Automated segmentation of CT temporal bone structures

George Liu[a]

Department of [a]Otolaryngology–Head and Neck Surgery and [b]Computer Science, Stanford University

Background

CardinalSim (https://cardinalsim.stanford.edu/) is a surgical simulator that uses 3D glasses, computerized rendering of temporal bone anatomy from loaded CT scans of the head, and a haptic feedback device to help train otolaryngology residents in surgery before they do it for real patients (Figures 1 and 2). One challenge is that the 3D rendering requires manual segmentation of relevant temporal bone anatomy, which may take up to hours, for each individualized CT scan surgical rehearsal.

Our approach to automated segmentation relies on fully convolutional neural networks (FCNs), a class of deep learning models specialized for computer vision and in particular semantic segmentation (i.e. class labeling) [1]. We are extending previously described FCN models, such as 3D U-Net [2] and V-Net [3], to our task of volumetric CT image segmentation of temporal bone structures.

Here I describe initial work to set up and begin training a 3D U-Net model to automatically segment 5 temporal bone structures: the carotid artery, facial nerve, sigmoid sinus, ossicles, and cochlea (Figure 3)..

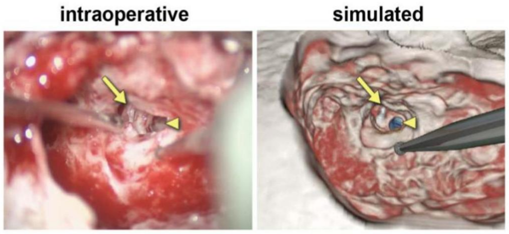

Figure 1. Intraoperative view (left) and surgical simulation (right) in CardinalSim of a tympanomastoid surgery (drilling behind the ear to get better access to the ear drum and middle ear).

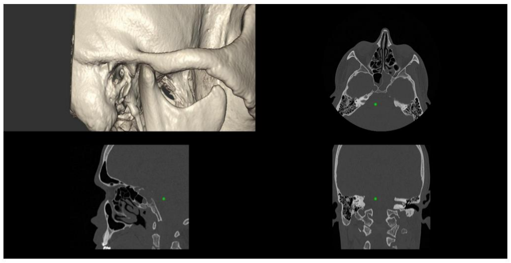

Figure 2. Realistic 3D rendering of lateral skull base anatomy in CardinalSim (top-left corner). Intensity-based thresholding of original CT scans in axial (top-right), sagittal (bottom-left), and coronal (bottom-right) views are used to produce 3D structure of bone.

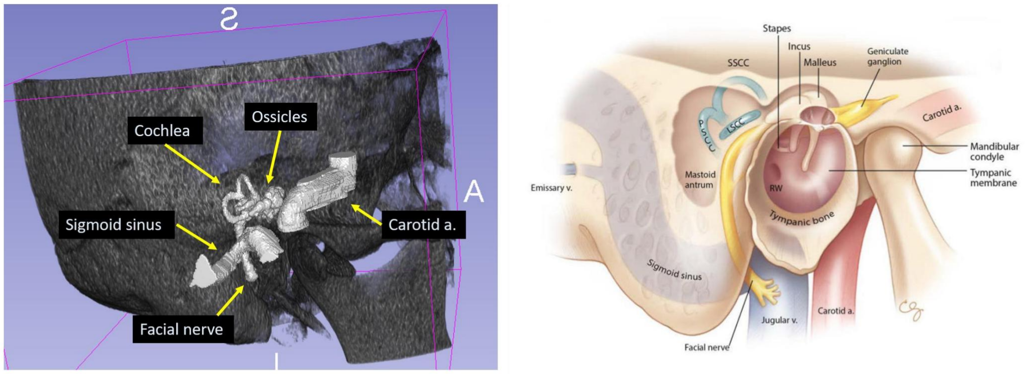

Figure 3. Example of labeled training data (left). The 3D rendering of an original CT scan is visualized in 3D Slicer (left) to show the anatomy of the right ear in one patient. Manual segmentations of five structures of interest (cochlea, ossicles, cochlea, sigmoid, sinus, and facial nerve) are and overlaid as white 3D-rendered structures within the original CT scan, and labeled (yellow arrows). Schematic of the anatomy of the right ear is shown for comparison (right).

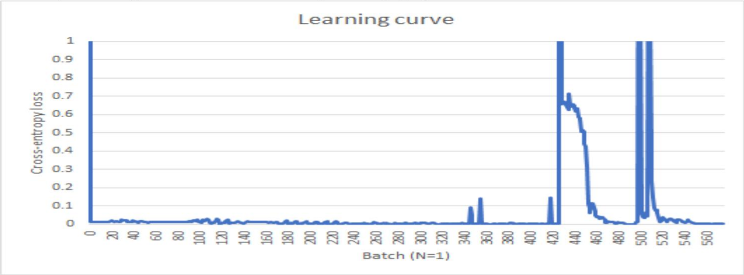

Figure 4. Learning curve after training 3D-Unet using stochastic gradient descent for all 576 tiles with carotid artery segmentation labels only in one CT scan image.

Conclusion

Based on assessment of my training data, I believe I have an imbalanced dataset where the negative samples (background pixels) overwhelmingly outnumber the number of foreground pixels (internal carotid artery, ossicles, cochlea, facial nerve, sigmoid sinus). I can address the issue by evenly sampling positive and negative samples, by creating each mini-batch of 4 tiles such that 2 tiles are foreground and 2 background.

In addition, I believe that structures in CT scans are stereotyped in terms of their absolute coordinates in a CT scan. This is because each patient's head is fixed in a CT scanner in a similar way. One way to take advantage of this global position data for segmentation is to input the whole image into the 3D-Unet (rather than just tiles). However, this will require more GPU memory. An alternative is to input the position coordinates of the image into a separate fully-connected neural network, and then skip-forward-feed that network to the final layer of the original 3D-Unet architecture for calculation of each pixel's class probabilities.

References

1. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation ppt. CVPR 2015 Proc IEEE Conf Comput Vis Pattern Recognit. 2015;39:3431–40.

2. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2016;9901 LNCS:424–32.

3. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proc - 2016 4th Int Conf 3D Vision, 3DV 2016. 2016;:565–71.

Acknowledgments

We thank the Blevins lab and Farzan Farnia for help with guiding this project.