

MedSAM-prior-guided coarse-to-fine CBCT synthesis from ZTE MRI

Jun-Seok Yun

Korea Institute of Industrial Technology

Ji Yun Lee

Yonsei University College of Dentistry

Sang-Sun Han

Yonsei University College of Dentistry

Hong-In Won

Korea Institute of Industrial Technology

Sanga Lee

Korea Institute of Industrial Technology

Min Su Kim

Korea Institute of Industrial Technology

Jong Pil Yun

`rebirth@kitech.re.kr`

Korea Institute of Industrial Technology

Chena Lee

Yonsei University College of Dentistry

Article

Keywords:

Posted Date: November 26th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8044755/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

MedSAM-prior-guided coarse-to-fine CBCT synthesis from ZTE MRI

Jun-Seok Yun¹, Ji Yun Lee², Sang-Sun Han^{2,3}, Hong-In Won¹, Sanga Lee^{1,4}, Min Su Kim¹, Jong Pil Yun^{1,4,*}, and Chena Lee^{2,3,*}

¹Korea Institute of Industrial Technology, Manufacturing AI Research Center, Incheon, 21999, Republic of Korea

²Yonsei University College of Dentistry, Department of Oral and Maxillofacial Radiology, Seoul, 03722, Republic of Korea

³Yonsei University, Institute for Innovative in Digital Healthcare, Seoul, 03722, Republic of Korea

⁴Chung-Ang University, Graduate School of Advanced Imaging Science, Multimedia & Film, Seoul, 06974, Republic of Korea

*rebirth@kitech.re.kr, chenalee@yuhs.ac

+these authors contributed equally to this work

ABSTRACT

Cone-beam computed tomography (CBCT) is valuable for assessing fine bony structures in dental and craniofacial imaging but entails ionizing radiation. Zero echo time (ZTE) MRI captures bone-related signals without radiation, yet it falls short of reproducing CBCT-level boundaries and textures. We propose a MedSAM-prior-guided coarse-to-fine framework to synthesize CBCT (sCBCT) from ZTE MRI while preserving low radiation benefits. A frozen MedSAM image encoder provides stable coarse anatomy; targeted low-rank adaptation (LoRA) confined to upper encoder blocks and a skipfusion decoder restore details. Using a single-institution dataset of 26 paired ZTE MRI–CBCT scans (21 training, 5 test), the proposed method reduced RMSE and MAE and increased PSNR and SSIM over a baseline, with consistent gains across all five test subjects. In clinical image assessment, real CBCT scored highest, and among synthetic images our method outperformed the baseline and a broader upper-layer adaptation model. Artifacts and noise were comparable or improved, whereas resolution remained lower than real CBCT. These results support combining a pretrained medical prior for coarse structure with data-aligned restoration of fine detail for sCBCT generation from ZTE MRI.

Introduction

With advances in deep learning, medical image synthesis has gained significant attention, particularly in generating synthetic CT (sCT) from Magnetic resonance imaging (MRI) for applications such as radiotherapy planning and multimodal image analysis^{1–3}. The success of these approaches strongly depends on the availability of high-quality paired MRI and CT data. Several studies have demonstrated the feasibility of deep learning-based MR-to-CT synthesis using convolutional neural networks and generative adversarial networks, showing promising results in both neuroimaging and radiotherapy domains^{4–7}. Recently, transformer-based vision backbones and promptable “foundation” models have rapidly entered medical imaging. For example, MedSAM⁸ adapts a segment anything model (SAM)⁹ image encoder to medical data, providing anatomically robust representations that are being leveraged beyond segmentation.

MRI has established itself as a major diagnostic tool in various medical fields and is increasingly considered a promising alternative to CT due to the absence of ionizing radiation. However, conventional MRI sequences are not well suited for bone imaging, since bone tissue contains tightly bound protons with very short relaxation times, resulting in poor signal detection. To address this limitation, novel sequences such as ultrashort echo time (UTE) and zero echo time (ZTE) have been developed. These so-called “bone MRI” techniques enable visualization of rapidly decaying signals, thereby improving the depiction of osseous structures^{10–13}. Cone-beam computed tomography (CBCT) has become an essential imaging modality in dentistry and craniofacial research due to its ability to provide high spatial resolution with relatively low radiation exposure. Its fine voxel size, often reaching the level of 100 micrometers in clinical devices, allows for detailed evaluation of trabecular bone patterns and small surrounding dental structures, making it particularly valuable for assessing microstructural changes and localized lesions^{14–16}. Given the high spatial resolution of CBCT and the bone-specific imaging capability of ZTE MRI, there is increasing interest in combining these modalities to enable accurate bone structure synthesis. In particular, generating CBCT-like images from ZTE MRI using deep learning methods holds potential for dental and craniofacial imaging applications. Such an approach could reduce radiation exposure while maintaining diagnostic accuracy in bone-related assessments. Therefore, this study was aimed at presenting an advanced model capable of generating high-definition, high-resolution synthetic CBCT

images that closely approximate true CBCT quality, based on the integration of MedSAM priors and LoRA. Ultimately, the development of this model was not merely intended for image enhancement but also aimed to offer a practical, radiation-free pathway for dental diagnosis using MRI.

Methods and material

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of our institution and conducted in accordance with the ethical standards of the Declaration of Helsinki and relevant national regulations. As this was a retrospective study using anonymized data, the requirement for informed consent was waived. All scan data were fully anonymized prior to analysis, and no identifiable patient information was included.

Sample data preparation

For this study, all patients who underwent both CBCT and MRI examination with interval less than 3 months in Yonsei University Dental Hospital were collected. Total 26 CBCT-MRI paired data set was collected and randomly divided into 21 training and 5 test sets. For CBCT data, Alphard 3030 unit (Asahi Roentgen, Kyoto, Japan) was used with following parameters: tube voltage, 90 kVp; tube current, 8 mA; exposure time, 17 s; field of view (FOV), 150 × 150 mm; and voxel size, 0.3 mm. All MRI data were obtained using the 3.0 T Pioneer (GE Healthcare, Waukesha, WI, USA) with a 21-channel head coil. Zero echo time (ZTE) sequences were acquired with the following parameters: TE/TR, 0/785 ms; flip angle, 4°; receiver bandwidth, 31.25 kHz; number of excitations (NEX), 2; FOV, 180 × 180 mm (supra-orbital rim to upper neck region); acquisition matrix, 260 × 260; voxel size, 0.35 mm; slice thickness, 1.0 mm; and scan time, 5 min. The MRI were registered to CBCT images and both scan data were resliced into 0.3 mm axial section using ITK-snap (ver. 4.4.0, www.itksnap.org). In total, 26,624 images (13,312 CBCT images and 13,312 MRI images) were used for this study.

MedSAM-prior-guided coarse-to-fine framework overview

This work employs the MedSAM image encoder with low-rank adaptation (LoRA)¹⁷ for parameter-efficient fine-tuning, followed by a decoder that maps a 3-slice MRI input to a 3-slice CT output at full resolution (512 × 512), as shown in Figure 1. Given $x \in \mathbb{R}^{B \times 3 \times 512 \times 512}$ (three consecutive MRI slices), the encoder extracts intermediate features from blocks {0, 2, 10, 11}, which are used as hierarchical skip connections to the decoder. The network produces $\hat{y} \in \mathbb{R}^{B \times 3 \times 512 \times 512}$, corresponding to three CT slices aligned with the input MRI triplet.

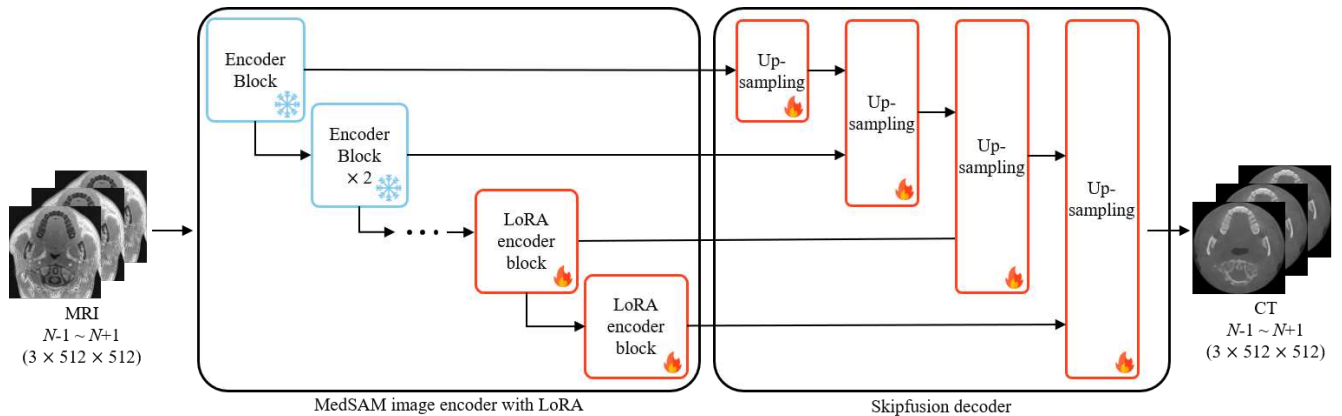


Figure 1. MedSAM-prior-guided coarse-to-fine framework overview. The MedSAM⁸ image encoder is frozen (snowflake) and supplies intermediate representations from selected blocks as hierarchical skip connections. Trainable components (flame icons) comprise the LoRA adapters attached to selected encoder blocks and the decoder.

Image encoder with LoRA

Image encoder.

SAM⁹ is trained on large-scale natural images for promptable segmentation and exhibits strong generalization to object masks. When transferred to medical imaging, however, performance often degrades due to domain shift and the lack of task-specific inductive biases. MedSAM⁸ alleviates this shift by pretraining the SAM image encoder on medical data, achieving strong zero-shot segmentation. Nevertheless, downstream tasks beyond segmentation remain misaligned with the pretraining objective

and therefore require additional adaptation. To address the lack of task-specific inductive biases while avoiding catastrophic forgetting of the medical prior, we adopt a parameter-efficient strategy that freezes the MedSAM encoder and adapts high-level representations.

LoRA adapters.

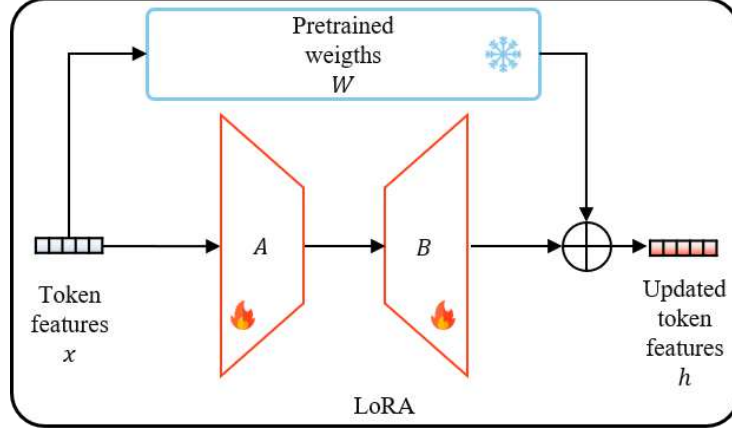


Figure 2. LoRA within a frozen MedSAM encoder block. Token features x flow through a frozen pretrained path in parallel with a trainable low-rank branch formed by a down-projection A and an up-projection B . Their residual sum yields the updated token features h . The snowflake icon denotes components that are not trained, whereas the flame icon denotes trainable components.

We adapt high-level representations while keeping the MedSAM encoder frozen by attaching LoRA¹⁷ modules to selected linear layers within the upper subset of the 12 encoder blocks. We denote by $x \in \mathbb{R}^{d_{in}}$ the token features entering a modified linear layer and by $h \in \mathbb{R}^{d_{out}}$ the updated token features. With the pretrained weights W frozen, the layer produces as follows,

$$h = Wx + \frac{\alpha}{r}B(Ax), \quad A \in \mathbb{R}^{r \times d_{in}}, B \in \mathbb{R}^{d_{out} \times r}, \quad (1)$$

Only the low-rank factors A (down-projection) and B (up-projection) are trainable. As illustrated in Fig. 2, each LoRA-augmented linear layer comprises two parallel paths—a frozen pretrained path and a low-rank adaptation path—whose residual sum yields the updated token features h as in Eq. (1). In practice, we attach LoRA to two linear sublayers within each selected encoder block: (i) the multi-head self-attention query–key–value projection and (ii) the attention output projection; the MLP first feed-forward layer remains unchanged by default. We set $r=8$ and $\alpha=16$, initialize A with Kaiming uniform initialization and B with zeros, and optionally apply dropout on the LoRA branch. This design preserves the medical prior encoded by MedSAM while providing sufficient capacity for task-specific adaptation with minimal additional parameters.

Skipfusion decoder

With the MedSAM encoder frozen, we regard its intermediate representations as a stable coarse anatomical prior and delegate fine-detail restoration to the decoder. To implement this design, we extract features from four encoder blocks (0, 2, 10, and 11). The early blocks (0 and 2) preserve global contours and organ layout, whereas the deeper blocks (10 and 11) are equipped with trainable LoRA adapters and encode sharp boundaries and subtle intensity transitions. This hierarchy reduces redundancy while providing sufficient coverage for CBCT synthesis from ZTE MRI.

Projection and alignment.

We first standardize the channel dimensionality and spatial alignment before fusion. Let $h_k \in \mathbb{R}^{B \times 768 \times H' \times W'}$ denote the encoder feature extracted at block $k \in \{0, 2, 10, 11\}$. Each feature map is projected to 256 channels using a 1×1 convolution, yielding $p_k \in \mathbb{R}^{B \times 256 \times H' \times W'}$. Although all extracted features share the same token grid (H', W') , before each fusion step we bilinearly align p_k to the current decoder stage resolution to maintain geometric consistency.

Top-down decoding with skipfusion.

As shown in Figure 1, decoding proceeds top-down: starting from p_{11} , we successively fuse p_{10} , p_2 , and p_0 , bilinearly aligning each skip to the stage resolution before concatenation. As detailed in Figure 3, each stage upsamples with a lightweight Upsample (UpPS) block and refines the concatenated feature with a residual block, with an additional 1×1 projection on the

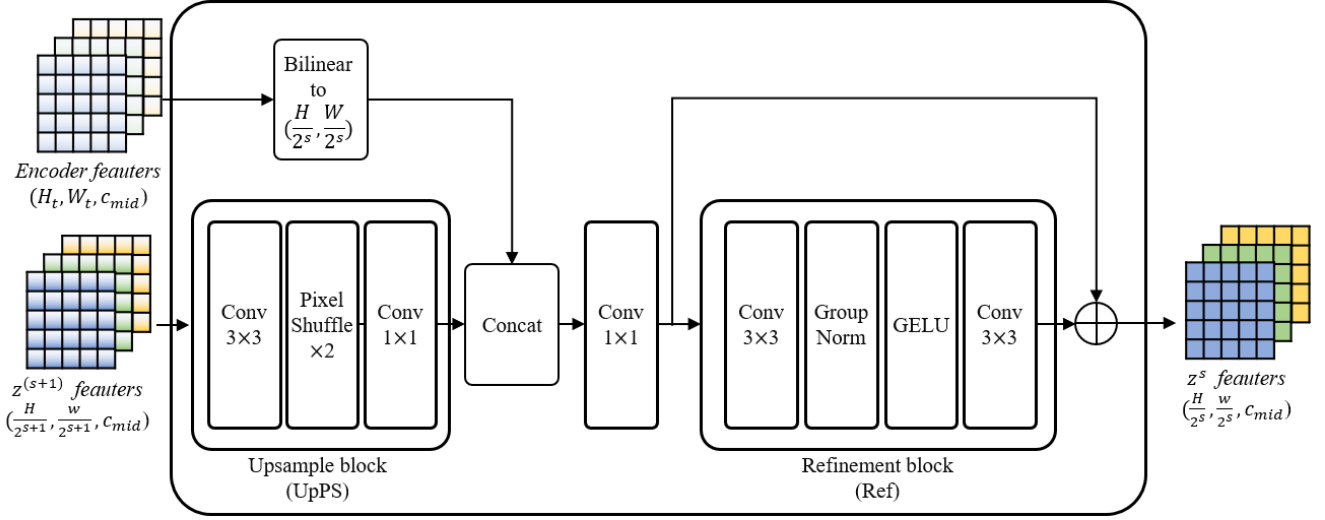


Figure 3. One stage of the Skipfusion decoder.

residual path to match channels. Spatial resolution doubles at every stage in the order $p_{11} \rightarrow p_{10} \rightarrow p_2 \rightarrow p_0$. Finally, a compact head produces the synthetic CT images. We use identity activation at the head; range mapping (e.g., tanh to $[-1, 1]$) is applied within the loss to maintain train–evaluation consistency.

Training configuration

Implementation.

We implemented the system in Python 3.11 using PyTorch 2.3.0 with CUDA 12.1 and cuDNN 8.3.2. Optimization used AdamW with weight decay 10^{-5} and AMSGrad. Each training sample comprised three consecutive ZTE MRI slices as input and the co-registered triplet of CBCT slices as the target, all at 512×512 resolution. We used 26 paired ZTE–CBCT cases in total: 21 subjects for training and 5 for testing. For reporting, metrics were computed per subject and averaged over the five test subjects.

Loss

Given the predicted triplet $\hat{y} \in \mathbb{R}^{B \times 3 \times H \times W}$ and the target y , the loss is defined as a weighted sum of reconstruction, edge-consistency, and structural similarity index (SSIM) losses, with slice weights $c = (1, 2, 1)$ to emphasize the center slice.

The reconstruction loss is the center-weighted sum of slice-wise L_1 errors. The reconstruction loss is defined as follows:

$$\mathcal{L}_{\text{rec}} = \sum_{k=1}^3 c_k \|\hat{y}_k - y_k\|_1, \quad (2)$$

where $\|\cdot\|_1$ denotes the mean per-pixel absolute deviation.

Edge consistency loss is enforced by minimizing a center-weighted, slice-wise Sobel-gradient discrepancy. The edge consistency loss is defined as follows :

$$\mathcal{L}_{\text{grad}} = \sum_{k=1}^3 c_k \|\nabla \hat{y}_k - \nabla y_k\|_1, \quad (3)$$

where ∇ denotes the 3×3 Sobel operators along x and y .

The SSIM loss is evaluated over a central circular region. The radius is set to $r = \min(H, W)/2$ to focus learning on anatomically informative areas and to downweight the large uniform background typical of ZTE–CBCT slices. The SSIM loss is defined as follows:

$$\mathcal{L}_{\text{ssim}} = \sum_{k=1}^3 c_k \left[1 - \text{SSIM}(S(\hat{y}_k), S(y_k)) \right], \quad (4)$$

where $S(\cdot)$ denotes scaling to $[0, 1]$ used for SSIM.

The total loss is a weighted sum:

$$\mathcal{L} = w_{\text{rec}} \mathcal{L}_{\text{rec}} + w_{\text{grad}} \mathcal{L}_{\text{grad}} + w_{\text{ssim}} \mathcal{L}_{\text{ssim}}. \quad (5)$$

In all experiments, we set $w_{\text{rec}} = 1.0$, $w_{\text{grad}} = 0.1$, and $w_{\text{ssim}} = 0.2$. These weights reflect complementary roles: the L_1 term anchors intensities, the gradient term aligns edge maps, and the SSIM term preserves local structure under contrast variation.

Training protocol

The MedSAM image encoder remains frozen during training. LoRA modules are attached to the upper encoder blocks (blocks 10–11). We train for 500 epochs with a per-GPU batch size of 7 and perform validation every 5 epochs. We set the LoRA rank to $r = 8$ and the scaling to $\alpha = 16$. The skipfusion decoder and LoRA parameters are optimized jointly with separate learning rates: 2.5×10^{-4} for the decoder and 5×10^{-5} for LoRA. The decoder uses Group Normalization with 16 groups.

Evaluation

Quantitative assessment metric

The mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and SSIM, were calculated for the sCBCT compared to the ground truth CBCT. The equation for each index was as follows;

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\text{sCBCT}_i - \text{CT}_i| \quad (6)$$

$$\text{PSNR} = 10 \times \log \left(\frac{f_{\max}^2}{\text{RMSE}^2} \right) \quad (7)$$

$$\text{SSIM} = \frac{(2\mu_{\hat{y}_i} \mu_{y_{i,k}} + c_1)(2\sigma_{\hat{y}_i y_{i,k}} + c_2)}{(\mu_{\hat{y}_i}^2 + \mu_{y_{i,k}}^2 + c_1)(\sigma_{\hat{y}_i}^2 + \sigma_{y_{i,k}}^2 + c_2)} \quad (8)$$

A lower MAE indicates better image accuracy, a higher PSNR reflects superior image quality, and an SSIM value closer to 1 denotes greater structural similarity to the reference image.

Clinical assessment metric

Two experts in oral and maxillofacial radiology evaluated both the ground truth CBCT and the synthesized CBCT (sCBCT) generated by the respective deep learning model, using a clinical image evaluation chart for CBCT. The chart consisted of three sections—artifact, noise, and contrast—and employed a three-point scoring system, as shown in Table 1. The evaluation procedure was conducted in blind condition. Intraclass correlation coefficient (ICC, two-way random-effects model, absolute agreement) was calculated to assess interobserver agreement for image quality scores between the two evaluators.

Results

Quantitative assessment

As shown in Table 2, applying LoRA to the top encoder blocks (blocks 10–11) consistently outperformed the non-LoRA baseline across all five subjects and four metrics (RMSE, MAE, PSNR, SSIM), indicating that lightly adapting high-level representations introduces CT-specific high-frequency detail while preserving the encoder prior. In contrast, expanding adaptation to four upper blocks (8–11) did not reduce RMSE and degraded MAE and SSIM, suggesting that broader adaptation can weaken the MedSAM prior and harm structural fidelity.

These trends align with the modeling assumption: a frozen encoder supplies a stable coarse anatomical scaffold, and a small amount of adaptation near the top layers is sufficient to align fine detail to the data distribution. Over-expanding the adaptive region likely increases representation drift and optimization sensitivity, yielding less faithful structures and lower SSIM.

Table 1. Clinical image evaluation chart for CBCT¹⁸.

		Score		
		3-point (Good)	2-point (Moderate)	1-point (Poor)
Artifact	1. Streak artifact 2. Ring artifact 3. Motion artifact			
	Subtotal	9		
Noise	1. Noise on anatomy structure 2. Noise on background			
	Subtotal	6		
Resolution	1. Clearness of enamel, dentin, pulp 2. Clearness of periodontal ligament space and lamina dura 3. Clearness of sinus floor 4. Clearness of trabecular bone pattern 5. Clearness of temporomandibular complex 6. Clearness of mandibular canal			
	Subtotal	18		
	Total	33		

Table 2. Subject-wise results and macro-averages. Best average per metric in **bold**.

ID	Baseline				LoRA (10,11)				LoRA (8,9,10,11)			
	RMSE	MAE	PSNR	SSIM	RMSE	MAE	PSNR	SSIM	RMSE	MAE	PSNR	SSIM
1	0.0969	7.1065	26.41	0.7830	0.0918	6.6518	26.87	0.7928	0.0960	7.1604	26.48	0.7562
4	0.1166	7.4891	24.85	0.7438	0.1153	7.3765	24.96	0.7503	0.1160	7.4810	24.87	0.7285
15	0.0906	5.8968	27.07	0.7947	0.0874	5.6339	27.35	0.8036	0.0905	5.9349	27.04	0.7817
19	0.0978	6.5915	26.34	0.7617	0.0956	6.4535	26.54	0.7688	0.0984	6.6887	26.28	0.7448
23	0.1007	6.6776	26.07	0.7736	0.0982	6.4604	26.30	0.7808	0.1009	6.7514	26.06	0.7560
Avg	0.1005	6.7523	26.14	0.7713	0.0976	6.5152	26.40	0.7792	0.1003	6.8032	26.14	0.7534

Clinical assessment

Clinical image evaluation showed the highest mean score for CBCT (29.5 ± 1.94 points), followed by synthetic CBCT images generated by LoRA (10,11) (24.2 ± 2.05 points), baseline model (23.2 ± 1.48 points), and LoRA (8,9,10,11) (22.8 ± 2.28 points), with only minor differences observed among the three models, as shown in Figure 4. The interobserver agreement for all image quality parameters showed a high level of reliability, with an ICC value of 0.89 (95% CI, 0.86–0.92).

In the categories of artifact, noise, and resolution, the sCBCT showed comparable scores to CBCT in artifact and even higher scores in noise reduction. However, the sCBCT demonstrated lower scores in the resolution category (Figure 5).

When the subitems within the resolution category were analyzed, all models showed the highest scores for the clearness of the sinus floor. LoRA (10,11) model achieved moderate or higher scores in the clearness of enamel, dentin, and pulp (2.2 points), while LoRA (8,9,10,11) also obtained a moderate score in the clearness of trabecular bone pattern (2.4 points). In contrast, the remaining subitems of sCBCT generally showed lower scores compared with CBCT, as shown in Figure 6.

Discussion

This study highlights the effectiveness of the MedSAM–prior–guided coarse-to-fine framework for generating synthetic CBCT images from ZTE MRI sequences. The proposed model successfully produced sCBCT images that closely resembled reference CBCT images, underscoring its potential utility in MRI-based dental diagnosis. In this MRI-to-CBCT synthesis approach, a frozen MedSAM encoder provides a stable anatomical prior, limited upper-layer LoRA adaptation (blocks 9–11) refines high-level representations, and a skip-fusion decoder restores fine structural details. Consequently, the LoRA (10,11) model demonstrated improved performance metrics—MAE, PSNR, and SSIM—compared with those reported in a previous study¹. Applying LoRA to blocks 10–11 consistently improved RMSE, MAE, PSNR, and SSIM over the non-adapted baseline across all five subjects, whereas broadening adaptation to blocks 8–11 did not reduce RMSE and degraded MAE and SSIM, indicating that small, targeted adaptation injects CT-specific high-frequency cues while preserving the prior. In blinded clinical scoring, the LoRA (10–11) variant achieved the highest overall quality among synthetic CBCT images, with artifacts comparable to CBCT and lower noise, though perceived resolution remained inferior to real CBCT. Future work will incorporate 3D

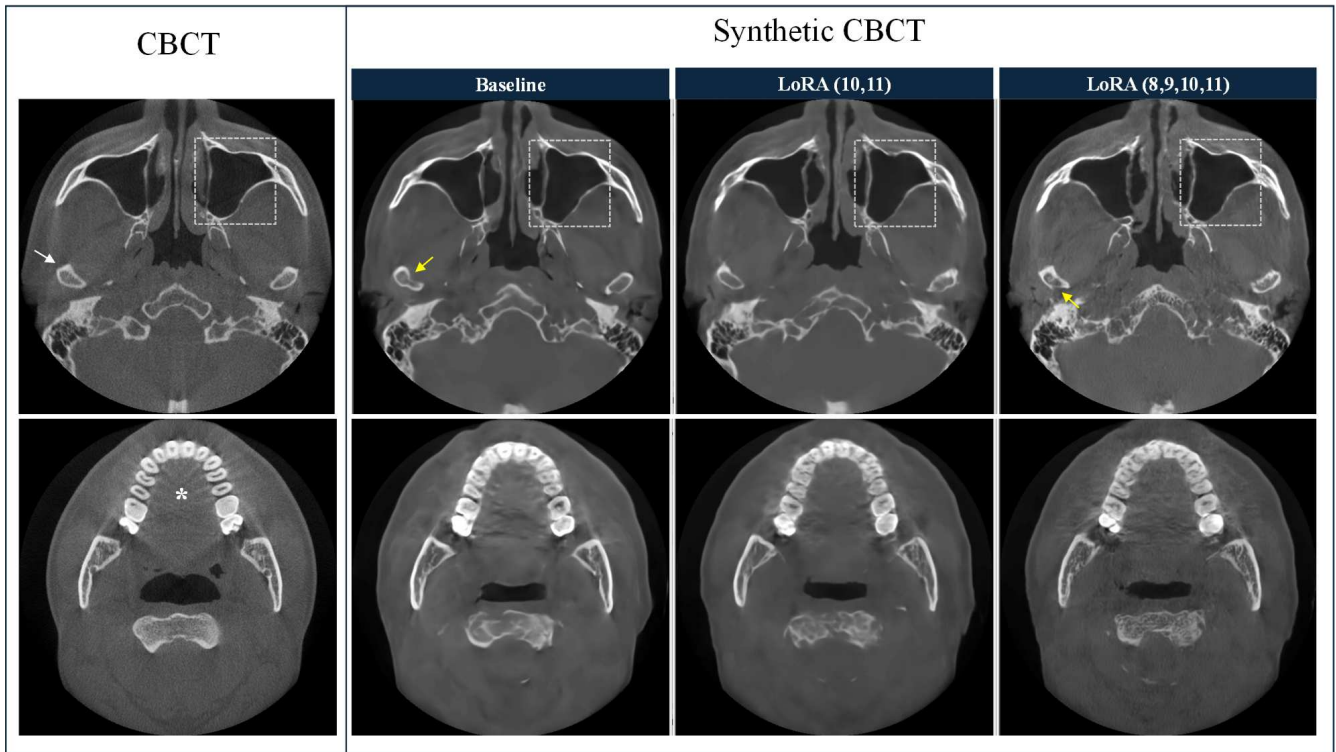


Figure 4. Axial images of the maxilla and mandible from CBCT and synthetic CBCT. The maxillary sinus (dotted box) is clearly visualized in the synthetic CBCT images, while the mandibular condyle (white arrow) shows a defect in the baseline and LoRA (8, 9, 10, 11) models (yellow arrow). A streak artifact observed in the CBCT (asterisk) is also present in the baseline and LoRA (8, 9, 10, 11) models but is absent in LoRA (10, 11).

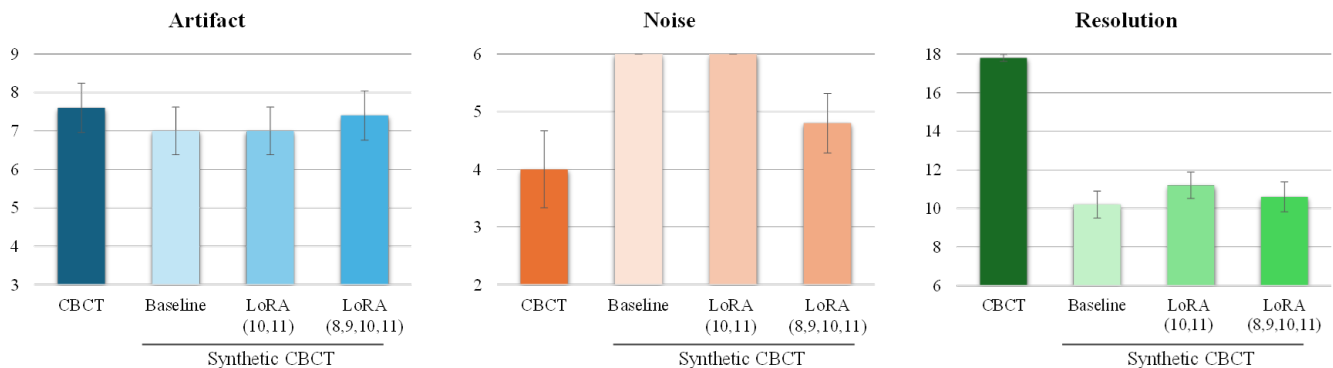


Figure 5. Comparison of CBCT and sCBCT scores for artifact, noise, and resolution. The sCBCT shows comparable scores to CBCT in the artifact category and demonstrates superior performance in noise reduction. However, sCBCT exhibits lower scores in the resolution assessment.

	CBCT	Synthetic CBCT		
		Baseline	LoRA (10,11)	LoRA (8,9,10,11)
Streak artifact	1.8	2.2	2.2	1.8
Ring artifact	3	3	3	3
Motion artifact	2.8	1.8	1.8	2.6
Noise on Anatomy structure	2	3	3	2.8
Noise on background	2	3	3	2
Cleanness of enamel, dentin, pulp	2.8	1.8	2.2	1.8
Cleanness of PDL space and lamina dura	3	1	1	1
Cleanness of sinus floor	3	2.8	2.8	2.8
Cleanness of trabecular bone pattern	3	1.6	1.6	2.4
Cleanness of TMJ complex	3	2	2	1.6
Cleanness of mandibular canal	3	1	1.6	1

Figure 6. Comparison of CBCT and sCBCT based on detailed items in the clinical evaluation chart. The sCBCT demonstrated improved scores for streak artifact, anatomical noise, and background. However, except for the sinus floor, sCBCT showed markedly lower scores in the clarity of anatomical structures.

or cross-slice temporal context, and prior-preserving super-resolution and artifact-mitigation modules to further narrow the resolution gap.

While existing literature predominantly emphasizes the use of conventional sequence, T1-weighted and T2-weighted images, for CT synthesis, limited research has explored the feasibility of using ZTE sequence¹⁹. In this study, we found in the performance of ZTE sequence, suggesting that modalities serve as effective inputs for sCBCT generation. This finding aligns with previous studies and implies flexibility in modality selection based on clinical availability.

As highlighted in multiple previous studies, synthesized images frequently showed blurring of normal anatomical boundaries^{1,20}). The present study aimed to mitigate this limitation and improve structural clarity. Overall, the synthesized CBCT images showed enhanced sharpness, with improved scores for streak artifacts, anatomical noise, and background. Larger anatomical structures, such as the sinus floor, received intermediate visibility scores. However, the clarity of fine anatomical details, including teeth, small surrounding structures, and fine bone patterns, remained lower. Blurred margins were still noticeable in certain regions, particularly along complex anatomical boundaries.

The baseline and LoRA (10,11) models produced overall well-balanced and structurally coherent images. Notably, the severe noise and artifacts commonly observed in conventional CBCT were not apparent, which may reflect the advantages of MRI-derived image characteristics contributing to improved image quality. In contrast, the LoRA (8,9,10,11) model exhibited relatively higher levels of noise and artifacts, comparable to those of conventional CBCT, and failed to fully reconstruct small and fine anatomical structures. Nevertheless, the LoRA (8,9,10,11) model generated images with less blurring and clearer bone margins, suggesting its potential applicability for clinical evaluation, particularly in dental diagnosis and treatment planning. These findings underscore the necessity for continued research into image pre-processing techniques aimed at further enhancing sharpness and clarity. Specifically, methods that minimize blurring while preserving fine anatomical details will be crucial for improving the quality of MRI-to-sCBCT images in clinical practice.

Conclusion

This study demonstrated that the MedSAM–prior–guided coarse-to-fine framework effectively generates high-quality synthetic CBCT images from ZTE MRI sequences. By leveraging a stable anatomical prior and targeted LoRA adaptation, the model produced sCBCT images with improved sharpness, reduced artifacts, and enhanced structural coherence. The results confirmed the framework’s capability to reproduce the fine detail and clarity characteristic of conventional CBCT, highlighting its potential for clinical translation. Although further optimization is required to fully capture small anatomical structures, the current findings suggest that continued refinement could enable MRI-based CBCT synthesis to serve as a practical diagnostic alternative. Future work will focus on enhancing resolution and preserving fine anatomical fidelity to achieve full clinical applicability.

Data availability

The data generated and analyzed during the current study are not publicly available due to privacy laws and policies in Korea, but are available from the corresponding author on reasonable request.

Author Contributions Statement

J.S.Y., H.I.W., S.L., M.K., and J.P.Y. designed the study and wrote/edited the manuscript. J.Y.L., S.S.H., and C.L. curated data and performed experiments, quantitative analysis, and blinded clinical evaluation. All authors reviewed the manuscript.

Acknowledgements

This research was supported by the Korea Institute of Industrial Technology (KITECH) under grant EH-25-0002, and by the Korea Planning & Evaluation Institute of Industrial Technology (KEIT) under grant RS-2024-00410959. No conflicts of interest are declared.

Funding

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government(MSIT) (RS-2023-00245224).

References

1. Choi, H. *et al.* Deep learning synthesis of cone-beam computed tomography from zero echo time magnetic resonance imaging. *Sci. Reports* **13**, 6031 (2023).
2. Sherwani, M. K. & Gopalakrishnan, S. A systematic literature review: deep learning techniques for synthetic medical image generation and their applications in radiotherapy. *Front. Radiol.* **4**, 1385742 (2024).
3. Bahloul, M. A. *et al.* Advancements in synthetic ct generation from mri: A review of techniques, and trends in radiation therapy planning. *J. Appl. Clin. Med. Phys.* **25**, e14499 (2024).
4. Han, X. Mr-based synthetic ct generation using a deep convolutional neural network method. *Med. physics* **44**, 1408–1419 (2017).
5. Nie, D. *et al.* Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomed. Eng.* **65**, 2720–2730 (2018).
6. Maspero, M. *et al.* Dose evaluation of fast synthetic-ct generation using a generative adversarial network for general pelvis mr-only radiotherapy. *Phys. Medicine & Biol.* **63**, 185001 (2018).
7. Dinkla, A. M. *et al.* Mr-only brain radiation therapy: dosimetric evaluation of synthetic cts generated by a dilated convolutional neural network. *Int. J. Radiat. Oncol. Biol. Phys.* **102**, 801–812 (2018).
8. Ma, J. *et al.* Segment anything in medical images. *Nat. Commun.* **15**, 654 (2024).
9. Kirillov, A. *et al.* Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4015–4026 (2023).
10. Gatehouse, P. & Bydder, G. Magnetic resonance imaging of short t2 components in tissue. *Clin. radiology* **58**, 1–19 (2003).
11. Weiger, M., Pruessmann, K. P. & Hennel, F. Mri with zero echo time: hard versus sweep pulse excitation. *Magn. resonance medicine* **66**, 379–389 (2011).
12. Grodzki, D. M., Jakob, P. M. & Heismann, B. Ultrashort echo time imaging using pointwise encoding time reduction with radial acquisition (petra). *Magn. resonance medicine* **67**, 510–518 (2012).
13. Wiesinger, F. *et al.* Zero te mr bone imaging in the head. *Magn. resonance medicine* **75**, 107–114 (2016).
14. Scarfe, W. C. & Farman, A. G. What is cone-beam ct and how does it work? *Dental Clin. North Am.* **52**, 707–730 (2008).
15. Pauwels, R. *et al.* Effective dose range for dental cone beam computed tomography scanners. *Eur. journal radiology* **81**, 267–271 (2012).
16. Tayman, M. A., Kamburoğlu, K., Ocak, M. & Özen, D. Effect of different voxel sizes on the accuracy of cbct measurements of trabecular bone microstructure: A comparative micro-ct study. *Imaging Sci. Dent.* **52**, 171 (2022).
17. Hu, E. J. *et al.* Lora: Low-rank adaptation of large language models. *ICLR* **1**, 3 (2022).
18. Ryu, K. *et al.* Multi-planar 2.5 d u-net for image quality enhancement of dental cone-beam ct. *Plos one* **18**, e0285608 (2023).
19. Kermani, M. Z. *et al.* Supervised versus unsupervised gan for pseudo-ct synthesis in brain mr-guided radiotherapy. *Phys. Eng. Sci. Medicine* 1–14 (2025).

20. Yuan, N. *et al.* Convolutional neural network enhancement of fast-scan low-dose cone-beam ct images for head and neck radiotherapy. *Phys. Medicine & Biol.* **65**, 035003 (2020).