Project Proposal

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

EN 601.656 Computer Integrated Surgery II

Andy Ding & Jessica Soong
2-25-2021

Table of Contents

Clinical Motivation

Operating in the temporal bone and lateral skull base is technically challenging. This region contains a complex geometry of nerves, arteries, veins, the end-organs for both hearing and balance, as well as the cranial nerves responsible for speech and swallowing.[1] To access this region, surgeons drill through varying densities of bone to identify surgical landmarks. In addition to the limited visibility of the surgical field and complex anatomical geometry in this space, critical anatomical structures are often within millimeters of each other.

Due to these conditions, temporal bone surgery poses a high risk of accidental damage to surrounding structures during free-hand procedures. For example, after cochlear implantation, cochlear implantation, 45% of patients experience changes in taste, with 20% of those patients having unresolved symptoms by the end of their follow-up period.[2] In more rare cases, patients also are at risk for facial paralysis due to accidental damage to the facial nerve.[3] Accidental damage to the brain or to the membrane surrounding the brain (dura) can lead to CSF leakage. Damage to the sigmoid sinus, which drains blood from the brain to the jugular vein, can lead to abnormal closure or even clotting of the sinus itself.[4]

One possible solution in mitigating accidental damage to surrounding structures is using a cooperative control robot intraoperatively. Previously, the Laboratory of Computational Sensing and Robotics (LCSR) has developed such a robot that holds on to the surgical drill, which the surgeon can freely control.[5] Robot-assisted surgery has the potential to reduce hand tremor and limit movement around sensitive structures, thereby increasing patient safety and improving long-term outcomes. However, a key dependency for realizing this technology in the operating room is providing meaningful information about patient anatomy so that the robot can safely guide the surgeon throughout the procedure. Effectively, this means highlighting important structures on patient CT imaging that can be registered to a robotic system.

Prior Work

Previous work in the LCSR has focused on segmenting CTs through registration methods (Figure 1).[6] With a manually segmented template CT, deformable registration methods can map or propagate template segmentations to target CTs that have not been segmented before. These segmentations can then be locally optimized to produce a final segmentation for the target CT. This segmentation propagation method achieves submillimeter accuracy for segmenting inner and middle ear structures, with an average surface distance of < 0.2 mm and almost 90% overlap with ground truth segmentations.

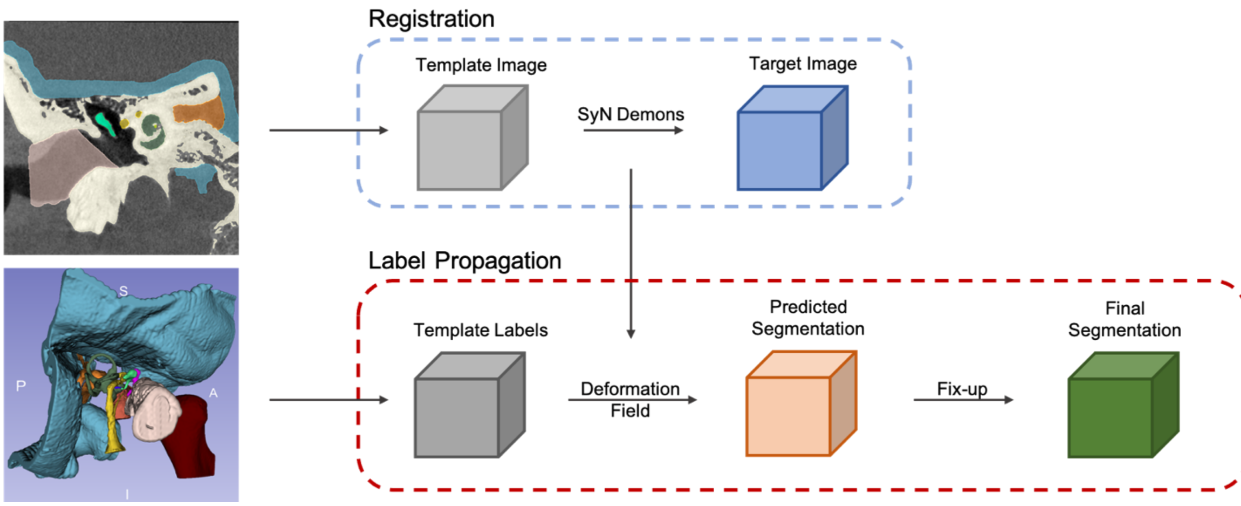

Figure 1. Pipeline of the segmentation propagation method for temporal bone CT segmentation

Goals & Significance

Significance:

Successful completion of this project will allow for more complete virtual safety barriers for robot-assisted temporal bone surgery. It can also be used to generate patient-specific segmentations as learning cases for junior otologists. Finally, this project has the potential to create the most complete dataset for model training and research.

Aside from the NIH OpenEar dataset, the dataset used for this project has the most complete segmentations of the temporal bone compared to any other group that has previously published in this area. In terms of anatomical boundaries of a mastoidectomy, which is the first step in virtually all temporal bone procedures, previous groups have only labeled one: the sigmoid sinus. This project's datasets not only label the sigmoid sinus, but also label the surrounding brain and the external auditory canal, which are the remaining mastoidectomy boundaries. By segmenting these areas, a cooperative control robotic system can then be able to apply virtual safety barriers to each of these boundaries, thereby providing for safe drilling throughout the procedure.

Broad Goals

- To evaluate state state-of-the-art deep learning models for semantic segmentation of the temporal bone.
- To build the largest comprehensively annotated temporal bone CT database to date.

General Experimental Setup

Our dataset consists of 23 manually segmented high resolution head CTs which have a voxel size of 0.1mm. The dimensions of the axial CT slices are 512x512 pixel^2 with an average z-stack of 494 images. The variability in the number of slices is due to differing patient anatomy and cropping parameters. All CT images are either of left or right temporal bone CTs with minor or no pathology and no prior temporal surgical procedures. The data has 17 labels: temporal bone, malleus, incus, stapes, vestibule and cochlea, vestibulocochlear nerve, superior vestibular nerve, inferior vestibular nerve, cochlear nerve, facial nerve, chorda tympani, ICA, sinus and dura, vestibular aqueduct, TMJ, and EAC, and foreign body (background).

For running experiments, MARCC and a local workstation will be used, both of which have specifications laid out in Table 1. The key takeaway from the setups is that both have state-of-the art GPUs with high amounts of VRAM suitable for training on 3D images.

Table 1: Training Rigs

	Temporary Setup	Big Rig
	Cloud: MARCC	**Baltimore Local Setup**
GPU	NVIDIA K80, V100	GeForce RTX 3090
VRAM	24 GB DDR5 / 32 GB DDR5	24 GB DDR6
OS	Ubuntu / CentOS	Ubuntu / Windows
RAM	16 to 64 GB	32 GB

Technical Approach

Since the dataset used is small for deep learning approaches, other methods must be explored to train the model on less data. The methods being explored will be semi-supervised pre-training and data generation. Then, three different models will be implemented, and their results compared.

Managing the Small Dataset

Data Generation

CT scans can range anywhere from 0.1 mm resolution to greater than 2mm resolution. Lower resolution CT scans of the temporal bone can be simulated by down sampling the data, then up sampling to the original dimensions. Furthermore, the data can be augmented to simulate different CT protocols, for example, ones optimized for soft tissue instead of bony structures.

Semi-Supervised Pre-Training

Using the segmentation propagation method described in the Prior Work section, the dataset can be expanded to 44 datasets at time of writing and is continuing to increase in size as more CT volumes are registered. Using these deformed volumes and segmentations preserves the overall quality of segmentations but may increase training bias due to similarity between deformed segmentations and the template segmentations from which they were derived. Importantly, these deformed segmentations can be used as good initial labels for their corresponding patient CT volumes to provide more ground truth images. Further exploration of data generation is being pursued in the form of statistical shape modeling of deformation fields generated via the segmentation propagation method. However, depending on the time constraints, more data may be synthesized through random deformable registration.

Models

nnUnet

nnUnet is a new benchmarking pipeline developed to standardize medical imaging.[7] It has top 33 leaderboard results for 53 different datasets, and effectively is a black box. It can be used to quickly establish a benchmark and there are 2D and 3D approaches available.

AH-Net

Anisotropic Hybrid Network (AH-Net) is a model that can leverage pre-trained 2D models and extrapolate the learned weights to apply to image volumes.[8] This reduces computational need for 3D neural network training by using pre-training on 2D images and leverages 3D convolutions to decrease redundant calculations that occur when doing slice-level 2D analysis. This model outperformed common architectures (UNet, 3D ResNet) for a similar application (segmentation of micro CTs on ears).[9] The model code is released on GitHub, although the dataset and dataloader functions will need to be coded by the group.

PWD-3D Net

PWD-3D net is a patch-based network that has been used to segment similar CT scans, albeit with different labels. It has shown good results with a high dice score for temporal bone segmentation (0.86) and low Hausdorff distance (0.755mm).[10] A concern when working with 3D data is always computational power. Since this uses balanced class patch-based sampling, the effective load is decreased since the network only operates on a small subset of the voxels at once.

Key Activities and Deliverables

The key activities and deliverables can be found in Table 2. The minimum, expected, and maximum activities and corresponding deliverables are laid out, and a Gantt Chart including the exact details and timeline is found in Figure 2.

Table 2: Activities and Corresponding Deliverables

	Activity	**Deliverable**
Minimum	Implementing nnUnet	Fully functioning model for CT segmentation.
	Training/validating nnUnet results on test data.	Internal validation report with ground truth segmentations.
Expected	Implementation of three different models for CT segmentation	Comparison report of different models for CT segmentation (including code).
	Training/validating the chosen three models on test data.	Accuracy and performance reports for CT segmentation models.
Maximum	Final manuscript preparation.	Submittable manuscript.
	Application of best model to unlabeled dataset.	High quality segmented temporal bone CT dataset using our segmentation models.
	Application of models to external dataset.	External validation reports with Western University's dataset.[10]

Dependencies

The project is mainly virtual so there are few physical dependencies. The dependencies are listed in Table 3. The project team members already have access to MARCC, so no such dependency exists for that. The annotations were finalized before any training began, and Dr. Unberath has agreed to be a deep learning mentor for the project, so two of the dependencies originally proposed have been resolved.

Table 3: Dependencies

Dependency	Need	Status	Follow-up	Contingency Plan	Deadline	Effect
Label/Annotation Finalization	Need Data to Train	100% Done.	Check-in with annotators EOD 2/11	Use unfinalized labels to debug/test with.	2/12	Can use preliminary labels (mostly done) to train/debug with
Workstation Arrival	Computational Power and Availability	Received Quote from Approved Vendor.	Check-in with JHU representative for purchase approval.	Continue to use gcloud/MARCC	2/27	If workstation purchase not approved, will continue to use MARCC/gcloud. Potentially, depending on GPU availability, it may take longer to train some implementations and reduce the overall # of network architectures implemented
Dr. Unberath Supervision Agreement	Need a Deep Learning Consultant for the project	100% Done. Dr. Unberath has agreed to be an advisor.	Meeting with Dr. Unberath every Friday	Max (technical consultant) has deep learning experience, continue project with him as lead.	2/12	May run into some issues if Max's expertise cannot help us through some issues, although the risk is low, since we are implementing methods that have been shown to work on similar datasets.

Timeline

The timeline for this project can be broken down into nine main tasks, outlined in Figure 2. This includes finalizing the data, setting up the environment, generating data, implementing three different networks, external dataset validation, applying the best model to an unlabeled dataset, and then preparing a manuscript to submit. The minimum deliverables are expected by 3/25/2021, expected deliverables by 4/11/2021, and final deliverables by 5/4/2021.

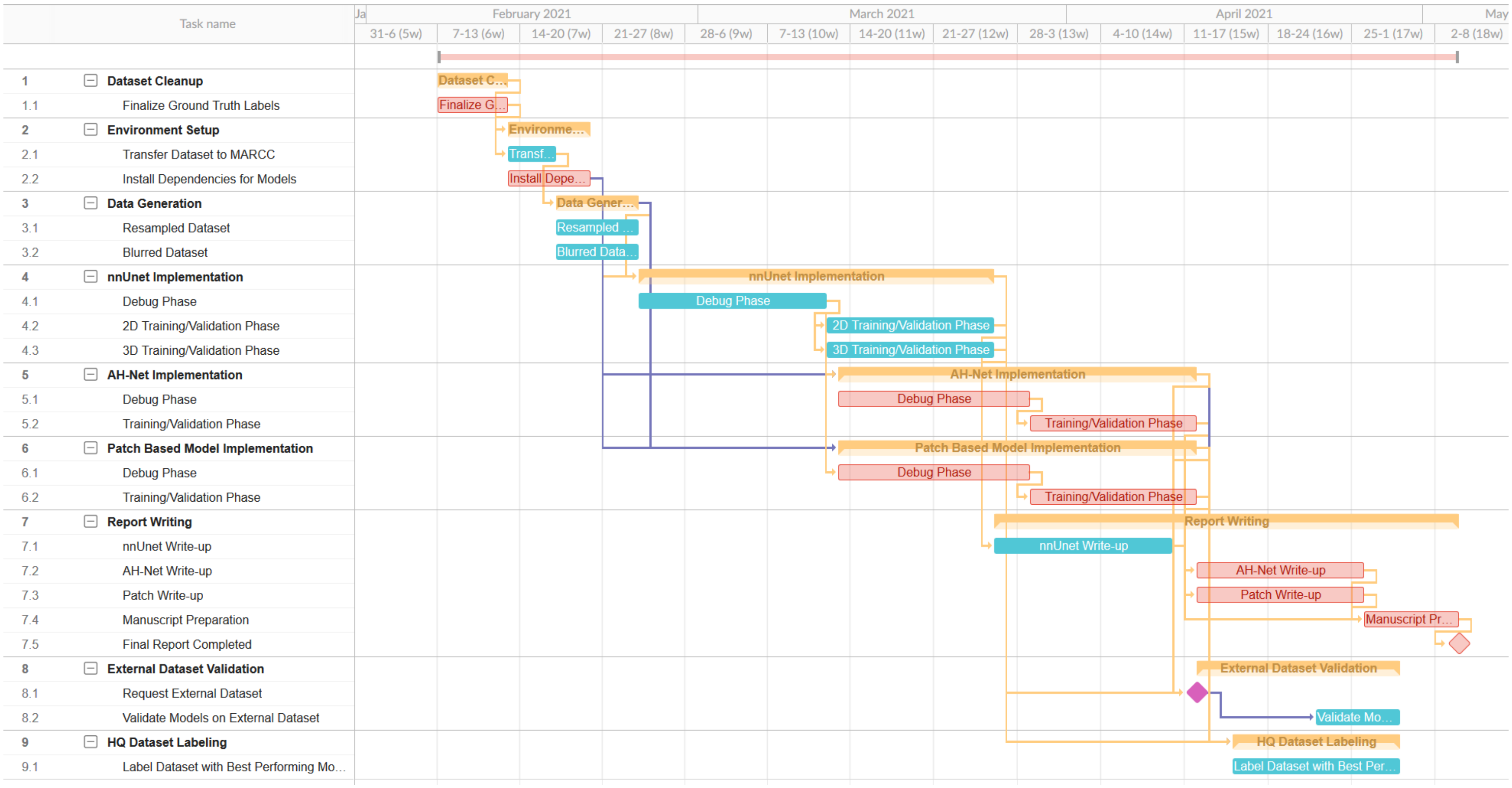

Figure 2: Project Timeline

Team Members & Roles

The team consist of:

- Jessica Soong (jsoong1@jhu.edu)
  MSE Student, LCSR, second-year
  Responsible for environment setup, patch-based model or AH-Net implementation and writeup, external dataset validation
- Andy Ding (andy.ding@jhmi.edu)
  MSE Student, LCSR & Department of Biomedical Engineering, first-year
  Responsible for environment setup, patch-based model or AH-Net implementation and writeup, dataset ground truth management

The team will also have some shared responsibilities, which are the data generation, nnUnet implementation and write-up, and the final manuscript preparation.

Mentors

The mentors consist of:

- **Dr. Russell Taylor** (rht@jhu.edu)
  Professor, Department of Computer Science
  Expertise in medical imaging, computer-integrated surgery.
- **Dr. Francis X. Creighton** (francis.creighton@jhmi.edu)
  Assistant Professor, Department of Otolaryngology
  Expertise in lateral skull base surgery.
- **Dr. Mathias Unberath** (unberath@jhu.edu)
  Professor, Department of Computer Science
  Expertise in medical imaging, deep learning.
- **Maxwell Zhaoshuo Li**: (zli122@jhu.edu)
  PhD Candidate, Department of Computer Science
  Expertise in deep learning, computer-integrated surgery, medical imaging.

Management Plan

Meetings:

At time of writing, there are weekly meetings with LCSR every Wednesday at 3 PM, as well as meetings with Dr. Unberath for the Models and Registration meeting every Friday at 2 PM. Jessica and Andy will meet on an ad hoc basis and consult with Max as needed.

Platforms:

Multiple platforms will be used for communication as well as documentation, file-sharing, and report writing.

- **Communication:** For communication, a slack channel with LCSR will be used. This will be supported with e-mail and Zoom meetings.
- **Code:** Code will be maintained on a private repository on GitHub.

- **Data & Filesharing**: The anonymized data will be shared through Hopkins OneDrive, which is secure and encrypted. Files such as the report and presentations will also be shared through OneDrive and the CIS II website.
- **Report Writing**: OneDrive will be used for basic report writing, and LaTeX (Overleaf) will be used for final manuscript preparation.

Reading List

S. Nikan *et al*., "PWD-3DNet: A Deep Learning-Based Fully-Automated Segmentation of Multiple Structures on Temporal Bone CT Scans," in *IEEE Transactions on Image Processing*, vol. 30, pp. 739-753, 2021, doi: 10.1109/TIP.2020.3038363.

Neves, C.A., Tran, E.D., Kessler, I.M. *et al.* Fully automated preoperative segmentation of temporal bone structures from clinical CT scans. *Sci Rep* **11,** 116 (2021). https://doi.org/10.1038/s41598-020-80619-0

li, Xiaoguang & Gong, Zhaopeng & Yin, Hongxia & Zhang, Hui & Wang, Zhenchang & Zhuo, li. (2020). A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images. Neural Networks. 124. 10.1016/j.neunet.2020.01.005.

Gibson, Eli et al. “NiftyNet: a Deep-Learning Platform for Medical Imaging.” Computer Methods and Programs in Biomedicine 158 (2018): 113–122. Crossref. Web.

Liu S. et al. (2018) 3D Anisotropic Hybrid Network: Transferring Convolutional Features from 2D Images to 3D Anisotropic Volumes. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11071. Springer, Cham. https://doi.org/10.1007/978-3-030-00934-2_94

References

1. Cousins VC. Lateral skull base surgery: a complicated pursuit? *The Journal of Laryngology & Otology*. 2008-03-01 2008;122(3):221-229. doi:10.1017/s0022215107000436

2. Lloyd S, Meerton L, Cuffa RD, Lavy J, Graham J. Taste change following cochlear implantation. *Cochlear Implants International*. 2007-12-01 2007;8(4):203-210. doi:10.1179/cim.2007.8.4.203

3. Fayad JN, Wanna GB, Micheletto JN, Parisier SC. Facial Nerve Paralysis Following Cochlear Implant Surgery. *The Laryngoscope*. 2003-08-01 2003;113(8):1344-1346. doi:10.1097/00005537-200308000-00014

4. Zanoletti E, Cazzador D, Faccioli C, Martini A, Mazzoni A. Closure of the sigmoid sinus in lateral skull base surgery. *Acta Otorhinolaryngol Ital*. 2014;34(3):184-188.

5. Razavi CR, Wilkening PR, Yin R, et al. Image-Guided Mastoidectomy with a Cooperatively Controlled ENT Microsurgery Robot. *Otolaryngology–Head and Neck Surgery*. 2019;161(5):852-855. doi:10.1177/0194599819861526

6. Sinha A, Leonard S, Reiter A, Ishii M, Taylor RH, Hager GD. Automatic segmentation and statistical shape modeling of the paranasal sinuses to estimate natural variations. SPIE; 2016:

7. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*. 2021-02-01 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z

8. Liu S, Xu D, Zhou SK, et al. 3D Anisotropic Hybrid Network: Transferring Convolutional Features from 2D Images to 3D Anisotropic Volumes. 2017;

9. Neves CA, Tran ED, Kessler IM, Blevins NH. Fully automated preoperative segmentation of temporal bone structures from clinical CT scans. *Scientific Reports*. 2021-12-01 2021;11(1)doi:10.1038/s41598-020-80619-0

10. Nikan S, Van Osch K, Bartling M, et al. PWD-3DNet: A Deep Learning-Based Fully-Automated Segmentation of Multiple Structures on Temporal Bone CT Scans. *IEEE Transactions on Image Processing*. 2021-01-01 2021;30:739-753. doi:10.1109/tip.2020.3038363