Oral-3D: Reconstructing the 3D Structure of Oral Cavity from Panoramic X-ray

Weinan Song [*], Yuan Liang [*], Jiawei Yang, Kun Wang, Lei He
Design Automation Laboratory, University of California, Los Angeles
wsong@ucla.edu, lhe@ee.ucla.edu

Abstract

Panoramic X-ray (PX) provides a 2D picture of the patient's mouth in a panoramic view to help dentists observe the invisible disease inside the gum. However, it provides limited 2D information compared with cone-beam computed tomography (CBCT), another dental imaging method that generates a 3D picture of the oral cavity but with more radiation dose and a higher price. Consequently, it is of great interest to reconstruct the 3D structure from a 2D X-ray image, which can greatly explore the application of X-ray imaging in dental surgeries. In this paper, we propose a framework, named *Oral-3D*, to reconstruct the 3D oral cavity from a single PX image and prior information of the dental arch. Specifically, we first train a generative model to learn the cross-dimension transformation from 2D to 3D. Then we restore the shape of the oral cavity with a deformation module with the dental arch curve, which can be obtained simply by taking a photo of the patient's mouth. To be noted, *Oral-3D* can restore both the density of bony tissues and the curved mandible surface. Experimental results show that *Oral-3D* can efficiently and effectively reconstruct the 3D oral structure and show critical information in clinical applications, *e.g.*, tooth pulling and dental implants. To the best of our knowledge, we are the first to explore this domain transformation problem between these two imaging methods.

Introduction

Extra-oral imaging techniques such as PX and CBCT are widely used in dental offices as examination methods before the treatment. Both methods can show detailed bone information, including the tooth, mandible, and maxilla, of the entire oral cavity. However, during the imaging process of PX, the X-ray tube moves around the patient's head and can only take a 2D panoramic picture. This has limited its application in the cases when the disease needs to be located. In comparison, CBCT can reconstruct the whole 3D structure of the lower head with divergent X-rays and provide abundant information about the health condition of the oral cavity. Nevertheless, the patient needs to take more radiation dose and pay a higher price during a CBCT scan. We summarize the characteristics of these two imaging methods in Table 1. We can see that although CBCT can provide more information in clinical applications (Momin et al. 2009), it generates 39.4× radiation (Brooks 2009) and takes 3.7× of the price (Petersen et al. 2014) on average than PX. This problem is especially evident for those sensitive to the radiation dose and the developing countries where people are unwilling to invest much in dental healthcare. Therefore, it is of great interest to directly reconstruct the 3D structure of the oral cavity from a PX image.

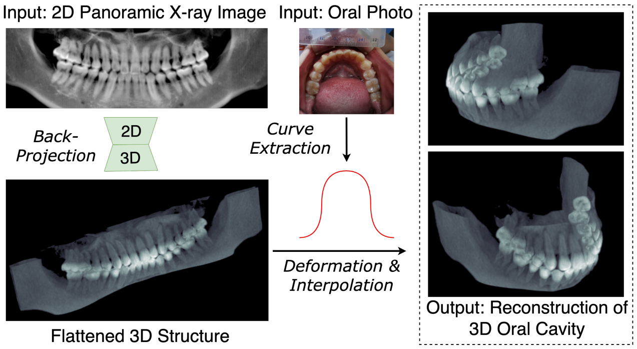
Input: 2D Panoramic X-ray Image
Input: Oral Photo
Back-Projection
2D
3D
Curve Extraction
Deformation & Interpolation
Flattened 3D Structure
Output: Reconstruction of 3D Oral Cavity

Figure 1: An overview of *Oral-3D*. We first back-project the panoramic image into a flattened 3D image of the oral cavity with a generative network, then deform the generation result into a curved plane according to the dental arch.

However, it is of great challenge to reconstruct a 3D object from a single 2D image due to the lack of spatial information in the rendering direction. Most works rely on additional information, such as shadow or prior shape of the object, to regularize the reconstruction result. Furthermore, this problem is more difficult for the oral cavity due to the complicated shape of the mandible and detailed density information of the teeth. To overcome such challenges, we propose a two-stage framework, named *Oral-3D*, to generate a high-resolution 3D structure of the oral cavity by decoupling the reconstruction process of shape and density. We first train a generation model to extract detailed density information from the 2D space, then restore the mandible shape with the prior knowledge of the dental arch. Although our method can not fully replace CBCT in the dental examination, we provide a compromise solution to obtain the 3D oral structure when only the PX is available.

[*] Equal Contributions

Imaging Method	Dimension	Cost	Radiation Dose	Diagnostic Accuracy	Wisdom Tooth	Tooth Decay	Implant	Orthodontics
CBCT	3D	€184.44	58.9-1025.4 μSv	94.8%	✓	✓	✓	✓
PX	2D	€49.29	5.5-22.0 μSv	83.3%	✓	✓		

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

Our work can be summarized as a combination of a single-view reconstruction problem and a cross-modality transformation problem, where the model should recover both the shape and density information of the target object from a single image at the same time. We show an overview of *Oral-3D* in Figure 1. At first, we train a generation network to learn the cross-dimension transformation that can back-project the 2D PX image 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, a flattened oral structure, into a curved plane to restore the original shape according to the dental arch. This prior knowledge effectively restricts the shape and location of the 3D structure and can be obtained in many ways, such as by fitting the β function with the width and depth of the mouth (Braun et al. 1998). To show the effectiveness of our framework, we first compare *Oral-3D* with other methods on synthesized images generated from a CBCT dataset. Then we evaluate the reconstruction results for some clinical cases to prove the feasibility of our method. Experimental results show that *Oral-3D* can reconstruct the 3D oral structure with high quality from a single panoramic X-ray image and keep the density information simultaneously. In conclusion, we make the following contributions:

- We are the first to explore the cross-modality transfer of images in different dimensions for dental imaging by deep learning. In addition to restoring the 3D shape and surface of the bone structure, our model can restore the density information simultaneously, which is of great help for dental diagnosis.
- We decouple the reconstruction process for density and shape recovery by proposing a deformation module that embeds a flattened 3D image into a curved plane. This has not been addressed in previous research and can significantly improve the reconstruction performance.
- We propose an automatic method to generate paired 2D and 3D images to train and evaluate the reconstruction models, where *Oral-3D* achieves relatively high performance and can show key features of some typical cases. Meanwhile, we propose a workflow to evaluate our model on a real-world dataset, which indicates the feasibility of clinical applications.

Related Work

Deep Learning for Oral Health

Deep learning has dramatically promoted the computer assistance system for dental healthcare by automatically learning feature representations from large amounts of data. For example, (Cui, Li, and Wang 2019) proposes an automatic method for instance-level segmentation and identification of teeth in the CBCT image. (Liang et al. 2020a) utilizes the smartphone to diagnose common dental disease with a detection model. (Imangaliyev et al. 2016) designs a classification model for red auto-fluorescence plaque images to assist in detecting dental caries and gum diseases. (Prajapati, Nagaraj, and Mitra 2017) uses transfer learning to classify three different oral diseases for X-ray images. Although these methods have improved oral healthcare service by providing intelligent assistance, the model needs to be trained with annotations on large datasets, which requires both professional knowledge and tedious labour. Compare with these works, our model helps dental healthcare without the supervision of labelled data, where the reconstruction is learned from the latent relationship between 2D and 3D images.

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 a medical examination or provide additional training samples without repetitive annotation work to augment the dataset. Most works take this as a pix-to-pix transfer problem, where the layout and the structure are consistent, but the colour distribution is changed after the transformation between images in different modalities. For example, as shown in Figure 2, (Costa et al. 2017) takes the vessel tree of eyes as a condition to synthesis new images for fundus photography. (Choi and Lee 2018) proposes a generation network to produce realistic structural MR images from florbetapir PET images. However, few studies 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 high-dimension information from the lower-dimension image. We only find two works that achieve a similar target to ours. Specifically, (Henzler et al. 2018) 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, (Ying et al. 2019) utilizes biplanar X-rays to extract 3D anatomical structures of body CT with adversarial training and reconstruction constraints. However, our problem is quite different from theirs as the PX image can not be synthesized only from the orthogonal projection over the corresponding CT. Besides, our task is more challenging due to the complicated structure of the oral cavity, where the model is required to restore more details of the teeth and mandible. (Liang et al. 2020b) proposes a very similar work to ours. However, their model requires pixel-wise annotation for each tooth, and they only reconstruct the

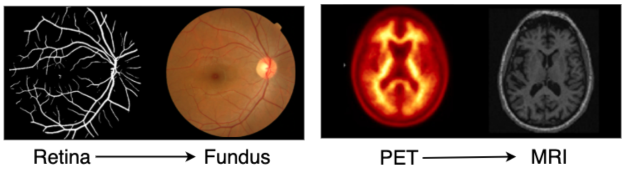

Figure 2: We show some examples of cross-modality transfer for Retina → Fundus (Costa et al. 2017) and PET → MRI (Choi and Lee 2018), where the source image and the target image usually contains consistent physiological structures although in different modalities.

shape for each tooth. In contrast, our model is unsupervised and can recover the entire oral structure with density distribution.

3D Reconstruction from 2D Image

The recent 3D reconstruction work from 2D images can be concluded as two categories: multi-view reconstruction and single-view reconstruction. For the first one, the method generally requires little prior knowledge about the 3D object as the images taken from multiple angles can restrict the reconstruction shape. For example, (Kolev, Brox, and Cremers 2012) computes the most probable 3D shape that gives rise to the observed colour information from a series of calibrated 2D images. (Choy et al. 2016) 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 inference the object shape. As such, (Yang et al. 2018) proposes a unified framework trained with a small amount of pose-annotated images to reconstruct a 3D object. (Wu et al. 2018) takes the adversarially learned shape priors as a regularizer to penalize the reconstruction model. However, the PX image can be either seen as a single-view image taken by a moving camera or a concatenate image blended with multiple views. In this paper, we take our problem as the first kind to decouple the reconstruction process for the bone density and the mandible shape. In the experiment, we also show that this can significantly promote performance over the multi-view reconstruction model both in quality and quantity.

Method

In this section, we introduce our framework that reconstructs a high-resolution 3D oral cavity from a 2D PX image. We choose to break this problem into two stages to recover more details of the bone density. We show the structure of *Oral-3D*, which consists of a back-projection module and a deformation module in Figure 3. The back-projection module develops from generative adversarial networks (GAN (Goodfellow et al. 2014)), where the generator is trained to learn a back-projection transformation by exploring the depth information contained in the X-ray image. The deformation module takes in the generated 3D image (Image c) from the back-projection module and the dental arch (Image e) to restore the curved shape of the mandible.

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 learns to output a fake image from a random vector to deceive the discriminator, while the discriminator tries to distinguish sampling data between real and fake images. As we aim to generate the consistent 3D content from the semantic information of the panoramic X-ray image, we utilize conditional GANs (Mirza and Osindero 2014) as the generative model to learn the back-projection transformation.

Objective Function To improve the generation quality and guarantee the stable training process, we use LSGAN (Mao et al. 2017) as the keystone to train the generator and discriminator. The adversarial loss can be summarized as:

$$\begin{aligned} Loss_D &= \mathbb{E}_y\left[(D(y)-1)^2\right] + \mathbb{E}_x\left[D(G(x))^2\right] \\ Loss_G &= \mathbb{E}_x\left[D(G(x))-1)^2\right], \end{aligned} \tag{1}$$

where *x* is the PX image and *y* is the flattened oral structure.

To maintain the structural consistency of the input and generation result, we also introduce the reconstruction loss and projection loss to improve the generation quality. These proposed loss functions can bring voxel-wise and plane-wise regularization to the generation network, which can be defined as:

$$\begin{aligned} Loss_R &= \mathbb{E}_{x,y}\left[(y-G(x))^2\right] \\ Loss_P &= \mathbb{E}_{x,y}\left[(P(y)-P(G(x)))^2\right], \end{aligned} \tag{2}$$

where the function $P()$ is achieved by orthogonal projections along each dimension of the generated 3D image. In summary, the total optimization problem can be concluded as:

$$\begin{aligned} D^* &= \arg\min_D Loss_D \\ G^* &= \arg\min_G \lambda_1 \cdot Loss_G + \lambda_2 \cdot Loss_R + \lambda_3 \cdot Loss_P. \end{aligned} \tag{3}$$

Generator During the X-ray imaging, the depth information can be reflected in the absorption of radiation through the bone. Therefore it is reasonable to extract the thickness of the tooth and the mandible from a PX image. Then the objective for the generator is to find a cross-dimension transformation G from 2D to 3D, which can be denoted as:

$$G : I^{2D}_{H\times W} \rightarrow I^{3D}_{H\times W\times D}, \tag{4}$$

where I^{2D} is the PX image with a size of $H \times W$ and I^{3D} is the flattened 3D structure with a size of $H \times W \times D$. In this paper, we utilize 2D convolution to retrieve the latent depth information. The 3D information is embedded into different channels of feature maps. As shown in Fig. 3, the encoding network decreases the resolution of feature maps but increases the number of feature channels, while the decoding network increases the resolution to generate a 3D object. The output voxel value is restricted to $(-1, 1)$ with a $tanh$ layer at the end.

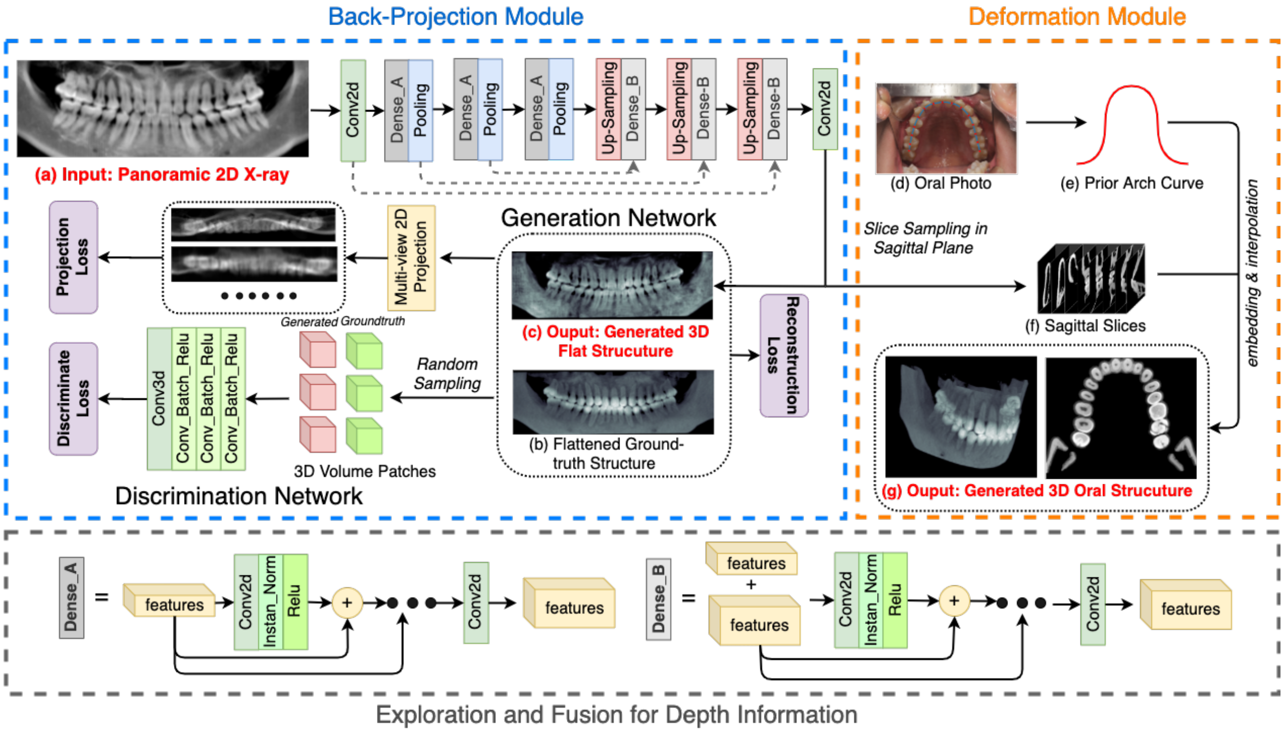

Figure 3: Our framework consists of two modules to decouple the recovery of bone density and the mandible shape. The back-projection module utilizes a generation network to restore the 3D density information from the 2D space, and the deformation module transforms the flattened 3D image into a curve plane according to the prior knowledge in the dental arch.

Dense Block Dense connections (Huang et al. 2017) have shown compelling advantages for feature extraction in deep neural networks. This architecture is especially efficient in forwarding 3D information as each channel of the output has a direct connection with intermediate feature maps. In the back-projection module, we utilize two kinds of dense blocks, noted as A and B, to extract depth information from the X-ray image. As shown at the bottom of Figure 3, the dense block A explores the depth information by increasing the channel number of feature maps. In contrast, the dense block B fuses feature maps from the $up - samplinglayer$ and the skip-connections but maintain the number of channels to forward the depth information. In the end, the number of stacked features in the output is equal to the depth of the generated 3D image.

Discriminator The discriminator has been frequently used in many generative models to improve the generation quality by introducing an instance-level loss. In the back-projection module, we adopt a patch discriminator introduced by (Isola et al. 2017) to improve the generation quality of tooth edges by learning high-frequency structures. We set the patch size as 70^3 and follow a similar structure in (Isola et al. 2017) but replace 2D convolution with 3D. The discrimination network ends with a $Sigmoid$ layer to predict the probability of the samples belonging to the real image. To be noted, we sample the same number of 3D patches at the same position from the paired of 3D images.

Deformation Module

With the generation of a 3D image from the back-projection module, the deformation model maps the flattened 3D structure into the curved space according to the arch curve to output the final reconstruction object. As shown in the right part of Figure 3, we propose a registration algorithm that can best restore the shape of the oral cavity and keep the recovered density information. We first sample the generated 3D image (Image c) into slices (Image f) in the sagittal plane, then interpolate these slices along the dental arch curve (Image e). To achieve this, we sample a number of points from the curve with equal distance and embed the slices into the curve. In the end, we interpolate the voxels between the neighbouring slices to output a smooth 3D image (Image g). For computation convenience, we combine these steps together and conclude it in Algorithm 1, where we assume that the generated 3D image and the bone model has the same height of H.

Experiment

Dataset

As grouped data of PX image, dental arch shape, and 3D oral structure of the same patient, especially in the same period, is hard to find, we first use synthesized data to evaluate the performance. We collect 100 CBCT scans from a major stomatological hospital in China and re-sample these 3D images into a size of $288 \times 256 \times 160$. The dataset is finally

Algorithm 1 Embedding and Interpolation

```
function REGISTER($Slices, W_{3D}, D_{3D}, curve$)
    $W, H, D \leftarrow$ SHAPE($Slices$)
    $OralImage \leftarrow$ ZEROS($W_{3D}, H, D_{3D}$)
    $SamplePoints \leftarrow$ SAMPLE($curve$)
    for $i = 0; i < W_{3D}; i++$ do
        for $j = 0; j < D_{3D}; j++$ do
            $id, dist \leftarrow$ DIST($(i, j), SamplePoints$)
            $Slice \leftarrow Slices[id, :, :]$
            $Slice \leftarrow$ INTERPOLATE($Slice, dist$)
            $OralImage[i, :, j] \leftarrow Slice$
        end for
    end for
    return $OralImage$
end function
```

normalized into a range of $(-1, 1)$ and split into a ratio of $3 : 1 : 1$ for training, validation, and testing.

An overview of preparing the synthesized data can be seen in Figure 4. We first obtain a 2D image in the axial plane by maximum intensity projection (MIP) (Image b) over the CBCT slices (Image a). Then we obtain the dental curve with a similar method as in (Yun et al. 2019) to estimate the curve function and boundaries of the dental arch. To generate the PX image (Image d), we simulate projection with the Beer-Lambert absorption-only model along the arch curve. This imaging process is similar to the way for a real PX machine, where the manufacturer usually improves the imaging quality by designing a trajectory of the camera to fit the mandible shape. Finally, we extract the 3D oral structure (Image e) by removing the unrelated tissues with the boundaries and generate the flattened 3D structure (Image c) by re-sampling along the arch curve.

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.
- **Dice:** In order to reflect the deformation of the reconstruction, we use dice coefficient between our reconstruction results and the groundtruth in a volume level of the oral cavity. The 3D volume of the oral cavity is obtained by setting a threshold (*e.g.,*-0.8 over the reconstruction result.
- **SSIM:** We use the structural similarity index (SSIM) (Wang et al. 2004) as the key criterion to quantify the performance of density recovery.SSIM considers the brightness, contrast and structure information at the same time and can match better the subjective evaluation of humans. It can effectively indicate the reconstruction quality and is widely used in other similar works, such as (Ying et al. 2019).
- **Overall:** To combine these three metrics together, we also define a score $S = (PSNR/20 + Dice + SSIM)/3$ to compare the overall performance of the 3D reconstruction.

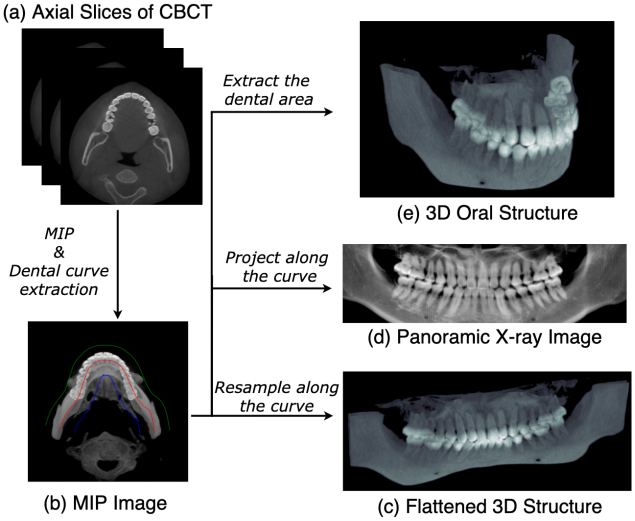

Figure 4: An overview of generating paired data for 3D oral structure and 2D panoramic X-ray is shown in this picture. We first get the MIP image from the CBCT scan to obtain the dental arch curve (red), and boundaries of the dental area (blue and green). Then we obtain the flattened oral structure, PX image, and the 3D oral structure by re-sampling, projection, and extraction, respectively.

Comparison Models

To show the effectiveness and efficiency of Oral-3D, we also compare our framework with other models that work on a similar problem:

- **Residual CNN:** An encoder-decoder network that has been introduced in (Henzler et al. 2018) to reconstruct the 3D model with a single X-ray.
- **GAN:** A generative model based on (Goodfellow et al. 2014) that takes the Res-CNN as the backbone for generator with reconstruction loss and the same discriminator as *Oral-3D*.
- **R2N2:** We transform our task into a multi-view reconstruction problem to train R2N2 (Choy et al. 2016) by taking the PX image as a composition of X-ray image taken from three different views.
- **Oral-3D (Auto-Encoder)** We remove the discriminative network in *Oral-3D* and keep the encoder-decoder network only in the back-projection module.

Training

All the experiment are trained by Adam optimizer (Kingma and Ba 2014) with a batch size of 1 for 300 epochs. The learning rate starts at 1×10^{-3} and decreases 10 times every 50 epochs. We use the validation data as the stop criterion, and all models converge after 300 epochs. For adversarial

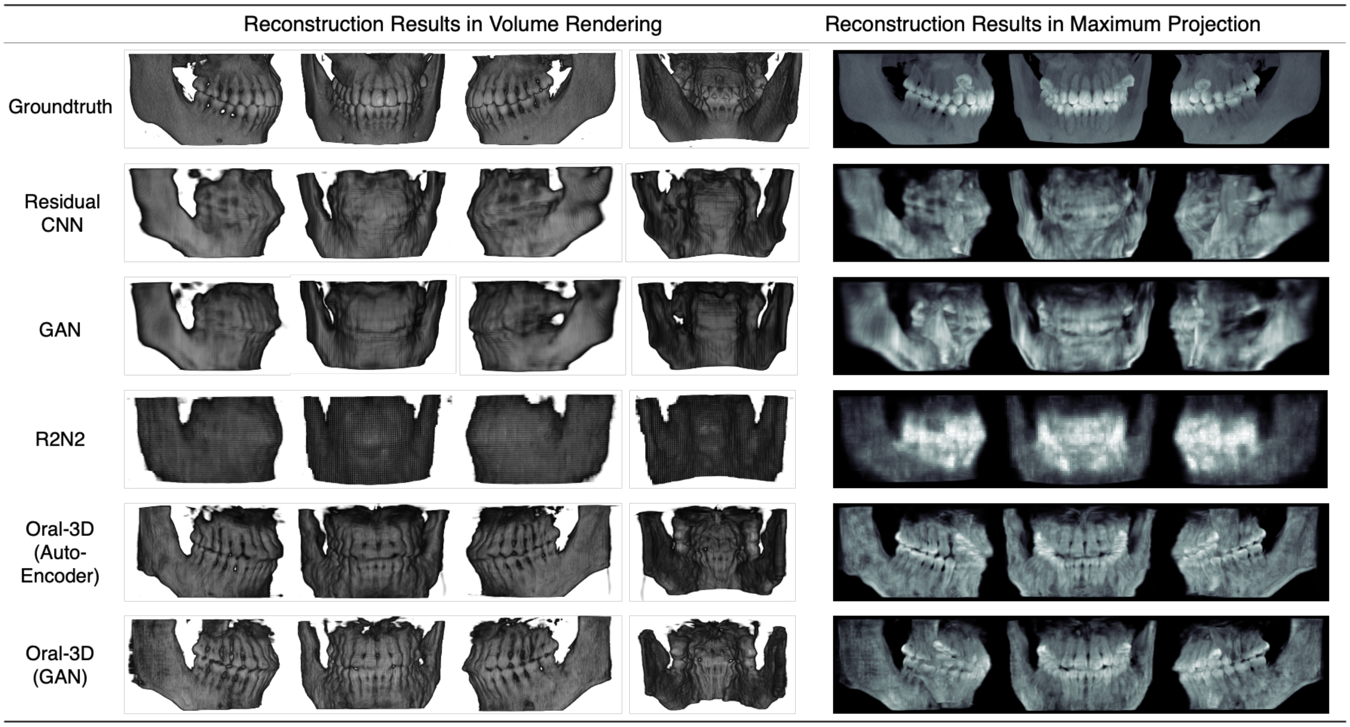
Reconstruction Results in Volume Rendering
Reconstruction Results in Maximum Projection
Groundtruth
Residual CNN
GAN
R2N2
Oral-3D (Auto-Encoder)
Oral-3D (GAN)

Figure 5: We show the qualitative comparison from different views and rendering ways in this picture. We can see that our method generates the best results with more detailed density and a more sharp surface.

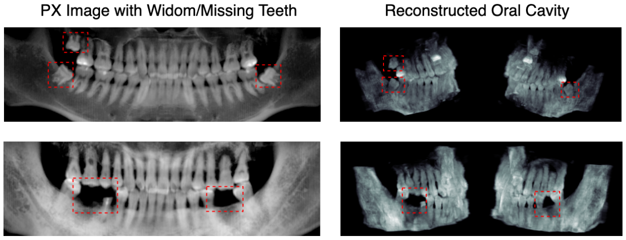

Figure 6: We show reconstruction results for patients with wisdom/missing teeth and mark the key features with red bounding boxes. We can see that our method can accurately locate these positions, which can be an important reference during the surgery.

networks, i.e. *Oral-3D* and GAN, we introduce the discriminative network after 100 epochs to alleviate the influence of discrimination loss at the beginning.

Results

In this section, we evaluate the reconstruction performance of *Oral-3D* from different perspectives. We first compare *Oral-3D* with other methods qualitatively and quantitatively. Then we show the results of special cases for some typical dental applications. In the end, we do clinical trials by evaluating our method on real-world images.

Comparison with Other Methods

We first show the results in Figure 5, where the volume rendering can show the reconstructed surface, and the maximum projection can indicate the restored density distribution. Then we summarize the evaluation metrics in Table 2 to compare with other methods. We can see that *Oral-3D* has the best performance over other models. Comparing *Oral-3D* with the Residual CNN and GAN, we can see the importance of decoupling the back-projection and deformation process. To be noted, R2N2 achieves the worst performance, where the model only learns the shape of the oral cavity but loses details of teeth. This has indicated the defect when converting the PX image as a collection of multi-view images. Additionally, we see that the *Oral-3D (Auto-Encoder)* has the closest performance to *Oral-3D*, although the latter has a more clear surface. This has proved the promotion brought by the adversarial loss.

Identification of Wisdom/Missing Teeth

In this paragraph, we show two of the most common cases in dental healthcare, *e.g.*, dental implants and tooth pulling, to see if *Oral-3D* can provide dentist useful reference. Both cases require locating the operation location before the surgery. In the first row of Figure 6, three wisdom teeth can be seen clearly on both sides in PX. These features also present in the two sides of the reconstruction results. In the second row, the patient misses two teeth on both sides of the mandible. While the missing place can also be located accurately in the reconstruction image.

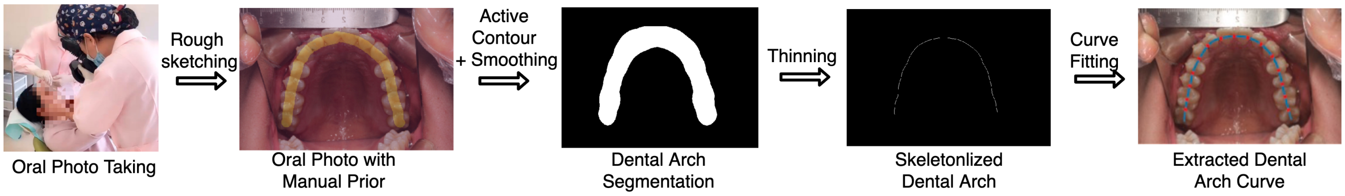
Oral Photo Taking
Rough sketching
Oral Photo with Manual Prior
Active Contour + Smoothing
Dental Arch Segmentation
Thinning
Skeletonlized Dental Arch
Curve Fitting
Extracted Dental Arch Curve

Figure 7: We show a workflow to apply *Oral-3D* to obtain the dental arch curve in real-world applications in this picture. We first take a picture of the patient's mouth and segment then dental area semi-automatically. Then we use a cubic function to the fit points sampled from the skeletonized image of the binary mask.

Method	View	Prior	D-Net	PSNR (dB)	SSIM (%)	Dice (%)	Overall
Residual CNN	1	No	No	17.46±9.58	72.90±2.09	57.95±7.43	73.54
GAN	1	No	Yes	17.71±1.04	69.96±1.91	57.80±7.76	73.78
R2N2	3	No	No	18.06±0.94	71.94±1.36	57.71±6.52	73.32
Oral-3D (Auto-Encoder)	1	Yes	No	19.04±0.85	76.78±1.65	69.68±4.98	80.56
Oral-3D (GAN)	1	Yes	Yes	**19.22±0.83**	**78.27**±1.74	**71.28**±4.69	**81.89**

Table 2: Quantitative evaluation of 3D reconstruction results.

	DL only	DL+PL	DL+RL	DL+RL+PL
PSNR	8.06	18.06	19.14	**19.22**
SSIM	46.61	73.0	**78.41**	78.27
Dice	35.50	64.53	70.89	**71.28**
Overall	40.79	75.95	81.66	**81.89**

Table 3: Evaluation results of different combination of discrimination loss (DL), reconstruction loss (RL), and projection loss (PL).

Dataset	PSNR	SSIM	Dice
Real	17.36±0.70	69.30±2.03	71.44±3.66
Synthesized	19.22±0.83	78.27±1.74	71.28±4.69

Table 4: Evaluation results on real-world images.

Ablation Study

To reveal the factors that influence the reconstruction quality of the generation network, we also do an ablation by changing the combination of the loss functions. As shown in Table 3, we see that the model shows the worst performance if trained only with the adversarial loss. This is mainly because the adversarial loss can not bring voxel-wise optimization. We can also see that the major improvement comes from the reconstruction loss, while the projection loss brings much less promotion, especially when trained with the reconstruction loss together. This is also reasonable as the reconstruction loss can supervise the generation network to learn more detailed information.

Clinical Trials

In the end, we evaluate *Oral-3D* on real-world data from 6 patients. The workflow of collecting dental arch information is shown in Figure 7. We use cycleGAN (Zhu et al. 2017) to alleviate the colour variance between the training and testing PX images. As shown in Table 4, the drop mainly comes from the PSNR and SSIM, which is because of the colour variance in different CBCT machines. From Figure 8 we can that although the quality decreases in density details, we can still identify each tooth in the reconstruction result.

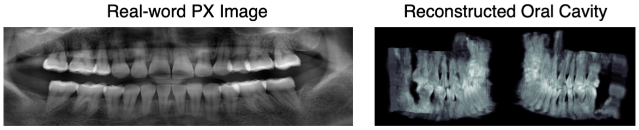

Figure 8: Although the quality decreases in density details for real-word PX, we can still identify each tooth in the reconstruction result.

Conclusion

In this paper, we propose a two-stage framework to reconstruct the 3D structure of the oral cavity from a single 2D PX image, where 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 deform the generated 3D image into a curved plane to restore the oral shape. We first use synthesized data to compare with different methods, then evaluate the model with real-world data to see the feasibility in clinical applications. Experimental results show that our model can recover both the shape and the density information in high resolution. We hope this work can help improve dental healthcare from a novel attitude.

Acknowledgements

We thank Dr. Liang Chengwen, Dr. Wang Bin, and Dr. Wang Weiqian from Hangzhou Stomatological Hospital for clinical suggestions and data collection. We also thank Prof. Wu Yingnian and Dr. Gao Ruiqi from UCLA for their academic advice.

References

Braun, S.; Hnat, W. P.; Fender, D. E.; and Legan, H. L. 1998. The form of the human dental arch. *The Angle Orthodontist* 68(1): 29–36.

Brooks, S. L. 2009. CBCT dosimetry: orthodontic considerations. In *Seminars in Orthodontics*, volume 15, 14–18. Elsevier.

Choi, H.; and Lee, D. S. 2018. Generation of structural MR images from amyloid PET: application to MR-less quantification. *Journal of Nuclear Medicine* 59(7): 1111–1117.

Choy, C. B.; Xu, D.; Gwak, J.; Chen, K.; and Savarese, S. 2016. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *European conference on computer vision*, 628–644. Springer.

Costa, P.; Galdran, A.; Meyer, M. I.; Abràmoff, M. D.; Niemeijer, M.; Mendonça, A. M.; and Campilho, A. 2017. Towards adversarial retinal image synthesis. *arXiv preprint arXiv:1701.08974* .

Cui, Z.; Li, C.; and Wang, W. 2019. ToothNet: automatic tooth instance segmentation and identification from cone beam CT images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6368–6377.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.

Henzler, P.; Rasche, V.; Ropinski, T.; and Ritschel, T. 2018. Single-image Tomography: 3D Volumes from 2D Cranial X-Rays. In *Computer Graphics Forum*, volume 37, 377–388. Wiley Online Library.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Imangaliyev, S.; van der Veen, M. H.; Volgenant, C. M.; Keijser, B. J.; Crielaard, W.; and Levin, E. 2016. Deep learning for classification of dental plaque images. In *International Workshop on Machine Learning, Optimization, and Big Data*, 407–410. Springer.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Kolev, K.; Brox, T.; and Cremers, D. 2012. Fast joint estimation of silhouettes and dense 3d geometry from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(3): 493–505.

Liang, Y.; Fan, H. W.; Fang, Z.; Miao, L.; Li, W.; Zhang, X.; Sun, W.; Wang, K.; He, L.; and Chen, X. 2020a. OralCam: Enabling Self-Examination and Awareness of Oral Health Using a Smartphone Camera. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

Liang, Y.; Song, W.; Yang, J.; Qiu, L.; Wang, K.; and He, L. 2020b. X2Teeth: 3D Teeth Reconstruction from a Single Panoramic Radiograph. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 400–409. Springer.

Mao, X.; Li, Q.; Xie, H.; Lau, R. Y.; Wang, Z.; and Paul Smolley, S. 2017. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2794–2802.

Mirza, M.; and Osindero, S. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* .

Momin, M. A.; Okochi, K.; Watanabe, H.; Imaizumi, A.; Omura, K.; Amagasa, T.; Okada, N.; Ohbayashi, N.; and Kurabayashi, T. 2009. 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): 75–81.

Petersen, L. B.; Olsen, K. R.; Christensen, J.; ; and Wenzel, A. 2014. Image and surgery-related costs comparing cone beam CT and panoramic imaging before removal of impacted mandibular third molars. *Dentomaxillofacial Radiology* 43(6): 20140001.

Prajapati, S. A.; Nagaraj, R.; and Mitra, S. 2017. Classification of dental diseases using CNN and transfer learning. In *2017 5th International Symposium on Computational and Business Intelligence (ISCBI)*, 70–74. IEEE.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13(4): 600–612.

Wu, J.; Zhang, C.; Zhang, X.; Zhang, Z.; Freeman, W. T.; and Tenenbaum, J. B. 2018. Learning shape priors for single-view 3d completion and reconstruction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 646–662.

Yang, G.; Cui, Y.; Belongie, S.; and Hariharan, B. 2018. Learning single-view 3d reconstruction with limited pose supervision. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 86–101.

Ying, X.; Guo, H.; Ma, K.; Wu, J.; Weng, Z.; and Zheng, Y. 2019. X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 10619–10628.

Yun, Z.; Yang, S.; Huang, E.; Zhao, L.; Yang, W.; and Feng, Q. 2019. Automatic reconstruction method for high-contrast panoramic image from dental cone-beam CT data. *Computer methods and programs in biomedicine* 175: 205–214.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.