3D Teeth Reconstruction from Panoramic Radiographs Using Neural Implicit Functions

Sihwa Park[1], Seongjun Kim[1], In-Seok Song[2], and Seung Jun Baek[1(✉)]

[1] Korea University, Seoul, South Korea
{sihwapark,iamsjune,sjbaek,densis}@korea.ac.kr
[2] Korea University Anam Hospital, Seoul, South Korea

Abstract. Panoramic radiography is a widely used imaging modality in dental practice and research. However, it only provides flattened 2D images, which limits the detailed assessment of dental structures. In this paper, we propose Occudent, a framework for 3D teeth reconstruction from panoramic radiographs using neural implicit functions, which, to the best of our knowledge, is the first work to do so. For a given point in 3D space, the implicit function estimates whether the point is occupied by a tooth, and thus implicitly determines the boundaries of 3D tooth shapes. Firstly, Occudent applies multi-label segmentation to the input panoramic radiograph. Next, tooth shape embeddings as well as tooth class embeddings are generated from the segmentation outputs, which are fed to the reconstruction network. A novel module called Conditional eXcitation (CX) is proposed in order to effectively incorporate the combined shape and class embeddings into the implicit function. The performance of Occudent is evaluated using both quantitative and qualitative measures. Importantly, Occudent is trained and validated with actual panoramic radiographs as input, distinct from recent works which used synthesized images. Experiments demonstrate the superiority of Occudent over state-of-the-art methods.

Keywords: Panoramic radiographs · 3D reconstruction · Teeth segmentation · Neural implicit function

1 Introduction

Panoramic radiography (panoramic X-ray, or PX) is a commonly used technique for dental examination and diagnosis. While PX produces 2D images from panoramic scanning, Cone-Beam Computed Tomography (CBCT) is an alternative imaging modality which provides 3D information on dental, oral, and maxillofacial structures. Despite providing more comprehensive information than

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_36.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14229, pp. 376–386, 2023.
https://doi.org/10.1007/978-3-031-43999-5_36

PX, CBCT is more expensive and exposes patients to a greater dose of radiation [3]. Thus, 3D teeth reconstruction from PX is of significant value, e.g., 3D visualization can aid clinicians with dental diagnosis and treatment planning. Other applications include treatment simulation and interactive virtual reality for dental education [12].

Previous 3D teeth reconstruction methods from 2D PX have relied on additional information such as tooth landmarks or tooth crown photographs. For example, [14] developed a model which uses landmarks on PX images to estimate 3D parametric models for tooth shapes, while [1] reconstructed a single tooth using a shape prior and reflectance model based on the corresponding crown photograph. Recent advances in deep neural networks have significantly impacted research on 3D teeth reconstruction. X2Teeth [13] performs 3D reconstruction of the entire set of teeth from PX based on 2D segmentation using convolutional neural networks. Oral-3D [22] generated 3D oral structures without supervised segmentation from PX using a GAN model [8]. Yet, those methods relied on synthesized images as input instead of real-world PX images, where the synthesized images are obtained from 2D projections of CBCT [27]. The 2D segmentation of teeth from PX is useful for 3D reconstruction in order to identify and isolate teeth individually. Prior studies on 2D teeth segmentation [11,28] focused on binary segmentation determining the presence of teeth. However, this information alone is insufficient for the construction of individual teeth. Instead, we leverage recent frameworks [17,21] on multi-label segmentation of PX into 32 classes including wisdom teeth.

In this paper, we propose *Occudent*, an end-to-end model to reconstruct 3D teeth from 2D PX images. Occudent consists of a multi-label 2D segmentation followed by 3D teeth reconstruction using *neural implicit functions* [15]. The function aims to learn the ***occu***pancy of ***dent***al structures, i.e., whether a point in space lies within the boundaries of 3D tooth shapes. Learning implicit functions is computationally advantageous over conventional encoder-decoder models outputting explicit 3D representations such as voxels, e.g., implicit models do not require large memory footprints to store and process voxels. Considering that 3D tooth shapes are characterized by tooth classes, we generate embeddings for tooth classes as well as segmented 2D tooth shapes. The combined class and shape embeddings are infused into the reconstruction network by a novel module called Conditional eXcitation (CX). CX performs learnable scaling of occupancy features conditioned on tooth class and shape embeddings. The performance of Occudent is evaluated with actual PX as input images, which differs from recent works using synthesized PX images [13,22]. Experiments show Occudent outperforms state-of-the-art baselines both quantitatively and qualitatively. The main contributions are summarized as follows: (1) the first use of a neural implicit function for 3D teeth reconstruction, (2) novel strategies to inject tooth class and shape information into implicit functions, (3) the superiority over existing baselines which is demonstrated with real-world PX images.

2 Methods

The proposed model, Occudent, consists of two main components: 2D teeth segmentation and 3D teeth reconstruction. The former performs the segmentation of 32 teeth from PX using UNet++ model [29]. The individually segmented tooth and the tooth class are subsequently passed to the latter for the reconstruction. The reconstruction process estimates the 3D representation of the tooth based on a neural implicit function. The overall architecture of Occudent is depicted in Fig. 1.

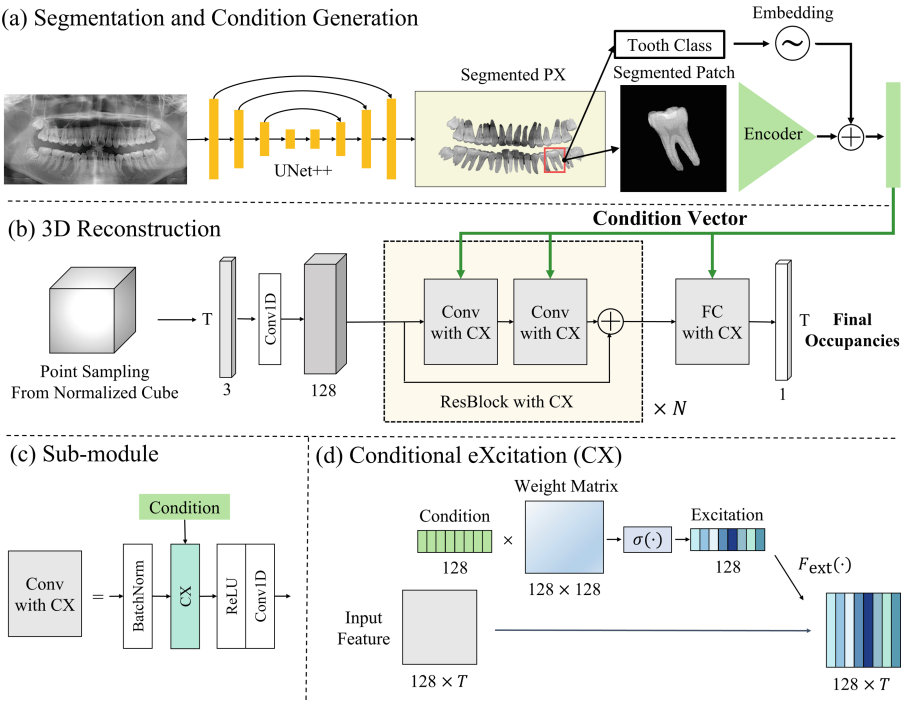

Fig. 1. (a) The PX image is segmented into 32 teeth classes using UNet++ model. For each tooth, a segmented patch is generated by cropping input PX with the predicted segmentation mask, which is subsequently encoded via an image encoder. The tooth class is encoded by an embedding layer. The patch and class embeddings are added together to produce the condition vector. (b) The reconstruction process consists of N ResBlocks to compute occupancy features of points sampled from 3D space. The condition vector from PX is processed by a Conditional eXcitation (CX) module incorporated in the ResBlocks. (c) The Conv with CX sub-module is composed of batch normalization, CX, ReLU, and a convolutional layer. The FC with CX is similar to the Conv with CX where Conv1D layer is replaced by fully connected layer. (d) CX injects condition information into the reconstruction network using excitation values. CX uses a trainable weight matrix to encode the condition vector into an excitation vector via a gating function. The input feature is scaled using the excitation vector through component-wise multiplication.

2.1 2D Teeth Segmentation

The teeth in input PX are segmented into 32 teeth classes. The 32 classes correspond to the traditional numbering of teeth, which includes incisors, canines, premolars and molars in both upper and lower jaws. We pose 2D teeth segmentation as a *multi-label segmentation* problem [17], since nearby teeth can overlap with each other in the PX image, i.e., a single pixel of the input image can be classified into two or more classes.

The input of the model is $H \times W$ size PX image. The segmentation output has dimension $C \times H \times W$, where channel dimension $C = 33$ represents the number of tooth classes: one class for the background and 32 classes for teeth similar to [17]. Hence, the $H \times W$ output at each channel is a segmentation output for each tooth class. The segmentation outputs are used to generate tooth patches for reconstruction, which is explained later in detail. For the segmentation, we adopt pre-trained UNet++ [29] as the base model. UNet++ is advantageous for medical image segmentation due to its modified skip pathways, which results in better performance compared to the vanilla UNet [20].

2.2 3D Teeth Reconstruction

Neural Implicit Representation. Typical representations of 3D shapes are point-based [7,19], voxel-based [4,26], or mesh-based methods [25]. These methods represent 3D shapes explicitly through a set of discrete points, vertices, and faces. Recently, implicit representation methods based on a continuous function which defines the boundary of 3D shapes have become increasingly popular [15,16,18]. Occupancy Networks [15] is a pioneering work which utilizes neural networks to approximate the implicit function of an object's occupancy. The term occupancy refers to whether a point in space lies in the interior or exterior of object boundaries. The occupancy function maps a 3D point to either 0 or 1, indicating the occupancy of the point. Let o_{A} denote the occupancy function for an object A as follows:

$$o_{\mathrm{A}} : \mathbb{R}^3 \rightarrow \{0, 1\} \tag{1}$$

In practice, o_{A} can be estimated only by a set of observations of object A, denoted by $\mathcal{X}_{\mathrm{A}}$. Examples of observations are projected images or point cloud data obtained from the object. Our objective is to estimate the occupancy function conditioned on $\mathcal{X}_{\mathrm{A}}$. Specifically, we would like to find function f_θ which estimates the occupancy probability of a point in 3D space based on $\mathcal{X}_{\mathrm{A}}$ [15]:

$$f_\theta : \mathbb{R}^3 \times \mathcal{X}_{\mathrm{A}} \rightarrow [0, 1] \tag{2}$$

Inspired by the aforementioned framework, we leverage segmented tooth patch and tooth class as observations denoted by condition vector $\boldsymbol{c}$. Specifically, the input to the function is a set of T randomly sampled locations within a unit cube, and the function outputs the occupancy probability of the input. Thus, the function is given by $f_\theta : (x, y, z, \boldsymbol{c}) \rightarrow [0, 1]$.

The model for f_θ is depicted in Fig. 1 (b). The sampled locations are projected to 128 dimensional feature vectors using 1D convolution. Next, the features are processed by a sequence of ResNet blocks followed by FC (fully connected) layers. Conditional vector $\boldsymbol{c}$ is used for each block through Conditional eXcitation (CX) which we will explain later.

Class-Specific Conditional Features. A distinctive feature of the tooth reconstruction task is that teeth with the same number share properties such as surface and root shapes. Hence, we propose to use tooth class information in combination with a segmented tooth patch from PX. The tooth class is processed by a learnable embedding layer which outputs a class embedding vector.

Next, we create a square patch of the tooth using the segmentation output as follows. A binary mask of the segmented tooth is generated by applying thresholding to the segmentation output. A tooth patch is created by cropping out the tooth region from the input PX, i.e., the binary mask is applied (bitwise AND) to the input PX to obtain the patch. The segmented tooth patch is subsequently encoded using a pre-trained ResNet18 model [9], which outputs a patch embedding vector. The patch and class embeddings are added to yield the condition vector for the reconstruction model. This process is depicted in Fig. 1 (a).

Our approach differs from previous approaches, such as Occupancy Networks [15] which uses only single-view images for 3D reconstruction. X2Teeth [13] also addresses the task of 3D teeth reconstruction from 2D PX. However, X2Teeth only uses segmented image features for the reconstruction. By contrast, Occudent leverages a class-specific encoding method to boost the reconstruction performance, as demonstrated in ablation analysis in Supplementary Materials.

Conditional eXcitation. To effectively inject 2D observations into the reconstruction network, we propose Conditional eXcitation (CX) inspired by Squeeze-and-Excitation Network (SENet) [10]. In SENet, excitation refers to scaling input features according to their importance. In Occudent, the concept of excitation is extended to incorporating conditional features into the network. Firstly, the condition vector is encoded into excitation vector $\boldsymbol{e}$. Next, the excitation is applied to input feature by scaling the feature components by $\boldsymbol{e}$. The CX procedure can be expressed as:

$$\boldsymbol{e} = \alpha \cdot \sigma(W\boldsymbol{c}), \tag{3}$$

$$\boldsymbol{y} = F_{\text{ext}}(\boldsymbol{e}, \boldsymbol{x}) \tag{4}$$

where $\boldsymbol{c}$ is the condition vector, σ is a gating function, W is a learnable weight matrix, α is a hyperparameter for the excitation result, and F_{ext} is the excitation function. We use sigmoid function for σ, and component-wise multiplication for the excitation, $F_{\text{ext}}(\boldsymbol{e}, \boldsymbol{x}) = \boldsymbol{e} \otimes \boldsymbol{x}$. The CX module is depicted in Fig. 1 (d). CX differs from SENet in that CX derives the excitation from the condition vector, whereas SENet derives it from input features. Our approach also differs from Occupancy Networks which used Conditional Batch Normalization (CBN) [5,6]

which combines conditioning with batch normalization. However, the conditioning process should be independent of input batches because those components serve different purposes in deep learning models. Thus, we propose to separate conditioning from batch normalization, as is done by CX.

3 Experiments

Dataset. The pre-training of the segmentation model was done with a dataset of 4000 PX images, sourced from 'The Open AI Dataset Project (AI-Hub, S. Korea)'. All data information can be accessed through 'AI-Hub (www.aihub.or.kr)'. The dataset consisted of two image sizes, 1976×976 and 2988×1468, which were resized to 256×768 to train the UNet++ model.

For the main experiments for reconstruction, we used a set of 39 PX images and matched CBCT images, obtained from Korea University Anam Hospital. This study was approved by the Institutional Review Board at Korea University (IRB number: 2020AN0410). The panoramic radiographs were of dimensions 1536×2860 and were resized to 600×1200 and randomly cropped of 592×1184 size for the segmentation training. The CBCT images were of size $768 \times 768 \times 576$, capturing cranial bones. The teeth labels for 2D PX and CBCT were manually annotated by two experienced annotators and subsequently verified by a board-certified dentist. To train and evaluate the model, the dataset was partitioned into training (30 cases), validation (2 cases), and testing (7 cases) subsets.

Implementation Details. For the pre-training of the segmentation model, we utilized a combination of cross-entropy and dice loss. For the main segmentation training, we used only dice loss. The segmentation and reconstruction models were trained separately. Following the completion of the segmentation model training, we fixed this model to predict its output for the reconstruction model.

Each 3D tooth label was fit in $144 \times 80 \times 80$ size tensor which was then regarded as $[-0.5, 0.5]^3$ normalized cube in 3D space. For the training of the neural implicit function, a set of $T = 100,000$ points was sampled from the unit cube. The preprocessing was consistent with that used in [23]. We trained all the other baseline models with these normalized cubes. For example, for 3D-R2N2 [4], we voxelized the cube to 128^3 size. For a fair comparison, the final meshes produced by each model were extracted and compared using four different metrics. The detailed configuration of our model is provided in Supplementary Materials.

Baselines. We considered several state-of-the-art models as baselines, including 3D-R2N2 [4], DeepRetrieval [13,24], Pix2Vox [26], PSGN [7], Occupancy Networks (OccNet) [15], and X2Teeth [13]. To adapt the 3D-R2N2 model to single-view reconstruction, we removed its LSTM component, following the approach in [15]. As for the DeepRetrieval method, we employed the same encoder architecture as 3D-R2N2, and utilized the encoded feature vector of the test image

Table 1. Comparison with baseline methods. The format of results is *mean*±*std* obtained from 10 repetitions of experiments.

Method	IoU	Chamfer-L_1	NC	Precision
3D-R2N2	0.585 ± 0.005	0.382 ± 0.008	0.617 ± 0.009	0.634 ± 0.010
PSGN	0.606 ± 0.015	0.342 ± 0.016	0.829 ± 0.012	0.737 ± 0.018
Pix2Vox	0.562 ± 0.005	0.388 ± 0.008	0.599 ± 0.007	0.664 ± 0.009
DeepRetrieval	0.564 ± 0.005	0.394 ± 0.006	0.824 ± 0.003	0.696 ± 0.005
X2Teeth	0.592 ± 0.006	0.361 ± 0.017	0.618 ± 0.002	0.670 ± 0.009
OccNet	0.611 ± 0.006	0.353 ± 0.008	0.872 ± 0.003	0.691 ± 0.011
Occudent (Ours)	$\mathbf{0.651 \pm 0.004}$	$\mathbf{0.298 \pm 0.006}$	$\mathbf{0.890 \pm 0.001}$	$\mathbf{0.739 \pm 0.008}$

query. Subsequently, we compared each encoded vector from the test image to the encoded vectors from the training set, and retrieved the tooth with the minimum Euclidean distance of encoded vectors from the training set.

Evaluation Metrics. The evaluation of the proposed method was conducted using the following metrics: volumetric Intersection over Union (IoU), Chamfer-L_1 distance, and Normal Consistency (NC), as outlined in prior work [15]. In addition, we used volumetric precision [13] as a metric given by $|D \cap G|/|D|$ where G denotes the ground-truth set of points occupied by the object, and D denotes the set of points predicted as the object.

Quantitative Comparison. Table 1 presents a quantitative comparison of the proposed model with several baseline models. The results demonstrate that Occudent surpasses the other methods across all the metrics, and the methods based on neural implicit functions (Occudent and OccNet) perform better compared to conventional encoder-decoder approaches, such as Pix2Vox and 3D-R2N2. The performance gap between Occudent and X2Teeth is presumably because real PX images are used as input data. X2Teeth used *synthesized* images generated from the 2D projections of CBCT in [27]. Thus, both the input 2D shape and the target 3D shape come from the same modality (CBCT). However, the distribution of real PX images may differ significantly from that of 2D-projected CBCT. Explicit methods can be more sensitive to such differences than implicit methods, because typically in explicit methods, input features are directly encoded and subsequently decoded to predict the target shapes [4,13,26]. Overall, the differences in the IoU performances among the baselines are somewhat small. This is because all the baselines are moderately successful in generating coarse tooth shapes. However, Occudent is significantly better at generating details such as root shapes, which will be shown in the subsequent section.

Qualitative Comparison. Figure 2A illustrates the qualitative results of the proposed method in generating 3D teeth mesh outputs. From our model, each tooth is generated, and generated teeth are combined along with an arch curve based on a beta function [2]. Figure 2A demonstrates that our proposed method generates the most similar-looking outputs compared to the ground truth. For instance, our model can reconstruct a plausible shape for all tooth types including detailed shapes of molar roots. 3D-R2N2 produces larger and less detailed tooth shapes. PSGN and OccNet are better at generating rough shapes than 3D-R2N2, however, lack in detailed root shapes.

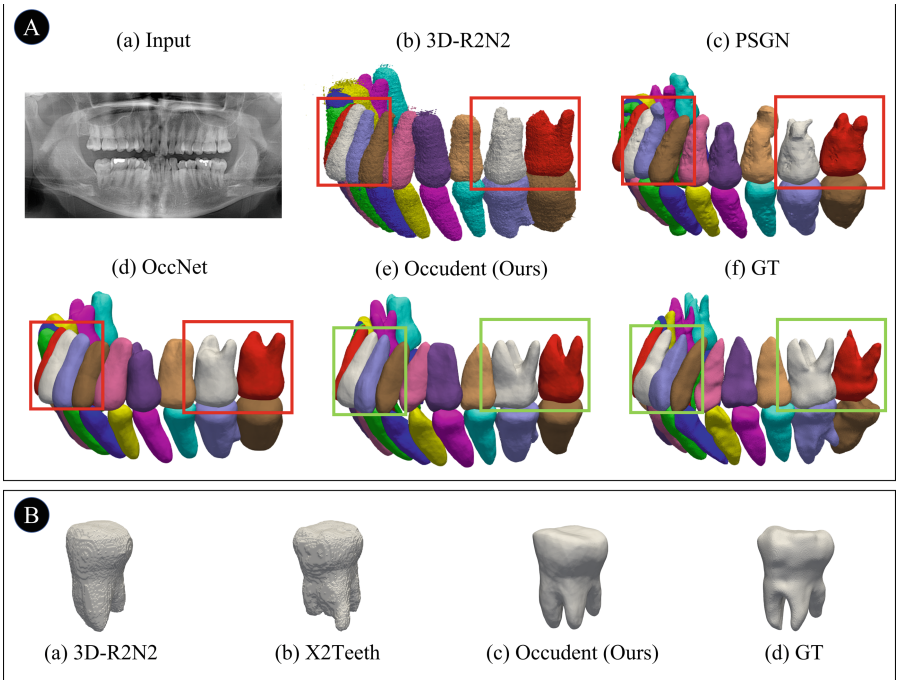

Fig. 2. Visual representation of sample outputs. Boxes are used to highlight the incisors and molars in the upper jaw.

As illustrated in Fig. 2B, Occudent produces a more refined mesh of tooth shape representation than voxel-based methods like 3D-R2N2 or X2Teeth. One of the limitations of voxel-based methods is that they heavily depend on the resolution of the output. For example, increasing the output size leads to an exponential increase in model size. By contrast, Occudent employs continuous neural implicit functions to represent shape boundaries, which enables us to generate smoother output and to be robust to the target size.

4 Conclusion

In this paper, we present a framework for 3D teeth reconstruction from a single PX. To the best of our knowledge, our method is the first to utilize a neural implicit function for 3D teeth reconstruction. The performance of our proposed framework is evaluated quantitatively and qualitatively, demonstrating its superiority over state-of-the-art techniques. Importantly, our framework is capable of accommodating two distinct modalities, PX, and CBCT. Our framework has the potential to be valuable in clinical practice and also can support virtual simulation or educational tools. In the future, further improvements can be made, such as incorporating additional imaging modalities or exploring neural architectures for more robust reconstruction.

Acknowledgements. This work was supported by the Korea Medical Device Development Fund grant funded by the Korea Government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711195279, RS-2021-KD000009); the National Research Foundation of Korea (NRF) Grant through the Ministry of Science and ICT (MSIT), Korea Government, under Grant 2022R1A5A1027646; the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1007215); the MSIT, Korea, under the ICT Creative Consilience program (IITP-2023-2020-0-01819) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation)

References

1. Abdelrehim, A.S., Farag, A.A., Shalaby, A.M., El-Melegy, M.T.: 2D-PCA shape models: application to 3D reconstruction of the human teeth from a single image. In: Menze, B., Langs, G., Montillo, A., Kelm, M., Müller, H., Tu, Z. (eds.) MCV 2013. LNCS, vol. 8331, pp. 44–52. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05530-5_5
2. Braun, S., Hnat, W.P., Fender, D.E., Legan, H.L.: The form of the human dental arch. Angle Orthod. **68**(1), 29–36 (1998)
3. Brooks, S.L.: CBCT dosimetry: orthodontic considerations. In: Seminars in Orthodontics, vol. 15, pp. 14–18. Elsevier (2009)
4. Choy, C.B., Xu, D., Gwak, J.Y., Chen, K., Savarese, S.: 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 628–644. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_38
5. De Vries, H., Strub, F., Mary, J., Larochelle, H., Pietquin, O., Courville, A.C.: Modulating early visual processing by language. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
6. Dumoulin, V., et al.: Adversarially learned inference. arXiv preprint: arXiv:1606.00704 (2016)
7. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3D object reconstruction from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 605–613 (2017)

8. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
9. 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, pp. 770–778 (2016)
10. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
11. Koch, T.L., Perslev, M., Igel, C., Brandt, S.S.: Accurate segmentation of dental panoramic radiographs with U-Nets. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 15–19. IEEE (2019)
12. Li, Y., et al.: The current situation and future prospects of simulators in dental education. J. Med. Internet Res. **23**(4), e23635 (2021)
13. 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., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 400–409. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_39
14. Mazzotta, L., Cozzani, M., Razionale, A., Mutinelli, S., Castaldo, A., Silvestrini-Biavati, A.: From 2D to 3D: construction of a 3D parametric model for detection of dental roots shape and position from a panoramic radiograph-a preliminary report. Int. J. Dent. **2013** (2013)
15. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: learning 3D reconstruction in function space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4460–4470 (2019)
16. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: representing scenes as neural radiance fields for view synthesis. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 405–421. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58452-8_24
17. Nader, R., Smorodin, A., De La Fourniere, N., Amouriq, Y., Autrusseau, F.: Automatic teeth segmentation on panoramic X-rays using deep neural networks. In: 2022 26th International Conference on Pattern Recognition (ICPR), pp. 4299–4305. IEEE (2022)
18. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
19. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660 (2017)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Silva, B., Pinheiro, L., Oliveira, L., Pithon, M.: A study on tooth segmentation and numbering using end-to-end deep neural networks. In: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 164–171. IEEE (2020)

22. Song, W., Liang, Y., Yang, J., Wang, K., He, L.: Oral-3D: reconstructing the 3D structure of oral cavity from panoramic x-ray. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, 2-9 February 2021, pp. 566–573. AAAI Press (2021). https://ojs.aaai.org/index.php/AAAI/article/view/16135
23. Stutz, D., Geiger, A.: Learning 3D shape completion under weak supervision. CoRR **abs/1805.07290** (2018). http://arxiv.org/abs/1805.07290
24. Tatarchenko, M., Richter, S.R., Ranftl, R., Li, Z., Koltun, V., Brox, T.: What do single-view 3D reconstruction networks learn? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3405–3414 (2019)
25. Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.-G.: Pixel2Mesh: generating 3D mesh models from single RGB images. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11215, pp. 55–71. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01252-6_4
26. Xie, H., Yao, H., Sun, X., Zhou, S., Zhang, S.: Pix2Vox: context-aware 3D reconstruction from single and multi-view images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2690–2698 (2019)
27. Yun, Z., Yang, S., Huang, E., Zhao, L., Yang, W., Feng, Q.: Automatic reconstruction method for high-contrast panoramic image from dental cone-beam CT data. Comput. Methods Programs Biomed. **175**, 205–214 (2019)
28. Zhao, Y., et al.: TSASNet: tooth segmentation on dental panoramic X-ray images by two-stage attention segmentation network. Knowl.-Based Syst. **206**, 106338 (2020)
29. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1