# Automated Segmentation of Temporal Bone CT Imaging for Robot-Assisted Microsurgery

Paper Seminar Report

EN 601.656 Computer Integrated Surgery II

Andy Ding

ading5@jhmi.edu

## Project Summary

Surgery within the temporal bone involves maneuvering around small and geometrically complex anatomy (Figure 1), which poses a high risk of accidental injury. This has motivated several groups to build cooperative control robotic systems for temporal bone surgery in an effort to reduce hand tremor, increase economy of motion, and ultimately augment surgeon skill in this space. This project aims to provide an automated system for creating high-quality segmentations of patient temporal bone CTs, which can be used to inform semiautonomous surgical systems about critical patient anatomy that should be avoided.

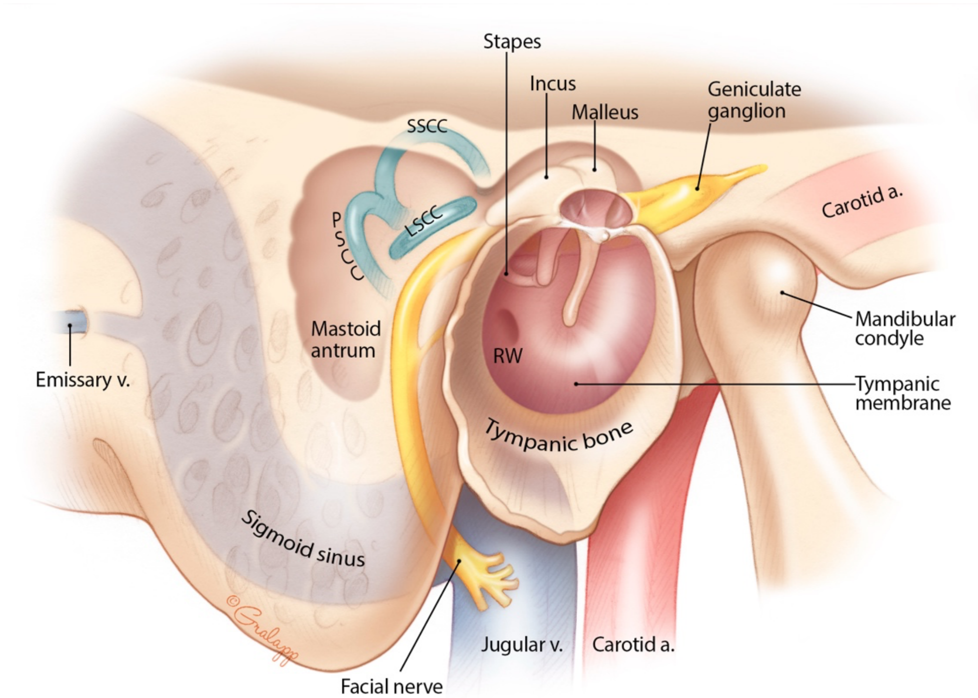



**Figure 1.** Relevant anatomy in and around the temporal bone[1]

## Paper

"Toward an automatic preoperative pipeline for image-guided temporal bone surgery" was chosen for its relevance to the project.[2] Although the motivation behind this paper is for pre-operative path planning in otologic procedures, the overall methods described are highly relevant and similar to this project's proposed strategies. This paper also provides strategies to improve segmentations that have not been previously reported for this project. Furthermore, this paper does a decent job of demonstrating potential pitfalls of using only UNet segmentation in such a complex and small anatomical area in CT

### *Summary and Key Results*

The summary of this paper is the development of an automatic temporal bone segmentation pipeline with UNet and probabilistic active shape modeling (PASM) for pre-operating path planning (e.g., cochlear access, superior semicircular canal approach, retrolabyrinthine access). Key results with this paper, which will be discussed in detail below, include segmentation accuracy evaluation of the author's UNet + PASM hybrid model against UNet or PASM alone, as well as pre-operative planning success with UNet + PASM hybrid compared to. UNet or PASM alone.

*Background*
Minimally invasive surgery (MIS) in the temporal bone procedures is increasing in popularity. Surgical procedures including cochlear implantation, vestibular schwannoma excision, and cholesteatoma removal all have evolved to include minimally invasive options for certain cases.[3] MIS, however, requires pre-operative planning, ideally with segmented CTs, to determine operating trajectories that minimize risk of damage to surrounding structures. To this end, the authors of this paper aim to create an automated system for segmenting temporal bone CTs for pre-operative planning.

***Segmentation Workflow***
24 CT scans of the temporal bone (resolution 0.2x0.2x0.4 $mm^3$) were manually segmented. The authors then implemented an ensemble 2D-UNet model to semantically segment CT scans slice-by-slice. Majority voting of the predicted segmentations then finalized the output of the UNet model. These predictions were used as initializations for a probabilistic active shape model (PASM) to provide anatomically accurate segmentations (Figure 2).

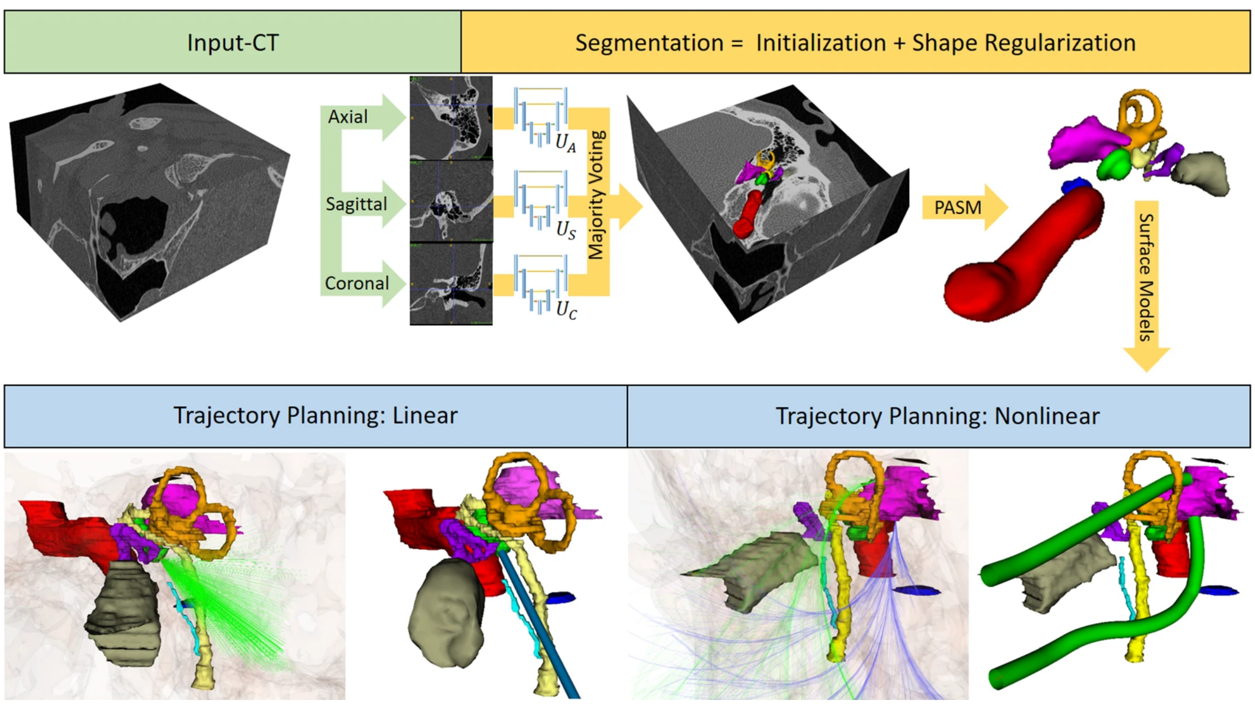


**Figure 2.** Segmentation workflow as described in the paper. Not shown is the separate UNet model for chorda tympani labeling.

Importantly, the authors noted that the chorda tympani was disproportionately smaller in volume than other structures—even the ossicles. Because of this, the authors specifically trained a separate ensemble 2D-UNet model to segment the chorda tympani and combined the results of this model with those of the aforementioned multi-class ensemble 2D-UNet.

***Shape Regularization with Probabilistic Active Shape Models***
Since UNet has not inherent knowledge of anatomical structure, the authors found that meshes formed from predicted labels exhibit small artifacts or small missing pieces. In particular, the internal carotid artery was broken in several pieces, the semicircular canals had holes in their arcs, the stapes was decayed, and the facial nerve had missing portions of its mastoid segment. Unlike UNet, active shape models (ASM) inherently respect the shape by restricting the segmentation to a trained shape atlas (Figure 3). As a variant of ASM, probabilistic active shape models (PASM) allow a more flexible adaptation by leaving the shape space if image features provide enough evidence.[4]

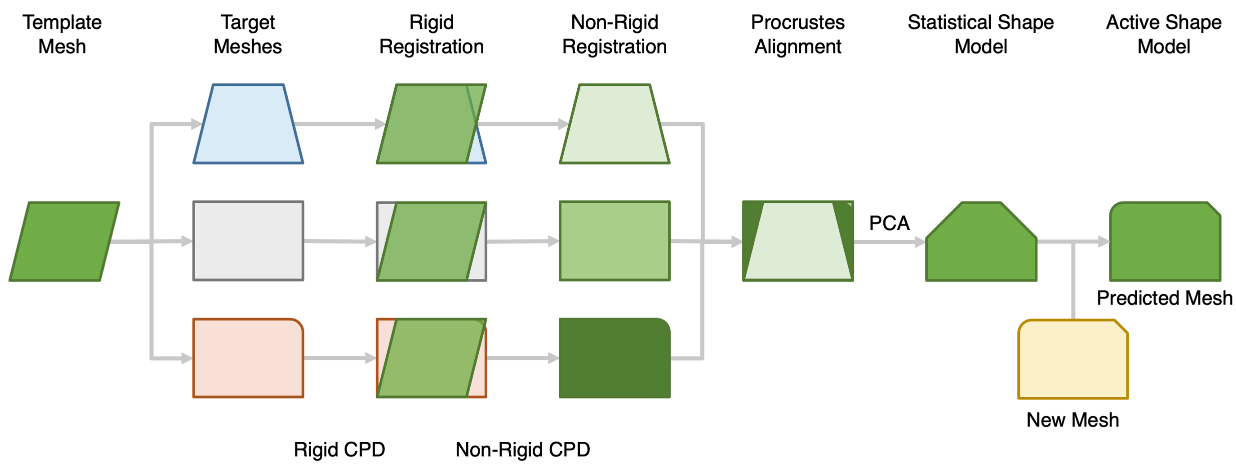



**Figure 3.** Workflow for building a statistical shape model and implementing active shape models.

By using the predictions from UNet as initializations for PASM, the authors were able to generate segmentations that looked more realistic overall (Figure 4).

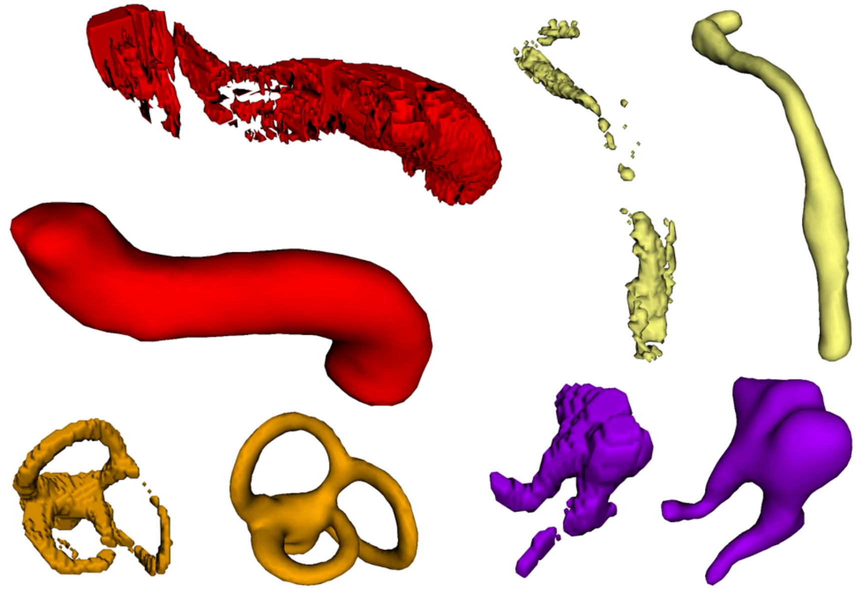


**Figure 4.** Representative fragmented structures from UNet and their regularized counterparts. Top left: Internal carotid artery; Top right: Facial nerve; Bottom left: Semicircular canals; Bottom right: Middle ear ossicles.

***Path Trajectory Planning***

The authors then used in-house path trajectory planning software to plan paths for cochlear access for cochlear implantation, superior semicircular canal approach for internal auditory canal access, and retrolabyrinthine approach for internal auditory canal access (Figure 5). Trajectories calculated using shape-regularized predictions were closer to ground truth paths than those calculated using PASM predictions alone. This ultimately resulted in a trajectory planning success rate identical to ground truth when using shape-regularized predictions. When using PASM for trajectory planning, however, the authors reported a success rate of only 66% (Table 1).

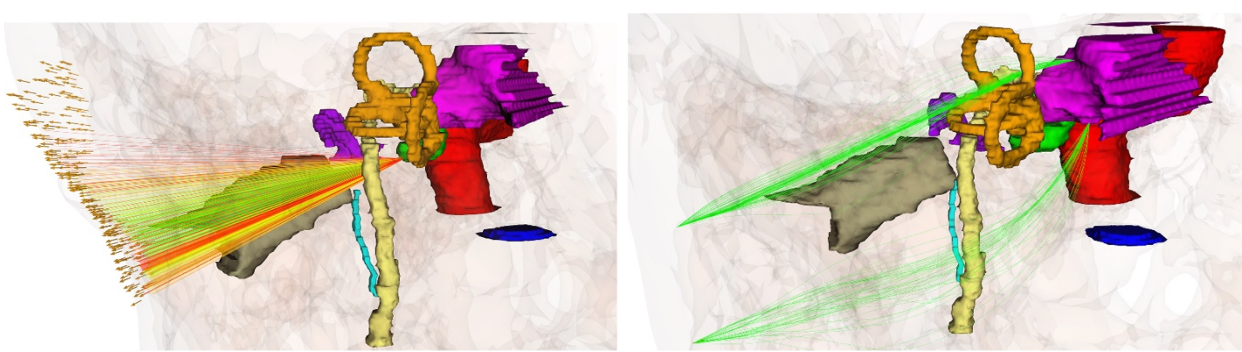


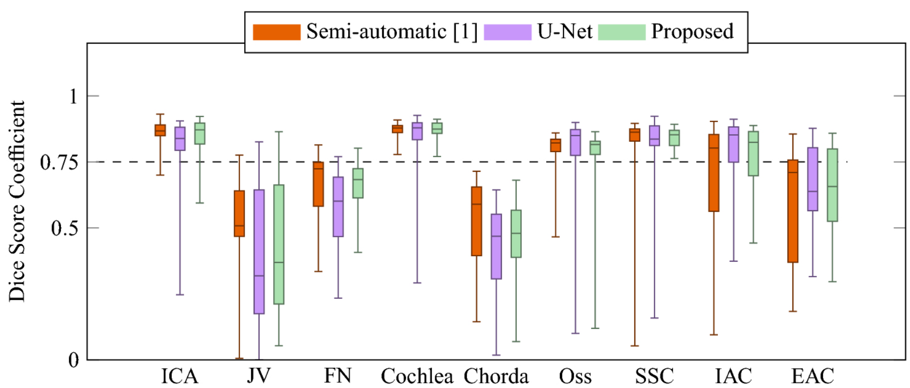
**Figure 5.** Left: Linear paths (colored lines) from start states at the surface of the bone(orange arrows) to the cochlea. Right: Nonlinear trajectories to the internal auditory canal that approximate two given start and target points.

**Table 1.** Trajectory comparisons between ground truth, PASM (labeled as "Semiautomatic"), and shape-regularized predictions (labeled as "Ours").

| | Success rate | | | Mean safety distance | | | Violation rate ($d < d_{min}$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Co** | **SSC** | **RL** | **Co (0.8)** | **SSC (1.5)** | **RL (1.5)** | **Co** | **SSC** | **RL** |
| Ground truth | 0.9 | 1.0 | 1.0 | 1.39 | 2.19 | 2.35 | – | – | – |
| Semiautomatic | 0.65 | 0.66 | 0.66 | 1.15 | 2.17 | 2.60 | 0.0 | 0.0 | 0.0 |
| Ours | 0.9 | 1.0 | 1.0 | 1.04 | 2.16 | 2.42 | 0.18 | 0.0 | 0.1 |

***Segmentation Accuracy Assessment***

When comparing shape-regularized UNet to UNet or PASM alone, the authors reported similar Dice similarity scores for each predicted label (Figure 6). Of note, jugular vein and chorda tympani predictions had particularly low Dice scores. The authors also reported the sensitivity (percent overlap with respect to ground truth) of their tested models for each label. While PASM was noted to have higher sensitivity than UNet or shape-regularized UNet, the authors contribute this to general oversegmentation with this technique.




| | IAC | JV | FN | Cochlea | Chorda | Oss | SSC | IAC | EAC |
|---|---|---|---|---|---|---|---|---|---|
| semi-automatic | 89.9 ± 04.0 | 73.8 ± 18.9 | 69.2 ± 17.0 | 87.1 ± 05.6 | 54.5 ± 17.0 | 83.2 ± 08.2 | 81.9 ± 07.4 | 86.3 ± 08.0 | 76.6 ± 16.0 |
| U-Net | 72.5 ± 21.3 | 30.3 ± 26.5 | 43.4 ± 16.8 | 80.9 ± 17.1 | 36.6 ± 20.4 | 74.8 ± 21.1 | 74.5 ± 16.7 | 72.7 ± 20.5 | 61.5 ± 22.9 |
| proposed | 83.8 ± 11.3 | 37.7 ± 29.2 | 62.7 ± 13.4 | 85.3 ± 06.6 | 39.3 ± 18.0 | 78.8 ± 17.0 | 79.8 ± 07.0 | 68.4 ± 17.2 | 60.9 ± 24.3 |

**Figure 6.** Dice scores for each model. Bottom table displays mean sensitivity (percent overlap with respect to ground truth).

***Conclusions***

The results of this paper suggest that semantic segmentation with UNet alone was not sufficient to create robust segmentations in this dataset. Adding shape regularization with PASM, however, results in anatomically accurate segmentations in the temporal bone. Furthermore, the UNet + PASM hybrid model performs similar to UNet or PASM alone with respect to Dice similarity scores. Despite the similarity in Dice scores, path planning using segments from the UNet + PASM hybrid model results in more realistic and reliable trajectories than PASM alone, making it more practical for clinical use.

## Paper Critiques

***Pros***

In general, this paper provides encouraging results for automatic temporal bone CT segmentation using deep learning methods. The paper clearly stated methods for implementing UNet and PASM together and provided an elegant solution for creating intact segments of temporal bone anatomy. Furthermore, the authors have made their code publicly available for testing and evaluation.

***Cons***

First, the authors do not have a true test set or dataset for external validation. Because of this, the authors have not evaluated the generalizability of their model. Furthermore, the sensitivity of some of the segments, particularly of the jugular vein and chorda tympani, are quite low. This suggests that either the hybrid model is labelling these anatomical structures inaccurately or is undersegmenting these structures. For our purposes, this provides suboptimal segments to implement virtual fixtures in a meaningful way. Finally, perhaps, the most concerning issue with this paper is that their figures for trajectory planning show a mixture of ground truth segments and predicted segments (Figure 7). This suggests that their trajectory planning experiments were not done using all predicted segments when testing the performance of a particular model. Trajectory planning experiments should ensure planning still succeeds when using all predicted segments, which would further demonstrate the accuracy of segmentations overall.

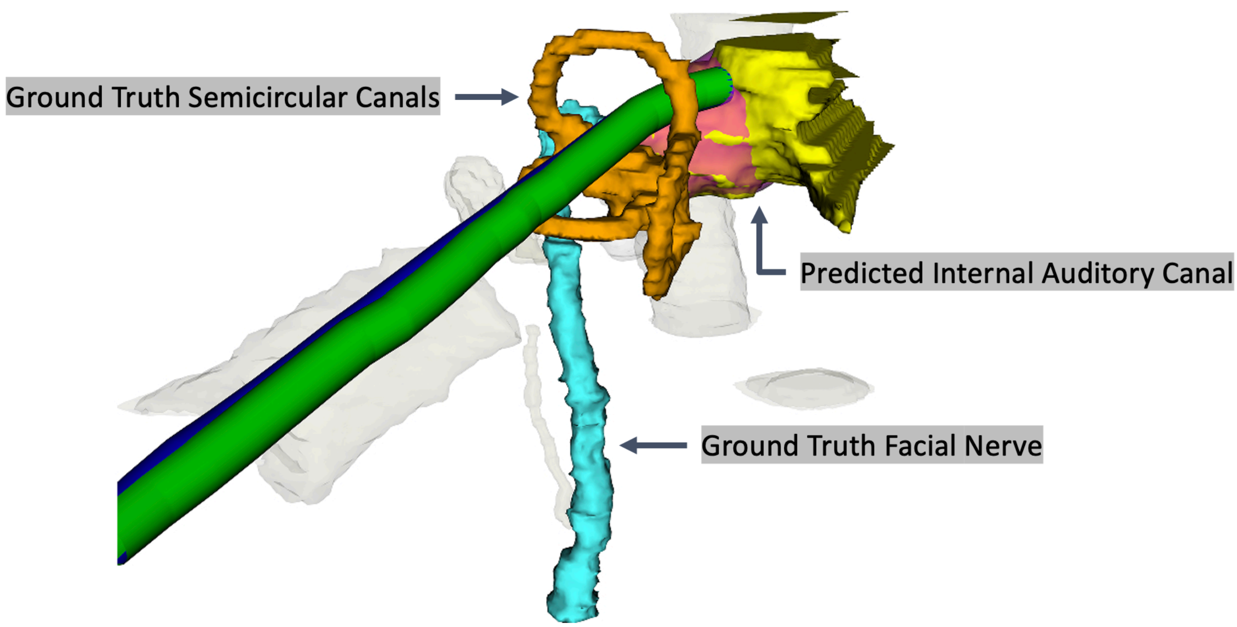



**Figure 7.** Non-linear trajectories through the superior semicircular canal to access the internal auditory canal. Blue path was generated from shape-regularized segments. Green path is ground truth.

***Potential Next Steps***

As discussed in the paper, the authors plan to integrate this model into a GUI application for surgeon usage. Given the critiques listed above, a logical next step would be to validate the model on an external dataset to demonstrate generalizability. Furthermore, Hausdorff distance calculations would help to further investigate segmentation accuracy.

**References**

1. Gralapp C. Intact Canal Wall Mastoidectomy. Stanford Medicine. 2021. 2021. https://otosurgeryatlas.stanford.edu/otologic-surgery-atlas/mastoidectomy/intact-canal-wall-mastoidectomy/
2. Fauser J, Stenin I, Bauer M, et al. Toward an automatic preoperative pipeline for image-guided temporal bone surgery. *International Journal of Computer Assisted Radiology and Surgery*. 2019-06-01 2019;14(6):967-976. doi:10.1007/s11548-019-01937-x
3. Stenin I, Hansen S, Becker M, et al. Minimally Invasive Multiport Surgery of the Lateral Skull Base. *BioMed Research International*. 2014/07/02 2014;2014:379295. doi:10.1155/2014/379295
4. Meike B, Matthias K, Georgios S. Segmentation of risk structures for otologic surgery using the Probabilistic Active Shape Model (PASM). 2014: