Reconstructing Teeth from a CT Scan

Renjie Wu , Xinyi Yu

Abstract—3D CT scans consist of a sequence of 2D scanned slices, each of which representes the tissue density by gray values. Since the density differences between teeth and mandible is relatively small, the task of performing 3D reconstruction from CT scans is more challenging than the usual segmentation work implemented for tissues with large differences in density. Thus, our work may provide more meaningful and effective assistance for clinical treatment of teeth. In order to reconstruct teeth from a CT scan, we proposed two methods. The one we fully discussed in the paper is 2D segmentation and 3D reconstruction. First, we annotate teeth labels as traning and tetsting dataset using Photoshop and other annotation tools artificially and then train a neural network to predict teeth segmentation in CT images. After segmentation, we extract edges in the segmented images, extract discrete points from the extracted edges and add the points into point cloud. Finally, reconstruct teeth with the point cloud files.
The other one method is similar to the first one. Also, use neural network to segment the teeth. Then convert the images back into Dicom files and use professional dental reconstruction tools to do the reconstruction, which is the difference lies in.

Index Terms—Computer Graphics, Teeth Reconstrution, CT Scan, Segmentation, Edge Extraction, Point Cloud

1 Introduction

NOWADAYS, quickly developing dentistry requires more useful and more accurate teeth models to help with the diagnosis. This results in CT scanning technology and CBCT scanning technology in the dental field. Since CT scan cannot focus on the teeth and thus is not accurate and convenient for teeth diagnosis. However, simplex oral scan just focuses on the teeth crown but ignores the teeth root and other nearby tissues, which is not comprehensive for doctor to get an overall view of the teeth, resulting in bad therapy. In order to overcome the shorts of the above two diagnosis methods, we try to segment teeth in the CT scan, including teeth crown and teeth root, and reconstruct a complete teeth model. Since CBCT enables lateral scanning of teeth while CT is good at vertical scanning, whether it is possible to combine CT and CBCT to achieve more effective functions becomes a hot topic.

A 3D CT scan consists of a sequence of 2D scanned slices, each of which records the tissue density of a layer of the human body and is represented by gray values. Segmentation tasks are easy to implement for tissues with large differences in density, such as bones and soft tissues. The difference in density between tooth and nearby tissues is relatively small, making the task of teeth segmention and reconstruction from CT scans is more challenging than the usual segmentation work. So our work on it may be more meaningful and helpful for clinical treatment of teeth. There is no such medical software for reconstructing the whole teeth model from a common CT scan, separating from bones. Hence, the single CT scan cannot play an completely important role in the clinical treatment of teeth.

Fig.1 shows the last results in different stages. Our code can be downloaded at https://github.com/Renjie-Woo/SEG/

SEG-1

December 25, 2018

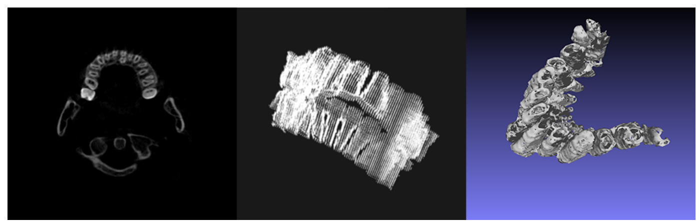

Fig. 1: Images from CT scans to 3D models

2 Methods

To solve the problem, we proposed two methods. They both have the segmentation part, the difference is the way of reconstruction. Using deep learning, the neural network is trained for segmentation in both two methods. In reconstruction part, the first method uses PCL and OpenCV Library to generate point cloud and reconstructs 3D model. However, the second method converts segmented images back into dicom files and use professional tools for reconstruction. We will fully discuss the first method in the paper on which our implementation bases.

2.1 Method 1

In order to do the teeth reconstruction, we proposed the method of 2D segmentation and 3D reconstruction.

With the purpose of accuracy, rapidity and convenience to split teeth from 2D slides, it is unrealistic to use human resources alone. So SegNet based on FCN which is used for segmentation of pictures provides us a good idea and a practical method to split teeth from CT scans. We design a network with reference to SegNet to split teeth parts. Then, in the reconstruction, we use 'canny' operator to extract edges from the segmented images of teeth in Matlab. Then extracte discrete points from the extracted edges and add them to point clouds layer by layer using PCL and OpenCV Library, generating a '.ply' file. Finally, we convert the point cloud to 3D teeth model.

2.1.1 Data Prepare

First step is data preparation. The source dataset we get is CT images in the form of ’.dcm’. Considering the bad operability of dicom files, we convert CT scans in the form of ’.dcm’ into ’.png’ files, which is more convenient to do the following work. After this, we use Photoshop and js-segment-annotator, a web tool for annotation, to annotate the teeth. Since artificial annotation is too time-consuming and is obviously unrealistic, a more efficient way for annotation is in urgent need, which introduces the neural network.

Fig.2 shows the process of data preparation.

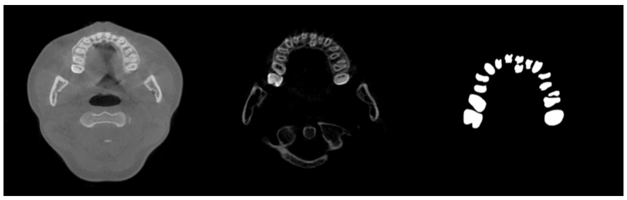

Fig. 2: Process of data preparation

2.1.2 Teeth Segmentation

We refer to SegNet to construct our teeth segmentation network. SegNet is based on FCN and they have very similar ideas, but the techniques used by Encoder and Decoder(Upsampling) are inconsistent. In addition, SegNet’s encoder uses the first 13 convolutional networks of VGG16. Each encoder layer corresponds to a decoder layer, and the output of the decoder is sent to soft-max classifier to generate class probability for each pixel independently.

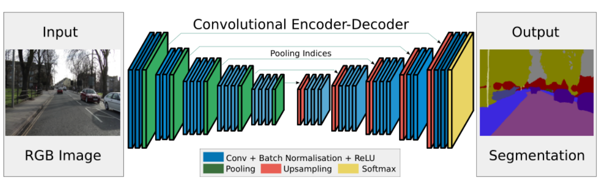

Fig. 3: SegNet architecture

Fig.3 shows the network architecture of SegNet. The left part of the Fig.3 is convolution feature extraction, whihc increases the receptive field by pooling and make the image smaller at the same time. This process is called the Encoder. On the right is the deconvolution and upsampling, after deconvolution makes image classification features are reproduced, upsampling restore to the original image size, the process is known as the Decoder. Finally, output the maximum of different classification through the Softmax and get the final segmentation image.

’SegNetTeeth’ is the natural network we designed for segmentation of teeth from CT scans, referring to ’SegNet’ and ’SegNetCMR’, which is used for ventricular segmentation. Both of them make use of the concept of convolution, used for low-dimensional feature extraction of high-dimensional data, and deconvolution, contrary to convolution, mentioned by FCN. Fig.4 and Fig.5 shows the schematic of convolution and deconvolution.

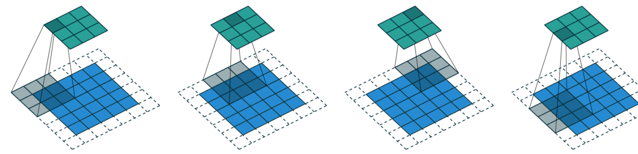

Fig. 4: Convolution layers to extract features

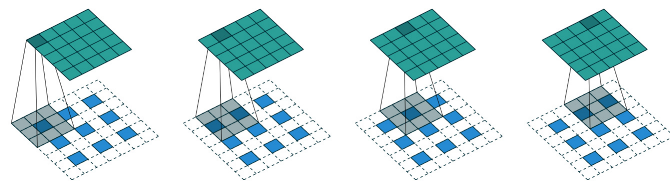

Fig. 5: Deconvolution layers to reduct images from extracted features

The graph of our network consists of five convolution layers which extract as many as 512 features from a input image in the format of ’768 $\times$ 768’ and the other five which use upsampling to reduct features to images in the same size.

Then we compared the generated images with labels we annotated manually, calculate and set the differences between them as the loss function. Fig.6 shows our whole graph of the neural network. Since we only have 300 pairs of images, we could only train them with lower expected accuracy. We set 1 as the batch size and the epoch is five, thus we totally train and test with 1500 steps.

Here, we use TensorBoard to achieve the visualization of training process. Fig.7 shows the accuracy during the training. Fig.8 shows the pictures which mix the input images and the input labels and the pictures is overlapped with high precision since the labels are annotated manually. Fig.9 shows pictures which mix the input images and labels generated by neural network with lower precision.

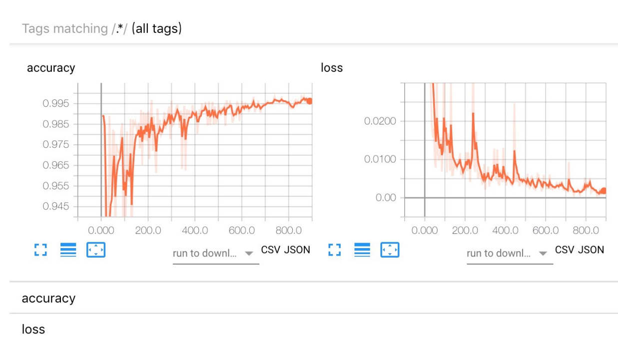

Fig. 7: Training accuracy

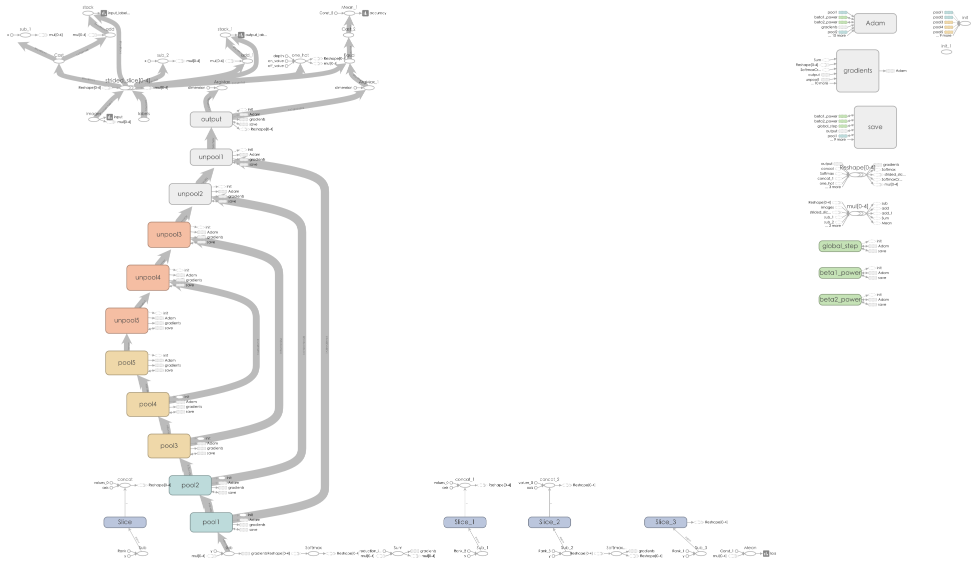

Fig. 6: Architecture graph of SegNetTeeth

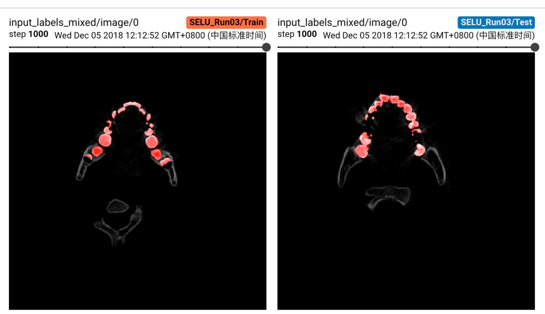

Fig. 8: Input images with input labels

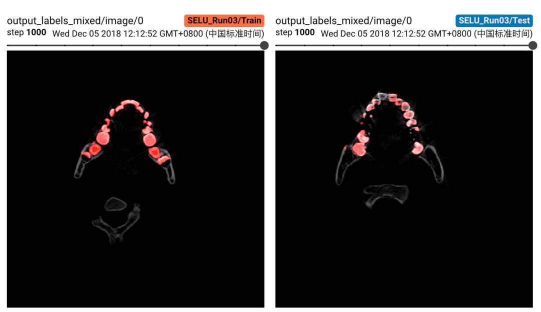

Fig. 9: Input images with generated labels

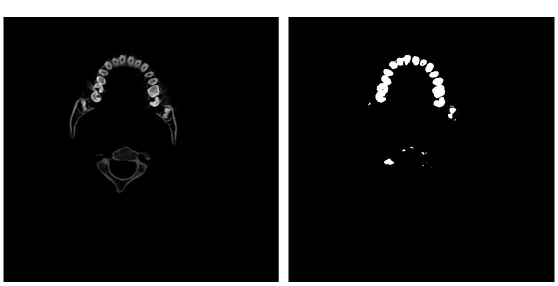

Fig. 10: Predicted label

After training we use the neural network for prediction and Fig.10 shows the comparision between images and generated labels. In general, most labels match the images well. However, they are not precise enough or even predict badly in some images because of the lack of the samples, since 300 pairs of samples are far from enough.

2.1.3 Edge Extraction

As we all know, teeth is a relative closed model. To reconstruct teeth model, we just focus on the outside surface of the teeth and don't need to take the inside condition of the teeth into consideration, such as the dental pulp, the inside part should be removed such that the important part can stand out. As for a segmented CT scan, what we focus on is the outline of the labels, indicating the necessity of edge extraction.

We use canny operator to extract edges from the segmented images of teeth in Matlab. The operator uses gaussian to smooth the image and then two thresholds to extract the edge of the image. The detailed algorithm is:

1) use gaussian filter to smooth the image and remove noises.

2) calculate the gradient intensity and direction of each pixel in the image.

3) apply Non-Maximum Suppression to eliminate the spurious response caused by edge detection.

4) apply Double Threshold detection to determine the real and potential edges.

5) edge detection is finally completed by inhibiting isolated weak edges.

We use the built-in function in Matlab to do the edge extraction with canny operator. Fig.11 shows the results of edge extraction. We can see that the almost all edges have been extracted.

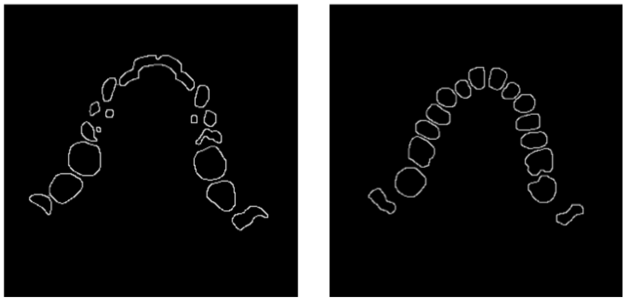

Fig. 11: Edge extraction

2.1.4 3D Reconstruction

In reconstruction part, we use PCL and OpenCV library to extract discrete points in the images after edge extraction and then generate point cloud. Then convert the point cloud into 3D mesh model.

The concrete method of generating point cloud is as follows: first, opencv is used to read the image after edge extraction and convert it into matrix format. Then traverse the matrix to read the element values, create a point for each non-zero value read, set its 3d coordinates according to the CT format, and finally add it to the point cloud. Repeat the above procedure for each image. Finally, we will obtain a ply model of teeth point cloud. Fig.12 shows the reconstructed teeth point cloud model.

After obtaining the point cloud, we use MeshLab to reconstruct the 3D mesh model. Fig.13 shows the 3D teeth model reconstructed by us. As you can see, this model is not particularly accurate, and we will explain this later. Our project goal is to segment and reconstruct the teeth and roots from other surrounding tissues in CT images. It can be clearly seen that the top tooth has three separate roots, which can also verify that our project goal has been basically achieved.

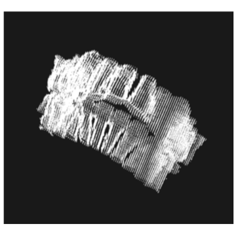

Fig. 12: Point cloud

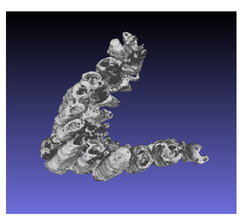

Fig. 13: 3D mesh model

2.2 Method 2

While implementing this method, we proposed another idea. Simalarly to the one above, we convert the dicom files to '.png' images and train a neural network to segment the teeth. After that, instead of extracting edges, generating point cloud and reconstructing 3D model, we convert the segmented images back into dicom files and then use professional dental reconstruction tools to do the reconstruction. Since the dental reconstruction technology is already mature, this method will lead to a more accurate 3D teeth model.

3 Discussion and Future Work

Fig.14 shows the results of the whole process, including CT scans, converted images, labels, extracted edges, point cloud and 3D model. It can be seen that we basicly achieve the goal of reconstructing teeth from a CT scan. However, because of the lack of data accuracy and quantity, the 3D model we finally get still has much room for improvement.

We apply the multilayer reconstruction to reconstruct the teeth but this method is not perfect. Since we deal with the CT images layer by layer, we just focus on the 2D plane instead of considering the whole teeth model reconstruction as a 3D problem. Due to the ignorance of z-axis, we split z axis with x and y axis and thus separate the layers, which may lead to a model unsmooth in the z axis.

In the future, we hope to find a better method which takes the z axis into consideration, not only focusing on the CT layers. It is believed that a more accurate teeth model will be reconstructed if all the three axis(x,y,z) are combined together.

4 Conclusion

In order to achieve the goal of reconstructing teeth from a CT scan, we propose the method of training neural network for teeth segmentation and then 3D reconstruction.

In terms of segmentation, we refer to SegNet to train the dataset of teeth. In the previous presentation, it can be seen that our network can basically separate the tooth from the surrounding tissues completely. However, due to the lack of professional dental knowledge, errors are inevitable in the early annotation of training data, making the segmentation results inaccurate. In addition, due to the lack of manpower and time, the number of labels is not large. The lack of data accuracy and quantity is the main reason for the inaccuracy of our reconstruction results. However, even if we only use a small amount of data for training, we can already produce relatively good results, which indicates that the reconstruction method of 3D teeth model from CT images is feasible. If the data accuracy and quantity can be increased in the later stage, there will be a lot of room for improvement.

In terms of reconstruction, we use OpenCV and PCL libraries to extract discrete points from the edge extracted images, convert them into point clouds, and then use MeshLab for 3D reconstruction.

Moreover, we propose another method of implementation. Similar to the method we have shown, first convert the Dicom file to PNG format and then segment the teeth, which can still be implemented using SegNet. The difference is that the segmented PNG file is then converted into a dicom file, and then the professional 3D reconstruction software is used to generate a 3D model of the dicom file. Because currently the technology of generating 3D models from CT images has been relatively mature, the reconstruction results should be more accurate than the "segmentation + reconstruction" method we apply.

Acknowledgments

Thank our teacher Bin Sheng and the teaching assistant Xiaoshuang Li for their patient guidance and help.

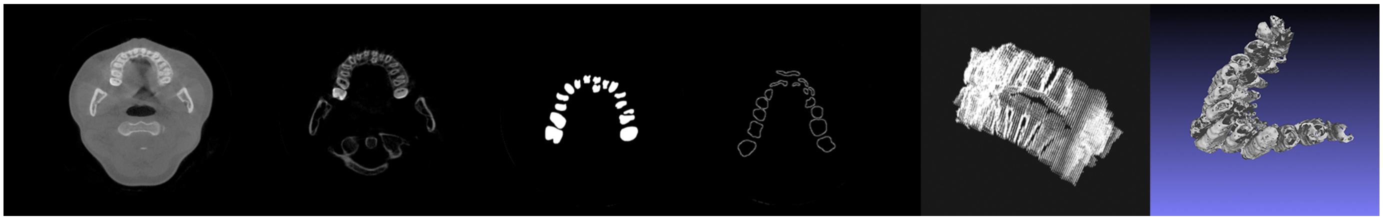

Fig. 14: Result of the whole process

REFERENCES

[1] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, Senior Member, IEEE *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*,
[2] Hui Gao, Oksam Chae *Touching Tooth Segmentation from CT Image Sequences Using Coupled Level Set Method*,
[3] JIANG Ting, LUO Sheng, ZHANG Rui *Automatic Reconstruction of Multi-layer Building 3D Contour Model From Airborne LiDAR Point Clouds*,

Renjie Wu (1997 -)
2016 SJTU Undergraduate
Computer Science and Technology
Belong to F1603305

Xinyi Yu has been studying in Shanghai Jiao Tong University since 2016. She is an undergraduate of School of Electronic Information and Electrical Engineering, majoring in Computer Science.