
ORAL-3D: RECONSTRUCTING THE 3D BONE STRUCTURE OF ORAL CAVITY FROM 2D PANORAMIC X-RAY

Weinan Song, Yuan Liang, Kun Wang, Lei He
Department of Electrical and Computer Engineering
University of California, Los Angeles, CA, USA
wsong@ucla.edu

March 27, 2020

ABSTRACT

Panoramic X-ray and Cone Beam Computed Tomography (CBCT) are two of the most general imaging methods in digital dentistry. While CBCT can provide higher-dimension information, the panoramic X-ray has the advantages of lower radiation dose and cost. Consequently, generating 3D information of bony tissues from the X-ray that can reflect dental diseases is of great interest. This technique can be even more helpful for developing areas where the CBCT is not always available due to the lack of screening machines or high screening cost. In this paper, we present *Oral-3D* to reconstruct the bone structure of oral cavity from a single panoramic X-ray image by taking advantage of some prior knowledge in oral structure, which conventionally can only be obtained by a 3D imaging method like CBCT. Specifically, we first train a generative network to back project the 2D X-ray image into 3D space, then restore the bone structure by registering the generated 3D image with the prior shape of the dental arch. To be noted, *Oral-3D* can restore both the density of bony tissues and the curved mandible surface. Experimental results show that our framework can reconstruct the 3D structure with significantly high quality. To the best of our knowledge, this is the first work that explores 3D reconstruction from a 2D image in dental health.

1 Introduction

Oral diseases are the most common diseases in the world and can affect people's health throughout their lifetime. This problem is more serious in developing countries due to the limited resource of dental service and insufficient awareness of oral health. Some oral diseases, *e.g.*, periodontitis, gingivitis, and dental plaque, can be directly inspected and diagnosed by eyes for a well-trained dentist. However, some dental therapy methods, *e.g.*, teeth implant/pulling, orthodontics, and dental filling, usually require the assistance of radiology imaging to observe the interior bony tissues of the oral cavity. Therefore, the imaging quality can significantly influence the accuracy of dental diagnosis, which has indicated the importance of the imaging methods in dental healthcare.

Panoramic X-ray is one of the most popular imaging methods in digital dentistry. The imaging machine can take a photo of all the teeth and most part of the mandible for the patient by moving the X-ray camera around the patient's head. It generates very little radiation as the exposure process only takes a few seconds, thus has been widely used in many dental clinics. Another popular imaging method is CBCT, which develops from traditional CT but reconstructs a 3D image for the oral cavity with divergent X-rays. Compared with panoramic X-ray, CBCT can provide a higher-dimension picture and show more details of teeth and mandible. Also it is becoming a necessary examination for endodontics, implantology, and orthodontics.

We summarize the characteristics of these two imaging methods in Table 2. Compared with panoramic X-ray, CBCT can provide more information for diagnosis and surgery, but the price and radiation also significantly increase at the same time. For example, CBCT generates $39.4\times$ radiation and takes $3.7\times$ of the price on average than panoramic X-ray. These factors have limited the popularization of CBCT into many small dental clinics, even if it can provide higher-dimension information over panoramic X-ray. This problem is more obvious in developing countries, where

Table 1: A comparison of CBCT and panoramic X-ray on common dental disease.

	CBCT	Panoramic X-ray
Image Dimension	3D	2D
Imaging Cost [1]	€ 184.44	€ 49.29
Radiation Dose [2]	58.9-1025.4 μ Sv	5.5-22.0 μ Sv
Diagnostic Accuracy [3]	94.8%	83.3%
Wisdom Tooth	✓	✓
Tooth Decay	✓	✓
Implant Planning	✓	✗
Orthodontics	✓	✗

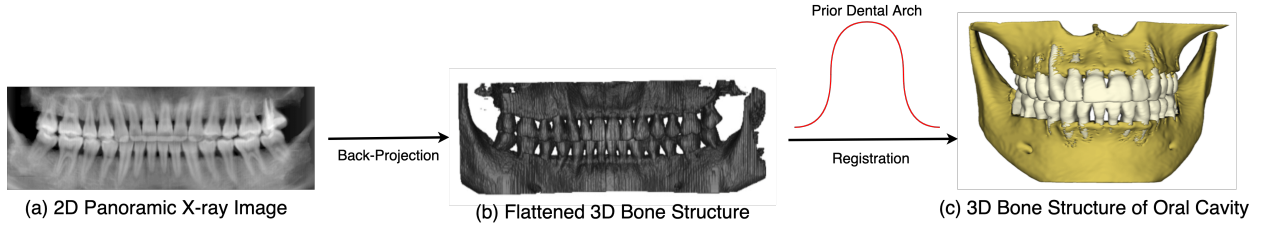


Figure 1: An overview of Oral-3D can be seen in this picture. We first back project the panoramic image into 3D space with a generative network, then we utilize the prior dental arch information to register the generated 3D image into a curved plane to reconstruct the bone structure.

people are not willing to invest too much in dental healthcare. To solve this problem, we propose to reconstruct the 3D bone structure of oral cavity from a single panoramic X-ray image with the help of some prior knowledge. It can enable dentists to inspect the patient’s teeth in a 3D view even if only panoramic X-ray images are available. We believe this work can contribute to the dental healthcare by lowering the examination requirement.

Single-view reconstruction has been a hot topic in recent years with the fast development of deep learning. Typically, the reconstruction models try to recover the shape and surface of the 3D object with a single-view image. They usually require some prior knowledge to regularize the reconstruction result. Our framework follows a similar principle as these works that take the shape of the dental arch as the prior information to constrain the oral shape. Additionally, as we also intend to recover the interior information of the 3D object, this can also be taken as cross-modality problem where a CT-like image is generated from an X-ray image.

We show an overview of our framework, named Oral-3D in Fig. 1, which is a two-step reconstruction work. In the first step, we train a generative network to learn the non-learning mapping function that can back project the 2D panoramic X-ray image (Image *a*) into 3D space, where the depth information of teeth and mandible can be learned automatically from the paired 2D and 3D images. In the second step, we register this generated 3D image (Image *b*), which can be seen as an embedding of the oral cavity into a curved plane, into the original shape to reconstruct the final bone structure according to the prior knowledge in the dental arch. This prior knowledge effectively restricts the shape and location of the bone structure and can be obtained from anytime of a CT scan of the corresponding patient or just taken from an averaged shape of the training cases. The paired panoramic X-ray image and its corresponding 3D bone structure are automatically generated from a CBCT dataset collected by a major stomatological hospital in China. Experimental results show that Oral-3D can reconstruct the oral structure with high quality from a single panoramic X-ray image and keep the density information at the same time. To summarize, we make the following contributions:

- We are the first to explore cross-modality transfer for dental imaging with deep learning, where the 3D bone structure of oral cavity can be reconstructed from a single panoramic image and some prior knowledge.
- In addition to generating the 3D shape and surface of the bone structure, we also restore the density information of bony tissues, which is of great help for dental diagnosis.
- We propose a registration method to embed a 3D image into a curved plane, which can significantly improve the reconstruction quality compared with single-step methods.
- Our reconstruction results can clearly reflect the abnormal dental structure, *e.g.*, missing teeth and wisdom teeth, in the 3D space. This has indicated the practicability of our method in clinical applications.
- We provide an automatic method to generate paired images for the panoramic X-ray image and the corresponding bone structure from a CBCT image, which can be explored in future research.

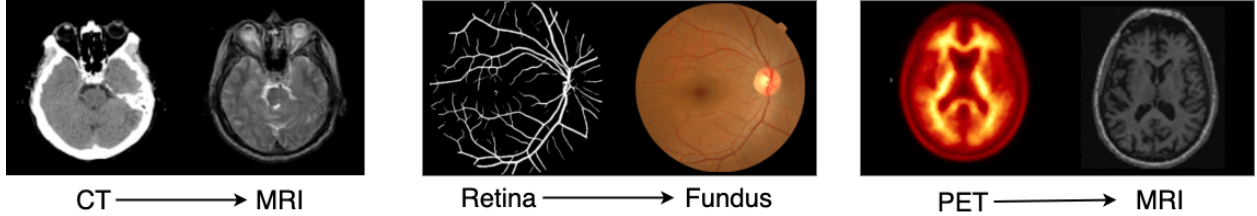


Figure 2: We show some examples of cross-modality transfer for same-dimension images: CT \rightarrow MRI [9], Retina \rightarrow Fundus [11], and PET \rightarrow MRI [12]. These works can be seen as a pix2pix framework and the source image usually contains more information than the target image. Compared with these work, our problem for oral cavity reconstruction is more challenging as we try to recover 3D information from a 2D image.

2 Related Work

2.1 Deep Learning for Oral Health

Deep learning has shown promising performance in the automated medical assistance system for healthcare by learning latent features from large amounts of datasets. It also plays an increasingly important role for oral health by improving dental examination efficiency with intelligent solutions. For example, [4] proposes an automatic method for instance-level segmentation and identification of teeth in the CBCT image. [5] trains a deep neural network to detect and diagnose dental caries from periapical radiographic images. [6] utilizes a U-Net structure to enhance the resolution of 2D CBCT slices. [7] designs a classification model for red auto-fluorescence plaque images to assist in detecting dental caries and gum diseases. [8] uses transfer learning to classify three different oral diseases for X-ray images. However, these methods have ignored the availability of imaging methods in dental healthcare, which is a practical problem in many developing countries due to the limited income and high cost of the imaging equipment. Therefore, we intend to improve oral healthcare service in a different view by lowering the examination requirement, where the 3D bone structure of oral cavity can be reconstructed without a CBCT machine.

2.2 Cross-Modality Transfer in Medical Imaging

The target of cross-modality transfer is to find a non-linear relationship between medical images in different modalities. It can help reduce the extra acquisition time and additional radiation in medical examination or provide additional training samples without repetitive annotation work to augment the dataset. Most works take this as a pix2pix problem, where the layout and the structure of objects are retained but the color distribution is changed when transforming the source image into a target space. For example, [9] utilizes CycleGAN [10] with unpaired data to synthesis CT slices from MRI slices for radiologists to observe abnormal anatomical lesions. [11] takes the vessel tree of eyes as a condition to synthesis new images for fundus photography. [12] proposes a generation network to produce realistic structural MR images from florbetapir PET images. Examples of cross-modality transfer are shown in Fig. 2.

However, only a few works have discussed the cross-modality transfer problem from a lower-dimension image to a higher-dimension one, which is more challenging as the model needs to infer the latent information in the lower-dimension image. To the best of our knowledge, there are only two papers that have done similar works with ours in medical imaging. The first one [13] uses an encoder-decoder network to reconstruct 3D skull volumes of 175 mammalian species from 2D cranial X-rays, but the result is subject to too much ambiguity. To improve the visual quality, [14] utilizes biplanar X-rays to extract 3D anatomical structures of body CT with adversarial training and reconstruction constraints. Differently, our framework improves the reconstruction quality with prior knowledge, where the back projection module only needs to learn depth information and the registration module is responsible for embedding the generated 3D image into a curved plane. We show that this scheme can significantly improve the reconstruction quality.

2.3 3D Reconstruction from 2D Image

Recent work of 3D Reconstruction from the 2D image can be summarized into two categories according to the views provided by the 2D image: multi-view reconstruction and single-view reconstruction. For the first one, multi-view images can provide shape information of the reconstruction object from multiple angles, thus requiring little prior knowledge as a restriction. For example, [15] computes the most probable 3D shape that gives rise to the observed

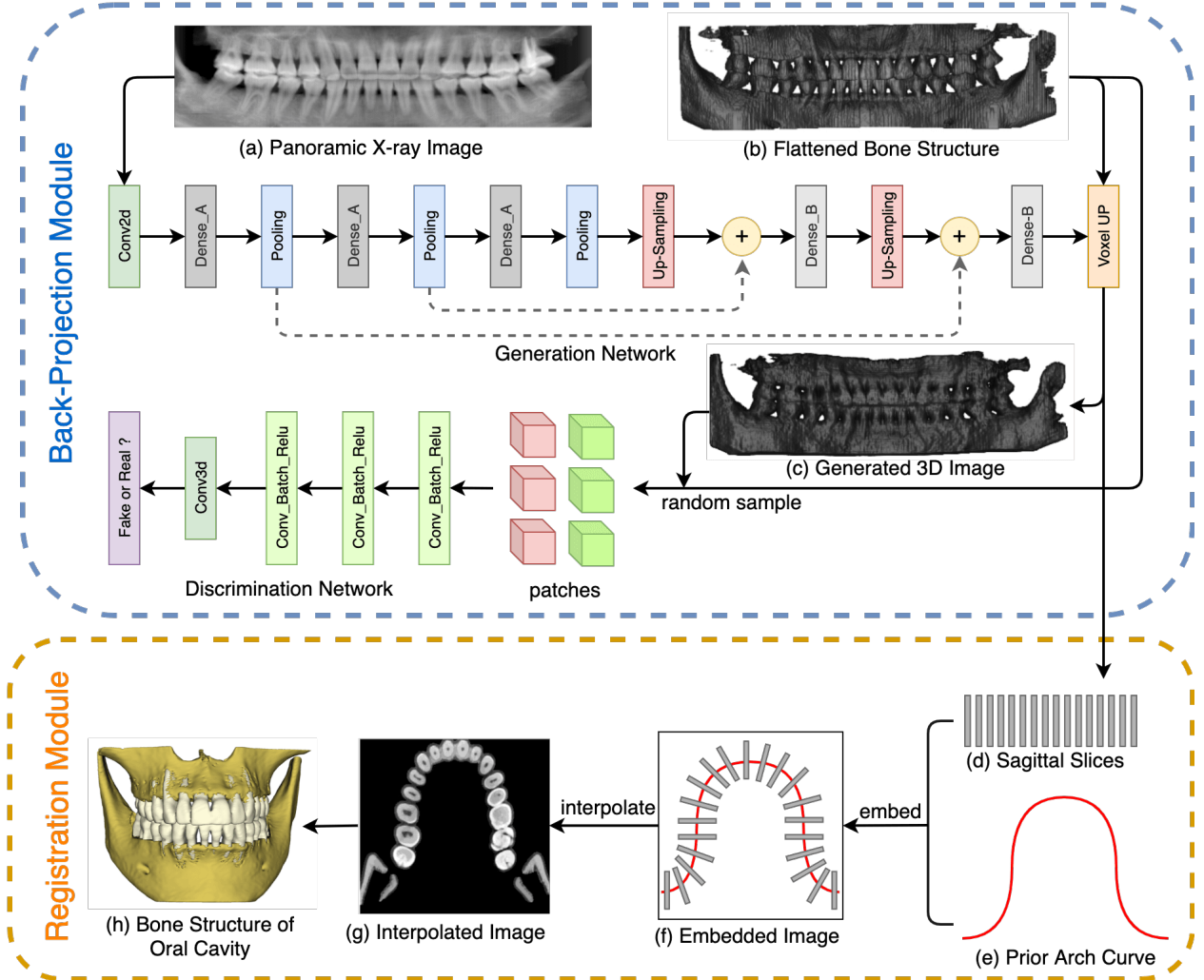


Figure 3: Our framework consists of a back-projection module to transform the 2D panoramic image into 3D space and a registration module to reconstruct the 3D structure of oral cavity.

color information from a series of calibrated 2D images. [16] learns the mapping function from arbitrary viewpoints to a 3D occupancy grid with a 3D recurrent neural network. As a comparison, reconstruction from a single-view image usually requires additional information, *e.g.*, prior shape, to infer the object shape. As such, [17] proposes a unified framework trained with a small amount of pose-annotated images to reconstruct a 3D object. [18] takes the adversarially learned shape priors as a regularizer to penalize the reconstruction model. [19] infers the 3D layout hypotheses of a room by geometric reasoning from the single spherical panorama before reconstruction. Specifically, as the panoramic image can be seen as taking a photo of an object with a moving camera, our work should belong to the second type. Following a similar way with models in single-view reconstruction, we take advantage of the dental arch shape as the prior knowledge to constraint the reconstruction result and alleviate the miss-registration.

3 Method

We show the detailed structure of *Oral-3D* in Fig. 3, which consists of a back-projection module and a registration module. The back-projection module is based on GAN, where the generator is trained to back project the 2D panoramic image into 3D space. The registration module takes in the generated 3D image (Image *c*) and the prior knowledge of the dental arch (Image *e*) to reconstruct the 3D oral structure (Image *h*).

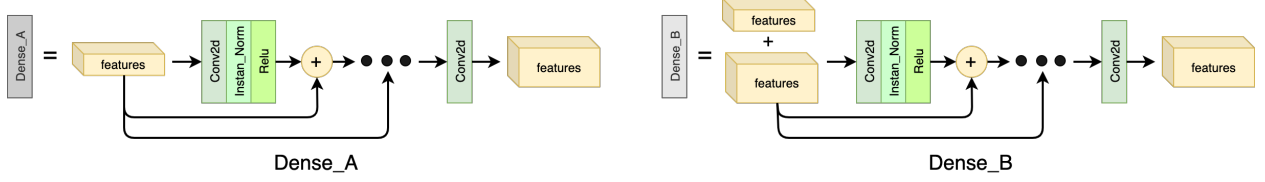


Figure 4: We show the two kinds of dense blocks for the generator in this picture, where $Dense_A$ is responsible for expanding depth information and $Dense_B$ fuses and forwards this information to the 3D output.

3.1 Back-Projection Module

GANs have proved to be an effective model to learn latent data distribution by training the generator G and the discriminator D in an adversarial way. The generator is expected to output x_{fake} that looks similar with the real image x_{real} from a random vector z sampled from a prior distribution $p(z)$. Also, the discriminator can be seen as a binary classifier that tries to distinguish x_{fake} and x_{real} . This min-max game can be expressed mathematically as in (1), where the generator learns to approximate the real data distribution $P_{real}(x)$ during the optimization process.

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{real}} [\log D(x)] + \mathbb{E}_{z \sim P_{noise}} [\log(1 - D(G(z)))] \quad (1)$$

Since we intend to generate the consistent 3D content from the semantic information of the panoramic X-ray image, we utilize conditional GANs [20] to learn the cross-dimension mapping for the generator. To improve the generation quality and guarantee the stable training process, we use LSGAN [21] as the keystone to train the generator and discriminator. To maintain the structural consistency of the input and generation result, we also introduce the mean square error of the generation as the reconstruction loss to the generator. Then the optimization problem in (1) changes to:

$$\begin{aligned} D^* &= \arg \min_D \mathbb{E}_y [(D(y) - 1)^2] + \mathbb{E}_x [D(G(x))^2] \\ G^* &= \arg \min_G \lambda_1 \cdot \mathbb{E}_x [D(G(x) - 1)^2] + \lambda_2 \cdot \mathbb{E}_{x,y} [(y - G(x))^2], \end{aligned} \quad (2)$$

where x is the panoramic image and y is the expanded 3D image.

Generator The objective of the generator is to learning a mapping function to transfer a 2D image into 3D space. Then the problem comes to finding a non-linear generator G denoted as:

$$G : I_{H \times W}^{2D} \rightarrow I_{H \times W \times D}^{3D}, \quad (3)$$

where I^{2D} is the panoramic image with a size of $H \times W$ and I^{3D} is the expanded 3D image with a size of $H \times W \times D$. In this problem, the pixel values of the panoramic X-ray image can reflect the absorption of radiation, thus making it reasonable to infer the depth information from the 2D image. Therefore, we utilize a 2D encoder-decoder network to learn the cross-dimension relationship, where the 3D information of the teeth are embedded into feature maps. As shown in Fig. 3, the encoding network decreases the resolution of features but increases the number of feature channels, while the decoding network increases the resolution of features and fuses them to reconstruct the 3D object. We utilize a 5×5 convolution with $stride = 2$ at the top of the generator to save memory and add a $tanh$ layer at the end to restrict the generated voxel value to $(-1, 1)$.

Dense Block Dense connection [22] has proved to be an efficient learning structure in deep neural networks. We thus embed dense blocks in the generation network to efficiently utilize the information from 2D panoramic X-ray images. Note that the dense blocks play two different roles in the encoder and decoder. In the encoding network, the dense layer learns to explore the depth information by doubling the channel number of feature maps. As shown in Fig. 4, for each $Dense_A$ in the encoding network with k -channel input feature maps, we set the dense block with three consecutive $conv - norm - relu$ layers with a growth rate of k . The $Dense_A$ layer outputs $2k$ -channel features in the end to expand the depth information over the input. In the decoding network, the dense block fuses the features up-sampled from the upper layer and the ones passed from the skip-connections. The $Dense_B$ layer outputs the same number of channels as the input to keep and forward the depth information.

Up-sampling For the up-sampling block in the decoding network of our generator, we utilize 3×3 transpose convolution with $stride = 2$ to increase the resolution of feature maps. For the $Voxel_up$ layer, we double the feature number at the same time to keep the shape of reconstruction results consistent with the ground truth.

Discriminator The discriminator can help the generator improve the generation quality by improving the classification ability for real and fake images. In our model, we adopt patch discriminator introduced by [10] to improve the generation quality of tooth edges by learning high-frequency structures in the flattened 3D image. We set the patch size as $70 \times 70 \times 70$ and follow a similar structure in [10] but with 3D convolution. As shown in the back-projection module in Fig. 3, we first random sample patches from both the real and generated images to be the input of the discriminator. Then we use three successive *conv - norm - relu* layers with *kernel* = 4, *stride* = 2, followed by a *conv - norm - relu* layer with *kernel* = 4, *stride* = 1, and a *conv - sigmoid* layer with *kernel* = 4, *stride* = 1 in the end to predict the probability of the samples belonging to the real image.

3.2 Registration Module

After obtaining the expanded 3D image from the generator, we register this flattened bone structure into a curved plane to reconstruct the final bone structure with the assistance of the prior knowledge about the dental arch. As shown in the registration module in Fig. 3, the whole process takes two steps. First, we cut the generated 3D image (Image *c*) into sagittal slices and embed them to the dental arch curve. This can be taken as a placement game about how to even put the sticks along a curve. To obtain this, we sample a series of points along the dental curve with equal distance and put the slices on these points in the normal direction (Image *f*). Second, we interpolate the voxels between the neighboring slices to output a smooth 3D structure (Image *h*). For computation convenience, we combine these steps together and show it in Algorithm 1, where we assume that the generated 3D image and the bone model has the same height of *H*.

Algorithm 1 Embedding and Interpolation

```

1: function REGISTER(Slices,  $W_{3D}$ ,  $D_{3D}$ , curve)
2:    $W, H, D \leftarrow \text{SHAPE}(\text{Slices})$ 
3:    $\text{OralImage} \leftarrow \text{ZEROS}(W_{3D}, H, D_{3D})$ 
4:    $\text{SamplePoints} \leftarrow \text{SAMPLE}(\text{curve})$ 
5:   for  $i = 0; i < W_{3D}; i ++$  do
6:     for  $j = 0; j < D_{3D}; j ++$  do
7:        $id, dist \leftarrow \text{MIN\_DISTANCE}((i, j), \text{SamplePoints})$ 
8:        $\text{OralImage}[i :, j] \leftarrow \text{INTERPOLATE}(\text{Slices}[id, :, :], dist)$ 
9:     end for
10:  end for
11:  return OralImage
12: end function

```

3.3 Generating Paired Images

In order to avoid the additional radiation for patients, we propose an automatic method to generate the paired images for panoramic X-ray and 3D bone structure of oral cavity from a CBCT image. The whole process is shown in Fig. 5.

Oral Structure Generation: To obtain the 3D bone structure from the CBCT image, we use global convex segmentation [23] to segment the bony tissues, which treats the bone segmentation as a minimization problem with a unique minimum. We use the Split Bergman method [24] to solve this and generate a binary mask of the bone structure. In the end, we apply this mask to the CBCT image to obtain the ground truth for the 3D structure.

Panoramic X-ray Generation: We follow the Beer-Lambert absorption-only model to synthesis the panoramic X-ray image from the CBCT image. Different from other methods that simulate the X-ray images from CT with parallel X-rays sent from a plane, we do the projection over a curved plane parallel to the dental arch with a similar method introduced by [25]. As shown in Fig. 5, we first obtain the dental arch mask (Image *c*) by maximum intensity projection (MIP) over the axial slices of CBCT (Image *a*). Then we skeletonize the binary mask and use a cubic spline to fit the skeleton image (Image *d*) to obtain the dental arch curve (the red line in Image *e*) and estimate the arch thickness. Finally, we sample $W + 1$ points along the arch curve with equal curve distance to transfer the curved plane into a multiple-fold plane for projection. Meanwhile, these sampled points and the curved slices parallel to the dental curve are kept as the prior knowledge of the dental arch and the ground truth of the expanded 3D image. The whole generation process can be seen from Image *a* - Image *f* in Fig. 5.

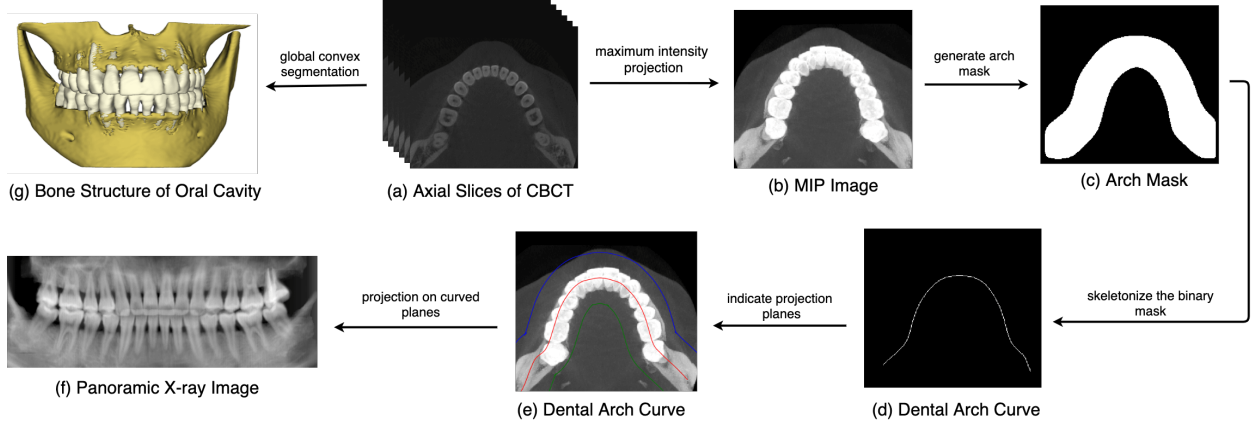


Figure 5: An overview of generating 3D bone structures of oral cavity and 2D panoramic images from CBCT is shown in this picture. The 3D oral structure is obtained by segmentation on CBCT with a global convex method. The 2D panoramic image is synthesized by projection over the estimated dental arch. In Image (e), the red line indicates the dental arch curve and the blue line and green line show the upper bound and lower bound of the projection space from an axial view.

4 Experiment

4.1 Dataset

The clinical dental CBCT images of 72 patients are collected from a major stomatological hospital in China. The images are scanned by NewTom VG, Italy, with a voxel size of 0.125mm. These images are further cropped by the radiologists to keep only the oral area. For convenience, we down-sample these images into a size of $288 \times 256 \times 160$ and normalize them into a range of $(-1, 1)$. Furthermore, we split the dataset with a ratio of 3 : 1 : 1 for training, validation, and testing. The panoramic X-ray image and the 3D bone structure of oral cavity are both automatically obtained from this CBCT dataset.

4.2 Evaluation Metrics

PSNR: Peak signal-to-noise ratio (PSNR) is often used to measure the difference between two signals. Compared with mean squared error, PSNR can be normalized by the signal range and expressed in terms of the logarithmic decibel scale. We take this to measure the density recovery of our models.

IOU: Intersection over union is a fundamental standard to evaluate the reconstructed shape. The binary mask of the reconstructed bone structure can be obtained by setting a threshold (*e.g.*, -0.5 in our experiment) to the reconstruction results.

SSIM: We also use the structure similarity index (SSIM) [26] as a key criterion of the reconstruction quality. Compared with the other two methods, SSIM considers the brightness, contrast and structure information at the same time and can match better the subjective evaluation of humans. It is widely used in many image transfer applications and can effectively indicate the reconstruction quality.

4.3 Training

We train the generator and the discriminator of the back-projection module by following a similar schema as in [21]. We utilize the Adam optimizer [27] with a learning rate of 1×10^{-3} and a batch size of 1. We use the validation data as the stop criterion and the model converges after 300 epochs. To show the influence brought by the different prior knowledge, we also calculate an averaged shape of the dental arch as another type of prior information for comparison, which is marked as Oral-3D (Avg) in the following paper.

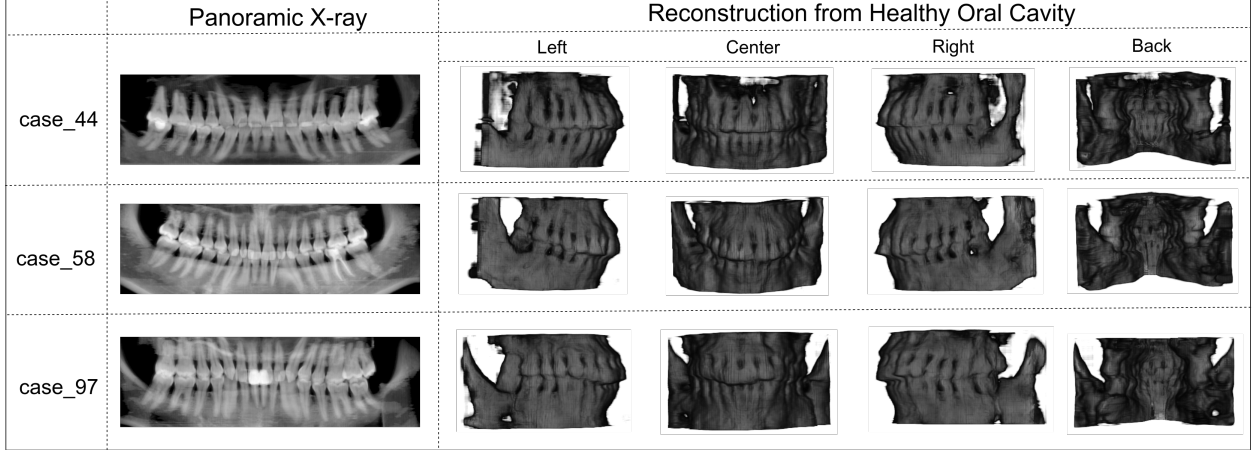


Figure 6: We show examples of reconstruction for a healthy cavity in this picture, where we can see that the surface of teeth and mandible can be clearly seen from multiple views.

4.4 Reconstruction Results

We show the reconstruction results under three scenes of oral cavity: healthy oral cavity, oral cavity with missing teeth, oral cavity with redundant teeth. The last two situations are two of the most general problem people can meet during a dental examination. These abnormalities can also indicate if our model is effective. Additionally, we give an insight view to see the photos sliced by different directions to observe the restored density information, which is an important inference for dental diagnosis.

Healthy Oral Cavity We first reconstruct the 3D structure for the healthy cases and show the results in Fig. 6. We can see that the structure surface can be seen clearly from different views.

Oral Cavity with Missing Teeth Then we reconstruct the cases with missing teeth to prove that our model is not just over-fitting. As shown in Fig. 7, the missing structures can be accurately reflected in the 3D structure.

Oral Cavity with Redundant Teeth Wisdom teeth usually grow in a vertical direction to the other teeth and are the most common redundant teeth for humans. As shown in Fig. 8, these redundant teeth can also be found in the reconstructed bone structure.

Density Information As our model can also indicate the density information of the bony tissues inside the oral cavity, we show some slices of our reconstructed results in Fig. 9. It is obvious that our framework can recover the density information with high quality, which is of great help for dental diagnosis.

4.5 Comparison with Baseline Models

To show the effectiveness of our method, we compare *Oral-3D* with several general models as baselines, such as the Residual CNN introduced by [13] and a condition GAN that directly back projects the panoramic X-ray image into the 3D bone structure. These networks are trained with the settings as *Oral-3D* except for the utilization of the prior knowledge in dental arch. We compare the reconstruction results from three views as following:

Quantitative Comparison We first compare the results of different methods by the metrics in Section 4.2. As shown in Table. 2, *Oral-3D* with individual dental arch information is significantly better than those without any prior knowledge, such as Residual CNN, where the model needs to learn both the shape and location information at the same time.

Qualitative Comparison of Restored Shape We first show the 3D bone structure in Fig. 10 from views of center, left, right, and back. By comparing results generated by different methods, we can find that *Oral-3D* can obtain a much clearer surface than the single-stage model. That is quite reasonable as our generation network only needs to

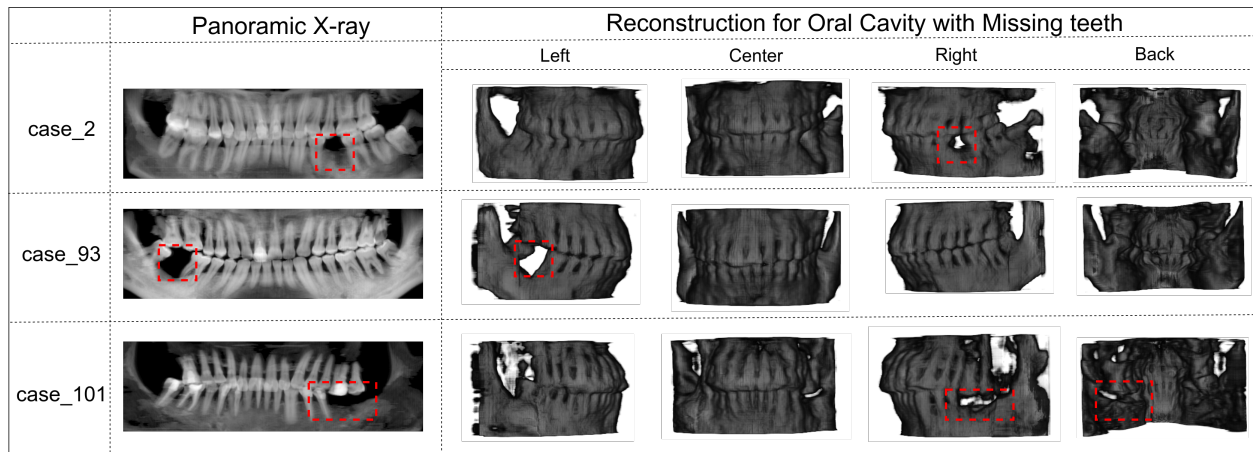


Figure 7: We mark the missing teeth in this picture with the bounding box, where we can see this information is also included in the reconstructed oral cavity.

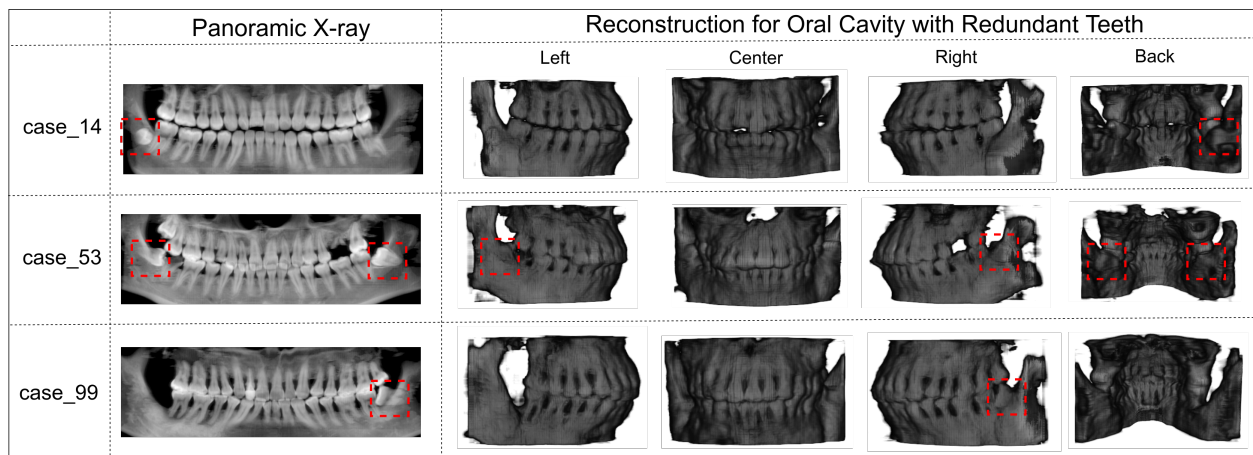


Figure 8: Reconstruction results for patients with wisdom teeth can be seen in this picture. We mark the wisdom teeth with the bounding box. And we can also these redundant teeth in the 3D structure.

Table 2: Evaluation on 3D Reconstruction

Method	Prior Info	IOU (%)	PSNR (dB)	SSIM (%)
Residual CNN	No	50.71	15.54	67.21
Dense CNN	No	47.90	14.80	64.93
GAN	No	49.01	15.45	55.17
Oral-3D	Avg	52.67	15.22	70.62
Oral-3D	Yes	63.77	17.36	75.54

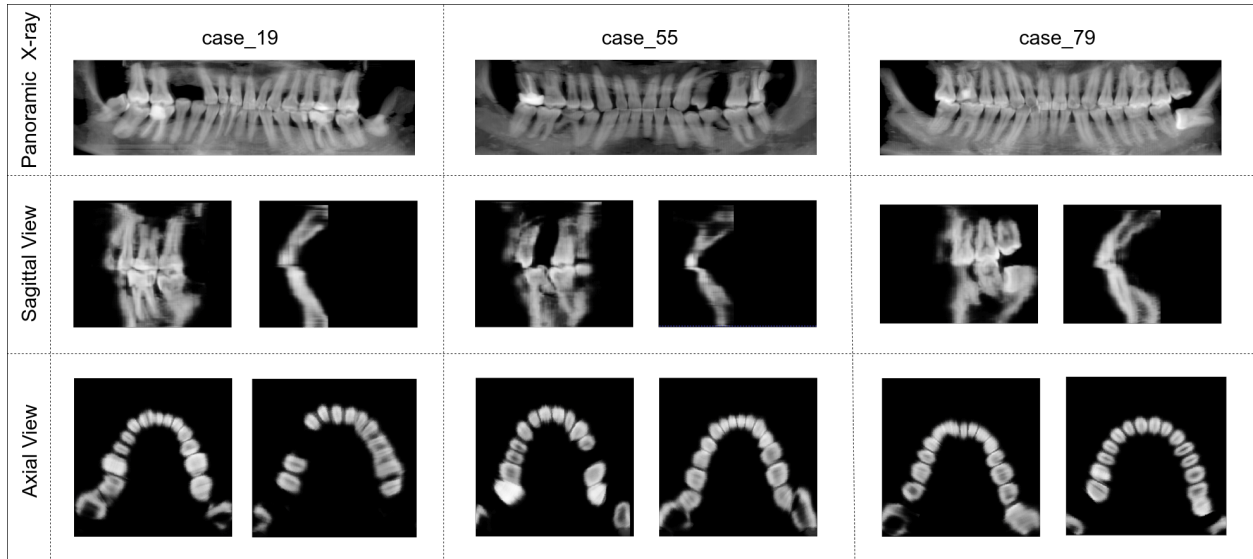


Figure 9: The reconstructed bone structure can restore the density information at the same time. Therefore, we can distinguish different bony tissues inside the oral cavity. This would provide useful information for dentists to find abnormality for each teeth.

learn the depth information. Specifically, we can see that there is a missing tooth in the upper right of the mouth in the ground truth. This information can be seen in both *Oral-3D* and *Oral-3D (Avg)*, but is not visible in the generation results of other methods. This can directly indicate the effectiveness of our reconstruction method.

Qualitative Comparison of Restored Density We also compare the restored density information by showing axial slices in Fig. 11, where we can see that our method can predict a much better result than these single-stage methods. By comparing *Oral-3D* with different prior information, we can also see that the primary contribution of the individual dental arch knowledge is to locate the position and restrict the shape of oral cavity.



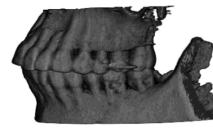
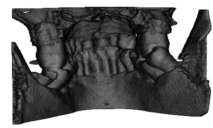
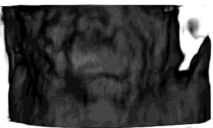
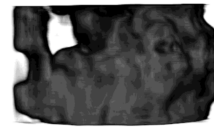
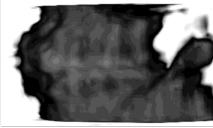
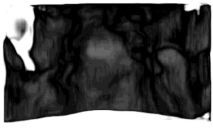
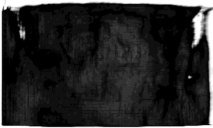
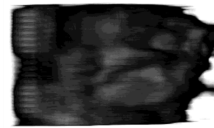
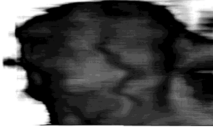



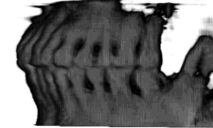
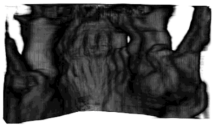
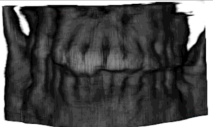


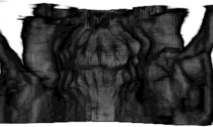
	Center	Left	Right	Back
Groundtruth				
Res-CNN				
GAN				
Oral-3D (Avg)				
Oral-3D				

Figure 10: Comparison of bone structures generated by different methods.

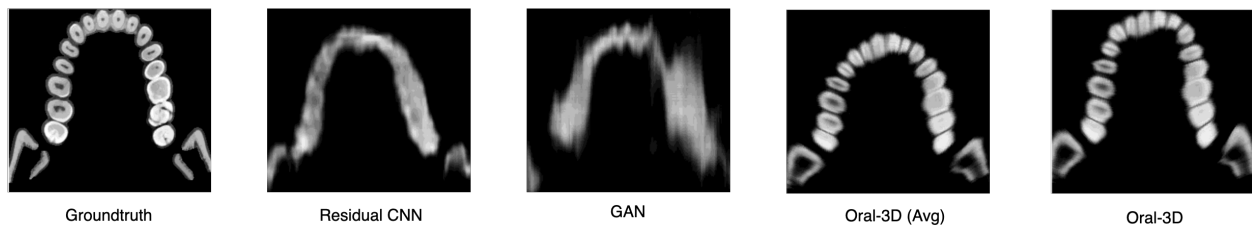


Figure 11: Comparison of axial slices generated by different methods.

5 Conclusion

In this paper, we propose a two-step framework to reconstruct the 3D bone structure of oral cavity from a single 2D panoramic X-ray image, where an averaged or individual shape information of the dental arch is provided as prior knowledge. We first utilize a generative model to back project the 2D image into 3D space, then take the generated 3D image for registration to reconstruct the 3D bone structure. Experimental results show that our model can restore both the shape and the density information with much higher quality than other methods. We hope this work can contribute to the oral healthcare from a novel attitude.

6 Acknowledgement

We greatly thank the doctors from Nanjing Stomatological Hospital, Medical School of Nanjing University to provide the dataset and helpful advice.

References

- [1] Petersen, L.B., Olsen, K.R., Christensen, J., Wenzel, A.: Image and surgery-related costs comparing cone beam ct and panoramic imaging before removal of impacted mandibular third molars. *Dentomaxillofacial Radiology* **43**(6) (2014) 20140001
- [2] Brooks, S.L.: Cbct dosimetry: orthodontic considerations. In: *Seminars in Orthodontics*. Volume 15., Elsevier (2009) 14–18
- [3] Momin, M.A., Okochi, K., Watanabe, H., Imaizumi, A., Omura, K., Amagasa, T., Okada, N., Ohbayashi, N., Kurabayashi, T.: Diagnostic accuracy of cone-beam ct in the assessment of mandibular invasion of lower gingival carcinoma: comparison with conventional panoramic radiography. *European journal of radiology* **72**(1) (2009) 75–81
- [4] Cui, Z., Li, C., Wang, W.: Toothnet: automatic tooth instance segmentation and identification from cone beam ct images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2019) 6368–6377
- [5] Lee, J.H., Kim, D.H., Jeong, S.N., Choi, S.H.: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of dentistry* **77** (2018) 106–111
- [6] Hatvani, J., Horváth, A., Michetti, J., Basarab, A., Kouamé, D., Gyöngy, M.: Deep learning-based super-resolution applied to dental computed tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences* **3**(2) (2018) 120–128
- [7] Imangaliyev, S., van der Veen, M.H., Volgenant, C.M., Keijsers, B.J., Crielaard, W., Levin, E.: Deep learning for classification of dental plaque images. In: *International Workshop on Machine Learning, Optimization, and Big Data*, Springer (2016) 407–410
- [8] Prajapati, S.A., Nagaraj, R., Mitra, S.: Classification of dental diseases using cnn and transfer learning. In: *2017 5th International Symposium on Computational and Business Intelligence (ISCBI)*, IEEE (2017) 70–74
- [9] Jin, C.B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X.: Deep ct to mr synthesis using paired and unpaired data. *Sensors* **19**(10) (2019) 2361
- [10] Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. (2017) 2223–2232
- [11] Costa, P., Galdran, A., Meyer, M.I., Abràmoff, M.D., Niemeijer, M., Mendonça, A.M., Campilho, A.: Towards adversarial retinal image synthesis. *arXiv preprint arXiv:1701.08974* (2017)
- [12] Choi, H., Lee, D.S.: Generation of structural mr images from amyloid pet: application to mr-less quantification. *Journal of Nuclear Medicine* **59**(7) (2018) 1111–1117
- [13] Henzler, P., Rasche, V., Ropinski, T., Ritschel, T.: Single-image tomography: 3d volumes from 2d cranial x-rays. In: *Computer Graphics Forum*. Volume 37., Wiley Online Library (2018) 377–388
- [14] Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2ct-gan: reconstructing ct from biplanar x-rays with generative adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2019) 10619–10628
- [15] Kolev, K., Brox, T., Cremers, D.: Fast joint estimation of silhouettes and dense 3d geometry from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(3) (2012) 493–505
- [16] Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S.: 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In: *European conference on computer vision*, Springer (2016) 628–644
- [17] Yang, G., Cui, Y., Belongie, S., Hariharan, B.: Learning single-view 3d reconstruction with limited pose supervision. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. (2018) 86–101
- [18] Wu, J., Zhang, C., Zhang, X., Zhang, Z., Freeman, W.T., Tenenbaum, J.B.: Learning shape priors for single-view 3d completion and reconstruction. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. (2018) 646–662
- [19] Fernandez-Labrador, C., Perez-Yus, A., Lopez-Nicolas, G., Guerrero, J.J.: Layouts from panoramic images with geometry and deep learning. *IEEE Robotics and Automation Letters* **3**(4) (2018) 3153–3160
- [20] Mirza, M., Osindero, S.: Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014)
- [21] Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. (2017) 2794–2802

- [22] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2017) 4700–4708
- [23] Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM journal on applied mathematics* **66**(5) (2006) 1632–1648
- [24] Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split bregman method: segmentation and surface reconstruction. *Journal of Scientific Computing* **45**(1-3) (2010) 272–293
- [25] 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. *Computer methods and programs in biomedicine* **175** (2019) 205–214
- [26] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4) (2004) 600–612
- [27] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)