

Tooth-Diffusion

Guided 3D CBCT Synthesis with Fine-Grained Tooth Conditioning

Said Djafar Said¹ Torkan Gholamalizadeh² Mostafa Mehdipour Ghazi¹

¹ Pioneer Centre for Artificial Intelligence, Department of Computer Science, University of Copenhagen

² Research and Development, 3Shape A/S, Copenhagen, Denmark

Motivation

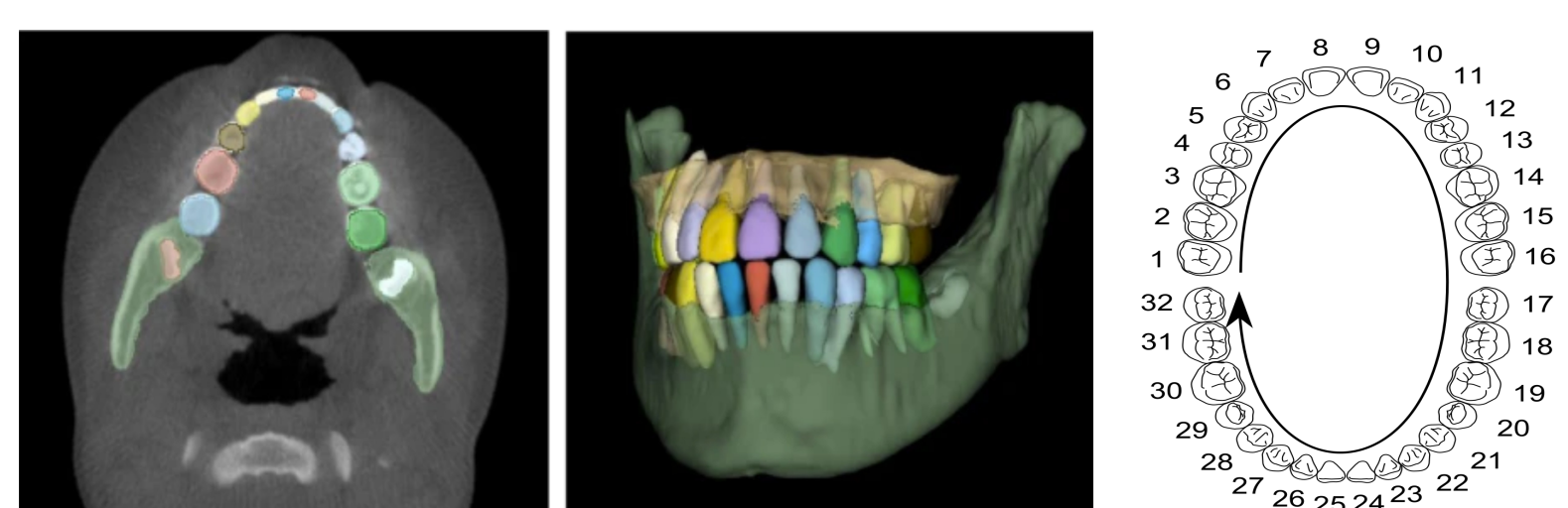
- CBCT is widely used for dental diagnosis, treatment planning, and education.
- Repeated scanning is costly, exposes patients to radiation, and is impractical for simulating alternative scenarios.
- CBCT synthesis offers an efficient way to augment data, support model training, and facilitate treatment planning.
- Most existing approaches either do not synthesize dental CBCT volumes or lack fine-grained anatomical control.

Objective

- Introduce a controllable diffusion model for 3D CBCT synthesis with tooth-level conditioning.
- Enable users to specify the presence/absence of individual teeth while preserving global anatomical coherence.

Dataset

- A curated labeled dataset consisting of 98 CBCT volumes.
- 90 volumes used for training/validation, and 8 unique patient scans reserved for testing.
- Covers edge cases such as full dentition, partial dentition, and scans with metal artifacts.
- Standardized to a $256 \times 256 \times 256$ voxel size (crop/pad), preserving the original resolution of $\sim 0.4 \text{ mm}^3$.
- Intensities normalized to $[1, 1]$ for stable training of wavelet diffusion models.
- Binary tooth labels (1–32) according to the Universal Numbering System (UNS), excluding supernumerary teeth.



A segmented dental CBCT and UNS-based relabeling

Methods

Tooth augmentation:

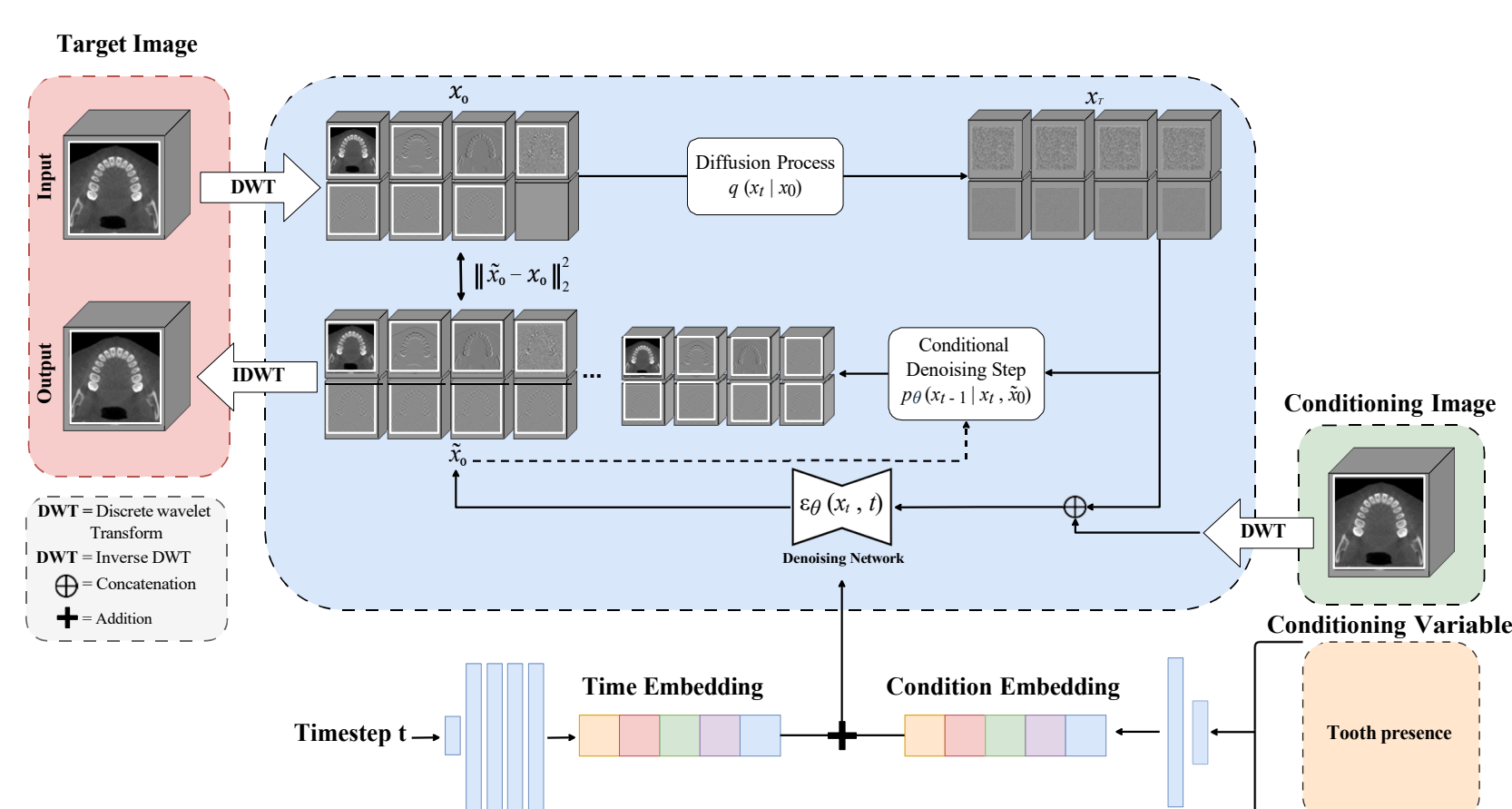
- Masking up to 50% of teeth in the conditioning scan for target reconstruction.
- Conditioning on the original scan while removing teeth from the target scan.
- Using an image-based inpainting pipeline to fill cavities with plausible intensities.
- Horizontal flips with relabeling to preserve tooth identity.

Masked L2 loss:

- A soft spatial mask from tooth segmentations emphasizes reconstructed regions instead of the background.
- The masked L2 loss is combined with the primary diffusion reconstruction loss:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Diffusion}} + \lambda \|M \odot (x - \hat{x})\|_2^2,$$

Conditional Diffusion Framework



The proposed model for conditional CBCT synthesis

Wavelet-based diffusion:

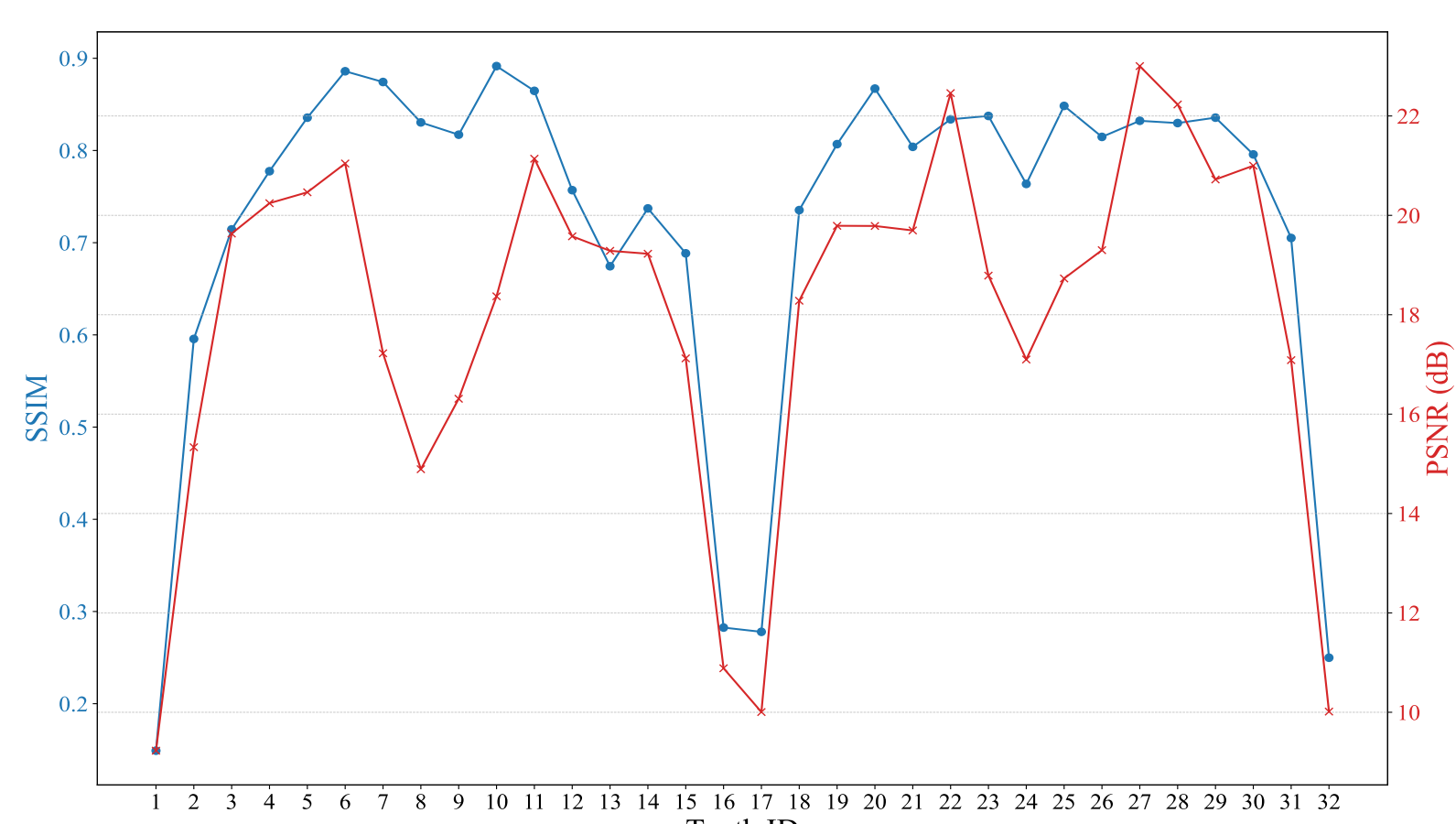
- Diffusion in the wavelet domain using a 3D Haar wavelet decomposition.
- Denoising performed in lower-resolution multiscale latent decompositions.
- Wavelet representation improves memory efficiency and preserves structural details compared to voxel space.

Tooth-based conditioning:

- A 32-dimensional binary condition vector encodes tooth presence or absence across the jaw.
- The vector is embedded, combined with timestep embeddings, and incorporated into the denoiser via FiLM.
- FiLM scales and shifts intermediate features according to tooth configuration, preserving global anatomy.

Tooth Addition Synthesis

- Reconstruction of simulated missing teeth in test scans with complete dentition.



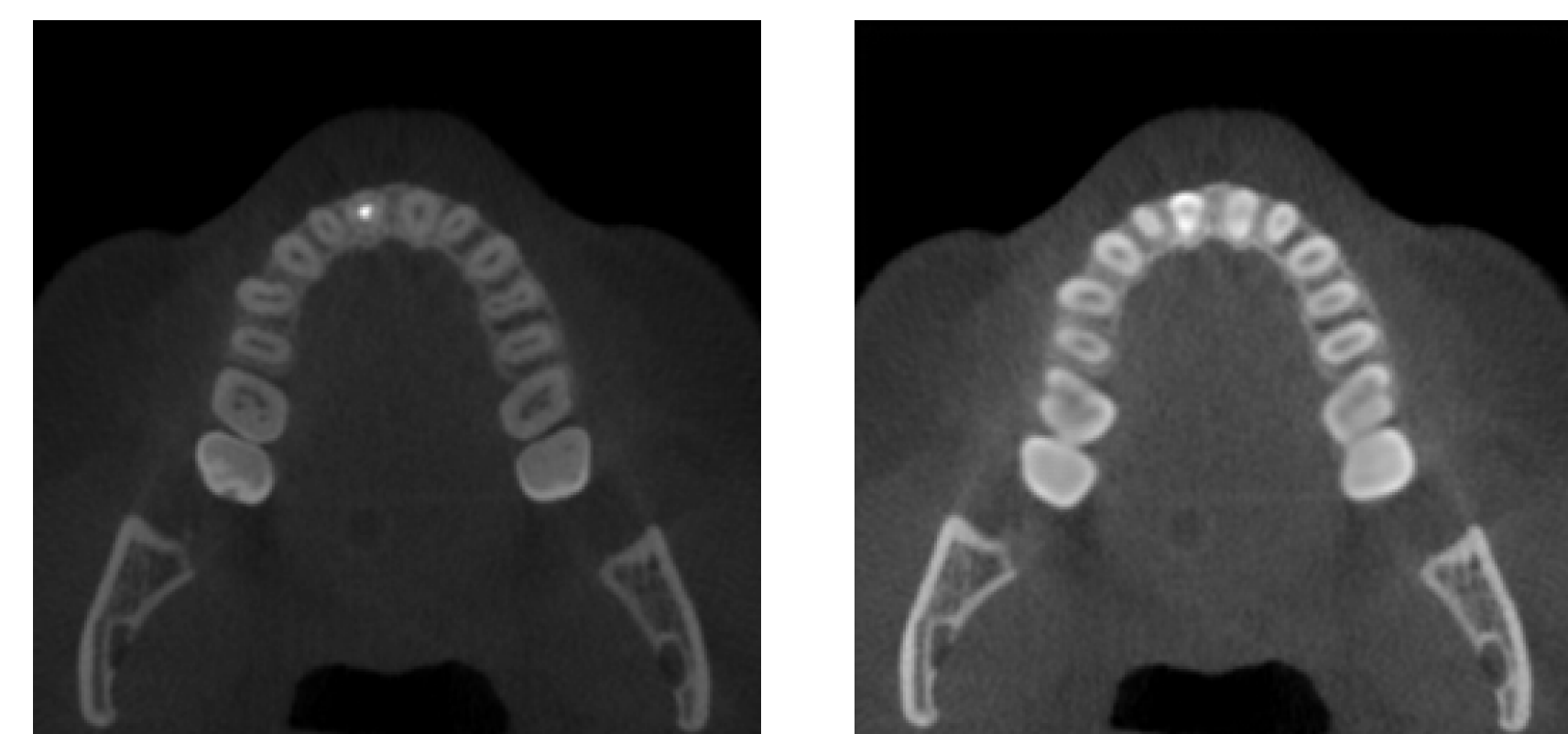
- High similarity between reconstructed/ground-truth teeth.
- Molars and wisdom teeth are less represented in the training set and show greater anatomical variability.

Tooth Removal Synthesis

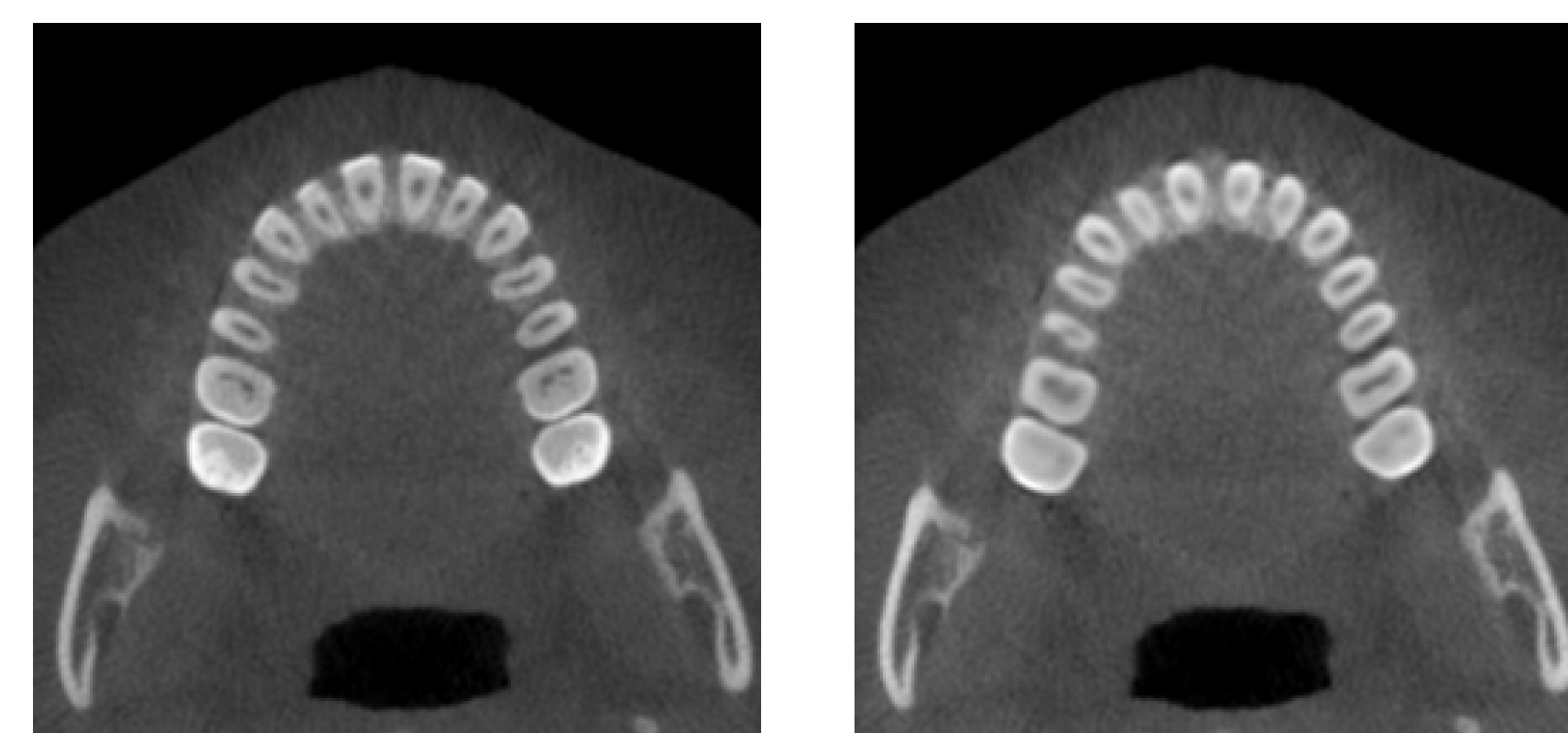
- Demonstrating the model's ability to synthesize realistic scans with specific teeth removed.
- Teeth were removed from test scans with full dentition, following common absence patterns in the training set.
- Generated scans were compared to real samples from the corresponding tooth-absence groups.

Missing Teeth	[1, 16]	[1, 16, 17, 32]	[16, 17, 18, 19]
FID Score	75.20	74.36	80.03

Full Dental Synthesis



Scan 1: Real vs. Synthetic



Scan 2: Real vs. Synthetic

- Demonstrating the model's ability to generate a complete dentition in held-out test scans with no teeth.
- Achieved an average SSIM of 0.91 and PSNR of 18.35 over the inpainted regions of real and synthetic scans.

Conclusion

- The model synthesizes anatomically coherent CBCTs while allowing individual teeth to be toggled on or off.
- Generated volumes preserve surrounding bone and tissue structures, even under varying tooth conditions.
- Visual comparisons confirm strong alignment between conditioning inputs and generated outputs.
- Fine-grained conditioning makes the results interpretable and actionable for dental workflows.

Future Directions

- Extend conditioning beyond presence/absence to include richer properties (restorations, artifacts, root/crown).
- Scale to larger and more diverse patient cohorts to improve generalization.
- Integrate clinically for treatment simulation and preoperative planning.

Acknowledgments

This project is supported by the Pioneer Centre for AI, funded by the Danish National Research Foundation (grant number P1).

References

- Cui et al., A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images, Nature Communications, 2022.
- Said et al., Tooth-Diffusion: Guided 3D CBCT synthesis with fine-grained tooth conditioning, arXiv:2508.14276, 2025.
- <https://github.com/djafar1/tooth-diffusion>.

