JOHNS HOPKINS UNIVERSITY

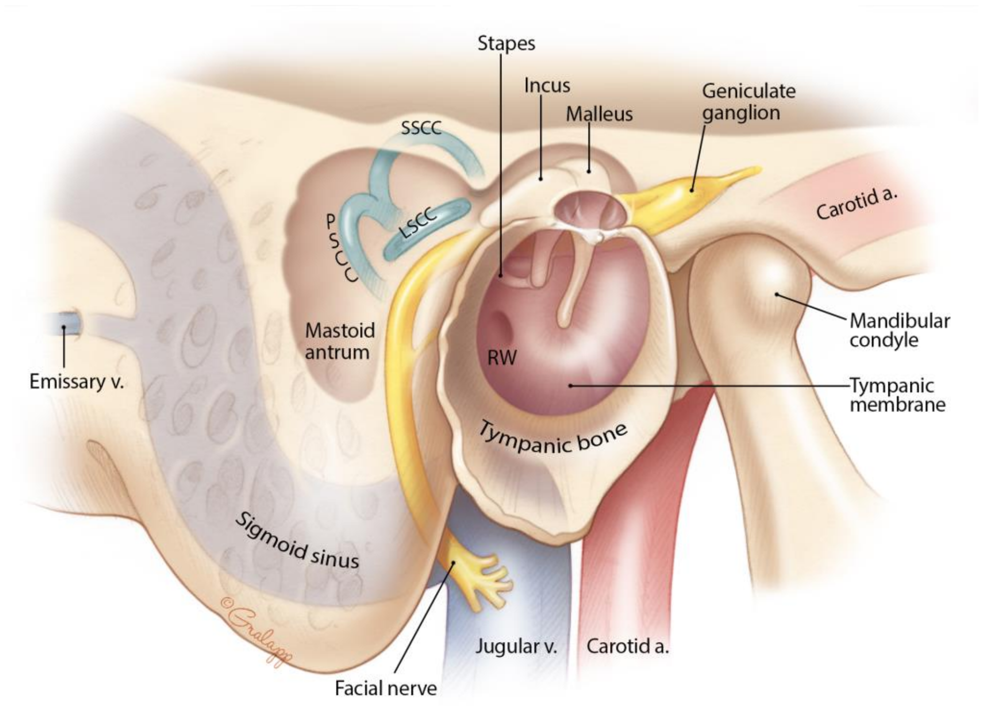
Automated Segmentation of Temporal Bone CT Imaging for Robot-Assisted Microsurgery

Group 4: Andy Ding, Jessica Soong

Mentors: Dr. Francis X. Creighton, Dr. Russell H. Taylor, Dr. Mathias Unberath, Max Zhaoshuo Li

Project Overview

- Problem
  - Temporal bone anatomy is geometrically complex with critical structures often within millimeters of each other. Surgery in this area poses a high risk of damage to anatomy.
- Goal
  - Develop an automated system for segmenting the temporal bone to help prevent intraoperative damage to critical anatomy.



Credit: Christine Gallup
https://otosurgeryatlas.stanford.edu/

Paper: Summary & Key Takeaways

PWD-3DNet: A Deep Learning-Based Fully-Automated Segmentation of Multiple Structures on Temporal Bone CT Scans

Soodeh Nikan, Kylen Van Osch, Mandolin Bartling, Daniel G. Allen, S. Alireza Rohani, Ben Connors, Sumit K. Agrawal, and Hanif M. Ladak, *Member, IEEE*

- **Summary**: Paper implements a patch-based network with balanced class sampling to segment temporal bone (& additional structures) from CT scans [1].
- **Key Takeaways**: At time of publishing, first 3D method designed for temporal bone segmentation. Has high dice score similarity scores and low Hausdorff distances for temporal bone structures improved due to overlapping of patches.
- **Relevance**: Network designed for temporal bone segmentation, designed with similar class imbalances in mind, as well as limited compute power.

[1] S. Nikan *et al.*

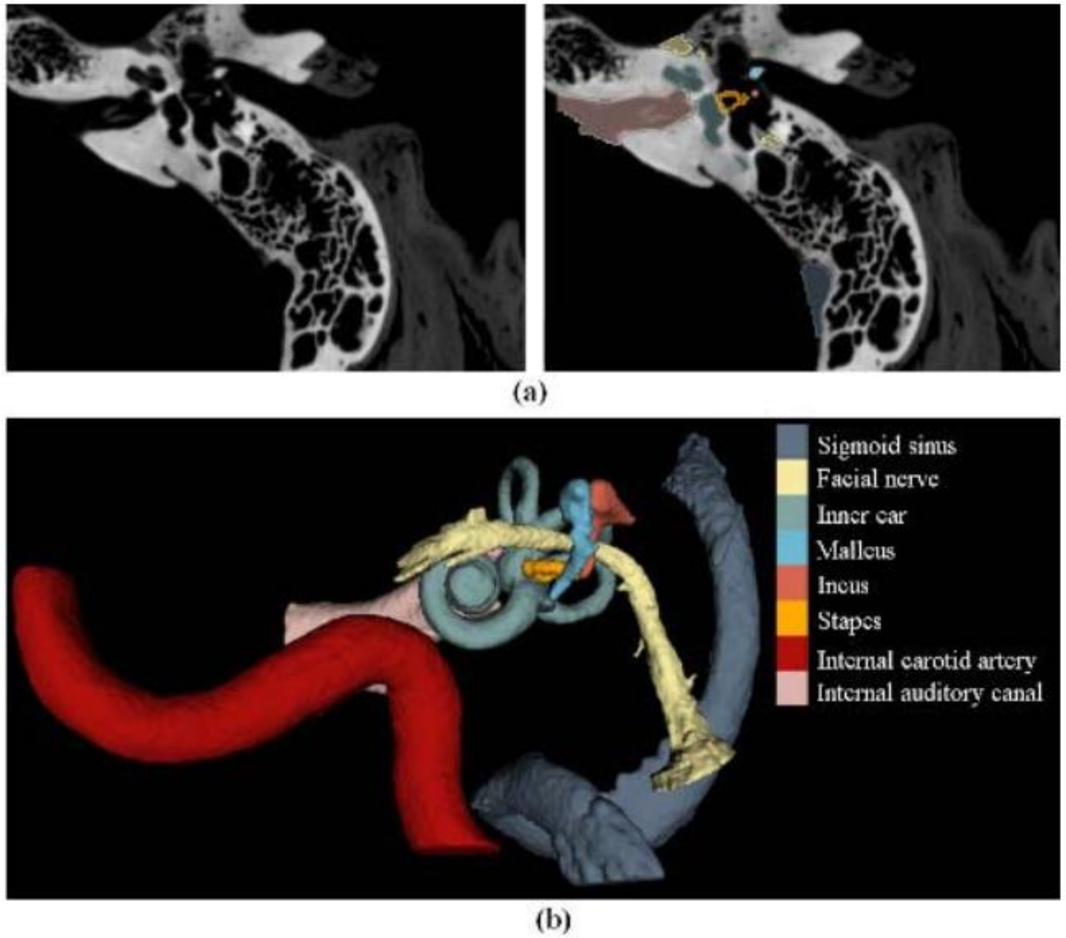
Paper: Motivations/Background

- Patient specific segmentation of temporal bone organs from CT scans are useful to improve feasibility of robotic surgery planning.
- Manual labeling is very time consuming and takes anywhere from 1 hour to 1 day depending on the image.
- Some previous algorithms implemented for this purpose use 2D U-Nets ensembled in some way [2].
  - This is inefficient since ensembles will have redundant convolutions that could be captured with 3D convolutions.
  - 2D convolutions also lose spatial context.
- Other 3D implementations often use down-sampling to fit the entire image on a GPU.
  - Since some structures are extremely small but still vital, this method cannot be used without completely losing the information

[2] Fauser, J., et al.

Paper: Dataset

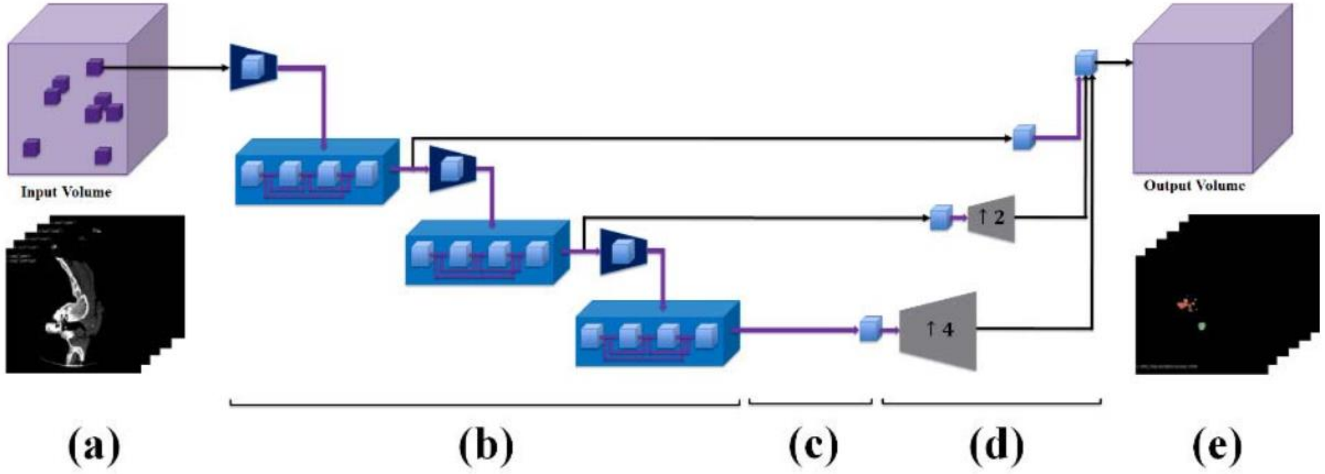
- 39 adult cadaveric temporal bone specimens used.
  - Labeled temporal bone structures using region growing and manual correction
- Total of 78 micro-CTs and clinical scans each (left and right ear)
- Train: 126 scans
- Validation: 14 scans
- Test: 18 unseen scans

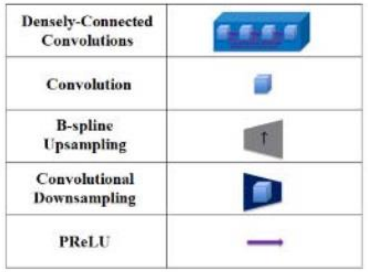


S. Nikan *et al., Figure 1*

Paper: Workflow

- An input volume is sampled 32 times using a balanced patch sampler.
- Using a structure similar to DenseVNet, the sub volumes go through a series of deeply connected residual layers and are upsampled at each level and concatenated before passing through the final layer to form the output volume
  - B-spline upsampling gradually decreases compression which decreased amount of artifacts
- Loss is calculated between prediction and labels, using a dice score sensitive to outliers.

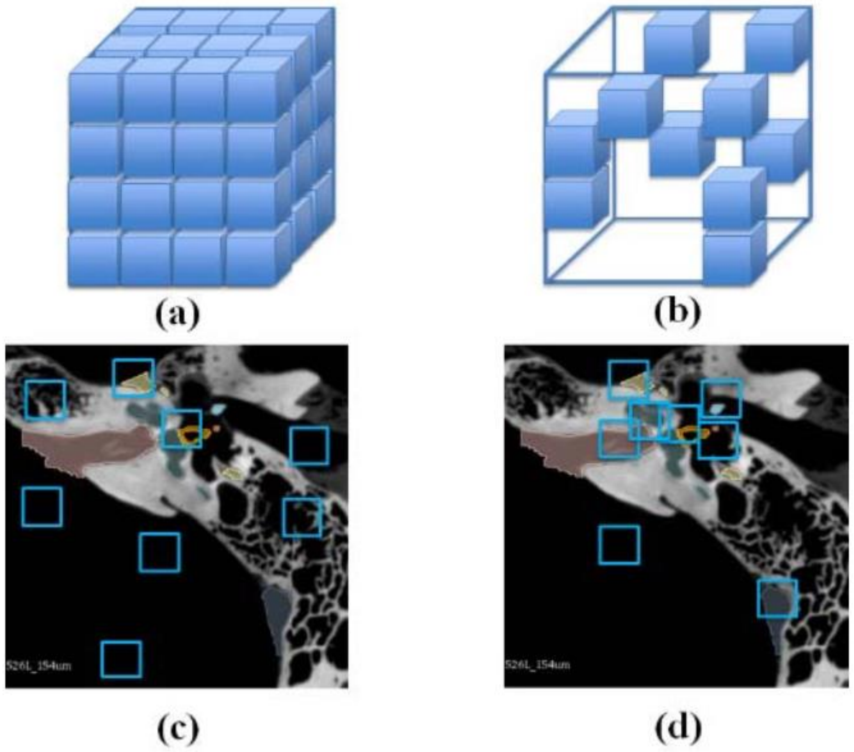




S. Nikan *et al., Figure 2*

Paper: Sampling Methods

- Some popular types of sampling methods are:
  - (a) Grid Window Sampling
    - Subsamples derived from sliding window, lots of redundant calculations
  - (b) Uniform Window Sampling
    - Slow since all feasible locations of image windows must be calculated, then randomly sampled (3D images, curse of dimensionality!)
  - (c) Background/Foreground Sampling
    - 50%/50% background and foreground class, still has imbalance
- Proposed Sampling Method:
  - **(d) Balanced Window Sampling**
    - Uses GT labels as sampler weights so each class is sampled with equal probability



S. Nikan *et al., Figure 3*

Paper: Data Synthesis

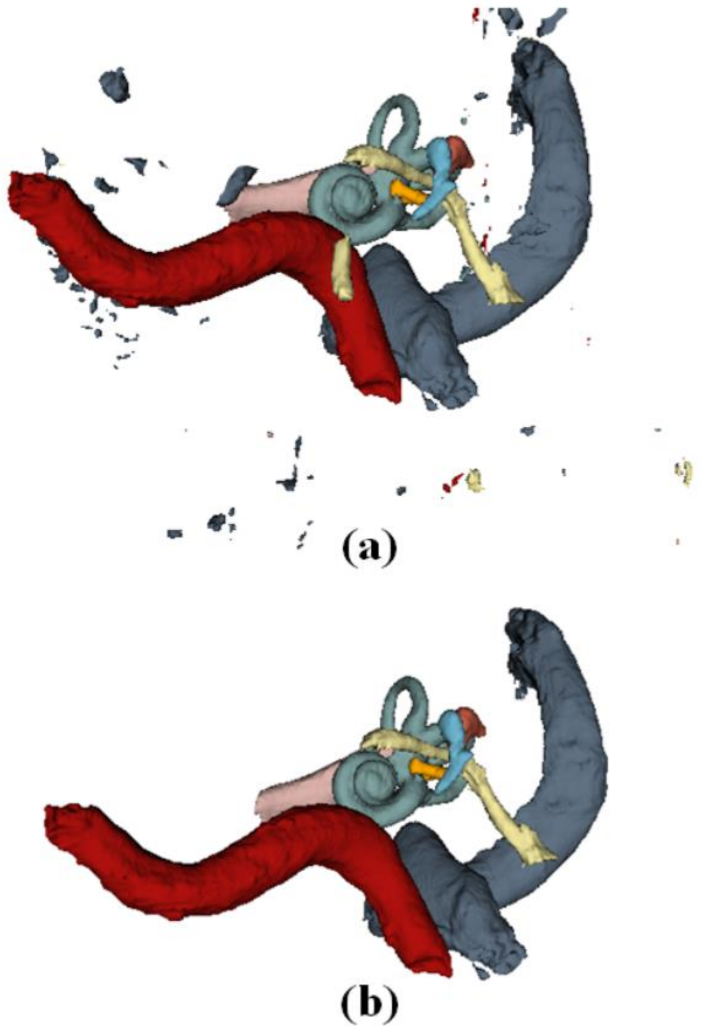
- Since the dataset is still small relative to other image datasets, augmentation was used to simulate a larger dataset.
  - Blurred Micro-CT:
    - Global gaussian filter applied to high-resolution CT images to allow the resulting model to segment various types of CT scans optimized for different tissues
  - Resampled Micro-CT
    - Images downsampled to larger slice thickness, then upsampled to original voxel size using B-splines to simulate CTs taken with a lower resolution.
  - Addition of clinical CTs
    - Micro-CTs not performed in human patients, only cadavers. Cadavers were scanned before Micro-CTs were performed to increase the dataset size.

Paper: Data Augmentation (On The Fly)

- Histogram standardization + Whitening Normalization performed based on training images.
- Random rotation of each orthogonal plane by [-10°, 10°]
- Spatial rescaling in the range of [0.9, 1.1] of original size.

Paper: Inference

- Patch based sampling is for training only and requires labels to sample.
- For inference, patches are sampled using the grid window sampling method, with amount of overlap as a hyperparameter.
  - More overlap means slower calculation, less overlap may mean decreased accuracy.
- Post-processing needed since the patch-based method results in noise and disconnected components.



S. Nikan *et al.*, *Figure 5*

Paper: Evaluation Metrics

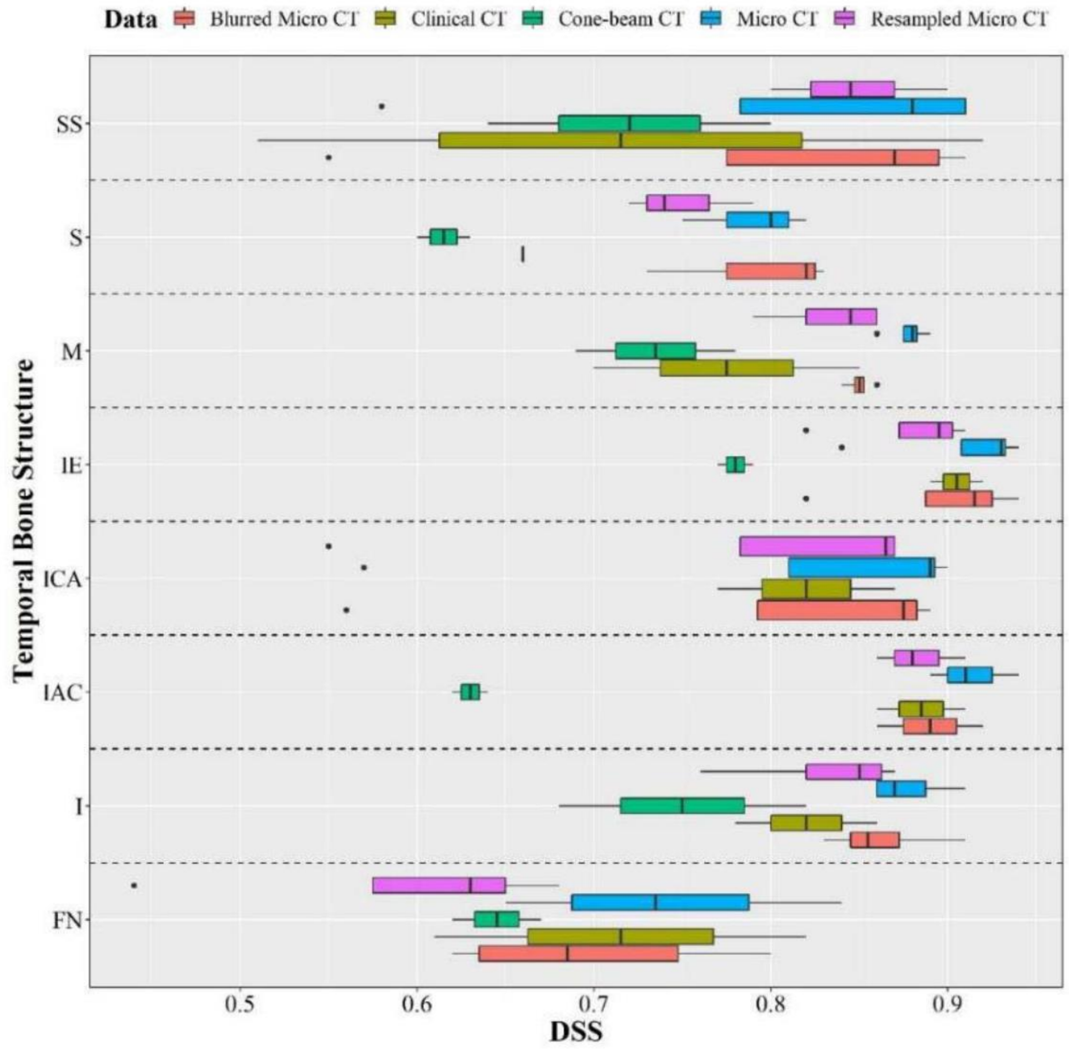
- Key Points:
  - Performance on Clinical CTs has a wide range of dice score.
  - Overall, the dice score metrics are above 0.8 except for the facial nerve and stapes.

TABLE II

AVERAGE DICE SIMILARITY SCORE (DSS), HAUSDORFF DISTANCES (MILLIMETER) AND JACCARD SCORE (JS) OF THE PROPOSED SEGMENTATION ALGORITHM (AUGMENTED NETWORK) FOR EIGHT TEMPORAL BONE STRUCTURES. STRUCTURES ARE SIGMOID SINUS (SS), FACIAL NERVE (FN), INNER EAR (IE), MALLEUS (M), INCUS (I), STAPES (S), INTERNAL CAROTID ARTERY (ICA) AND INTERNAL AUDITORY CANAL (IAC)

	SS	FN	IE	M	I	S	ICA	IAC
DSS	0.86	$\boxed{0.74}$	0.90	0.84	0.85	$\boxed{0.77}$	0.81	0.89
HD	1.91	1.23	0.27	0.26	0.28	0.28	1.96	0.62
JS	0.75	0.59	0.82	0.72	0.74	0.63	0.68	0.80

S. Nikan *et al., Table II*



S. Nikan *et al., Figure 7*

Paper: Conclusions

- Experiments on patch-size show that a patch size of 144x144x144 is optimal.
- Experiments on volume of overlap show that 32x32x32 is optimal.
- Balanced window sampling is the most appropriate sampling technique for temporal bone segmentation due to the large class imbalances.
  - Authors credit the outperformance of other state of the art models due to this.
- Synthesis/Augmentation of data is key to making the data generalizable on a small dataset.
  - When tested on a held-out dataset from another institution, an average dice score of 0.64 was obtained.

Paper: Critiques (Pros)

- Code available on github with docker file.
- Data available through written request.
- Paper's training methods were well explained, model explained with limited introduction of mathematical notation (generally appreciated by clinicians)
- Extremely thorough experiments:
  - Had comparisions of performance on augmented/non augmented data for experiments
  - Experiments to find optimal patch size
  - Experiments to find optimal volume of overlap for inference
  - Contains comparison of proposed network with established segmentation methods (DeepMedic, DenseVNet) which could serve as a comparison for window sampling
    - DeepMedic uses foreground/background window sampling

Paper: Critiques (Cons)

- Code poorly documented.
- Inference method not clearly explained.
  - What was done with overlapping sub-sampled pixels with different predicted labels?
- Dataset after augmentation was primarily micro-CTs, which are not done on living subjects.
  - For increasing feasibility of robotic surgery, it would be better if the study could be done as closely as possible to the same workflow in a clinical setting.
  - Augmentation of a flat gaussian blur is not a good method for simulating another CT data optimized for softer tissues.
- Frames 0.64 DSC on external validation as sign that the algorithm can generalize well.

Paper: Potential Next Steps

- Study the efficacy of the algorithm on scans containing abnormal pathology.
  - Validate algorithm by augmenting pediatric CT scans from samples of patients with cochlear implants or other abnormal pathology
- Add additional structures on temporal CTs.
- Add multi-institutional data to the training set to help model generalize better.

Project Relevance

- One of the goals of the project is to compare different methods for temporal bone segmentation and their performance on our dataset.
- Potentially PWD-3D Net will be a good model to compare results with since it shows promising results.

References

1. 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.
2. Fauser, J., et al. (2019). "Toward an automatic preoperative pipeline for image-guided temporal bone surgery." International Journal of Computer Assisted Radiology and Surgery **14**(6): 967-976.

Image References

- Temporal Bone Anatomy. https://otosurgeryatlas.stanford.edu/wp-content/uploads/2020/06/7a-1.jpg. Accessed March 8, 2021

Thank you!

Questions?