The Transactions of the Korea Information Processing Society
TKIPS Vol.13, No.7, pp.326~334

eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2024.13.7.326

3DentAI: U-Nets for 3D Oral Structure Reconstruction from Panoramic X-rays

Anusree P.Sunilkumar† · Seong Yong Moon†† · Wonsang You†††

ABSTRACT

Extra-oral imaging techniques such as Panoramic X-rays (PXs) and Cone Beam Computed Tomography (CBCT) are the most preferred imaging modalities in dental clinics owing to its patient convenience during imaging as well as their ability to visualize entire teeth information. PXs are preferred for routine clinical treatments and CBCTs for complex surgeries and implant treatments. However, PXs are limited by the lack of third dimensional spatial information whereas CBCTs inflict high radiation exposure to patient. When a PX is already available, it is beneficial to reconstruct the 3D oral structure from the PX to avoid further expenses and radiation dose. In this paper, we propose 3DentAI – an U-Net based deep learning framework for 3D reconstruction of oral structure from a PX image. Our framework consists of three module – a reconstruction module based on attention U-Net for estimating depth from a PX image, a realignment module for aligning the predicted flattened volume to the shape of jaw using a predefined focal trough and ray data, and lastly a refinement module based on 3D U-Net for interpolating the missing information to obtain a smooth representation of oral cavity. Synthetic PXs obtained from CBCT by ray tracing and rendering were used to train the networks without the need of paired PX and CBCT datasets. Our method, trained and tested on a diverse datasets of 600 patients, achieved superior performance to GAN-based models even with low computational complexity.

Keywords : Attention U-Net, CBCT, Panoramic X-ray, Focal Trough, 3D Reconstruction

3DentAI: 파노라마 X-ray로부터 3차원 구강구조 복원을 위한 U-Nets

Anusree P.Sunilkumar† · 문 성 용†† · 유 원 상†††

요 약

파노라마 X-ray (PX) 및 Cone Beam Computed Tomography (CBCT)와 같은 구강 영상 기술은 영상 촬영 시 환자의 편의성과 전체 치아 정보를 시각화할 수 있는 능력으로 인해 치과 진료소에서 가장 선호되는 영상 기법이다. PX는 일상적인 임상 치료에 선호되고, CBCT는 복잡한 수술 및 임플란트 치료에 선호된다. 그러나 PX는 3차원 공간정보가 부족하다는 한계가 있는 반면 CBCT는 환자에게 높은 방사선 노출을 초래한다. PX가 이미 사용 가능한 경우 PX로부터 3D강 구강구조를 복원함으로써 추가 비용을 줄이고 방사선량을 피할 수 있다. 본 논문에서는 PX 이미지로부터 구강구조의 3차원 복원을 위한 U-Net 기반 딥러닝 프레임워크인 3DentAI를 제안한다. 제안된 프레임워크는 PX 이미지에서 깊이를 추정하기 위한 Attention U-Net 기반 재구성 모듈, 사전 정의된 초점 골 및 광선 데이터를 사용하여 예측된 편평 볼륨을 턱 모양에 정렬하기 위한 재정렬 모듈과, 구강의 원활한 표현을 얻기 위해 누락된 정보를 보간하는 3D U-Net 기반 개선 모듈의 세 가지 모듈로 구성된다. 네트워크를 훈련하기 위해, 쌍을 이루는 PX 및 CBCT 데이터셋 대신에 광선 추적 및 렌더링을 통해 CBCT로 부터 합성한 PX 데이터를 사용하였다. 600명의 환자로 구성된 다양한 데이터셋으로 모델을 훈련한 결과, 낮은 계산 복잡도에도 GAN 기반 모델에 비해 우수한 성능을 보였다.

키워드 : Attention U-Net, CBCT, 파노라마 엑스레이, 초점 영역, 3D-복원

1. Introduction

Panoramic X-rays (PXs) are widely utilized in dental clinics owing to its ability to capture entire maxillofacial region in a single 2D radiograph [1]. PXs can be used for routine clinical treatments such as assessment of cavities, bone loss etc. They are easy to perform and imparts less radiation dose on the patient. However, their usage and

※ 본 연구는 2022년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구(2022R1 F1A1075204), 4단계 두뇌한국21 사업(4단계 BK21 사업) 및 지자체-대학 협력기반 지역혁신사업(2022RIS-004), 중소기업벤처부의 재원으로 수행된 2021년도 창업성장기술개발사업(S3228660)의 연구결과로 수행되었음.
† 비 회 원 : 선문대학교 정보통신공학과 석사과정
†† 비 회 원 : 조선대학교 치과대학 구강악안면외과 교수
††† 정 회 원 : 선문대학교 정보통신공학과 부교수
Manuscript Received : June 13, 2024
Accepted : June 19, 2024
* Corresponding Author : Wonsang You(wyou@kaist.ac.kr)

efficiency is limited by its lack of dimension, noises and blurring caused by superimposition of adjacent structures, metal artifacts as well as magnification [2]. On the other hand, Cone Beam Computed Tomography(CBCTs) are preferred for performing complex surgeries and implant treatment because of its ability to visualize three-dimensional information with better resolution [3, 4]. Still, they inflict high radiation dose on patients and is costly to perform because of which they cannot be used often [5].

To overcome such limitations, it can be beneficial to reconstruct 3D oral cavities from available PX images. These 3D models offer variety of applications ranging from patient education, virtual surgical simulations [6], to assist in treatment planning. They doesn't aim to entirely replace the necessity of CBCT scans. Nevertheless, it can supplement necessary information for treatment planning and helps postponement of an additional scanning.

Previous studies used either auto-encoder or GAN [7] models as well as NeRF [8] based models for 3D reconstruction using PX image. Auto-encoder and GAN based models [9, 10] were trained using synthetic PX and 3D flattened volumes of the jaw, instead of real world PX. Synthesized PX demonstrates one-to-one correspondence with the 3D flattened views and this enables the model to learn effectively. Besides, obtaining paired PX-CBCT data from the same patient is challenging in a clinical situation. However a major disadvantage of this approach is that information is lost during jaw flattening as well as PX synthesis. Moreover, the flow of information between the encoder and decoder in auto-encoder models are not efficient. Additionally, the flattened jaw needs to be realigned to the jaw shape using any prior information such as dental arch. This necessitates additional experiments to find a suitable dental arch shape either by employing additional X-rays or by creating a template arch based on available dataset.

NeRF-based models [11, 12], on the other hand, uses principles of PX imaging to simulate the synthetic PX as well as to train the 3D reconstruction. These models to revert the PX synthesis process during 3D reconstruction, however, the model needs to be aware of PX synthesis process. A major disadvantage of these methods is the high computational power associated with these models. NeBLa proposed by Park *et al.* [12] requires around 80 GB of GPU memory for training.

To overcome such limitations, we propose a novel 3D reconstruction pipeline based on Attention U-Net which can predict the depth information from a PX image. The realignment of the predicted depth is performed using a common focal trough and ray data which was predetermined during synthesis process. The use of focal trough enable us to limit the region to be reconstructed, whereas the rays restore the original structure. To reduce the loss of information we included a refinement module based on 3D U-Net which can help interpolate the lost information from flattening. By training the models standalone, we were also successful in reducing the memory requirement. Experiments showed that our model gives superior performance over state-of-the-art baseline models.

Our contributions in this paper are: First, we propose a novel pipeline for 3D reconstruction and refinement of oral cavity. Second, we define a focal trough to limit the reconstructed area. Third, our synthetic PXs are formed by detecting and reshaping the jaw region for all patients to a common dimension which provides us with a uniformly spaced ray data that can be used for all patients irrespective of the jaw variations. This also eliminates the need of extracting a suitable dental arch.

The reminder of this paper are organized as follows. Section 2 discusses the related works. Our proposed methodology is outlined in section 3. The section 4 demonstrates the dataset generation, qualitative results and quantitative analysis. Finally section 5 and 6, discusses the limitations, future directions and conclusions from our research.

2. Related works

2.1 Cross-Modality Transfer

Cross-Modality transfer in medical imaging is highly discussed topic and majorly concentrates on translation between Magnetic Resonance Imaging (MRI)/ Computed Tomography (CT)/ Positron Emission Tomography (PET). This is especially beneficial when one particular modality lacks some crucial information offered by another. For instance, MRI-CT translation facilitates combining the structural precision of CT with soft tissue contrast in MRI.

Burgos *et al.* [13] proposed a multi-atlas information propagation scheme that perform organ segmentation and translates structural MRI images to pseudo-X-ray CT data. Similarly, Huang et al. [14] used dictionary learning for super resolution and cross-modality transfer with both paired and unpaired dataset.

2.2 X-ray based 3D Reconstructions

Cross Modality transfer has been explored a lot in translation between same dimensional data, however, those involving lower to higher dimensions are relatively scarce. This stems from the complexity in estimating the lost information during the 2D imaging. After the introduction of generative models such as GAN, there has been a growing interest towards X-ray based 3D reconstruction. Nevertheless, most of the works in this domain primarily focused on full-view or sparse-view reconstructions. Henzler *et al.* [15] used an encoder-decoder model to reconstruct 3D skull volumes of mammalian species from their 2D cranial X-rays. Another study by Ying *et al.* [16] proposed X2CT-GAN to reconstruct CT scans from bi-planar chest X-rays.

Single-view 3D reconstruction are challenging owing to the limited information available from the single viewpoint. This necessitates the availability of some prior knowledge, for example, dental arches in case of 3D oral cavity reconstruction. In dental domain, 3D reconstruction rely on PX images. This approach offer an advantage because panoramic images can be considered as multiple views stitched together to create a flattened view. Pioneer works for Oral 3D reconstruction, hence, focused on creating flattened 3D volumes from CBCT and used PX images to predict this depth using generative models such as auto-encoders and GANs. Finally the predicted depth volumes are aligned to a predefined template arch. A significant challenge for training these models are the necessity of paired datasets from same patient. Considering the practical difficulty of obtaining such dataset, most of the works resort to synthetic PX images generated from the CBCT scans. This ensures efficient learning since the model can now learn from consistent representation of data in both modalities.

Liang *et al.* [9] in their work X2Teeth, proposed auto-encoder based 3D reconstruction based on teeth-localization and single shape estimation. They considered teeth as multiple objects to estimate the shape of each tooth by encorporating teeth segmentation maps to facilitate patch sampling and localization. In a similar approach, Liang *et al.* [6] proposed OralViewer, an interactive application aimed at 3D visualization of surgical procedures for patient education. The reconstructed 3D tooth are aligned to a predefined dental arch and subsequently combined with gum and jaw bone models to obtain a complete representation of oral cavity. On the other hand, Song *et al.* [10] proposed Oral 3D, an encode-decoder GAN model to convert 2D PX image to 3D oral representation. Unlike previous methods, they were able to restore the jaw as a whole by employing flattened CBCT slices and then aligning it back to the predefined template arch.

The major disadvantage of these methods is that the final results after the jaw alignment were not smooth due to the lost information from flattening and depth prediction. Besides, the use of auto-encoder models cannot consistently prioritize the region to be reconstructed, especially the teeth.

Very recently, 3D reconstruction using NeRF-based models has been in demand, and its application could be seen in dental imaging as well. NeRF models aims to reverse the image acquisition process of PX by leveraging the rotational principles of panoramic radiography. This means NeRF will be aware of the PX synthesis process. To facilitate this, synthetic PX generation makes use of a simulated panoramic acquisition geometry with ray tracing and rendering making the parameters available to NeRF during the training process.

Song *et al.* [11] proposed a novel model based on NeRF to reconstruct 3D jaw from the projection data from panoramic scans. Their framework composed of a multi-head prediction, dynamic sampling and adaptive rendering uses the NeRF principles and tries to minimize the rendering loss to facilitate model learning. Dynamic sampling of rays produces multiple sample points along that view for the model to learn a smooth intensity distribution in 3D space. These points are encoded into high-dimensional embeddings which can be further used to predict the beam of voxel intensities. These predicted points are rendered into a single pixel value which is then compared with input PX to minimize the loss function. Their method evaluated using 80 CBCT scans demontsrtaed superior performance over its predecessor Oral-3D.

In a similar vein, Park *et al.* [12] proposed NeBLa, a NeRF-based oral cavity reconstruction by using an intermediate representation called SimPX, which a synthetic PX simulated from CBCT scans. The SimPXs were created by simulating PX acquisition geometry for ray tracing and rendering. Additionally, unpaired image-to-image translation was performed to bridge the gap between real and synthetic PX images. This facilitates the real PX to be used during inference. Consequently, their method completely eliminated the need of paired dataset for training. The key advantage of their method is that no prior

information is needed to guide the reconstruction process. Their model was also able to outperform existing methodologies. However, their model is limited by the high computational cost.

3. Methodology

The complete of our proposed methodology is shown in Fig. 1. It comprises of three modules: A reconstruction module modeled after as an attention U-net estimates the depth from input PX image resulting a flattened 3D view of the jaw. Next a realignment module aligns the output flattened 3D volume to the shape of the jaw. The realignment is guided by the predefined focal trough mask. Finally, the refinement module based on 3D U-net interpolates missing information in the realigned output to obtain a smooth final representation of the 3D oral cavity.

The attention U-net is a variation of conventional U-net architecture, with added attention modules before each decoder modules. This enables the model to learn the essential features from the input improving the contrast of final output. The attention mechanism creates attention maps by suitably highlighting critical regions that requires greater emphasis during reconstruction, in this case, the teeth.

The model outputs a flattened volume of size 96x 128x256. Each convolution uses a 3x3 kernel with a stride of 1, and is followed by ReLU activation function. The

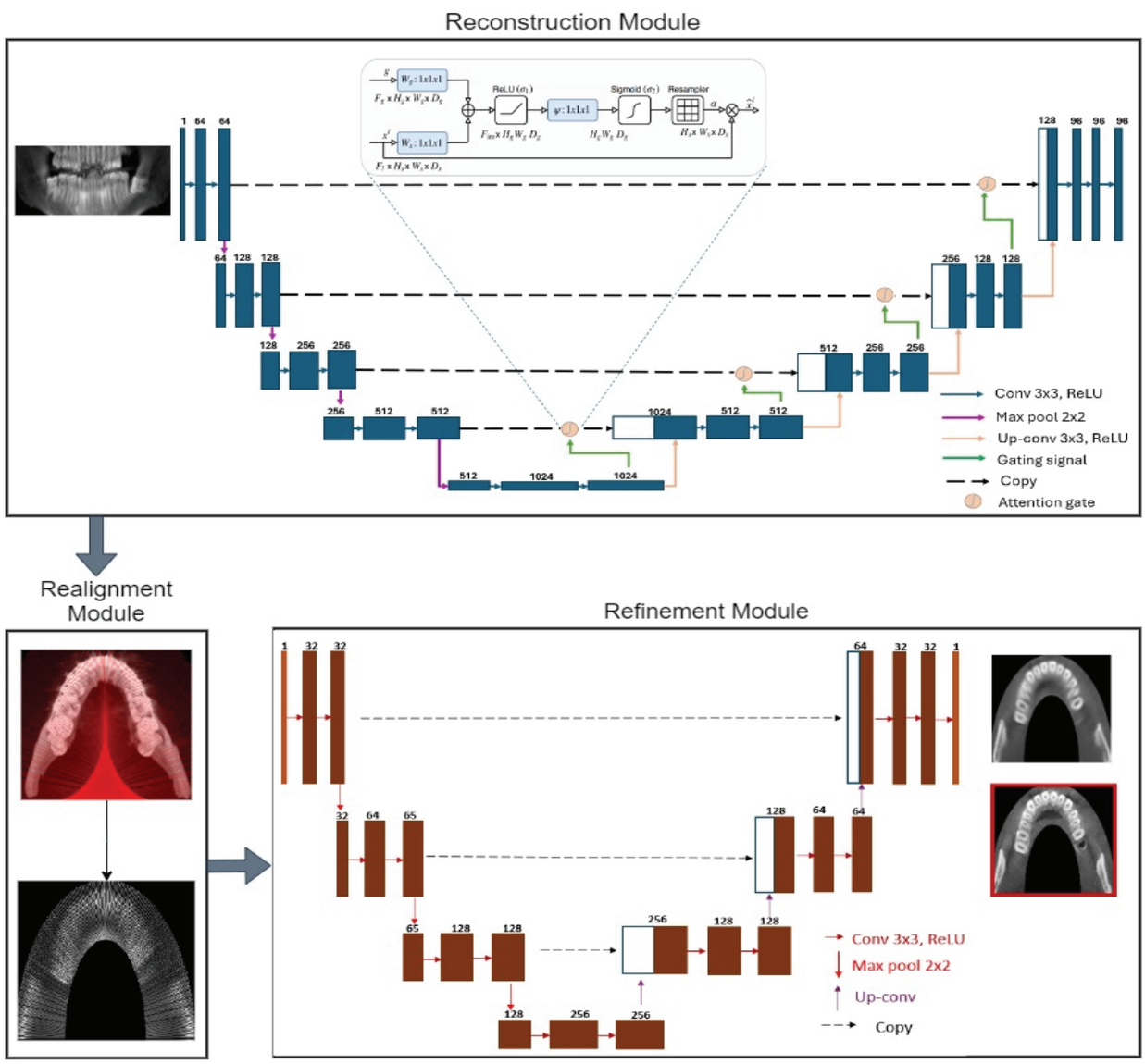

Fig. 1. Complete Workflow of the Proposed Framework, Consisting of Reconstruction, Realignment, and Refinement Modules

training utilizes L2 Loss function and an Adam optimizer with parameters, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, along with a learning rate of 1×10^{-4}. The models was set to stop early after 15 epochs if the model doesn't learn anymore.

The realignment module undo the flattening process performed during dataset preparation with the help of focal trough mask and ray data. It retraces the original location of the pixels along the rays and assign it back to its position in CBCT domain. Thus the final output will have a dimension of 128x256x256. Subsequently, the refinement module further enhance the quality of the realigned output to generate the final reconstructed volume. Reconstruction module utilizes the flattened CBCT volumes as ground truth whereas refinement module uses the reshaped ROIs. To optimize the memory utilization both modules are trained standalone.

4. Experiments

4.1 Data collection

The 623 clinical dental head CBCT scans (408 Females and 215 Males, Mean age: 50, Deviation: 26) used in this study were collected from Chosun School of Dentistry in Gwangju, South Korea. Scanning utilized CS900 scanner from Carestream Health and the Planmeca VISOG7. The images were formatted as 16-bit DICOM slices with a resolution ranging from 0.25mm to 0.5mm. All data were collected with proper patient consent and following the ethical guidelines outlined by The Code of Ethics of the World Medical Association. Approval for data collection was obtained from the Institutional Review Board (IRB).

4.2 Dataset preparation

The dataset preparation is outlined in Fig. 2. Initially, all CBCT scans are resampled to a common pixel spacing of 0.3mm. Subsequently. the CBCT data is preprocessed to obtain a contour box plot enclosing the jaw region. This forms the ROI for 3D reconstruction. The ROI is cropped out from CBCT and then reshaped to a uniform dimension of 256x256x128 (DxWxH). Similar to Park *et al.* [12], a panoramic acquisition geometry is modeled and the ray tracing is performed along the defined trajectory to produce synthetic PX images of shape 256x128. Unlike Park *et al.*[12], we have used an elliptical trajectory and rays are considered tangent to this trajectory. The number of sampled pixels along the rays are limited to a horseshoe shaped focal region enclosing the jaw. Consequently, these pixels were aligned sequentially to a rectangular grid to generate a flattened jaw of size 96x128x256 (DxWxH). The model training utilizes both the flattened CBCTs as well as the reshaped ROIs, in conjunction with the synthesized PX images. Out of 623 CBCT scans, 600 patient data was generated using this approach and the data was split in a ratio 8:1:1 for training, validation and testing, respectively. Training was performed using NVIDIA A6000 GPUs with 50GB memory.

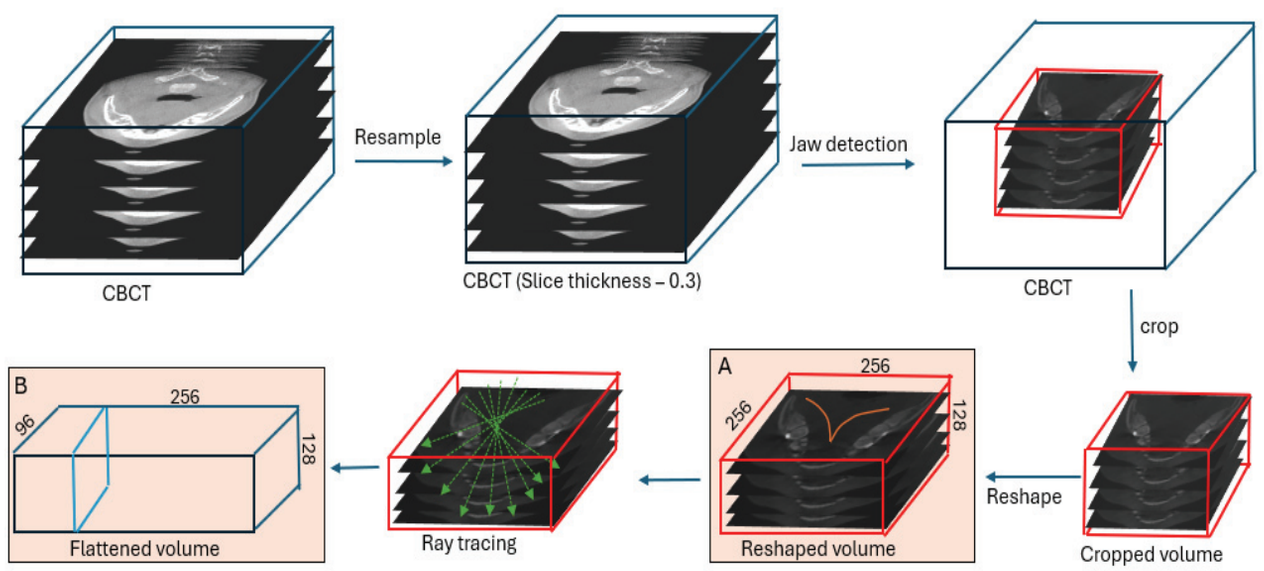

Fig. 2. Data Preprocessing for Model Training

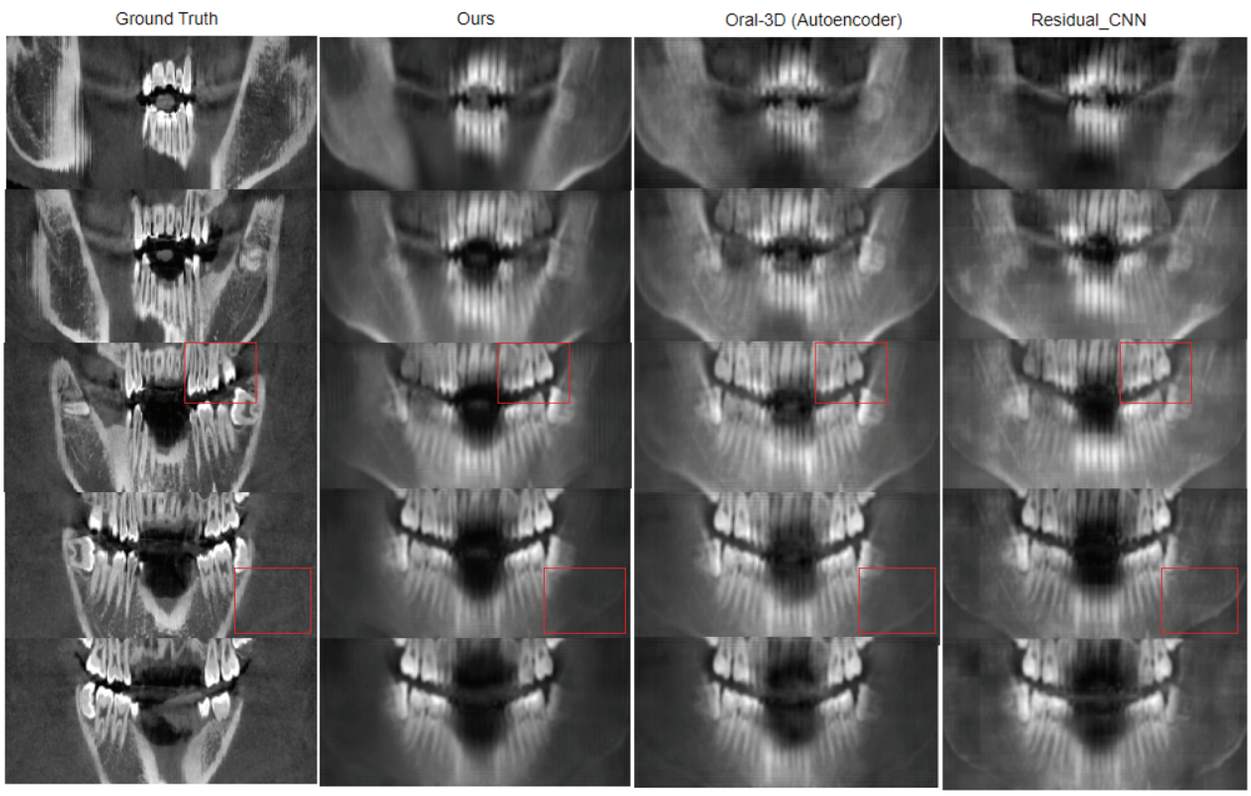

Fig. 3. Qualitative Results from Reconstruction Module. Our Results Demonstrate Better Contrast Compared to the Baseline Models. Ours and Oral-3D has Better Definition of Jaw Edges Compared to Residual-CNN

4.3 Qualitative Results

In this section, we will assess the qualitative results of our model. Our comparison is performed based on two baseline models taken from previous papers. First one is based on Oral-3D by Song *et al.* [10] and the second model is a Residual-CNN taken from Henzler *et al.* [15]. However, for [10] we have only considered the auto-encoder used in the generation module. Similarly for [15], we have not taken bi-planar views, instead used same synthetic PXs as input. These models were trained as part of the reconstruction module under identical training conditions as attention U-net. The results from the reconstruction module is shown in Fig. 3. Although the results does not demonstrate significant changes, there is notable contrast difference among the reconstructed slices. Additionally, Oral-3D and our model has a better distinction for jaw edges than residual CNN.

Fig. 4 depicts the outputs from the refinement modules. These outputs are visualized using Maximal Intensity Projections (MIP) projections along coronal and sagittal viewpoints. Our model were successful in providing smooth representation of texture for both jaw bone and teeth. However, the metallic wires with slight intensity changes are not distinguishable. Additionally, our model was also tested on a real PX. The results shows that the model does not provide a similar performance with real PX, however, it is worth notable that some of the teeth especially in the upper jaw are much more distinguishable from each other.

4.4 Quantitative Analysis

In addition to the qualitative comparison, reconstruction accuracy was measured using two metrics: Peak Signal-to-Noise ratio (PSNR) and Structural Similarity. PSNR measures the quality of reconstructed images, as a ratio between the maximum achievable pixel value and the mean squared error (MSE) between the ground truth image as well as the reconstructed images. It is usually measured in decibels (dB). PSNR can be calculated as follows:

$$PSNR = 20\log_{10}(2^k) - 10\log_{10}(MSE), \qquad (1)$$

where k is the number of bits per pixel and MSE is calculated as:

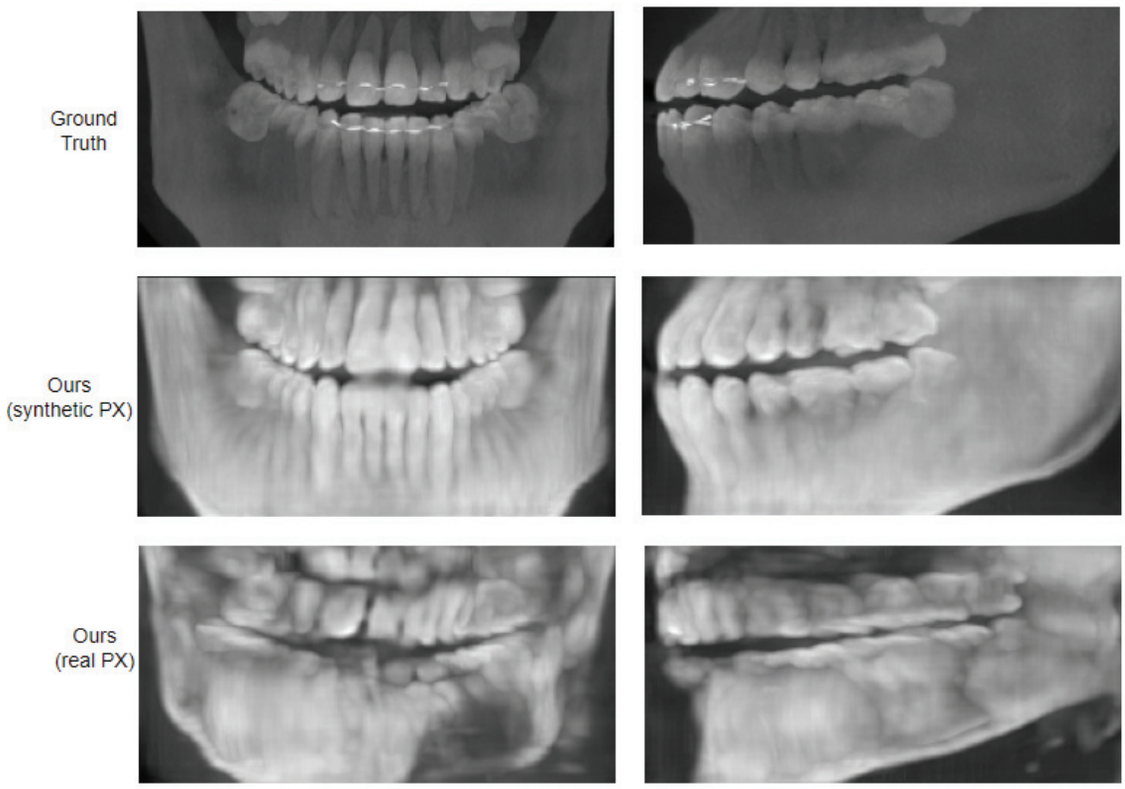

Fig. 4. Qualitative Results from the Refinement Module: using Synthetic PX vs. Real PX.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(I(i) - K(i))^2, \quad (2)$$

where n is the total number of pixels in the image, I(i) is the intensity of i^{th} pixel in the original image and K(i) is the value of i^{th} pixel in the reconstructed image. SSIM values ranges between −1 and 1, with 1 indicating perfect similarity and −1 indicating complete dissimilarity.

As summarized in Table 1, the evaluation of our method demonstrates its superior performance compared to both Oral 3D and Residual CNN. The reconstruction results achieved a higher PSNR of 18.18, indicating better image quality compared to the other models. Furthermore, the addition of the refinement module significantly enhanced this performance, boosting the PSNR to 20. This improvement highlights the effectiveness of the refinement module in enhancing the reconstruction quality. The considerable shift in both metrics for the output from the refinement module underscores its substantial impact on improving the overall reconstruction accuracy and quality. Even though Park et al. [12] achieved a higher PSNR of 21.68, it is worth notable that their memory requirement was comparatively higher tha ours. The computational memory required was around 80GB and their training utilized NVIDIA A100 and A6000 GPUs. Besides, we were not able to reproduce their method because of this high memory requirement and the PSNR value mentioned above is based on their experiments.

Table 1: Quantitative Evaluation Results

Models	PSNR	SSIM
Reconstruction Module		
Residual CNN	17.30 ± 0.13	0.35 ± 0.04
Oral-3D(auto-encoder)	17.70 ± 0.13	0.36 ± 0.04
Ours	18.18 ± 0.19	0.38 ± 0.05
Refinement Module		
Ours(3D U-net)	20.05 ± 0.40	0.69 ± 0.03

5. Limitations and Future directions

Even though our method demonstrated impressive performance it has several limitations. Because we have utilized synthetic PX for model training, it is not able to provide comparable performance with real world PX data. This discrepancy stems from the substantial gap between

real and synthetic dataset in terms of intensity distribution and contrast. However, this can be effectively minimized by adding an image-to-image translation between real and synthetic PX images during inference. This can also help the model to generalize well to unseen data.

Another critical limitation is that our model cannot accurately distinguish high intensity regions such as metallic implants. Metallic implants usually exhibit significantly higher intensity values depending on the tube current and exposure rate applied during CBCT acquisition. These peak values affect the normalization process and can limit the model ability to learn textures. It necessitates a trade-off, leading the model to compromise on intensity differentiation. One possible solution is to suppress the effect of metal implants during image acquisition so that they doesn't occupy peak intensity regions. However, considering the difficulty in obtaining real medical data, it would be beneficial to develop an anomaly-aware learnable normalization optimized to medical imaging.

Moreover, the current outputs does not guarantee its utility in a clinical practice due to the standardization of patient geometry irrespective of the jaw variations. This is also a target area of improvement for our future works. NeRF based models have achieved superior performance irresepctive of their high computational complexity. Hence, a computational efficient NeRF model can proposed to mitigate such issues.

The findings of this study highlights the need of further refinement to our model to improve its robustness and generalizability. Future works will focus on addressing these challenges to ensure that our model can be employed in clinical practices.

6. Conclusion

In this paper, we have proposed a novel workflow for the 3D reconstruction of oral cavity by using single PX image. Our proposed method which consists of reconstruction, realignment and refinement modules based on U-net architectures, outperformed the baseline auto-encoder and residual CNN models quantitatively as well as qualitatively. Also, our model required less computational power than NeRF-based models for training and evaluation. Our experiments showed that the computational memory requirement was comparatively lesser than NeRF model, even when the training was performed in a pipeline.

Our model's limitations in replicating similar performance with real PX images will be effectively handled in our future works. Also, considering the diverse patient anatomy, we also aim to modify our methodology to accommodate individual anatomical variations.

References

[1] A. G. Farman, Panoramic Corporation (Eds.), "Panoramic radiology: Seminars on maxillofacial imaging and interpretation," *Springer*, Berlin; New York, 2007. OCLC: ocn123114795.

[2] R. Izzetti, M. Nisi, G. Aringhieri, L. Crocetti, F. Graziani, and C. Nardi, "Basic knowledge and new advances in panoramic radiography imaging techniques: A narrative review on what dentists and radiologists should know," *Applied Sciences*, Vol.11, No.17, pp.7858, 2021.

[3] S. C. White and M. J. Pharoah (Eds.), "Oral radiology: Principles and interpretation," edition 7. ed., *Elsevier*, St. Louis, Missouri, 2014. OCLC: 862758150.

[4] R. Jacobs, B. Salmon, M. Codari, B. Hassan, and M. M. Bornstein, "Cone beam computed tomography in implant dentistry: Recommendations for clinical use," *BMC Oral Health*, Vol.18, pp.88, 2018.

[5] A. Al-Okshi, M. Nilsson, A. Petersson, M. Wiese, and C. Lindh, "Using GafChromic film to estimate the effective dose from dental cone beam CT and panoramic radiography," *Dentomaxillofacial Radiology*, Vol.42, No.추가, pp.20120343, 2013.

[6] Y. Liang et al., "OralViewer: 3D Demonstration of Dental Surgeries for Patient Education with Oral Cavity Reconstruction from a 2D Panoramic X-ray," *26th International Conference on Intelligent User Interfaces,* Association for Computing Machinery, New York, NY, USA, pp.553-563, 2021.

[7] I. Goodfellow et al., "Generative adversarial nets, in Advances in Neural Information Processing Systems (NIPS)," Montreal, Canada, pp.2672-2680, 2014.

[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in: A. Vedaldi, H. Bischof, T. Brox, J.- M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, pp.405-421, 2020.

[9] Y. Liang, W. Song, J. Yang, L. Qiu, K. Wang, and L. He, "X2Teeth: 3D Teeth Reconstruction from a Single Panoramic Radiograph," in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Springer International Publishing, Cham, pp.400-409, 2020.

[10] W. Song, Y. Liang, J. Yang, K. Wang, and L. He, "Oral-3D: Reconstructing the 3D structure of oral cavity from panoramic X-ray," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.35, No.1, pp.566-573, 2021.

[11] W. Song, H. Zheng, D. Tu, C. Liang, and L. He, "Oral-3Dv2: 3D oral reconstruction from panoramic x-ray imaging with implicit neural representation," *arXiv preprint arXiv: 2303.12123*, 2023.

[12] S. Park, S. Kim, D. Kwon, Y. Jang, I.-S. Song, and S. Baek, "NeBLa: Neural beer-lambert for 3D reconstruction of oral structures from panoramic radiographs," *arXiv e-prints, arXiv-2304*, 2023.

[13] N. Burgos et al., "Iterative framework for the joint segmentation and CT synthesis of MR images: Application to MRI-only radiotherapy treatment planning," *Physics in Medicine & Biology*, Vol.62, No.11, pp.4237-4253, 2017.

[14] Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding," In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, pp.5787-5796, 2017.

[15] P. Henzler, V. Rasche, T. Ropinski, and T. Ritschel, "Single-image tomography: 3d volumes from 2d cranial x-rays," *Computer Graphics Forum,* Vol.37, No.2, pp.377-388, 2018.

[16] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.10619-10628, 2019.

Anusree P.Sunilkumar

https://orcid.org/0009-0006-9381-3618

e-mail : anusreepandath@gmail.com

She received a Bachelors degree in Computer Science and Engineering from A.P.J. Abdul Kalam Technological University in 2019. She is pursuing her Masters in Information and Communication Engineering at SunMoon Univ. since 2022. From 2019 to 2021, she was a Systems Engineer at Infosys, India. Since 2022, she has been a Research Assistant with the Artificial Intelligence and Image Processing Laboratory, at Sun Moon University, South Korea.

Seong Yong Moon

https://orcid.org/0000-0002-7513-4404

e-mail : msygood@chosun.ac.kr

He received the B.S. degree in dentistry from Chosun University and the Ph.D. degree in dentistry from Chonnan University. He is currently a Professor with the Department of Oral and Maxillofacial Surgery, College of Dentistry, Chosun University. He is also the CEO of HT Core, a VR/AR-based dental simulation company.

Wonsang You

https://orcid.org/0000-0002-6806-7135

e-mail : wyou@kaist.ac.kr

He received an M.Sc. degree in computer vision and image processing from the Korea Advanced Institute of Science and Technology (KAIST) in 2008, and a Ph.D. degree in brain imaging data analysis from Otto-von-Guericke University Magdeburg in 2013. From 2009 to 2012, he worked as a research assistant at the Leibniz Institute for Neurobiology, Magdeburg, Germany. From 2013 to 2019, he worked as a staff scientist at the Center for the Developing Brain in Children's National Hospital, Washington, DC. He is currently an Associate Professor with the Department of Information and Communication Engineering at Sun Moon University, South Korea. He is also the director of the Artificial Intelligence and Image Processing Laboratory (AIIP Lab). His research interests include computer vision, image analysis, deep learning, and medical imaging.