# Curvature-Enhanced Implicit Function Network for High-quality Tooth Model Generation from CBCT Images

Yu Fang[1,2,5], Zhiming Cui[1], Lei Ma[1], Lanzhuju Mei[1,2,5], Bojun Zhang[3], Yue Zhao[4], Zhihao Jiang[2], Yiqiang Zhan[5], Yongsheng Pan[1], Min Zhu[3], and Dinggang Shen[1,5(✉)]

[1] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
dgshen@shanghaitech.edu.cn
[2] School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[3] Shanghai Ninth Peoples Hospital, Shanghai Jiao Tong University, Shanghai, China
[4] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
[5] Shanghai United Imaging Intelligence Co. Ltd., Shanghai, China

**Abstract.** In digital dentistry, high-quality tooth models are essential for dental diagnosis and treatment. 3D CBCT images and intra-oral scanning models are widely used in dental clinics to obtain tooth models. However, CBCT image is volumetric data often with limited resolution (about 0.3–1.0 mm spacing), while intra-oral scanning model is high-resolution tooth crown surface (about 0.03 mm spacing) without root information. Hence, dentists usually scan and combine these two modalities of data to build high-quality tooth models, which is time-consuming and easily affected by various patient conditions or acquisition artifacts. To address this problem, we propose a learning-based framework to generate high-quality tooth models with both fine-grained tooth crown details and root information only from CBCT images. Specifically, we first introduce a tooth segmentation network to extract individual teeth from CBCT images. Then, we utilize an implicit function network to generate tooth models at arbitrary resolution in a continuous learning space. Moreover, to capture fine-grained crown details, we further explore a curvature enhancement module in our framework. Experimental results show that our proposed framework outperforms other state-of-the-art methods quantitatively and qualitatively, demonstrating the effectiveness of our method and its potential applicability in clinical practice.

## 1 Introduction

With the development of computer-aided techniques, digital dentistry has been widely used in dental clinics for diagnosis [15], restoration [10], and treatment planning [6,17]. In these systems, the acquisition of high-quality 3D tooth models is essential to assist dentists in extracting [14], implanting [7], or rearranging


L. Wang et al. (Eds.): MICCAI 2022, LNCS 13435, pp. 225–234, 2022.
https://doi.org/10.1007/978-3-031-16443-9_22

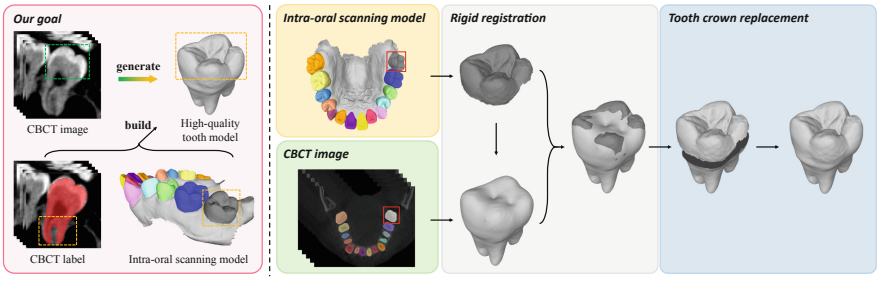


**Fig. 1.** Left: Overview of generating high-quality tooth model from CBCT images only. Our proposal is to use the tooth crown surfaces obtained from intra-oral scanning models to guide the training process of high-resolution tooth model generation from CBCT images. Right: Overview of high-quality tooth model building from CBCT image and intra-oral scanning model.

teeth [9]. In this regard, segmenting individual teeth from cone-beam computed tomography (CBCT) images [3,4] (the 3D volumetric data of all oral tissues) is a long-standing topic and has achieved promising results. However, due to the imaging techniques and radiation exposure, the spatial resolution in CBCT images is relatively low (about 0.3–1.0 mm spacing), which limits capturing of the tooth crown details. Thus, dentists usually rely on intra-oral scanning models (high-precision tooth crown surface without root information), to analyze occlusion relations of upper and lower jaws [16]. But it is time-consuming to collect both modalities of data, and is easily affected by various patient conditions or acquisition artifacts. In this situation, it is of great significance for developing a framework to generate high-quality tooth models with fine-grained tooth crown details and root information only from CBCT images.

To effectively reconstruct 3D shapes with high resolution, implicit function networks [2,13] have achieved outstanding performance in 3D synthetic datasets for shape recovery, completion, and super-resolution. Their advantage is the ability to handle different objects in a continuous learning space. Unfortunately, most of these methods are designed to capture general shapes, so the predictions tend to be over-smooth, thus ignoring many important geometric details. In the meantime, compared with clean 3D models in the synthetic datasets, 3D models derived from CBCT images or intra-oral scanning models usually introduce more noises from real-world clinical scenarios. Hence, it is extremely challenging to recover high-quality tooth models by segmenting teeth directly from CBCT images, especially on tooth crowns with rich geometric details.

In this study, to tackle the above limitations, we propose a novel curvature-enhanced implicit function network for high-quality tooth model generation from CBCT images. Our key idea is to combine the commonly used CNN-based segmentation network with an implicit function network to generate 3D tooth models with fine-grained geometric details. Specifically, given a 3D CBCT image, we first utilize a segmentation network to segment individual teeth, and represent

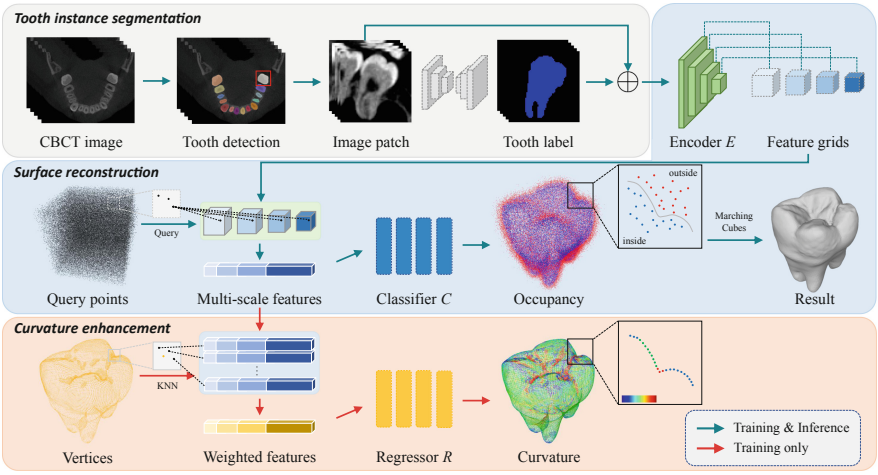


**Fig. 2.** The overview of our framework. The green arrows indicate the flows for training and inference, and the red arrows are for training only. (Color figure online)

the high-resolution output in the voxel space. Then, we introduce an implicit function network to generate high-quality 3D tooth models at arbitrary resolution in a continuous space. Particularly, to retain tooth crown surfaces with fine-grained details, a curvature enhancement module is proposed to predict local shape properties, which in return guides the implicit function network to reproduce plausible tooth shapes. Note that the ground truth, i.e., high-quality tooth models, in the training stage is built by merging respective surfaces from the CBCT image and intra-oral scanning model as shown in Fig. 1.

## 2 Method

The overview of our framework to generate high-quality tooth models from CBCT images is shown in Fig. 2, including a *tooth instance segmentation* module, a *surface reconstruction* module, and a *curvature enhancement* module, as detailed below.

### 2.1 High-quality Tooth Model Building

In this study, we combine the two modalities of data together (i.e., 3D CBCT images and intra-oral scanning models) to build high-quality tooth models, where root surfaces and crown surfaces are produced from the paired CBCT image and the intra-oral scanning model, respectively. As shown in Fig. 1, the ground truth building process is composed of three steps. First, given a CBCT image and its paired intra-oral scanning model, we manually delineate tooth model on the CBCT image, and crown model on the intra-oral scanning model. Then, as the paired tooth model and crown model are scanned from the same patient, we

directly apply the rigid ICP [1] algorithm to align these two models. Finally, we remove the tooth crown surface on the tooth model, and use Screened Poisson surface reconstruction [8] to merge the remaining root surface with the crown surface extracted from the intra-oral scanning model. In this way, the generated tooth model is of high quality with fine-grained tooth crown details.

### 2.2 Tooth Instance Segmentation

With the built high-quality 3D tooth models, we utilize them as ground truth to supervise the tooth instance segmentation network. The main purpose of this step is to produce the label of each tooth, which can effectively remove the background of soft tissues. In this work, we apply the typical method, HMG-Net [5], with state-of-the-art performance for tooth instance segmentation from CBCT images. The key idea of HMG-Net is to first detect all teeth in the CBCT images, and then apply tooth-level segmentation to delineate each tooth. We define $\mathcal{L}_{seg}$ with the cross-entropy loss and Dice loss to supervise the segmentation network.

As shown in Fig. 2, although the method can achieve promising performance on tooth segmentation, details especially on tooth crowns are usually lost due to limited resolution of CBCT images. Hence, we then take the predicted tooth label and the cropped image patch as input to the surface reconstruction module, to generate high-quality tooth models with rich geometric details.

### 2.3 Surface Reconstruction

In the surface reconstruction module, inspired by the work in 3D model completion [2], we introduce the implicit function strategy with multi-scale encoding and shape decoding, to preserve fine-grained surface details by reconstructing 3D tooth models at arbitrary resolution in a continuous space.

**Multi-scale Encoding.** With the tooth-level input $X$ (i.e., cropped image patch and tooth label), we first employ an encoder $E$ to extract multi-scale features $\{F_1, F_2, \ldots, F_L\}$ from different convolutional layers. Note that the first feature map $F_1$ is the input $X$. In this way, the feature maps at early stages can capture local information, while the feature maps at late stages contain global information. And all the feature maps from different layers preserve 3D volumetric structures aligned with the input data, which is defined as:

$$E(X) := F_1, F_2, \ldots, F_L. \tag{1}$$

**Shape Decoding.** As multi-scale features are discrete in grids, given a query point $p \in \mathbb{R}^3$ in the continuous space, we can obtain its feature by trilinear interpolation. Moreover, to encode more local neighborhood information for shape decoding, we extract features at the location of the query point $p$, and additionally at surrounding points in a distance $d$ ($d = 0.05$ in this paper) along

the Cartesian axes. At last, we integrate the point-wise features from the multi-scale features $\{F_1, F_2, \ldots, F_L\}$ with different receptive fields, which is denoted as $\{F_1(p), F_2(p), \ldots, F_L(p)\}$. The integrated point-wise features are then fed into a point-wise decoder $\mathcal{D}$, parameterized by a fully connected neural network, to predict the corresponding occupancy value (i.e., inside or outside the surface):

$$\hat{o_p} = \mathcal{D}(F_1(p), F_2(p), \ldots, F_L(p)) \in \{0, 1\}, \tag{2}$$

where $\{0, 1\}$ denotes the query point being outside or inside the surface, respectively. We employ the BCE loss $\mathcal{L}_{occ}$ to supervise the learning process. Note that, to robustly train the network, we sample a number of query points in the continuous space, using the sampling strategy described in Sect. 3.2.

### 2.4 Curvature Enhancement

Since our method is defined in the continuous space, it is capable of describing a surface at arbitrary resolution. However, it still cannot effectively produce tooth crown surfaces with fine-grained details, for the reason that only the binary occupancy classification on query points cannot faithfully learn the changes of local shape properties (e.g., surface curvature). Thus, we further propose a branch to predict the curvature of each query point.

Specifically, we first extract the vertices $V = \{v_1, v_2, \ldots, v_T; v_t \in \mathbb{R}^3\}$ on the ground truth tooth surface, and compute their corresponding curvature values $\{c_1, c_2, \ldots, c_T; c_t \in \mathbb{R}\}$. For each $v_t \in V$, we extract the features from its $K$ nearest query points $\{p'_1, p'_2, \ldots, p'_K\}$ ($K = 5$ in this paper), and obtain curvature features $F_c(v_t)$ by Inverse Distance Weighting (IDW) [12], which is defined as

$$F_c(v_t) = \sum_{k=1}^{K} \frac{1/D(p'_k, v_t)}{\sum_{k=1}^{K} 1/D(p'_k, v_t)} F(p'_k), \tag{3}$$

where $D$ denotes the Euclidean distance. The curvature features are then fed into the point-wise regressor $R$, parameterized by a fully connected neural network, to predict the curvature value of $v_t$. And we use the smooth L1 error $\mathcal{L}_{cur}$ to supervise the curvature enhancement module.

Finally, the overall loss with multiple supervision is computed by:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \mathcal{L}_{occ} + \mathcal{L}_{cur}. \tag{4}$$

$\mathcal{L}_{seg}$ and $\mathcal{L}_{occ}$ refer to the loss functions of the tooth segmentation network and the implicit function network, respectively.

## 3 Experiments

### 3.1 Dataset and Evaluation Metrics

To evaluate the performance of our proposed method, we collect a dataset with 50 patients in dental clinics. Each subject has a 3D CBCT image (with 0.4 mm

spacing) and paired intra-oral scanning model (with 0.03 mm spacing). To build the ground truth of our dataset, 3 experts are first employed to manually delineate the tooth labels and crown labels on CBCT images and intra-oral scanning models, respectively. Then, we merge the two modalities of data to obtain the high-quality tooth models (see details in Sect. 2.1). In this study, 50 samples are randomly divided into 3 subsets, using 20 for training, 10 for validation, and the remaining 20 samples for testing. Note that only CBCT images are fed into the framework to generate high-quality tooth models.

To quantitatively analyze the performance of our method, we report the following four metrics, including Intersection over Union (IoU), Chamfer-L2, Normal Consistency (Normals), and occupancy accuracy (OccAcc). IoU measures the similarity between two volumes, and Chamfer-L2 is the metric to measure a bidirectional distance between two surfaces. Normals is first proposed by OccNet [13] to measure the normal consistency between two surfaces. We define OccAcc as an additional accuracy metric to evaluate the occupancy prediction.

### 3.2 Sampling Strategies

To approximate the continuous query space, we briefly introduce the sampling strategies during network training and inference stages. In the training stage, the most intuitive way is to sample points around the ground truth tooth surface within a small distance. Specifically, we first sample points on the ground truth tooth surface, and then add random displacements with two Gaussian distributions, where their deviations are $\sigma_1 = 0.02$ and $\sigma_2 = 0.1$, respectively. Thus, the sample points within $\sigma_1$ can capture fine-grained surface details, and the sample points within $\sigma_2$ can cover the entire geometric space. In the network training, we sample 50K points from each of the two distributions. In the inference stage, since the ground truth tooth model is not available, we uniformly sample points along each axis in the continuous query space. Note that the retrieval resolution is determined by the density of query points on each axis. In our experiments, to obtain the tooth model with rich geometry information, especially on the tooth crowns, we query output with a resolution of $256^3$, which is about $4^3$ times larger than the image patch cropped from the original 3D CBCT image.

### 3.3 Implementation Details

Our framework is built on a PyTorch platform with an NVIDIA Tesla V100S GPU. The encoder is composed of four blocks, including one Convolution(Conv, with a $3 \times 3 \times 3$ kernel and a $1 \times 1 \times 1$ padding)-ReLU-Batch Normalization(BN) block and three MaxPooling-Conv-ReLU-Conv-ReLU-BN blocks. And the network architectures of the point-wise decoder $C$ and the point-wise regressor $R$ are the same, which include four fully connected layers with ReLU. We use the Adam optimizer with a learning rate of 0.0001, divided by 10 for every 20 epochs. In the testing stage, with the predicted occupancy of each query point (i.e., inside or outside the surface), we apply the traditional Marching Cubes algorithm [11] to generate high-quality tooth models.

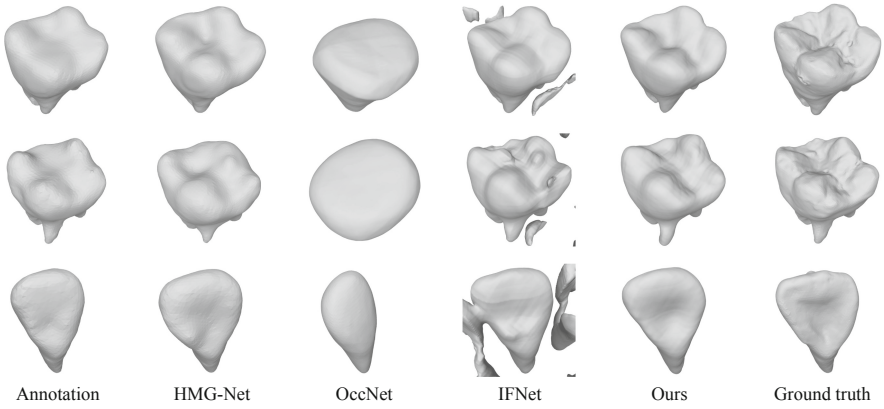


**Fig. 3.** Qualitative results of three typical cases generated by different methods.

**Table 1.** Quantitative comparison with different methods for high-quality tooth model generation.

| Method | OccAcc ↑ | IoU ↑ | Chamfer-L2 ↓ | Normals ↑ |
|---|---|---|---|---|
| HMG-Net | 75.95 | 77.10 | 6.30e-4 | 95.09 |
| OccNet | 61.41 | 54.04 | 2.92e-3 | 85.08 |
| IF-Net | 77.47 | 65.52 | 9.34e-3 | 88.67 |
| Ours | **79.70** | **83.03** | **3.00e-4** | **96.25** |

### 3.4 Comparison with Other Methods

Recently, many methods have been proposed to reconstruct 3D shapes with implicit functions, including the Occupancy Networks (OccNet) [13], and the Implicit Feature Networks (IF-Net) [2]. We also implement the baseline segmentation method (i.e., HMG-Net [5]) with high-quality tooth model supervision.

The quantitative results of different methods are presented in Table 1. It can be found that our approach achieves the best performance across all metrics. Moreover, the accuracy of directly applying the implicit function based methods on tooth surfaces (i.e., OccNet and IF-Net) is relatively low. The main reason is that there is a large domain gap between the input CBCT images and output tooth surfaces. Directly applying surface reconstruction algorithms cannot effectively address the distribution shift problem.

In order to analyze the advantage of our algorithm more comprehensively, we further provide visual comparison of three typical examples in Fig. 3. It can be found that, the tooth instance segmentation method (i.e., HMG-Net) can generate promising tooth shapes, but many details, especially on the tooth crowns, are missed due to limited resolution of CBCT images. For the implicit function based methods, we observe that the tooth surfaces produced by OccNet are too smooth with only global structures. With the multi-scale

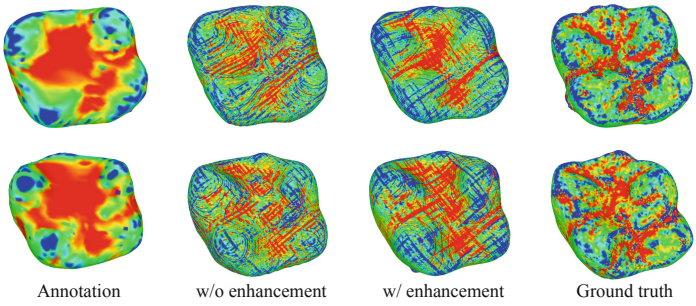


**Fig. 4.** Investigation on surface curvatures. The figure shows the comparison of our framework without and with curvature enhancement for surface details. Different colors are used to show curvature distribution on the surface.

**Table 2.** Statistical results for analyzing different components in our framework. "H-IF" denotes the joint learning of HMG-Net and IF-Net.

| Method | OccAcc ↑ | IoU ↑ | Chamfer-L2 ↓ | Normals ↑ |
|---|---|---|---|---|
| HMG-Net | 75.95 | 77.10 | 6.30e-4 | 95.09 |
| H-IF | 78.67 | 82.74 | **2.88e-4** | 95.87 |
| Ours | **79.70** | **83.03** | 3.00e-4 | **96.25** |

feature extraction scheme, IF-Net generates better results with fine-grained details. However, many noises are introduced due to large domain gap between CBCT images and tooth models. Notably, our method matches better with the ground truth, where global structures and local details can be successfully retained, indicating the effectiveness of both the image-to-surface tooth model generation scheme and the curvature enhancement in this specific task.

### 3.5 Ablation Study

To validate the effectiveness of each component in our method, we conduct several ablation experiments by gradually augmenting the baseline network.

**Surface Reconstruction.** Based on the baseline tooth instance segmentation network (i.e., HMG-Net), we add another branch, surface reconstruction by the implicit function learning, to enhance the network capability on global tooth structures, and denote it as H-IF. The results are shown in Table 2. It can be found that the accuracy is consistently improved in terms of all metrics. Particularly, the surface distance error (i.e., Chamfer-L2) is greatly reduced ($6.30e{-}4$ vs. $2.88e{-}4$), demonstrating that the implicit function strategy can more effectively learn global tooth structures.

**Curvature Enhancement.** To validate the effectiveness of the curvature enhancement module, we include this module based on H-IF as our final framework. As shown in Table 2, the statistical results are improved, especially on the metric of surface normals (95.87% vs. 96.25%). Note that surface normals are usually sensitive to the geometric details on surfaces. Moreover, we further present the visual comparison of curvature distributions in Fig. 4. The first column is the curvature distributions computed of manual annotations in original CBCT images, where many geometric details, especially on the crowns, are missed due to limited CBCT resolution. The second and third columns are the tooth models produced by our method without and with the curvature enhancement module. It can be seen that, with the curvature enhancement (3rd column), our framework can produce local shape properties more clearly, and the curvature distribution matches with the ground truth more consistently.

## 4 Conclusion

In this paper, we propose an effective framework to generate high-quality tooth models from CBCT images. Our framework first introduces a tooth instance segmentation network to segment individual teeth coarsely, and then learns to generate high-quality tooth models under implicit function learning. Meanwhile, a curvature enhancement module is further proposed to guide the surface reconstruction. Experimental results on the real-patient dataset demonstrate that our proposed method outperforms other state-of-the-art methods, showing its potential to be applied in dental clinics.

## References

1. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor Fusion IV: Control Paradigms and Data Structures, vol. 1611, pp. 586–606. International Society for Optics and Photonics (1992)
2. Chibane, J., Alldieck, T., Pons-Moll, G.: Implicit functions in feature space for 3d shape reconstruction and completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6970–6981 (2020)
3. Chung, M., et al.: Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation. Comput. Biol. Med. **120**, 103720 (2020)
4. Cui, Z., Li, C., Wang, W.: Toothnet: automatic tooth instance segmentation and identification from cone beam ct images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6368–6377 (2019)
5. Cui, Z., et al.: Hierarchical morphology-guided tooth instance segmentation from CBCT images. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 150–162. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_12
6. Ewers, R., et al.: Computer-aided navigation in dental implantology: 7 years of clinical experience. J. Oral Maxillof. Surg. **62**(3), 329–334 (2004)
7. Hoffmann, O., Zafiropoulos, G.G.: Tooth-implant connection: a review. J. Oral Implantol. **38**(2), 194–200 (2012)

8. Kazhdan, M., Hoppe, H.: Screened poisson surface reconstruction. ACM Trans. Graph. **32**(3), 1–13 (2013)
9. Li, Y., Jacox, L.A., Little, S.H., Ko, C.C.: Orthodontic tooth movement: the biology and clinical implications. Kaohsiung J. Med. Sci. **34**(4), 207–214 (2018)
10. Liu, C., Guo, J., Gao, J., Yu, H.: Computer-assisted tooth preparation template and predesigned restoration: a digital workflow. Int. J. Computeriz. Dentist. **23**(4), 351–362 (2020)
11. Lorensen, W.E., Cline, H.E.: Marching cubes: a high resolution 3d surface construction algorithm. ACM Siggraph Comput. Graph. **21**(4), 163–169 (1987)
12. Lu, G.Y., Wong, D.W.: An adaptive inverse-distance weighting spatial interpolation technique. Comput. Geosci. **34**(9), 1044–1055 (2008)
13. 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)
14. Orlowska, M., Jozwiak, R., Regulski, P.: Virtual tooth extraction from cone beam computed tomography scans. In: Augustyniak, P., Maniewski, R., Tadeusiewicz, R. (eds.) PCBBE 2017. AISC, vol. 647, pp. 275–285. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-66905-2_24
15. Paredes, V., Gandia, J.L., Cibrián, R.: Digital diagnosis records in orthodontics. An overview. Med. Oral Patol. Oral Cir. Bucal. **11**(1), E88–E93 (2006)
16. Roberta, T., Federico, M., Federica, B., Antonietta, C.M., Sergio, B., Ugo, C.: Study of the potential cytotoxicity of dental impression materials. Toxicol. In vitro **17**(5–6), 657–662 (2003)
17. Van Der Meer, W.J., Vissink, A., Ng, Y.L., Gulabivala, K.: 3d computer aided treatment planning in endodontics. J. Dentist. **45**, 67–72 (2016)