

# GOPEN ACCESS

**Citation:** Ryu K, Lee C, Han Y, Pang S, Kim YH, Choi C, et al. (2023) Multi-planar 2.5D U-Net for image quality enhancement of dental cone-beam CT. PLoS ONE 18(5): e0285608. https://doi.org/ 10.1371/journal.pone.0285608

**Editor:** Minsoo Chun, Chung-Ang University Gwangmyeong Hospital, REPUBLIC OF KOREA

Received: January 17, 2023

Accepted: April 26, 2023

Published: May 11, 2023

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the <u>Creative</u> Commons CC0 public domain dedication.

Data Availability Statement: Data cannot be shared publicly because of patient privacy. Data are available from the Yonsei University Dental Hospital Ethics Committee (contact via IRB@yuhs. ac) for researchers who meet the criteria for access to confidential data.

**Funding:** Sang-Sun Han was funded by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2022R1A2B5B01002517) The funders had no role in study design, data collection and analysis,

#### **RESEARCH ARTICLE**

# Multi-planar 2.5D U-Net for image quality enhancement of dental cone-beam CT

Kanghyun Ryu<sup>1</sup>°, Chena Lee<sub>0</sub><sup>2°</sup>, Yoseob Han<sup>3,4</sup>, Subeen Pang<sup>5</sup>, Young Hyun Kim<sup>6</sup>, Chanyeol Choi<sup>7</sup>, Ikbeom Jang<sup>8,9‡</sup>\*, Sang-Sun Han<sub>0</sub><sup>2‡</sup>\*

Artificial Intelligence and Robotics Institute, Korea Institute of Science and Technology, Seoul, South Korea, 2 Department of Oral and Maxillofacial Radiology, Yonsei University College of Dentistry, Seoul, South Korea, 3 College of Information Technology, Department of Electronic Engineering, IT Convergence Major, Soongsil University, Seoul, Korea, 4 Department of Radiology, Harvard Medical School, Boston, MA, United States of America, 5 Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States of America, 6 Department of R&D Performance Evaluation, Korea Health Industry Development Institute (KHIDI), Cheongju, South Korea, 7 Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United States of America,
 8 Division of Computer Engineering, Hankuk University of Foreign Studies, Yongin, South Korea,
 9 Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America

• These authors contributed equally to this work.

‡ IJ and SSH also contributed equally to this work.

\* SSHAN@yuhs.ac.kr (SSH); ijang@hufs.ac.kr (IJ)

# Abstract

Cone-beam computed tomography (CBCT) can provide 3D images of a targeted area with the advantage of lower dosage than multidetector computed tomography (MDCT; also simply referred to as CT). However, in CBCT, due to the cone-shaped geometry of the X-ray source and the absence of post-patient collimation, the presence of more scattering rays deteriorates the image quality compared with MDCT. CBCT is commonly used in dental clinics, and image artifacts negatively affect the radiology workflow and diagnosis. Studies have attempted to eliminate image artifacts and improve image quality; however, a vast majority of that work sacrificed structural details of the image. The current study presents a novel approach to reduce image artifacts while preserving details and sharpness in the original CBCT image for precise diagnostic purposes. We used MDCT images as reference highquality images. Pairs of CBCT and MDCT scans were collected retrospectively at a university hospital, followed by co-registration between the CBCT and MDCT images. A contextual loss-optimized multi-planar 2.5D U-Net was proposed. Images corrected using this model were evaluated quantitatively and qualitatively by dental clinicians. The quantitative metrics showed superior quality in output images compared to the original CBCT. In the qualitative evaluation, the generated images presented significantly higher scores for artifacts, noise, resolution, and overall image quality. This proposed novel approach for noise and artifact reduction with sharpness preservation in CBCT suggests the potential of this method for diagnostic imaging.

decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

# Introduction

Cone-beam computed tomography (CBCT) is widely used in the dental field for purposes ranging from disease diagnosis to preoperative simulation and surgical guide construction [1, 2]. CBCT can provide three-dimensional (3D) images of the targeted area with the advantage of lower dosage than multidetector computed tomography (MDCT) [3, 4]. However, in CBCT, due to the cone-shaped geometry of the X-ray source and the absence of post-patient collimation, more scattering rays deteriorate the image quality than is the case with MDCT. The low radiation dose also causes under-sampling of the signal, which results in more noise and artifacts [4, 5].

Deep learning (DL) based approaches for CBCT image improvement have recently emerged as a possible viable solution in clinics. While prior studies have demonstrated the possibility of improving CBCT through the use of DL by reducing artifacts and noise and by standardizing pixel intensity [6–12], DL inference is still limited by obscured anatomic fine details and blurred edges in the image [7, 13]. These factors may limit the diagnostic utility of CBCT for intricate features such as the teeth, alveolar bone pattern, sinuses, or the temporomandibular joint (TMJ) complex. Thus, developing a network that minimizes artifacts and noise while maintaining fine anatomic details is necessary for potential diagnostic usage.

Automated methods for 3D image synthesis, quality enhancement, and segmentation tasks are becoming increasingly important in the biomedical field. Various recent network architectures which demonstrated good performance in tooth segmentation (e.g., 3.5D U-Net) could also be explored for cone-beam artifact correction [14]. A common approach is to expand the U-Net structure to work with volumetric data by utilizing 3D convolutions [15]. However, this approach requires a huge memory and long training time. Additionally, medical images are often anisotropic and inconsistent in volume resolution across participants. This problem can sometimes be resolved by resampling the volumes to be isotropic [16, 17]. However, these methods only interpolate the data and often result in blurry images. DL methods therefore are often applied to 2D slice images, but these slice images do not contain information on the full 3D data, which makes the tasks challenging. One solution for incorporating information from the 3D surroundings is to train the model on orthogonal patches extracted from axial, sagittal and coronal views [18–20]; this approach is referred to as "multi-planar" in this paper. For instance, a multi-planar U-Net would predict a value for an intersecting voxel from three orthogonal slice images. Another remedy is to expand a 2D U-Net model to accept 3-channel data and supply the model with a slice and its adjacent slices (e.g., slices #N-1, #N, and #N+1 instead of slice #N) [7, 14]; the technique is referred to as "2.5D" in this paper. This method is particularly useful in our study when the input and output images are slightly misaligned.

Previous DL-based studies by Yuan et al. [7] and Chen et al. [8] presented a two-dimensional U-Net with L1 loss and focused on correcting the pixel values as Hounsfield units (HUs) for head and neck tumor localization to plan radiation therapy [6–8]. However, training a U-Net with a conventional loss function has limitations in that it computes pixel-wise error while assuming a precise alignment between the input (CBCT) and output images (MDCT). The problem is that their perfect alignment is quite difficult in the real world because MDCT scans are usually performed lying down, while CBCT scans are performed sitting or standing, which leads to differences in topology and the relative positions of the maxillary-mandibular and facial soft tissue between the two images. A slight misalignment can easily propagate to exacerbate blurring. In particular, pixel-wise L1 loss is known to be insufficient for considering critical local features such as minute details, resulting in blurry images [21].

We introduced and applied several advanced techniques to resolve the aforementioned issues. First, we built an automated data preparation process involving intensive image

registration and erroneous background masking, which aimed to minimize errors in image registration. Second, we leveraged the two recent techniques introduced above—namely, the multi-planar and 2.5D methods—to efficiently train the model with limited hardware resources while allowing slight misalignments between the images. Finally, we introduced a novel loss function by incorporating a recently-introduced contextual loss [21] into the conventional L1 loss to enhance the image quality and correct for artifacts while minimizing the effect of potential large misalignments.

The images corrected by the network were evaluated using three commonly used image quality metrics: the mean absolute error (MAE), normalized root-mean-square deviation (NRMSE), and structural similarity index (SSIM). The MAE ( $\ell$ 1-norm) of pixel values has been frequently used to measure image quality for a long time. The MAE is based on the mean squared error (MSE;  $\ell$ 2-norm) and has been shown to outperform the MSE in image restoration tasks [22]. The NRMSE relates the RMSE to the observed range of the variable through normalization. The NRMSE is more sensitive to large errors than the MAE because the errors are squared before they are averaged. Although these MSE-based metrics are the simplest of all fidelity metrics, they have been criticized for their limited correlations with human perceptions of image quality [23]. The SSIM is a "perceptual" quality measure that considers image degradation as the perceived change in structural information, while also incorporating luminance masking and contrast masking phenomena [24]. Various aspects of the corrected images were also visually evaluated by expert clinicians.

We found that the proposed Contextual loss-Optimized Multi-Planar 2.5D U-Net (COM-PUNet) significantly improved CBCT image quality, as validated by the three quantitative metrics described above. Clinical experts found that the resulting CBCT images showed fewer artifacts and less noise, without undue sacrifice of the sharpness and structural details in the original image, consistent with dental diagnostic purposes. Therefore, these novel procedures using deep neural networks may have applications in enhancing diagnostics in clinical practice.

# Material and methods

We propose an approach to correct image artifacts in clinical CBCT images by using the corresponding MDCT scans as ground-truth high-quality images. Therefore, paired images of CBCT and MDCT from the same subjects were used in this study. The scanned images of anthropomorphic phantoms and the patient CBCT and MDCT data were both used for DL model development. This study was approved by the institutional review board (IRB) of Yonsei University Dental Hospital (IRB no. 2-2022-0024). The requirement for informed consent was waived due to the retrospective nature of this study, and all patient data were anonymized. The overall study workflow is described in Fig 1.

## Data collection

**Clinical patient data.** A clinical sample was retrospectively collected for training and testing the proposed model. In total, 30 CBCT-MDCT image pairs were retrieved from patients who underwent both CBCT and MDCT examinations at Yonsei University Dental Hospital between August 1, 2018 and August 31, 2021. MDCT and CBCT scans were taken for various purposes by different clinical departments, such as MDCT for sinusitis evaluation and CBCT for dental implant surgery. All image pairs were taken within a 2-month of interval, and image pairs with significant changes due to major surgery were excluded. The mean age of the subjects was 42.3 (± 19.0) years. Nine male and 21 female patients were included. All patients had



more than one metallic restoration, including crowns, implant fixtures, inlays, bridges, and fixed retainers. No patients had an orthodontic bracket attached.

We used CBCT data obtained from two scanners manufactured by different vendors to improve the generalizability of the trained model. A set of 10 images was obtained using Rayscan Alpha Plus (Ray Co., Hwaseong, Korea) and a set of 20 images using Alphard 3030 (Asahi Roentgen Ind. Co. Ltd., Kyoto, Japan). All MDCT data were acquired with Optima CT520 (GE Healthcare, Chicago, IL, USA). The CBCT and MDCT images were taken with patients in different postures. The MDCT scans were taken with the patients lying down, whereas the CBCT scans were taken with the patients standing or sitting. Patients also often held a biteblock in the mouth during the CBCT examination. Therefore, the topology and relative position of the maxillary-mandibular and facial soft tissue were different between the two images, making image registration challenging (described in "Image Registration" in the Methods section). The scanner model and imaging parameters are described in Table 1. The clinical patient data were randomly split into training (40%; n = 12) and test (60%; n = 18) datasets.

Anthropomorphic phantom data. There were two obstacles in our method of correcting clinical CBCT images by using the corresponding MDCT as the ground truth high-quality image. First, it was difficult to perfectly register the two volumes due to differences in topology and the relative position of the maxillary-mandibular and facial soft tissue. Second, not every clinical MDCT image showed perfect image quality, often due to motion artifacts. We prospectively collected a phantom sample to alleviate these issues. Six pairs of CBCT-MDCT data were prepared with an anthropomorphic head phantom. CBCT images were acquired with Rayscan Alpha Plus (Ray Co. Ltd, Hwaseong, South Korea), and scanning was performed six times with different head positions by rotating 60° on the axial plane. For MDCT images, the radiation dose and pixel size were adjusted to obtain high-resolution images different from routine clinical images. The unit used was Sensation 64 (Siemens Medical Solutions, For-chheim, Germany). All phantom data were used as the training set.

**Data preparation.** We built an automated data preparation process involving intensive image registration and erroneous background masking, which aimed to minimize errors in image registration. As illustrated in Table 1, each CBCT and MDCT pair had a different

	Mı	Multi-detector computed tomography			
	Phantom data	Pati	Patient data		
Number of data sets	6	30	30		
Unit model	Sensation 64	Optima CT520			
Tube voltage	120 kVp	120 kVp			
Tube current	480 mAs	120 mAs			
Pitch	0.8	0.9	0.9		
Pixel size	0.37 mm	0.45 mm	0.45 mm		
Slice thickness	0.60 mm	0.63 mm	0.63 mm		
Display field of view	200 mm	200 mm	200 mm		
CTDI <sub>vol</sub>	46.92 mGy	14.44 mGy	14.44 mGy		
	C	Cone-beam computed tomography			
	Phantom data	Patient data			
Number of data sets	6	10	20		
Unit model	Rayscan Alpha Plus	Rayscan Alpha Plus	Alphard 3030		
Tube voltage	80 kVp	80 kVp	80 kVp		
Tube current	12 mA	12 mA	8 mA		
Scan time	14 seconds	14 seconds	17 seconds		
Voxel size	0.4 mm	0.4 mm	0.6 mm		
Field of view	160 × 100 mm	160 × 100 mm	$154 \times 154 \text{ mm}$		
DAP	1863 mGycm <sup>2</sup>	1863 mGycm <sup>2</sup>	2743 mGycm <sup>2</sup>		

	Table 1.	Exposure condi	itions for imag	e data acquisition
--	----------	----------------	-----------------	--------------------

CTDI<sub>vol</sub>, computed tomography dose index; DAP, dose area product

https://doi.org/10.1371/journal.pone.0285608.t001

resolution (pixel/voxel size and slice thickness), field of view (FOV), and orientation. Thus, an alignment (or registration) method is required to prepare the dataset. For this process, a 3D array of volume data was established from the DICOM images of CBCT and MDCT.

**Image adjustment.** The image orientation of MDCT was adjusted in terms of the rightanterior-superior orientation to match the CBCT scans. Next the MDCT image volumes were resampled to match the voxel size of the corresponding CBCT image data. For voxel intensity alignment, the background of CBCT was masked using the DIPY toolbox's median *otsu* technique (Fig 1) [25]. The voxel intensities of CBCT were then matched with those of MDCT, converting into Hounsfield units (HUs). For both MDCT and CBCT, an intensity threshold of -1200 to 3071 HU was used, and data outside of this range were discarded.

**Image registration.** Two techniques were used to register the MDCT volume image to the matching CBCT (Fig 2). Because the FOV of MDCT was larger than that of the CBCT data, the additional region was eliminated from the MDCT data for effective registration. In detail, the overlapping block (or ROI in Fig 2) between the MDCT and CBCT scans was obtained by the following process. 1) The CBCT volumetric image was registered to the MDCT scan through robust and fast rigid-transform-based registration. 2) The ROI of the MDCT volume was determined by identifying the overlapping block (cuboid) between the MDCT and CBCT scans. 3) Any ROI that did not correspond to this overlapping block was then masked, as demonstrated in Fig 2.

The alignment of MDCT to CBCT was then performed using affine registration (rotation, translation, and scaling) and fine-tuned using deformable registration. The Mattes mutual information metric [26] was used as the metric. Throughout this process, registration procedures from advanced normalization tools were used [27]. Finally, the CBCT and MDCT volume image pairings were prepared with the same orientation, resolution, and FOV.



Fig 2. The data preparation process. CBCT images were generated by masking out the background signal noise (a). MDCT images were resampled to match the resolution of the CBCT images and then prepared using region-of-interest (ROI) masking and a two-step registration process utilizing affine and non-rigid registration to match the CBCT images (b). The final set of CBCT and MDCT image pairs was used for training and testing (c). CBCT, cone-beam computed tomography; MDCT, multidetector computed tomography.

https://doi.org/10.1371/journal.pone.0285608.g002

# Deep learning model

We developed a novel model, Contextual loss-Optimized Multi-Planar 2.5D U-Net (COMPU-Net), by applying advanced techniques to resolve the issues of 2D and 3D U-Nets. We leveraged the recent 2.5D and multi-planar techniques to efficiently train the model with limited hardware resources while allowing a slight misalignment between the images. On top of that, a novel loss function was introduced by incorporating a recently-introduced contextual loss function along with conventional L1 loss to enhance the image quality and correct for artifacts while minimizing the effect of potential large misalignments. The network sliced each of the CBCT and MDCT volumes into multiple slices in three orthogonal directions: axial, coronal, and sagittal. The network was then trained using image slices in all three directions, referred to as multi-planar U-Net (Fig 3A). Each U-Net received three consecutive image slices (e.g., slices #N-1, #N, and #N+1) at once (reflecting the term "2.5D") and generated a single corrected slice corresponding to the central image slice (e.g., slice #N). Then, during the inference phase, the outputs from the three different directions of the image were averaged to obtain a more robust estimate. The U-Net contained four encoder blocks and four decoder blocks (Fig 3B), and we used the pretrained ResNet34 [28, 29] (on ImageNet) as an encoder of each U-Net. In the decoder block, an attention gating method was employed for effective training [30]. For the network training, we introduced an additional loss function (contextual loss) to the conventional L1 loss function. Contextual loss is a recently introduced technique that uses the similarity between features rather than the pixel-wise distance function and is known to be effective for misaligned input-ground truth pairs [31]. This was thought to be advantageous



**Fig 3.** Schematic of COMPUNet (multi-planar 2.5D U-Net), which comprises three single-planar 2.5D U-Nets. (a) The architecture of a single-planar 2.5D U-Net with a ResNet-34 encoder and (b) attention modules.  $x^l$ : input features,  $\alpha$ : attention coefficients, g: gating signal collected from a coarser scale.

for our study since misregistration errors between CBCT and MDCT images can result in blurring of the fine details of teeth, tooth-supporting structures, and trabecular bone.

**Training strategy.** The model proposed in this study was trained over 100 epochs with the Adam optimizer [32] and a learning rate of 0.001 to minimize the proposed loss (i.e., a combination of contextual and L1 loss). We randomly applied image augmentation to each image during neural network training to virtually increase the number of training examples. In each training iteration, we performed three types of augmentation.

- Random flips on the horizontal and vertical axes.
- Random rotations between -90° and 90°.
- Cropping at a random location to 240×240 size.

PyTorch 1.8.0 [33] was used for implementation and training. The training dataset included 18 pairs of CBCT–MDCT data, including 6 phantoms and 12 patients. To compare the model performance, two additional models (Model 1 and Model 2) were trained. Those models were introduced in previous studies [7], and the model in this study can be described as follows:

- Model 1: A single-planar 2.5D U-Net with L1 loss
- Model 2: A multi-planar 2.5D U-Net with L1 loss
- COMPUNet: A multi-planar 2.5D U-Net with L1 and contextual losses

**Performance evaluation.** *Quantitative evaluation*. The image quality of each network's output was evaluated using quantitative image assessment indices. The assessment was conducted on 18 test data sets using three image quality metrics: NRMSE, SSIM, and MAE. For each volume in the test-set  $y = \{y_1, y_2, ..., y_N\}$  and  $\hat{y} = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}_N\}$ , the three metrics are defined as follows:

$$NRMSE = \frac{\sqrt{\sum_{i=1}^{N} (\hat{y}_i - y_{i,k})^2}}{\hat{y}_i^{max} - \hat{y}_i^{min}}$$
(1)

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

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_{i,k}|$$
(3)

where *y* and  $\hat{y}$  are MDCT and DL outputs, respectively. *N* denotes the number of paired data and *k* denotes the DL models used.  $\mu_y$  and  $\mu_{\hat{y}}$  are the averages of *y* and  $\hat{y}$ , respectively,  $\sigma_y^2$  and  $\sigma_{\hat{y}}^2$  are the variances of *y* and  $\hat{y}$ , respectively, and  $\sigma_{\hat{y}y}$  is the covariance of *y* and  $\hat{y}$ .  $c_1 = (k_1L)^2$ and  $c_2 = (k_2L)^2$  are used to stabilize division with a weak denominator, where L is the dynamic range of the pixel values and  $k_1 = 0.01$  and  $k_2 = 0.03$  are used. A smaller NRMSE value, a smaller MAE value, and a larger SSIM value indicate superior image quality.

*Qualitative validation*. For the 18 original CBCT (oCBCT) and corresponding predicted (or corrected) CBCT (pCBCT) images, the evaluation was conducted using a PACS viewer (Zetta, Tae-young, South Korea) in random order. Two radiologists made an evaluation using a modified version of the clinical CBCT image evaluation chart provided by the Korean Academy of Oral and Maxillofacial Radiology (Table 2) [13]. The evaluation procedure was conducted individually in blind condition. The clinical CBCT image evaluation chart was composed of

	Severe (1)	Moderate (2)	Few (3)
[Artifacts]			
1. Streak artifact			
1–1. Mandibular ramus area			
1–2. Zygomatic process area			
2. Ring artifact			
3. Blurring edge and shading			
3–1. Mandibular angle area			
[Noise]			
4. Noise on mouth floor			
5. Noise on maxillary sinus			
	Unclear	Moderately defined (2)	Clear (3)
	(1)		
[Resolution]			
6. Enamel, dentin, pulp			
(upper molar and lower molar)			
7. PDL space and lamina dura			
(upper molar and lower molar)			
8. Sinus floor			
9. Bone pattern			
9–1. Mastoid air cell bone structure			
9–2. Trabecular bone of mandible			
10. TMJ complex			
[Overall grade of image]			
	No diagnostic value (0	) Poor (1) Moderate	(2) Good (3

#### Table 2. Clinical CBCT image evaluation chart.

CBCT, cone-beam computed tomography; PDL, periodontal ligament; TMJ, temporomandibular joint.

https://doi.org/10.1371/journal.pone.0285608.t002

three criteria (artifacts, noise, and contrast) of 10 evaluation items. Individual items could be scored as 1 = poor quality, 2 = moderate, and 3 = good quality. The artifact criterion was assessed using 3 items with a perfect score of 9 points. The noise criterion was evaluated using 2 items and the contrast criterion was assessed using 5 items. Each item received a score between 0 and 3 and the maximum total score was 30 points. The overall image grade was also assessed as 0 = image quality with no diagnostic value, 1 = poor but feasible to diagnose, 2 = moderate, and 3 = good image.

### Statistical analysis

For a quantitative evaluation of the model proposed in this study, Model 1, and Model 2, the mean values of NMRSE, SSIM, and MAE were compared among the images generated with the three models and the oCBCT using analysis of variance with a 95% confidence interval. When there was statistical significance, the metrics of the oCBCT were compared to those of all three models using the *t*-test while correcting for multiple comparisons. For the qualitative evaluation, the interobserver reliability was assessed using the Cohen's kappa coefficient ( $\kappa$ ). Scores were compared between the oCBCT and pCBCT using the Wilcoxon signed-rank test according to the artifacts, noise, and resolution criteria.

# Results

The NRMSE, SSIM, and MAE values of the pCBCT of the proposed model, COMPUNet, were all significantly improved compared with the oCBCT. The pCBCT of the proposed model also showed significantly better performance than that of Model 1 and Model 2, in terms of NRMSE, SSIM, and MAE values (Fig 4). Fig 5 shows that the fine bone details were only preserved in the pCBCT of the COMPUNet, unlike the other models.



**Fig 4. Quantitative evaluation through an ablation study.** NRMSE (a), SSIM (b), and MAE (c) for comparisons between the original (oCBCT) and predicted CBCT (pCBCT), showing that the best-performing model was the proposed model compared to Model 1 and Model 2. (\* p < 0.005).

The interobserver reliability for the qualitative evaluation showed good to excellent agreement ( $\kappa = 0.87$ , *P*<0.05). For the artifacts, noise, and contrast criteria, the scores were significantly different between the oCBCT and pCBCT of COMPUNet. The total score (i.e., the sum of items for each criterion) was also significantly higher for pCBCT than for the oCBCT (Table 3). For the individual evaluation items, most items showed higher scores in the pCBCT, except for item 6 (enamel, dentin, pulp) and 9 (bone pattern) in the contrast section, which showed almost the same score (Fig 6). There were more images with good grades in the pCBCT of COMPUNet than in the oCBCT. Moderate and poor grades were presented more frequently for the oCBCT images than for the pCBCT images (Figs 7 and 8).

# Discussion

With the growth of digital dentistry and the use of 3D images, the scope of CBCT application has grown significantly during the last decade. However, due to pronounced artifacts, its use



Fig 5. The trabecular bone pattern (arrow) in the same image slides. The fine bone details are best preserved in the predicted CBCT of the proposed model, COMPUNet, compared to that of Model 1 and Model 2.

https://doi.org/10.1371/journal.pone.0285608.g005

	Original CBCT	Predicted CBCT	P-value
Artifacts (9 points)	$6.14 \pm 0.79$	$7.91 \pm 0.79$	0.000*
Noise (6 points)	$3.95 \pm 0.56$	$5.28 \pm 0.41$	0.000*
Resolution (15 points)	$14.09 \pm 1.19$	$14.64 \pm 0.77$	0.001*
Total (30 points)	24.18 ± 1.91	27.83 ± 1.55	0.000*

#### Table 3. Scores from the qualitative evaluation.

\* Wilcoxon signed-rank test, *p*-value < 0.05.

https://doi.org/10.1371/journal.pone.0285608.t003

has been limited, impairing the diagnostic utility and precision of 3D models for surgical simulation [34]. Although numerous deep learning methods have tackled the challenge of removing artifacts in CBCT, their usage in dentistry has posed several challenges due to the difficulty of preparing training datasets (i.e., aligning MDCT and CBCT data acquired from diverse scanning conditions), and the loss of fine details due to blurring. The present study is the first to attempt to solve these issues using a multi-step registration process and novel network architecture with a multi-planar 2.5D U-Net-based network and a carefully designed loss function.

By employing the multi-planar method, it was possible to eliminate artifacts that were difficult to correct with a single-planar network. The present study used three 2.5D single-planar U-Nets, trained in the axial, coronal, and sagittal directions, respectively, and averaged across the three generated volumes. This was advantageous for minimizing streaking artifacts, which can be difficult to distinguish in one orientation but are readily discernible in the other two. Some artifacts were more visible in certain orientations than the others (often depending on



Fig 6. Mean scores of individual items in the clinical CBCT evaluation chart. Items for the artifacts and noise criteria showed improvements in the pCBCT images, while those for the resolution criterion improved less.

https://doi.org/10.1371/journal.pone.0285608.g006



**Fig 7. Original CBCT with an overall image grade of poor and the corresponding predicted CBCT image.** (a) The maxilla in an axial view (b), the anterior teeth region in a sagittal view, and (c) the temporomandibular region in a parasagittal view show reduced artifacts and noise in the COMPUNet CBCT compared to the original CBCT images.

the anatomical structure present in the orientation). By complementing one another, the multi-planar network removed artifacts more efficiently than the single-planar network. As a result, the NRMSE decreased from 0.1455 to 0.1450, the SSIM increased from 0.8228 to 0.8365, and the MAE decreased from 142.6 to 138.1.

By incorporating a contextual loss term, which is computed on a per-instance basis on the feature map error, it was possible to recover blurred fine features compared with the conventional L1 loss alone. Our loss function was advantageous in the presence of misalignment as it made no assumptions that the input image and the target image were perfectly aligned. It extracted contextual features from each image using a pretrained VGG19 network, located corresponding features, and analyzed their similarity. As a result, the SSIM increased (from 0.8365 to 0.8461) and the error values decreased (NRMSE from 0.1450 to 0.1410, MAE from 138.1 to 131.6). As seen in Fig 6, the CBCT images' artifacts and noise were reduced, but the intricate and fine structural details were maintained. Additionally, experts confirmed that the suggested method preserved fine details well, as the ratings for fine details (referred to as resolution) were higher in the pCBCT than in the oCBCT.

The clinical evaluation demonstrated that the proposed CBCT enhancement improved the visibility of the sinus floor and the TMJ complex, which are critical structures for diagnosis. Since the majority of the upper teeth protrude into the maxillary sinus, identifying the



Fig 8. Distribution of overall image quality evaluation grades. The proportion of CBCT with poor grades decreases (sky blue) and those with good grades increases (dark blue) in the predicted CBCT (pCBCT) compared to the original CBCT (oCBCT) for both observers.

https://doi.org/10.1371/journal.pone.0285608.g008

maxillary floor on CBCT prior to tooth extraction is critical for avoiding sinus surgery [35]. Additionally, the distance between the tooth root and the sinus floor impacts the effectiveness of dental implant surgery [36], and dental clinicians have expressed interest in being able to clearly define the sinus floor on CBCT. For the examination of TMJ disease, CBCT is preferred over MDCT as a general strategy for the current clinical situation [37]. Thus, image enhancement, particularly of certain anatomic regions in our investigation, has significant clinical implications.

While the current work demonstrates an intriguing technical development in deep learning for improving CBCT images, the study's scope is limited. We have demonstrated that the proposed technique preserves fine details on CBCT, but does not improve them. More specifically, the increase in the score for resolution was less than the increase in the scores for artifacts and noise. This is mostly due to the fact that the tooth structures (enamel, dentin, and pulp) and bone patterns (mastoid air cell and trabecular bone) stayed constant rather than improving. Although the model effectively reduced various artifacts, including beam hardening artifacts and metal artifacts, they were not completely removed—especially when the original quality

was quite poor (e.g., Fig 8) or when the artifact patterns were not common in the training dataset. It should be kept in mind that we designed the loss function of the model so that the texture and fine details in CBCT would not vanish. We observed that artifacts could be further removed by adjusting weights to each term in the loss function; however, this resulted in removing the CBCT texture and making it look artificial around the soft tissue in many cases. Therefore, the model can be easily tuned to remove artifacts more aggressively as needed.

More thorough dataset preparation may make it possible to improve the fine details. Due to the study's retrospective design, the clinical CBCT and MDCT images were taken with different patient postures. Specifically, some patients bit on the bite-block while standing or sitting during the CBCT examination. In contrast, the MDCT scans were taken with the patient lying down. As standing or sitting alters the relative position and topology of the maxillary-mandibular and facial soft tissue shapes, perfect registration between the pairs of images was difficult. More rigorous dataset preparation, such as realistic phantom scanning or dedicated image acquisition design, may be advantageous for further improving fine details. We will also continue to collect more CBCT-MDCT pairs to obtain a larger training dataset. This research direction might have clinical value for analyzing bone morphology and trabecular structure, which are critical for dental surgical procedures in order to determine the ideal placement of dental implants and monitor the bone healing process. Through the examination of bone patterns, CBCT images have proven their efficacy in predicting the prognosis of dental surgery [38, 39].

# Conclusion

In conclusion, we proposed a novel technique for improving the quality of CBCT images for dental diagnostic and treatment planning purposes. The significant improvement in CBCT image quality was validated by comparing the derived CBCT images to the original CBCT images using NMRSE, SSIM, and MAE values. Clinical experts additionally evaluated the resulting CBCT images as having fewer artifacts and less noise, while keeping the high resolution and sharpness of the original image. The DL method validated in the current study is suggested as a tool with practical applicability in actual clinical practice.

# **Author Contributions**

Conceptualization: Chena Lee, Young Hyun Kim, Chanyeol Choi, Sang-Sun Han.
Data curation: Chena Lee, Young Hyun Kim, Chanyeol Choi.
Formal analysis: Kanghyun Ryu, Ikbeom Jang.
Funding acquisition: Sang-Sun Han.
Investigation: Chena Lee, Yoseob Han, Subeen Pang.
Methodology: Kanghyun Ryu, Chena Lee, Yoseob Han, Subeen Pang, Ikbeom Jang.
Project administration: Chanyeol Choi, Sang-Sun Han.
Resources: Kanghyun Ryu.
Software: Kanghyun Ryu.
Supervision: Yoseob Han, Subeen Pang, Chanyeol Choi, Sang-Sun Han.
Validation: Chena Lee, Young Hyun Kim, Chanyeol Choi, Ikbeom Jang, Sang-Sun Han.

Writing – original draft: Kanghyun Ryu, Chena Lee, Ikbeom Jang.

Writing – review & editing: Kanghyun Ryu, Chena Lee, Yoseob Han, Subeen Pang, Young Hyun Kim, Chanyeol Choi, Ikbeom Jang, Sang-Sun Han.

#### References

- Vaddi A, Villagran S, Muttanahally KS, Tadinada A. Evaluation of available height, location, and patency
  of the ostium for sinus augmentation from an implant treatment planning perspective. Imaging Science
  Dentistry. 2021; 51(3):243–250. https://doi.org/10.5624/isd.20200218 PMID: 34621651
- Safi Y, Amid R, Zadbin F, Ghazizadeh Ahsaie M, Mortazavi H. The occurrence of dental implant malpositioning and related factors: A cross-sectional cone-beam computed tomography survey. Imaging Science Dentistry. 2021; 51(3):251–260. https://doi.org/10.5624/isd.20200331 PMID: 34621652
- Lee C, Yoon J, Han SS, Na JY, Lee JH, Kim YH, et al. Dose assessment in dental cone-beam computed tomography: Comparison of optically stimulated luminescence dosimetry with Monte Carlo method. PLoS One. 2020; 15(3):e0219103. https://doi.org/10.1371/journal.pone.0219103 PMID: 32231373
- Nardi C, Talamonti C, Pallotta S, Saletti P, Calistri L, Cordopatri C, et al. Head and neck effective dose and quantitative assessment of image quality: a study to compare cone beam CT and multislice spiral CT. Dentomaxillofacial Radiology. 2017; 46(7):20170030. https://doi.org/10.1259/dmfr.20170030 PMID: 28511564
- Pauwels R, Beinsberger J, Stamatakis H, Tsiklakis K, Walker A, Bosmans H, et al. Comparison of spatial and contrast resolution for cone-beam computed tomography scanners. Oral surgery, oral medicine, oral pathology and oral radiology, 114(1), 127–135. <u>https://doi.org/10.1016/j.oooo.2012.01.020</u> PMID: 22727102
- Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. Physics in medicine and biology, 64(12), 125002. <u>https://doi.org/10.1088/1361-6560/ab22f9</u> PMID: 31108465
- Yuan N, Dyer B, Rao S, Chen Q, Benedict S, Shang L, et al. Convolutional neural network enhancement of fast-scan low-dose cone-beam CT images for head and neck radiotherapy. Physics in medicine and biology, 65(3), 035003. https://doi.org/10.1088/1361-6560/ab6240 PMID: 31842014
- Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. Medical Physics. 2020; 47(3):1115–1125. https://doi.org/10.1002/mp.13978 PMID: 31853974
- 9. Xie S, Yang C, Zhang Z, Haibo L. Scatter artifacts removal using learning-based method for CBCT in IGRT system. IEEE Access.2018; 6; 78031–78037
- Zhang Y, Yue N, Su MY, Liu B, Ding Y, Zhou Y, et al. Improving CBCT quality to CT level using deep learning with generative adversarial network. Medical Physics. 2021; 48(6): 2816–2826. https://doi.org/ 10.1002/mp.14624 PMID: 33259647
- Wu W, Qu J, Cai J, Yang R. Multiresolution residual deep neural network for improving pelvic CBCT image quality. Medical Physics. 2022; 49(3):1522–1534. https://doi.org/10.1002/mp.15460 PMID: 35034367
- Gao L, Xie K, Wu X, <Lu Z, Li C, Sun J, et al. Generating synthetic CT from low-dose cone-beam CT by using generative adversarial networks for adaptive radiotherapy. Radiation oncology (London, England), 16(1), 202. https://doi.org/10.1186/s13014-021-01928-w
- Choi H, Yun JP, Lee A, Han SS, Kim SW, Lee C. Deep learning synthesis of cone-beam computed tomography from zero echo time magnetic resonance imaging. Sci Rep. 2023 Apr 13; 13(1):6031. https://doi.org/10.1038/s41598-023-33288-8 PMID: 37055501
- Hsu K, Yuh DY, Lin SC, Lyu PS, Pan GX, Zhuang YC, et al. Improving performance of deep learning models using 3.5 D U-Net via majority voting for tooth segmentation on cone beam computed tomography. Scientific Reports. 2022; 17 (1):19809.
- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. InMedical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19 2016 (pp. 424–432). Springer International Publishing.
- Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 34(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694 PMID: 25494501

- Guizard N, Coupé P, Fonov VS, Manjón JV, Arnold DL, Collins DL. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. NeuroImage Clinical. 2015; 8; 376–389. https://doi.org/ 10.1016/j.nicl.2015.05.001 PMID: 26106563
- Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging. 2016, 11; 35(5):1285–98. <u>https://doi.org/10.1109/TMI.2016.2528162</u> PMID: 26886976
- Roth HR, Le L, Jiamin L, Jiahua Y, Ari S, Kevin C, et al. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE transactions on medical imaging*. 2015; 35(5): 1170–1181. https://doi.org/10.1109/TMI.2015.2482920 PMID: 26441412
- Zhou H, Cao M, Min Y, Yoon S, Kishan A, Ruan D. Ensemble learning and tensor regularization for cone-beam computed tomography-based pelvic organ segmentation. Medical Physics. 2022; 49 (3):1660–1672. https://doi.org/10.1002/mp.15475 PMID: 35061244
- Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. IEEE conference on computer vision and pattern recognition. 2017; 1125–1134.
- Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 2016; 3(1), 47–57.
- 23. Girod B. What's wrong with mean-squared error? Cambridge, MA: MIT Press. 1993.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *EEE transactions on image processing: a publication of the IEEE Signal Processing Society*. 2004; 13(4): 600–612. https://doi.org/10.1109/tip.2003.819861 PMID: 15376593
- Garyfallidis E, Brett M, Amirbekian B, Rokem A, van der Walt S, Descoteaux M, et al. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics*. 2014; 21(8): 8. <u>https://doi.org/10.</u> 3389/fninf.2014.00008 PMID: 24600385
- Mattes D, Haynor DR, Vesselle H, Lewellyn TK, Eubank W. Nonrigid multimodality image registration. Medical imaging. 2001; 4322: 1609–1620.
- Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). The Insight Journal. 2009; 2 (365): 1–35.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. IEEE conference on computer vision and pattern recognition. 2009; 248–255.
- Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. 2018; 11.
- Mechrez R, Talmi I, Zelnik-Manor L. The contextual loss for image transformation with non-aligned data. In Proceedings of the European conference on computer vision. 2018; 768–783.
- Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014; 22.
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, highperformance deep learning library. Advances in neural information processing systems. 2019;32.
- Lee C, Lee A, Choi YJ, Jeon KJ, Kim YH, Han SS. Accuracy of digital model generated from CT data with metal artifact reduction algorithm. Scientific Reports. 2021; 11(1):10332. <u>https://doi.org/10.1038/</u> s41598-021-89298-x PMID: 33990637
- 35. Shrestha B, Shrestha R, Lu H, Mai Z, Chen L, Chen Z, et al. Relationship of the maxillary posterior teeth and maxillary sinus floor in different skeletal growth patterns: A cone-beam computed tomographic study of 1600 roots. Imaging Science Dentistry. 2022; 52(1): 19–25. <u>https://doi.org/10.5624/isd.</u> 20210145 PMID: 35387101
- Choi YJ, Kim YH, Han SS, Jung UW, Lee C, Lee A, et al. Alveolar bone height according to the anatomical relationship between the maxillary molar and sinus. Journal of periodontal & implant science. 2020; 50(1): 38–47. https://doi.org/10.5051/jpis.2020.50.1.38
- Jain S, Choudhary K, Nagi R, Shukla S, Kaur N, Grover D. New evolution of cone-beam computed tomography in dentistry: Combining digital technologies. Imaging Science Dentistry. 2019; 49(3): 179– 190. https://doi.org/10.5624/isd.2019.49.3.179 PMID: 31583200
- Vadiati Saberi B, Khosravifard N, Nooshmand K, Dalili Kajan Z, Ghaffari M. Fractal analysis of the trabecular bone pattern in the presence/absence of metal artifact-producing objects: Comparison of conebeam computed tomography with panoramic and periapical radiography. Dentomaxillofacial Radiology. 2021; 50(6): 20200559. https://doi.org/10.1259/dmfr.20200559 PMID: 33705225

 Shahmirzadi S, Maghsoodi-Zahedi T, Saadat S, Demirturk Kocasarac H, Rezvan M, Katkar RA, et al. Assessment of alveolar bone changes in response to minimally invasive periodontal surgery: A conebeam computed tomographic evaluation. Imaging Sci Dent. 2023 Mar; 53(1):1–9. https://doi.org/10. 5624/isd.20220053 PMID: 37006791