3DPX: Progressive 2D-to-3D Oral Image Reconstruction with Hybrid MLP-CNN Networks

Xiaoshuang Li[1,2*][0000-0002-0773-6462], Mingyuan Meng[1,2*][0000−0002−9562−1613], Zimo Huang[2][0009-0005-5808-6251], Lei Bi[1,2][0000-0001-9759-0200], Eduardo Delamare[2][0000-0002-4866-1946], Dagan Feng[2][0000-0002-3381-214X], Bin Sheng[1(✉)][0000-0001-8510-2556], and Jinman Kim[2][0000-0001-5960-1060]

[1] Shanghai Jiao Tong University, Shanghai 200240, China
shengbin@sjtu.edu.cn
[2] The University of Sydney, Sydney NSW 2000, Australia

Abstract. Panoramic X-ray (PX) is a prevalent modality in dental practice for its wide availability and low cost. However, as a 2D projection image, PX does not contain 3D anatomical information, and therefore has limited use in dental applications that can benefit from 3D information, e.g., tooth angular misalignment detection and classification. Reconstructing 3D structures directly from 2D PX has recently been explored to address limitations with existing methods primarily reliant on Convolutional Neural Networks (CNNs) for direct 2D-to-3D mapping. These methods, however, are unable to correctly infer depth-axis spatial information. In addition, they are limited by the intrinsic locality of convolution operations, as the convolution kernels only capture the information of immediate neighborhood pixels. In this study, we propose a progressive hybrid Multilayer Perceptron (MLP)-CNN pyramid network (3DPX) for 2D-to-3D oral PX reconstruction. We introduce a progressive reconstruction strategy, where 3D images are progressively reconstructed in the 3DPX with guidance imposed on the intermediate reconstruction result at each pyramid level. Further, motivated by the recent advancement of MLPs that show promise in capturing fine-grained long-range dependency, our 3DPX integrates MLPs and CNNs to improve the semantic understanding during reconstruction. Extensive experiments on two large datasets involving 464 studies demonstrate that our 3DPX outperforms state-of-the-art 2D-to-3D oral reconstruction methods, including standalone MLP and transformers, in reconstruction quality, and also improves the performance of downstream angular misalignment classification tasks.

Keywords: Progressive Reconstruction, MLPs, Oral Panoramic X-ray.

1 Introduction

Panoramic X-ray (PX), an extra-oral imaging technique, is widely used in dental practices for diagnostic, assessment and monitoring purposes [1–4]. It generates stretched 2-dimensional (2D) images of the entire maxillomandibular area, enabling the evaluation of both teeth and jawbone, by rotating an X-ray emitter around the patient's head

* Xiaoshuang Li and Mingyuan Meng contributed equally to this work.

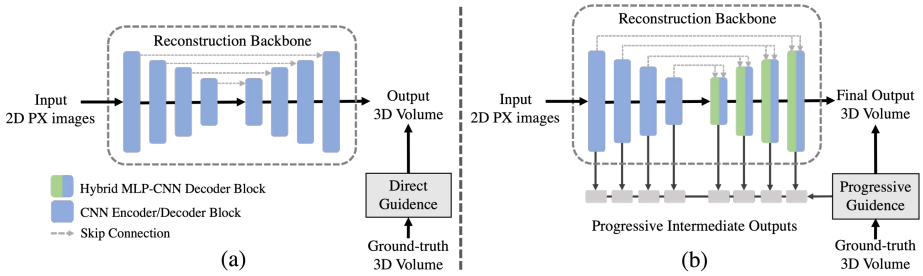

Fig. 1. A conceptual illustration of (a) regular encoder-decoder direct reconstruction architecture, and (b) our hybrid CNN-MLP progressive reconstruction architecture used in the 3DPX for 2D-to-3D panoramic X-ray reconstruction.

along a curved trajectory while capturing X-ray projections of the anatomical structures. When compared to other dental imaging modalities such as cone-beam computed tomography (CBCT), magnetic resonance imaging (MRI), and ultrasonography, PX has the advantages in lower cost, patient convenience, and lower radiation dose [5]. However, as a flat projection 2D image, PX lacks 3D anatomical information, which impedes accurate disease interpretation [6–8] and potential downstream tasks such as lesion segmentation, disease classification, and angular misalignment detection [9]. Recently, there has been great research interests in reconstructing 3D structures from 2D oral PX [5, 10], with the aim of extracting 3D information from standard 2D PX images.

Existing 2D-to-3D oral PX reconstruction methods primarily rely on using Convolutional Neural Networks (CNNs) to directly predict 3D structures from 2D PX images using a 2D encoder-decoder network architecture. By using 2D convolutional layers on 3D image data, the depth information of 3D images is processed as the feature channels in the 2D convolutional layers. Specifically, at the encoder side, the network gradually reduces the image width/height and infers the depth information by successively downsampling the feature size and increasing the feature channels. At the decoder side, the network gradually recovers the width and height information with the aid of skip connections while maintaining the number of feature channels. During this process, 2D CNNs have counter-intuitive advantages over 3D CNNs, as 2D convolutional kernels leverage all depth information in the feature channels at the same time, while 3D convolutional layers only leverage part of depth information within the 3D convolution kernels such that leads to an incomplete usage of the inferred depth information. Song et al. [5] developed Oral-3D, a Generative Adversarial Network (GAN) model with a Residual CNN generator. Liang et al. [10] proposed a CNN architecture to firstly segment the PX images and then generate voxelized teeth based on the segmentation masks. Nevertheless, existing CNN-based oral PX reconstruction methods still have two key unsolved limitations:

Firstly, existing 2D-to-3D reconstruction methods employ CNNs to directly map 2D PX images to 3D image volumes (Fig. 1a). However, this simple mapping underestimates the complexity of 2D-to-3D reconstruction and cannot fully handle the difficulties in inferring depth-axis spatial information from 2D images with only height and width axes. Further, the intermediate feature maps are not fully leveraged such that the reconstruction results often lack details and tend to generate artifacts.

Secondly, CNN-based methods are limited by the intrinsic locality of convolution operations. Although 2D CNN can leverage all the depth information, its receptive fields on the height and width axes are still limited. Multilayer Perceptrons (MLPs) have demonstrated strong capabilities in capturing fine-grained long-range dependence among high-resolution image details [11]. However, to capture long-range dependence with low computation complexity, MLPs decouple the spatial and channel dimensions and process them independently, such that has limited capabilities in preserving the coherence on the depth (i.e., channel) dimension when processing spatial information on the height and width dimensions.

In this study, we propose 3DPX, a progressive hybrid MLP-CNN pyramid network for 2D-to-3D oral PX reconstruction to overcome the two limitations mentioned above. Our 3DPX exten ds the definition of MLPs with a hybrid MLP-CNN module with a 2-step skip connection to reinforce the coherence of depth information. When compared to the current state-of-the-art methods, we introduce the following contributions: (i) we propose a progressive reconstruction strategy where the 3D images are progressively reconstructed in our 3DPX with progressive guidance imposed on the intermediate reconstruction results at each pyramid level (Fig. 1b), thus resulting in more fine-grained reconstruction; and (ii) Our 3DPX integrates the advantages of MLPs and CNNs, such that allows to capture long-range visual dependence and small subtle details, thus improving semantic understanding during reconstruction. Extensive experiments on public (Cui et al. [12]) and our private CBCT datasets demonstrate that our 3DPX can outperform state-of-the-art 2D-to-3D oral reconstruction methods in reconstruction quality and in downstream angular misalignment classification tasks.

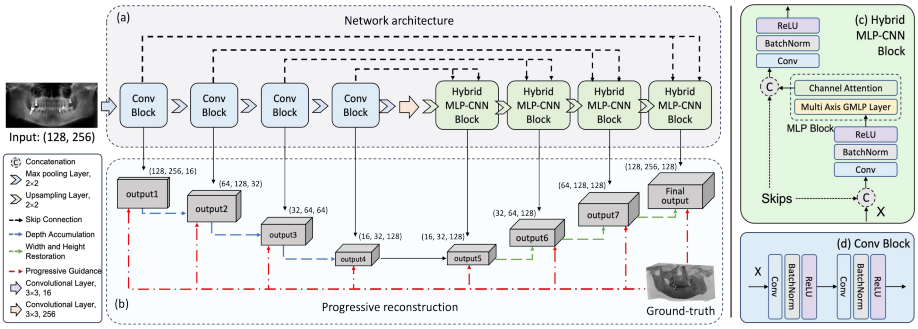

Fig. 2. The workflow of 3DPX for 2D-to-3D PX reconstruction. (a) Network architecture. (b) Progressive reconstruction process. (c) Hybrid MLP-CNN Block combining MLP and convolutional layers with a 2-step skip connection to preserve the coherence of depth information in 2D-to-3D reconstruction tasks.

2 Method

We propose 3DPX to reconstruct flattened 3D structure from a single PX image, with the intention of enhancing the PX image analysis in the downstream tasks. Fig. 2 shows the workflow of 3DPX. It takes a PX image (128×256) as input and reconstructs the corresponding 3D flattened structure (128×256×128). The reconstructed results are

used to perform 3D-enhanced 2D PX analysis. Our 3DPX introduces a progressive guided reconstruction (PGR) strategy with Hybrid MLP-CNN Blocks (HB).

3DPX is based on a customized U-Net structure. As is shown in Fig. 2, it consists of an encoder branch with four convolutional blocks, and a decoder branch with four Hybrid MLP-CNN Blocks. The hybrid MLP-CNN Block is applied only in the decoder branch. All decoder blocks have the same output channel size of 128, which corresponding to the depth dimension size.

2.1 Progressive Guided Reconstruction

In this section, we introduce the PGR strategy for 2D-to-3D PX reconstruction and its differences from the concept of deep supervision [13, 14]. Deep supervision adds auxiliary supervision signals to the intermediate network layers to facilitate the training convergence [15], where the intermediate outputs are used to calculate the losses but not used to compose the final prediction. In contrast, 2D-to-3D reconstruction presents a different scenario. Firstly, all feature maps of 3DPX are trained to imitate the 3D reconstruction successively at different scales, instead of learning low-level or high-level semantic features representation. Secondly, applying the progressive guidance with 3D ground truth on the intermediate outputs makes the feature maps into step-by-step reconstruction, where the intermediate reconstruction outputs (i.e., the feature maps) are directly used to facilitate the next step of reconstruction.

Let B be a given encoder or decoder block in 3DPX, our progressive reconstruction strategy introduces multiple guidance by applying penalties L on the intermediate output of B. Specifically, the penalty $L = L_{SSE}(f, Y)$ where $f = B(X)$ denotes the intermediate feature map, Y denotes the scaled label, and L_{SSE} represents the Error Sum of Square (SSE) loss. At each reconstruction stage i, Y_i was reshaped from the label into the size of f_i for a proper comparison, denoted with red arrows in Fig.2. The guided penalty is formulated as:

$$\begin{aligned} L_0(X) &= L_{SSE}(B_0(X), Y_0) \\ L_1(X) &= L_{SSE}(B_1 \circ B_0(X), Y_1) \\ &\cdots \\ L_i(X) &= L_{SSE}(B_i \circ B_{i-1} \circ \cdots \circ B_0(X), Y_i). \end{aligned} \tag{1}$$

As intermediate reconstructions progressively improve and approach the final output, the intermediate penalties should accordingly have progressive weight to emphasize the guidance close to the output layer and downplay the role of guidance close to the input layer. To achieve this, a set of hyper-parameters $\boldsymbol{\alpha}$ is set on $L_i(X)$ and the final training loss function for progressively guided reconstruction is L_{PR} is formulated as

$$L_{PR} = \sum_{i=0}^{n-1} \alpha_i \cdot L_i(X). \tag{2}$$

where n is the number of the encoder and decoder blocks. In the experiment, we set $\alpha_i = 2^{n-1-i}$ and for $i \leq 2$, α_i is empirical set to 0 to get the best reconstruction quality. Please refers to Supplementary Table. S1 for details about the ablative study of $\boldsymbol{\alpha}$.

2.2 3DPX with Hybrid MLP-CNN Block

The architecture of 3DPX is illustrated in Fig. 2. It's a specially designed U-Net structure consists of 2D convolutional (Conv) blocks in the encoder and Hybrid MLP-CNN blocks in the decoder. The encoder starts with a single convolutional layer that maps the single-channel input to a feature space with size of 16. The details of Conv block are shown in Fig. 2(d), which consists of two convolutional layers followed by Batch Normalization (BN) and ReLU activation. The output feature size in Fig. 2(a) is shown in an order of width, height, and feature channel. The encoder increases the size of depth channel to 128 and decrease the width and height to [16, 32]. The encoder and decoder are connected by a bottleneck convolutional layer that increases the feature channel size to 256. The Hybrid MLP-CNN block depicted in Fig. 2(c) combines MLP block proposed by Tu et al [16] and convolutional layer together. Before MLP block, the first convolutional layer fuses the features that come from the former layer and the skip connection. In the MLP block, the multi-axis gated MLP layer [16] enables effective interactions between different feature spatial dimensions and capture both local and long-range dependencies of the input features. Then, following [16], a channel attention mechanism is used to weight the importance of different feature channels and improve the concentration on some channels while suppressing on others. The second convolutional layer do the same movement with another skip connection to integrate the long-range attention information and recover the depth coherence. The output channel size of all these layers is 128. Max pooling layers and upsampling layers with a kernel size of 2×2 are used between adjacent blocks for downsampling and upsampling.

2.3 Evaluation Metric

The 2D-to-3D reconstruction task was evaluated with Peak signal-to-noise ratio (PSNR/dB), Structure similarity index (SSIM/%) and Dice similarity coefficient (DSC/%); these are standard metrics to assess image reconstruction results. DSC score was computed by extracting jaw bones from soft tissues. The bone volume was obtained by applying a threshold to the density representation. During the experiments, we set the threshold to the mean density of the ground-truth flattened 3D structure.

We further evaluated the reconstructed results via downstream application, which is on 3D-enhanced angular misalignment classification task. The classification performance was evaluated by accuracy, precision, recall, and F1 score.

3 Experiments and Results

3.1 Dataset and Augmentation

We used 464 CBCT scans where 91 scans were released by Cui et al. [12], and 373 were from our private dataset from the School of Stomatology, Wuhan University (WHU). Cui et al. data were scanned in routine clinical care, where patient required dental treatments such as orthodontics, dental implants, or restoration. The images were acquired at 400×400, with a varying height of ~280 pixels, at an interslice distance

of $0.4 \times 0.4 \times 0.4\ mm^3$. For our private CBCT dataset, the original resolutions were 512×512, with a varying height of ~512 pixels, at an interslice distance of $0.3 \times 0.3 \times 0.3\ mm^3$.

For PX reconstruction, dental arch curves were manually marked on all CBCTs under the guidance of an experienced dentist. The synthesized PX is reconstructed by projecting the CBCT along the dental arch trajectory, in the depth direction with a 0.2 mm unit size, encompassing a depth range of 40 mm and a height of 100 mm, with the width matching the length of the curve, typically around 200 mm. This region was reformatted into a flat 3D structure using the curved planar reformat method. In clinical practice, the misplacement of patients' head causes angular misalignment of captured images. We augmented our dataset to acquire angular misalignment PX and corresponding flattened 3D structure by vertically and laterally rotating the CBCT scans according to Kwon *et al.* [17].

Following data augmentation, we obtained 2922 pairs of PX images with dimensions of [128, 256] and corresponding 3D flattened structures with dimensions of [128, 256, 128]. These pairs were divided into a training set consisting of 2060 samples, an evaluation set consisting of 412 sample, and a test set consisting of 450 sample.

Table 1. The comparison results between the proposed 3DPX and existing 2D-to-3D reconstruction models. Methods marked with ~ were customized from existing methods and/or with our proposed innovations to fit the 2D-to-3D reconstruction scenario. Best results are **bolded** and second-best underlined.

Architecture	U-Net based			Residual CNN based		
	PSNR	DSC	SSIM	PSNR	DSC	SSIM
Transformer (~UNETR[18])	14.76	60.57	60.3	-	-	-
CNN (~U-Net[19])	14.76	62.22	67.72	15.21	61.17	70.58
CNN GAN (Oral-3D[5])	14.69	62.61	68.97	15.26	60.75	68.47
Hybrid MLP-CNN (~)	14.99	62.2	68.55	15.42	<u>61.64</u>	<u>72.25</u>
Hybrid MLP-CNN GAN (~)	15.11	<u>63.7</u>	68.42	15.23	60.96	69.02
Progressive Hybrid MLP-CNN GAN(~)	<u>15.51</u>	63.21	<u>72.17</u>	<u>15.45</u>	60.69	71.22
Progressive Hybrid MLP-CNN (**3DPX**)	**15.84**	**63.72**	**74.09**	**15.73**	**62.01**	**73.45**

3.2 Experiments

The proposed 3DPX is compared with existing 2D-to-3D reconstruction models with the reconstruction results in Table 1. Pure CNN-based customized U-Net achieved SSIM of 67.72% and DSC of 62.22% on bone segmentation. The transformer-based customized U-Net (UNETR), however, deteriorated the reconstructions quality greatly, resulting in a decrease of 2% in DSC and more than 7% in SSIM. Compared to the U-Net, Residual CNN exceled in retaining the gradient throughout convolution blocks due to its residual connection and therefore achieved a higher SSIM of 70.58%. The integration of GAN training strategy, hybrid MLP-CNN blocks, and progressive guidance, improved the reconstruction results. Specifically, the GAN strategy elevated the SSIM to 68.97% for the U-Net but decreased it to 68.47% for the Residual CNN in Oral-3D. The introduction of the Hybrid MLP-CNN blocks enhanced both basic models, notably increasing the SSIM for the Residual CNN to 72.25% and the DSC to 61.64%. The

Hybrid MLP-CNN introduces both the expansion of the receptive field and a detrimental effect on depth coherence. Without progressive guidance, only a 2-step skip connection provided the depth restoration for the U-Net. However, for Residual CNN, both the skip and the residual connections helped to maintain depth coherence. Despite this, the GAN strategy failed to deliver improvement with the presence of hybrid MLP-CNN blocks, both with and without progressive guidance. With progressive intermediate guidance, 3DPX significantly outperformed the U-Net by over 6% in terms of SSIM, 7% in terms of DSC, and 1.1% in PSNR. In contrast, we only observed a moderate improvement with Residual CNN, where SSIM was increased by 2%. When progressive guidance sufficiently restored intermediate depth coherence, the residual connection lost its efficacy and failed to surpass the more straightforward designed counterpart.

Table 2. Ablative study of MLP and Hybrid CNN-MLP Blocks (HB) and PGR. Best results are **bolded** and second-best results are underlined.

Method	PSNR(dB)	DSC(%)	SSIM(%)
Baseline (customized U-Net)	14.76	62.22	67.72
Baseline + 3D Decoder	14.63	**64.32**	62.82
Baseline + MLP Decoder	14.67	59.72	60.09
Baseline + HB Decoder	15.51	63.21	71.21
Baseline + HB Decoder + PGR (3DPX)	**15.84**	63.72	**74.09**

Table 2 shows the results of the ablation study. The proposed Hybrid MLP-CNN Block was more effective than MLP or the 3D convolutional blocks. The results of the 3D decoder U-Net showed a decrease when compared to the 2D customized U-Net on SSIM and PSNR. As the 3D decoder focused on the local area within the 3D convolutional kernel, the local continuity of the area with high bone density was improved according to DSC at 64.32%. Standalone MLP decoder deteriorated the spatial continuity of reconstructions on depth channel, resulting in a decrease of 3% in DSC and more than 7% in SSIM. Hybrid CNN-MLP blocks restored and further improved the reconstruction quality based on the MLP blocks to an SSIM of 71.21% and a DSC of 63.21%. With the progressive intermediate guidance, our 3DPX achieved the highest score on all three metrics.

We further conducted downstream experiments on angular misalignment classification task for the PX images. The comparison happens between a 2D ResNet and joint

Table 3. The comparison results of 2D ResNet and joint 3DPX on PX angular misalignment classification task with binary and 5 categories.

Task	Method	Acc (%)	Precision	Recall	F1 score
2-class	2D	85.8	0.795	0.733	0.757
	3DPX	**88.4**	**0.822**	**0.822**	**0.822**
5-class	2D	88.9	0.88	**0.885**	**0.881**
	3DPX	**89.8**	**0.895**	0.873	0.88

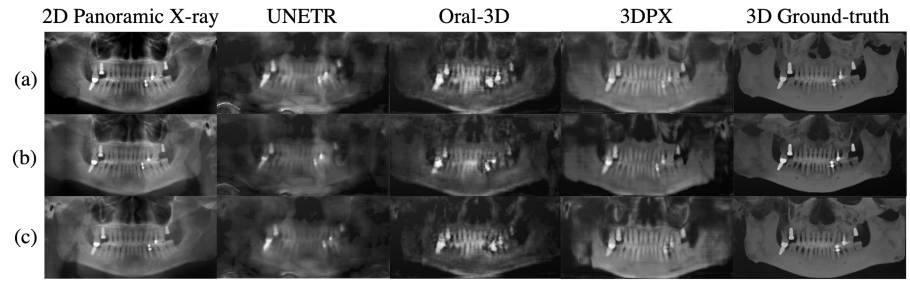

Fig. 3. Visualizations of reconstructed 3D flattened structure from PX images from comparison methods. Two types of angular misalignment augmentation are depicted, (a) regular PX capturing angle, (b) PX with left rotation misalignment, and (c) PX with right rotation misalignment.

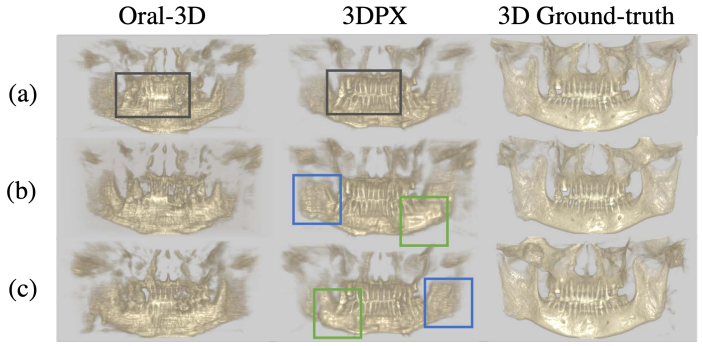

Fig. 4. Volume rendering of reconstructed flattened 3D structure of Oral-3D, 3DPX and the ground-truth. Two types of misalignment augmentation are depicted, (a) regular PX capturing angle, (b) PX with left rotation misalignment, and (c) PX with right rotation misalignment.

2D-3D ResNet designed for pure PX and 3D-enhanced PX classification. Improvement in accuracy and other metrics is observed on both binary and 5-category angular misalignment classification. Table 3 provides more detailed results about the downstream classification task results.

Fig. 3 visualizes the reconstructed flattened 3D structure against the comparison methods on three angles during capturing PX of a patient. The UNETR was unable to generate detailed tooth structure and clear boundary between bone and soft tissue and produced incoherent anatomic structure especially in the depth channels. Fig. 4 illustrates 3D volume rendering of the same patient with Oral-3D and 3DPX. The regions highlighted by the black box highlights that the 3DPX generated relatively fine-grained details for the anterior teeth and tooth implants. Additionally, it produced clearer expansions of one side of the jawbone ramus in the blue box and the other side of the jawbone body in the green box, which were caused by rotation misalignment during PX capturing. Although parts of the anatomy are missing for both methods, the important tooth sections are well preserved in 3DPX.

4 Conclusion

In this study, we propose a progressive hybrid Multilayer Perceptron (MLP)-CNN pyramid network (3DPX) for 2D-to-3D oral PX reconstruction. Extensive experiments with two large datasets involving 464 studies demonstrate that our 3DPX outperforms state-of-the-art 2D-to-3D oral reconstruction methods, including standalone MLP and transformers, in reconstruction quality, and also improves the performance of downstream angular misalignment classification tasks.

Acknowledgments. This study was funded by National Natural Science Foundation of China (62272298), Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102) and China Scholarship Council (202006230239).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Katsumata, A.: Deep learning and artificial intelligence in dental diagnostic imaging. Japanese Dental Science Review. 59, 329–333 (2023). https://doi.org/10.1016/j.jdsr.2023.09.004.
2. The use of dental radiographs: Update and recommendations - ScienceDirect, https://www.sciencedirect.com/science/article/pii/S0002817714643221, last accessed 2023/11/29.
3. Shahidi, S., Zamiri, B., Abolvardi, M., Akhlaghian, M., Paknahad, M.: Comparison of Dental Panoramic Radiography and CBCT for Measuring Vertical Bone Height in Different Horizontal Locations of Posterior Mandibular Alveolar Process. J Dent (Shiraz). 19, 83–91 (2018).
4. Różyło-Kalinowska, I.: Panoramic radiography in dentistry. Clin Dent Rev. 5, 26 (2021). https://doi.org/10.1007/s41894-021-00111-4.
5. Song, W., Liang, Y., Yang, J., Wang, K., He, L.: Oral-3D: Reconstructing the 3D Structure of Oral Cavity from Panoramic X-ray. Proceedings of the AAAI Conference on Artificial Intelligence. 35, 566–573 (2021). https://doi.org/10.1609/aaai.v35i1.16135.
6. Applied Sciences | Free Full-Text | Basic Knowledge and New Advances in Panoramic Radiography Imaging Techniques: A Narrative Review on What Dentists and Radiologists Should Know, https://www.mdpi.com/2076-3417/11/17/7858, last accessed 2023/11/29.
7. Estrela, C., Bueno, M.R., Azevedo, B.C., Azevedo, J.R., Pécora, J.D.: A new periapical index based on cone beam computed tomography. Journal of endodontics. 34, 1325–1331 (2008).
8. Accuracy of Cone-Beam Computed Tomography and Periapical Radiography in Detecting Small Periapical Lesions - ScienceDirect, https://www.sciencedirect.com/science/article/pii/S0099239912002828, last accessed 2023/11/14.
9. Delamare, E., Fu, X., Huang, Z., Kim, J.: Panoramic imaging errors in machine learning model development: a systematic review. Dentomaxillofac Radiol. https://doi.org/10.1093/dmfr/twae002.
10. Liang, Y., Song, W., Yang, J., Qiu, L., Wang, K., He, L.: X2Teeth: 3D Teeth Reconstruction from a Single Panoramic Radiograph. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D.,

Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., and Joskowicz, L. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 400–409. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_39.

11. Meng, M., Xue, Y., Feng, D., Bi, L., Kim, J.: Full-resolution MLPs Empower Medical Dense Prediction, http://arxiv.org/abs/2311.16707, (2023). https://doi.org/10.48550/arXiv.2311.16707.
12. Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., Jiang, C., Sun, Y., Ma, L., Huang, J., Liu, Y., Zhao, Y., Lian, C., Ding, Z., Zhu, M., Shen, D.: A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat Commun. 13, 2096 (2022). https://doi.org/10.1038/s41467-022-29637-2.
13. Wang, L., Lee, C.-Y., Tu, Z., Lazebnik, S.: Training Deeper Convolutional Networks with Deep Supervision, http://arxiv.org/abs/1505.02496, (2015). https://doi.org/10.48550/arXiv.1505.02496.
14. Deep Supervision with Intermediate Concepts | IEEE Journals & Magazine | IEEE Xplore, https://ieeexplore.ieee.org/abstract/document/8434117, last accessed 2024/03/07.
15. Contrastive Deep Supervision | SpringerLink, https://link.springer.com/chapter/10.1007/978-3-031-19809-0_1, last accessed 2024/03/07.
16. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y.: MAXIM: Multi-Axis MLP for Image Processing. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022).
17. Kwon, T., Choi, D., Hwang, J., Lee, T., Lee, I., Cho, S.: Panoramic dental tomosynthesis imaging by use of CBCT projection data. Sci Rep. 13, 8817 (2023). https://doi.org/10.1038/s41598-023-35805-1.
18. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: UNETR: Transformers for 3D Medical Image Segmentation. Presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (2022).
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., and Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015).