UNIVERSITY OF COPENHAGEN

Segmentation of 3D jaw images with MultiPlanar UNet

Peidi Xu

Department of Computer Science

Contact Information:
Universitetsparken 2
2100 København Ø

Phone: +45 91865390
Email: peidi@di.ku.dk

Abstract

Segmenting 3D medical data is a challenging task due to small amount of available training data and large computational overload. Applying 3D convolutions directly to large 3D images will easily make memory overflow, so that these models are usually trained on small patches. However, this will result in limited field of view and loss of global information. Therefore, we use 2D model proposed by Perslev et al.[2] to segment the 3D jaw data while preserving as much 3D spatial information as possible by generating views from different perspectives. The result on such 2D UNet shows better performance in our dataset with few available training data. The segmented probability map is then passed as input to marching cube algorithms for 3D rending, which could further serve as meshing tools.

Introduction

An encoder-decoder structure of Convolutional Neural Networks has been widely used on semantic segmentation tasks, among which the most successful one is the UNet structure that was originally proposed for cell images. The architecture uses skip connection to include the high-resolution feature maps in encoding path to include more fine-grained information. The variation of such model, e.g. 3D UNet is a straight forward way to segment a 3D data like CT scans and has shown its state-of-the-art performance on 3D medical image segmentation although increasingly more complicated models have been proposed on semantic segmentation on general case, e.g. DeepLabV3+. However, due to the memory overload, 3D models are usually trained from patches at the cost of global information. Our experiments shows the Multiplanar UNet model proposed by Perslev et al. [2] allows the 2D UNet model to learn representative semantic information of the 3D image volume with only few annotated examples.

Main Objectives

1. Segment 3D jaw data from CT scan of the whole head
2. Verify the effectiveness of Multi-Planar UNet on 3D image data over 3D UNet
3. Generate meshes from the segmentation results for further finite element analysis

Materials and Methods

The way we usually segment 3D data is by generating multiple parallel slices from the 3D volume which are then passed to a 2D model to do the segmentation of each slice. The final volume is then reconstructed by 3D iterative reconstruction.

This way of reconstruction may lose volumetric spatial information because all the slices are viewed from the same direction. Therefore, the intuition behind MultiPlanar Unet is to generate several views from different perspectives, which would result in multiple volumes from several views. The results on each view are then combined in weighted average way by a fusion model, which is simply a network with one linear layer that learns the weight of each view.

Mathematical Section

The core part of MultiPlanar UNet is the encoder-decoder architecture, where the output of each view is in the same dimension of the input $R^{|V|*c}$ with c classes, a pixel-wise loss function is then applied for back-propagation, where we use a combination of cross-entropy loss and dice loss here.

$$L_{CE}(y,p) = -\sum_{i=1}^{c} y_i \log p_i \tag{1}$$

$$L_{DICE}(y,p) = 1 - 2\frac{\sum_i^c y_i p_i}{\sum_i^c y_i + \sum_i^c p_i} \tag{2}$$

where the final loss is a weighted combination of the two losses over all voxel locations v in the output segmentation volume.

$$L = \sum_{v} L_{CE}(y,p) + 0.3 * \sum_{v} L_{DICE}(y,p) \tag{3}$$

The result by each view from MultiPlanar UNet is combined by a fusion model which learns the weight when summing up.

$$y = \sum_{i=1}^{c} (W \odot x)_i + b \tag{4}$$

where $x \in R^{n*c}$ and $W \in R^{n*c}$, the bias term $b \in R^c$ and n is the number of views. $\odot$ denotes element-wise multiplication. The model is trained over all voxel locations $|V|$.

Results

Figure 1 and Table 1 show both a visualization and numerical results of both MultiPlanar UNet and 3D UNet, where we could see that MultiPlanar UNet gives generally better results with much smoother results, but more missing roots on some of the teeth. With less training images, we believe this is due to less overfitting with much fewer parameters than 3D UNet.

Table 1: Performance by MultiPlanarUNet and 3D UNet

	MPUNet w.o. fusion	MPUNet w. fusion	3DUNet
Dice score	86.97	87.45	80.24

We further ran 3D rendering over the segmentation output with marching cubes algorithm, which extract 2D surface mesh from a 3D discrete scalar field. Meshes are generated by Triangulations in regions with a probability above 0.5 [1]. The results look really promising, and we expect that these can be transferred to finite element model for further simulation needs.

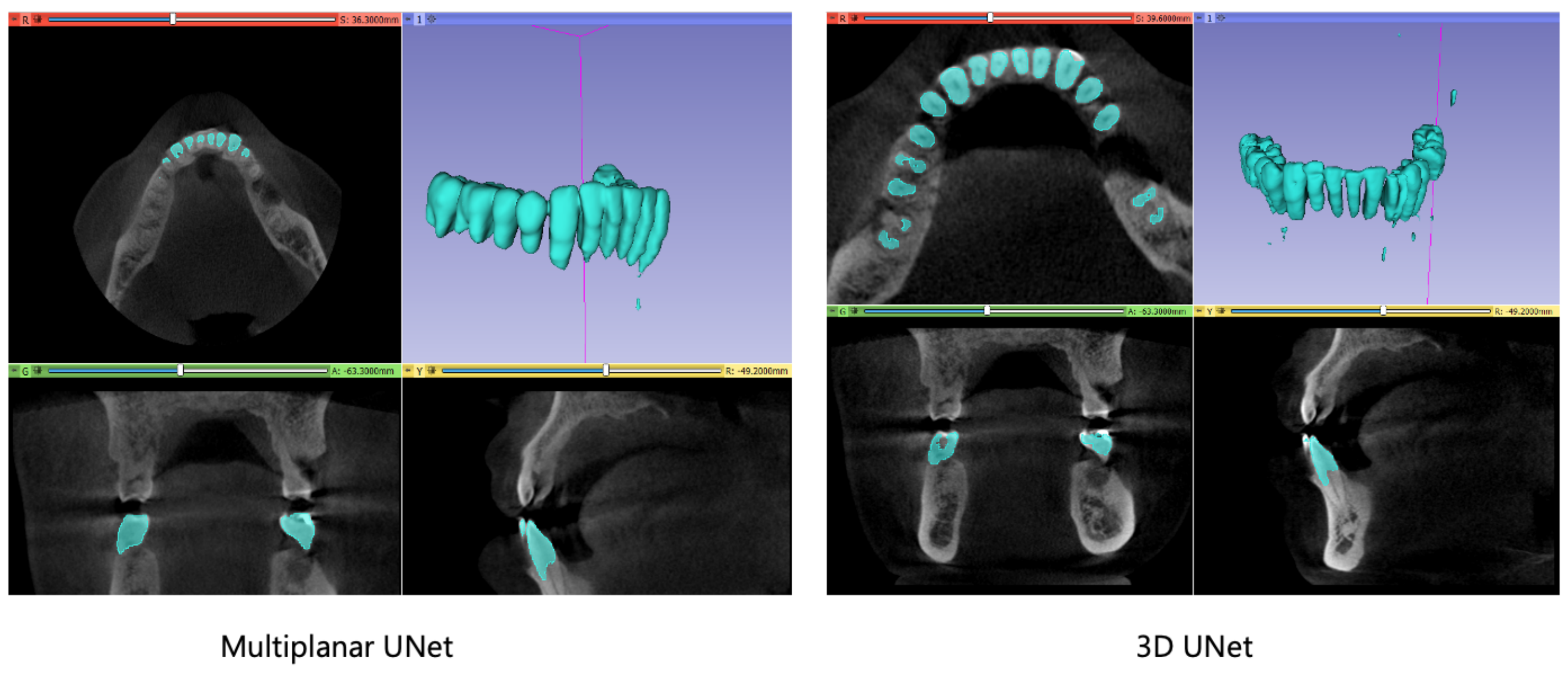

Figure 1: Segmentation results by MultiPlanar UNet and 3D UNet

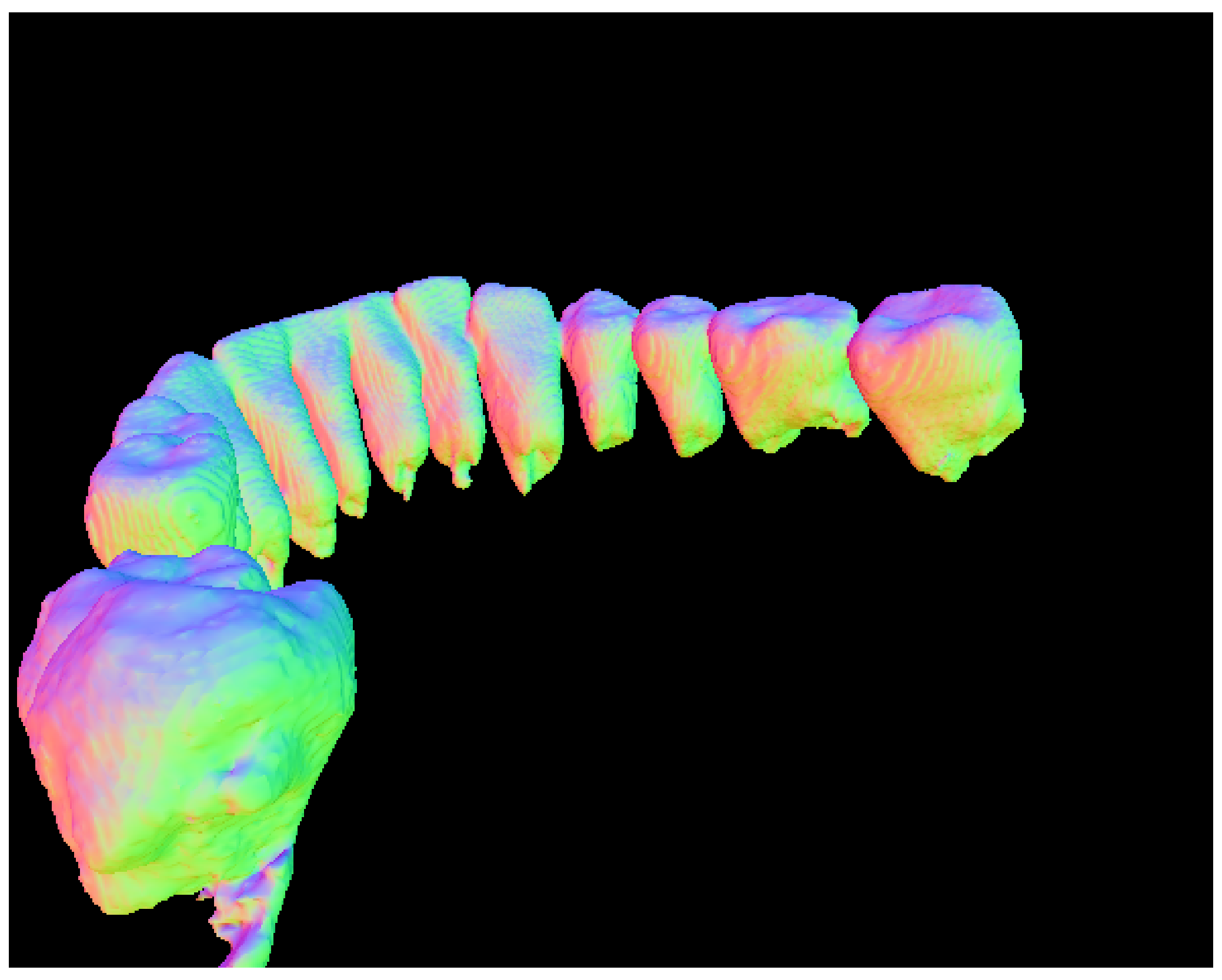

Figure 2: 3D rendering over segmentation map by marching cube algorithm

Conclusions

- MultiPlanar UNet is a novel way to segment 3D medical images, which gives promising results on our jaw data, except that there are some roots missing roots in some teeth.
- The result by 3D UNet is more noisy but also more inclusive as it can better predict the roots.
- MultiPlanar UNet gives generally better results than 3D UNet, which we believe is due to less overfitting with much fewer parameters when we don't have enough data
- Marching cubes provides great 3D rendering for semantic segmentation results, with the potential for future mesh tool analysis.

Forthcoming Research

Even though MultiPlanr UNet generates very promising segmentations over our jaw data, there are definitely still rooms for improvement. MultiPlanar UNet is quite flexible in that that it can be fit into any 2D segmentation model, so any architectural modifications such as squeeze and attention on 2D convolutions could be applied for improvement [3]. It is also possible to totally change the UNet to a more recent model to better predict roots.

References

[1] Marching Cubes. A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. New York: Association for Computing Machinery*, pages 163–69, 1987.

[2] Mathias Perslev, Erik Bjørnager Dam, Akshay Pai, and Christian Igel. One network to segment them all: A general, lightweight system for accurate 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 30–38. Springer, 2019.

[3] Zilong Zhong, Zhong Qiu Lin, Rene Bidart, Xiaodan Hu, Ibrahim Ben Daya, Zhifeng Li, Wei-Shi Zheng, Jonathan Li, and Alexander Wong. Squeeze-and-attention networks for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13065–13074, 2020.