187	—	0.663 ± 0.126	0.717 ± 0.087	0.714 ± 0.108
ILL4	$n = 26$	$n = 7$	$n = 0$	$n = 27$	$n = 27$	$n = 28$
	0.276 ± 0.178	0.205 ± 0.186	0.247 ± 0.133	0.670 ± 0.120	0.711 ± 0.098	0.685 ± 0.092
ILL5	$n = 41$	$n = 3$	$n = 3$	$n = 33$	$n = 37$	$n = 38$
	—	—	—	—	—	—
ILR1	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$
	—	—	—	—	—	—
ILR2	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$	$n = 0$
	0.264 ± 0.168	0.193 ± 0.222	0.278 ± 0.080	0.638 ± 0.135	0.703 ± 0.071	0.576 ± 0.117
ILR3	$n = 26$	$n = 8$	$n = 3$	$n = 26$	$n = 28$	$n = 27$
	0.300 ± 0.203	0.150 ± 0.162	0.413 ± 0.000	0.667 ± 0.129	0.765 ± 0.047	0.729 ± 0.077
ILR4	$n = 29$	$n = 8$	$n = 1$	$n = 29$	$n = 32$	$n = 31$
	0.277 ± 0.189	0.022 ± 0.039	—	0.691 ± 0.106	0.717 ± 0.068	0.686 ± 0.101
ILR5	$n = 40$	$n = 7$	$n = 0$	$n = 38$	$n = 41$	$n = 42$