Automated Segmentation of the Eustachian Tube – A Deep Learning Platform

CIS-II Final Report
Team #3
May 1, 2022

Mentors:
Dr. Francis Creighton
Dr. Manish Sahu
Prof. Mathias Unberath
Prof. Russell Taylor

Team Members:
Ameen Amanian
Chanha Kim
Yuliang Xiao

ABSTRACT
Eustachian tube dysfunction (ETD) results from impairment in middle ear ventilation and pressure regulation. Treatment of this condition entail medical and surgical management such as eustachian tube dilation. However, with the current clinical tools such as patient-reported surveys and tympanograms, clinicians often face a diagnostic dilemma with classifying which patients may have ETD. Obtaining a 3-dimensional mesh of the eustachian tube and differentiating variances between patients with ETD vs. no ETD would be of clinical utility. However, manual segmentation is time-consuming and prone to inter-reader variability. This project aimed to develop a deep learning framework for automated segmentation of the eustachian tube on computed tomography images. We employed nnUNet, a state-of-the-art deep learning neural network proven to be transferrable within multiple domains in medical image segmentation. The preliminary results are promising in obtaining accurate segmentations of the bony and nasal portion of the eustachian tube. However, the middle (i.e., cartilaginous portion) is poorly predicted secondary to difficulty in deciphering the ground truth using CT images. With 9 testing images, the overall dice similarity coefficient for the eustachian tube, internal carotid artery, and torus tubarius was 0.649, 0.891, and 0.735 respectively. Future studies will assess the benefit of integrating MRI and registration-label propagation techniques to further improve the ground truth segmentation prior to training via the deep learning framework.

KEYWORDS (5): Eustachian tube, medical image segmentation, deep learning, image registration, prediction

1. INTRODUCTION

Eustachian tube dysfunction (ETD) results from impairment in middle ear ventilation and pressure regulation [1], [2]. As a result, patients experience a range of symptoms ranging from ear pain, pressure, cracking, to difficulty hearing which has a significant impact on a patient's quality of life [1]. Eustachian tube dilation is a procedure approved for the surgical management of ETD [2]. However, its proximity to certain critical structures such as the internal carotid artery pose a risk of injury associated with this procedure [2]. Therefore, to better treat this pathology, a better anatomical understanding of this intricate structure as well as surrounding anatomical structures is needed. Computed Tomography (CT) scans are one of a series of modalities that allow clinicians to visualize the path of this structure. Manual segmentation of medical images proposes a solution to visualize the three-dimensional nature of a structure for better visualization; however, it is time consuming, prone to inter-reader variability, and difficult to translate into the clinical setting. Deep learning has paved the way for automated segmentation of an anatomical structure meanwhile achieving significant accuracy. Therefore, we aim to assess the utility of and develop a deep learning pipeline to perform automated segmentation of the eustachian tube (ET), define near-by critical structures, and establish the first pipeline that can be translated into the clinical realm for the treatment of ETD.

2. CLINICAL PROBLEM

The current diagnostic tools such as a tympanogram and patient-reported surveys for classifying patients with ETD lack clinical utility [3]. A framework that provides a 3-dimensional representation of the eustachian tube and identifies features that distinguish a patient with ETD vs. one without any pathology could have the potential to solve the clinicians' dilemma pertaining to the diagnosis and management of ETD.

3. METHODS

3.1 Data preparation and annotation

Deidentified and de-faced computed tomography (CT) images were obtained from the Department of Otolaryngology - Head and Neck Surgery at Johns Hopkins University. Scans with any eustachian tube pathology were excluded. The resolution of the images was 512x512xN where N refers to the number of slices in the axial direction. Two sets of annotations were prepared: 1) eustachian tube 2) eustachian tube and surrounding structures (internal carotid artery, torus tubarius). The structures were manually segmented on 3D slicer in a slice-by-slice manner by an Otolaryngology - Head and Neck Surgery senior resident physician and then verified by an Otology and Lateral Skull Base surgeon.

3.2 Image Registration

To make the data compatible with the deep learning frameworks (e.g., nnU-Net), the raw images are required to be co-aligned. We utilized ANTsPy, a Python library which includes blazing-fast IO, registration, segmentation, statistical learning, and visualization functionalities [4]. First, a random image was chosen as a template and then the remaining images with registered to the template. The 'forward' deformation field was then applied to each image within the dataset to ensure they were co-aligned with the template. Overall, this task was employed to confirm that the images have the same rotation, angle, and spacing.

3.3 Deep Learning Framework

In recent years, deep convolutional networks have outperformed the state-of-the-art in many visual recognition tasks. In biomedical image processing, convolutional neural networks have been used for semantic segmentation tasks, where instead of assigning a single label to an entire image, a localization process is performed where a class label is assigned to each pixel. Two deep learning frameworks utilized in this paper included i) nnU-Net and ii) VoxelMorph. They both utilize the U-Net convolutional neural networks architecture which can yield precise segmentations from very few training images. The architecture of U-Net is symmetric and consists of two major parts: the first half is called the contracting path, which is constituted by the general convolutional process, and the second half is an expansive path, which is constituted by transposed 2d convolutional layers (upsampling technique). Upsampling occurs along this latter path to recover the location information pertaining to each pixel. Thus, the output of U-Net is a complete high-resolution image in which all the pixels are classified.

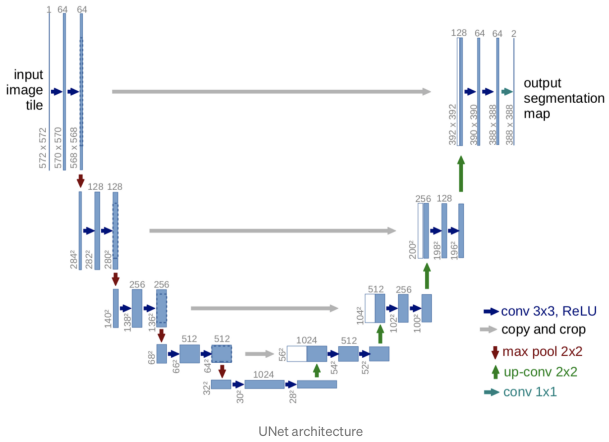

Figure 1 *- Architecture of U-Net*

i. nnU-Net

The nnU-net algorithm is the first segmentation method designed to deal with dataset diversity found in the medical image segmentation domain [5]. For our project, we focused on using nnU-net as the basis for our deep learning model for semantic segmentation of CT images. nnU-net first uses its novel heuristic rule to determine the data-dependent hyperparameters, or data fingerprints, to automatically ingest the training data set. Then, the blueprint parameters (such as loss function, and network architecture) and inferred parameters (such as image resampling and batch size) along with the data fingerprint generate the pipeline fingerprints. Then, the pipeline fingerprints produce network training for 2D, 3D, and 3D-Cascade U-Net using the hyperparameters determined so far. Then, with post-processing and an optional ensembling strategy, (i.e., assigning weights to each of the models and combining them together) the best configuration will be used by nnU-net to produce the final prediction. In summary, the heuristic rule determines the hyperparameters based on the training data properties, and then the pipeline fingerprint is generated which produces the best network training model.

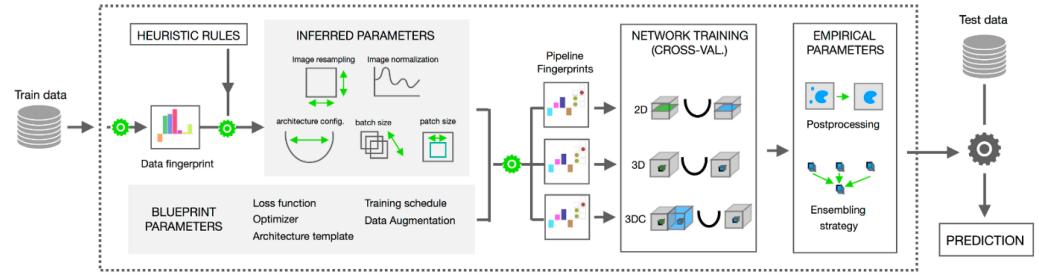

Figure 2 *- nnU-Net Configuration*

The two main motivations behind using nnU-Net are first its ability to handle a wide variety of target structures. In other words, unlike other deep learning models, nnU-Net is not a specialized solution for a certain type of data set, but rather a very generalized deep learning algorithm that not only proved to handle a variety of data sets, but also surpass most existing approaches for data segmentation tasks. Second, its self-configuring ability allows us to efficiently train and use the model for which the result can serve as a benchmark to be improved upon if the training is not successful.

Our project workflow utilizing nnUNet is shown below, which include data preparation, training with different configurations, and ET-specific trained-model evaluation.

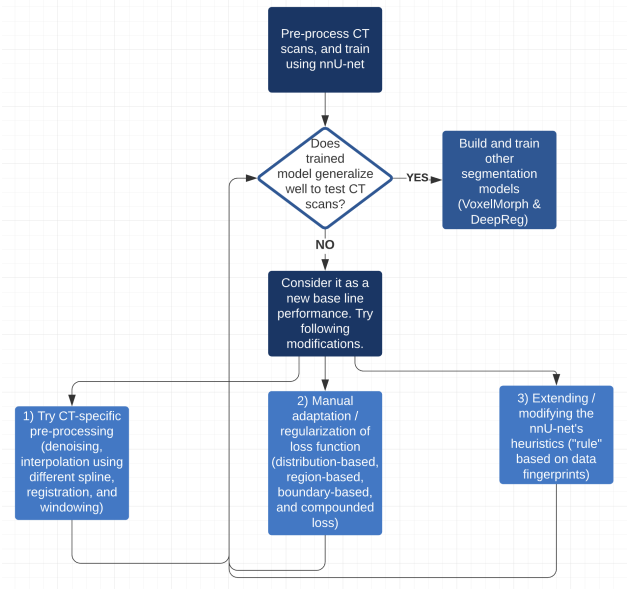

Figure 3 *- Workflow for the Deep Learning Framework*

ii. Voxelmorph

VoxelMorph is a deformable registration deep learning framework that aims to learn the deformation field function from input CT scans with or without ground-truth labels (unsupervised vs. semi-supervised) [6]. The registration result can then be used for label propagation to warp the labels of one CT scan to another to obtain the desired segmentation. VoxelMorph introduces a novel registration method that learns a parametrized registration

function from a collection of volumes using the U-Net architecture. This parametrized registration function can compute the deformation field given input volumes, which allows pixel-to-pixel mapping from one image to another. Thus, VoxelMorph has a single global function optimization through gradients which can be applied to any pair of images in the dataset, instead of optimizing one pair of images at a time as used in traditional registration algorithms.

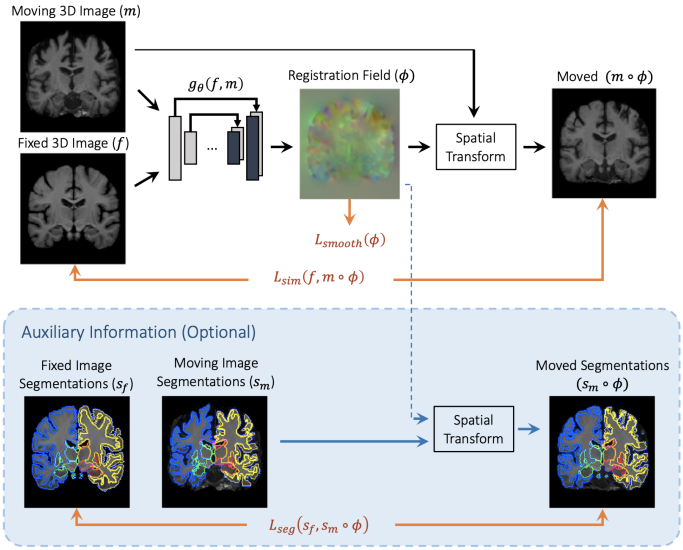

Figure 4 *- VoxelMorph Overview Method - We learn parameters θ for a function $g_\theta(f, m)$, and register a 3D volume 'm' to a second, fixed volume 'f'. During training, we warp 'm' with 'φ' using a spatial transformer function. Optionally, auxiliary information such as anatomical segmentations s_f, s_m can be leveraged during training (blue box).*

Our project workflow utilizing VoxelMorph is shown below which includes manual preprocessing of data, template creation, and training and evaluation of two methods 1) unsupervised model and 2) semi-supervised model.

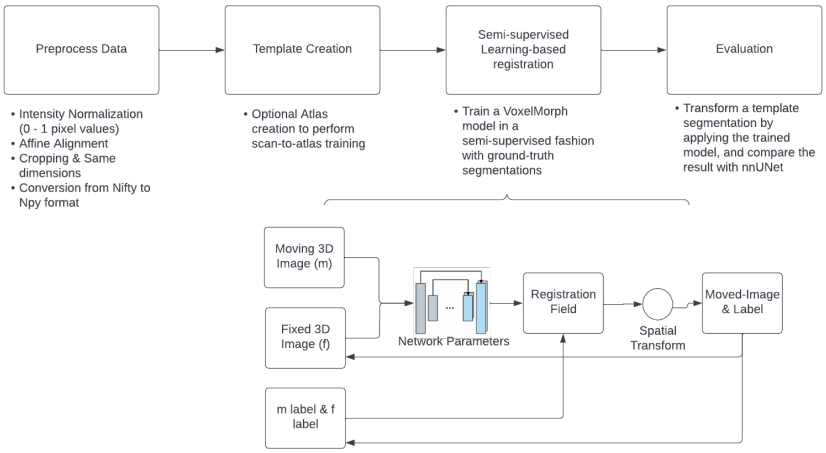

Figure 5 *- Registration-Segmentation Pipeline Utilizing the VoxelMorph Architecture*

3.4 Validation

One measure used for image segmentation-prediction model validation includes the dice similarity coefficient (DSC), a scoring system which measures volumetric overlap between two

images [7]. However, as the eustachian tube is a very thin structure, the conventional use of DSC is not appropriate for our project due to the eustachian tube's natural structure. First, we utilized the heat map to visualize where nnUNet was having challenges in providing predictions. Second, we calculated the Average Hausdorff Distance (AHD) and Weighted Hausdorff Distance (WHD) to quantify metrics that would be most appropriate for the eustachian tube.

A) Heat Map

To visualize where nnUNet is facing challenges in its predictions, we constructed a heat map. For this, we computed the closest distance between each vertex of the prediction mesh to the ground truth mesh, and then visualized the heat map based on this closest distance. An example is shown below which demonstrates the model has less accuracy when predicting the segmentation for the nasal portion of the eustachian tube.

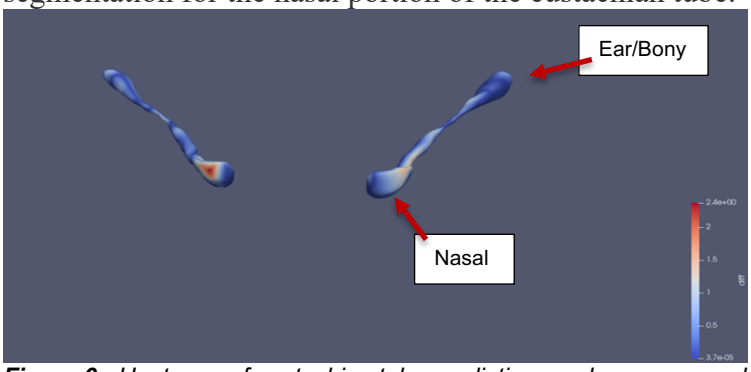

Figure 6 *- Heat map of eustachian tube prediction mesh as compared with the ground truth.*

B) Mean Hausdorff Distance

The mean surface distance, d_{mean}, is the distance between the the surface (S) and the reference surface (S_{ref}) where d(S, S_{ref}) is the mean of distances between every surface voxel in S and the closest surface voxel in S_{ref}, while d(S_{ref}, S) is computed in a similar way [8].

$$d_{mean} = \frac{1}{2}[\underline{d}(S, S_{ref}) + \underline{d}(S_{ref}, S)]$$

C) Weighted Hausdorff Distance

The finding from the heat map identified the nasal and ear end of the eustachian tube as most important in the treatment of ETD. Therefore, to develop a metric that would weigh different sections of the tube based on importance, we have begun research into the weighted hausdorff distance. The maximum Hausdorff distance (HD) is the maximum distance of a set to the nearest point in the other set. More formally, the maximum Hausdorff distance from set X to set Y is a max-min function. Weighted hausdorff distance (WHD) is similar to the maximum HD; however, it is based on the probability map for the region of interest [9]. Larger weight will be deemed more significant and vice versa. The purpose for using WHD is to make the clinician focus on the anatomical parts that they deem important within the clinical domain. Equations are shown in figure 7 below.

$$d_{\mathrm{WH}}(p, Y) = \frac{1}{\mathcal{S} + \epsilon} \sum_{x \in \Omega} p_x \min_{y \in Y} d(x, y) +$$
$$\frac{1}{|Y|} \sum_{y \in Y} \underset{x \in \Omega}{M_\alpha} \left[p_x d(x, y) + (1 - p_x) d_{max} \right]$$

where

$$\mathcal{S} = \sum_{x \in \Omega} p_x,$$

$$\underset{a \in A}{M_\alpha} [f(a)] = \left(\frac{1}{|A|} \sum_{a \in A} f^\alpha(a) \right)^{\frac{1}{\alpha}}$$

Figure 7 *- Equations pertaining to calculation of the Weighted Hausdorff Distance.*

4. RESULTS

i. nnUNet

In the nnU-Net experiment, we used 22 CT volumes for training (17 for training and 5 for validation) and 9 for testing. We tested multiple configurations of nnUNet (e.g., 2D U-Net, 3D-full resolution, 3D-low resolution, and 3D Cascade); however, we found that the 3D full resolution configuration was most optimal in predicting the ET both from a visualization and quantitative perspective. As a result, we will present the results in this report focusing on the 3D full resolution model. Furthermore, during training, it was observed that the validation loss and global evaluation metric converged around 50-75 epochs; therefore, the number of epochs was reduced from 1000 to 100 to improve efficiency. Finally, 5-fold cross-validation was performed during model training.

From the learning that we obtained from the background reading, we investigated whether adding more segmentations pertaining to structures surrounding the ET could increase the prediction accuracy. Therefore, we added the Torus Tubarius and Internal Carotid Artery (ICA) to the dataset (figure 8).

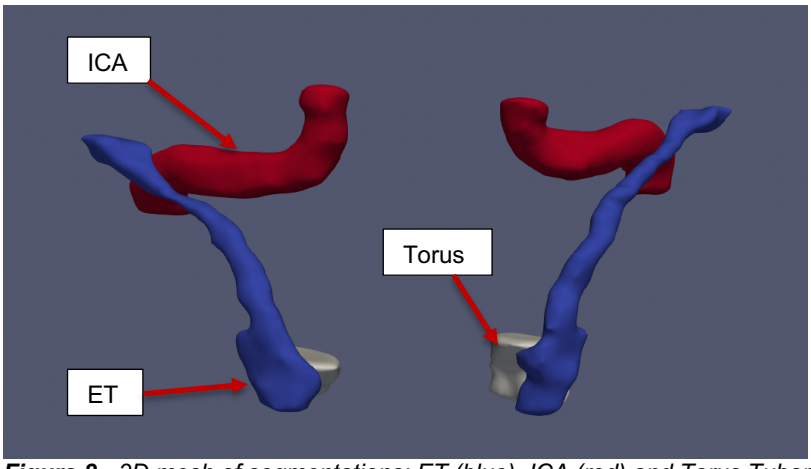

Figure 8 *- 3D mesh of segmentations: ET (blue), ICA (red) and Torus Tubarius (gray)*

The evaluation metric included the DSC and AHD for the ET only and ET with surrounding anatomical structures combination via two sets of loss functions 1) cross entropy (CE) with dice (DC) loss and 2) focal + DC loss (Table 1).

	CE + DC Loss		Focal + DC Loss	
	DSC	AHD	DSC	AHD
ET	0.529	0.613		
ET	0.556	0.601	0.649	0.454
ICA	0.855	0.278	0.891	0.205
Torus	0.731	0.389	0.735	0.402

Table 1 *- Performance Metrics for Two Dataset Configurations 1) ET only 2) ET + Surrounding Structures*

From the table above, when adding surrounding anatomical structures and utilizing the CE+DC loss, it was seen there was a marginal improvement in ET accuracy. There may be two possible reasons. Firstly, the number of training images is small. The model may need more voxel information of different ETs during training to make a good prediction on the test data. Second, the cross-entropy loss does not perform well when training on long, thin, linear structures such as the ET. Cross entropy aims to compute the probability from one probability distribution to another distribution. During training, the linear structure has a high probability to be divided into several small segments and recognized as the wrong structure resulting in poor predictions. As a result, to account for the suboptimal performance of cross-entropy loss, we assessed the use of the DC + focal loss function during training. Focal loss adds a smoothing term to cross-entropy to aid the model in focusing on the misclassified classes [10]. The respective DSC and AHD values can be seen in the second part of table 1 (Focal + DC loss).

Compared to figure 8, there is an improvement in the ET prediction from the bony and nasal aspect, which is expected according to the definition of focal loss. However, there were some predictions that were not continuous within the middle portion of the ET (figure 9). Reasons for this may include an insufficient training dataset or inaccurate ground truth segmentations for the 'middle' segment of the ET given the difficulty to visualize this portion on CT images.

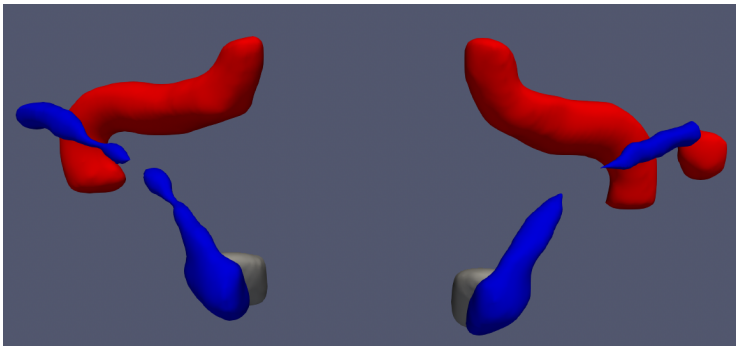

Figure 9 *- 3D mesh of segmentations: ET (blue), ICA (red) and Torus(gray)*

Weighted Hausdorff Distance

Currently, we have developed a method to generate a probability map for the ET. According to the number of nodes in the skeleton, we split them into three segments with the ratio provided by the user and then assign a probability to each section. Then, we compute the closest distance from the mesh to the skeleton to give the probability back to each vertex in the original segmentation. A preliminary sample is shown in figure 10 where the nasal, middle, and ear end of the ET is weighted differently.

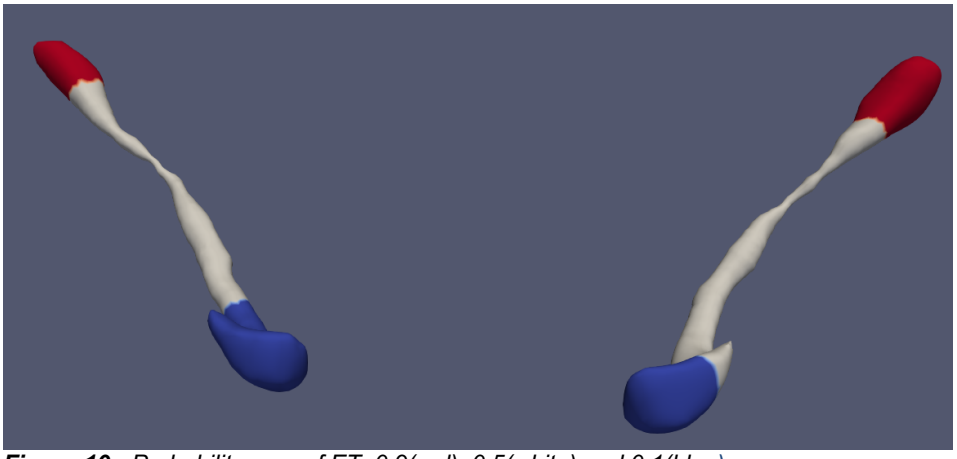

Figure 10 *- Probability map of ET: 0.9(red), 0.5(white) and 0.1(blue)*

Due to the time constraints, the final step of implementing WHD is still in progress. In future consultation with our clinical mentor, we will finalize and apply the probability map to the WHD equation and then compute the hausdorff distance between the prediction and ground truth to help users check the performance of the model.

ii. VoxelMorph

To date, we have developed a script for performing unsupervised registration-label propagation in VoxelMorph. For this task, 19 images were used for training, 5 for validation, and 4 for testing.

Data Preprocessing

The CT scans are in NIfTI format with dimensions of (512, 512, z) where z varies from 180 to 310. First, to facilitate training via convolutional neural networks, all pixel values have been normalized to be in the 0-1 range. Second, the z dimension was cropped to 256 if $z > 256$ and padded to 256 if $z < 256$ to be compatible with U-Net's down-sampling process. Third, the image was resized to (256,256,128) using spline interpolation with an order of 0 to avoid exceeding the memory limit of the GPU. Lastly, each image was converted to a numpy array from the original NIfTI format to meet the training requirements of VoxelMorph (figure 11).

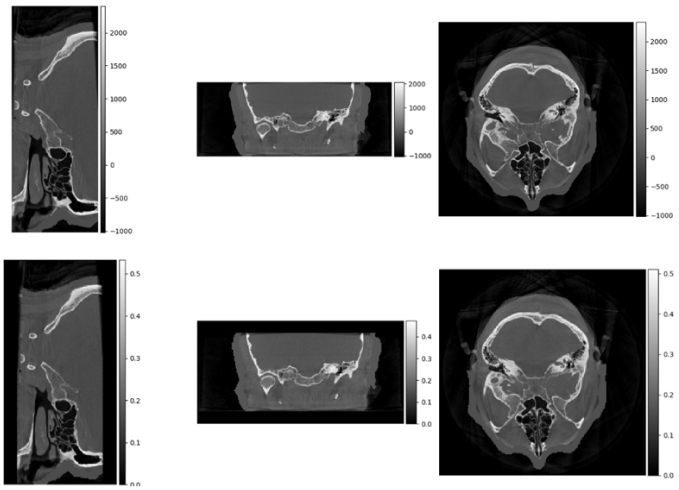

Figure 11 *- CT volume after normalization, padding / cropping, and down-sampling with 0 order spline interpolation*

Network Training

The training, network architecture, and loss parameters utilized for the unsupervised Voxelmorph training are shown below. Training duration was approximately 10 hours at 6 minutes per epoch.

Training		*Network Architecture*		*Loss Parameters*	
batch size	2	List of U-Net encoder filter	[16, 32, 32, 32]	image reconstruction loss	MSE
epochs	100	List of U-Net decoder filter	[32, 32, 32, 32, 32, 16, 16]	weight of gradient loss (lambda)	0.01
steps per epoch	171	Number of integration steps	7	image noise parameter (sigma)	1
learning rate	0.0001	Flow downsample factor	2		

Table 2 *- Training, Network Architecture, and Loss Parameters for 'Unsupervised' VoxelMorph*

Evaluation

Below, registrations results are visualized for 5 test pairs in the sequence of moving, fixed, and moved images. Qualitatively, the warped images (moved) do not look similar to their corresponding fixed images. Ideally, a good registration result would result in near identical moved and fixed images. From looking at the deformation fields flow, which is visualized as RGB in each direction, the first three pairs have localized reasonably well to where the deformation occurs (figure 12). However, for the last two pairs, deformation occurs where there should be little to no deformations.

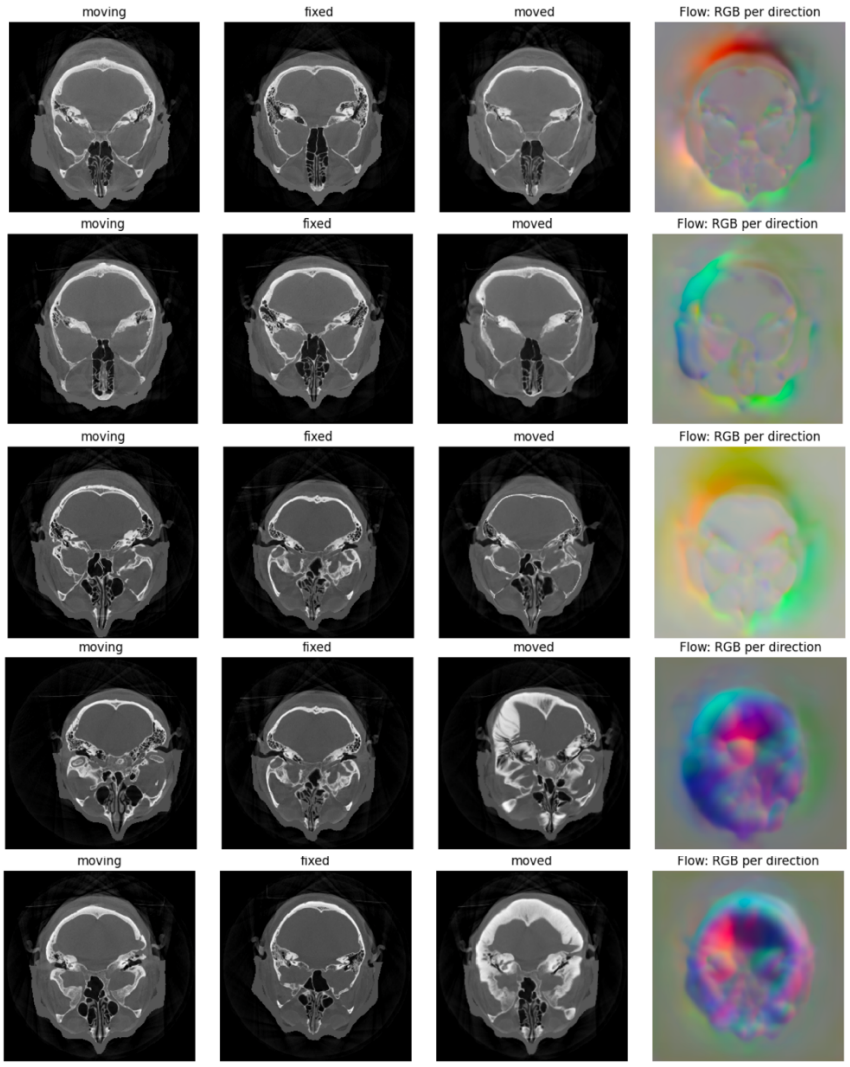

Figure 12 *- Registration Results for 5-paired images (Sequence: Moving, Fixed, Moved, RGB Deformation Field)*

For the quantitative evaluation, all combination pairs of test images were warped and the average dice score for the propagated ET label was computed as shown below.

	Average Dice Score (Background + ET)	Average Dice ET	Highest Dice ET
DSC	0.5050	0.0100	0.2247

Given the ET dice results and visualizations of flow, it is very likely that the trained model is underfitting or not able to reach convergence due to both the complex structures of CT scans and high levels of deformity, with current training parameters. Since the registration results were poor, it is expected that the label propagation is inaccurate as shown through the average dice ET of 0.01.

Overall, ET is a very small structure on a CT scan. On average, about ~700 / 8,400,000 pixels are labeled 1 as ET where the rest of the pixel is labeled 0 as background. This amounts to less than 0.01% of the total volume that ET takes up on the CT volume, which can impose difficulties

as VoxelMorph's unsupervised algorithm gives equal weights to every pixel (i.e., both ET and background) during its optimization. However, there are many steps we will take in the future to try to improve the registration process which include fine-tuning hyperparameters, template creation to lower the learning complexity, and using a semi-supervised approach where the ground-truth labels guide the training to focus on a region of interest.

5. SIGNIFICANCE

This deep learning framework provides an automated segmentation of the eustachian tube with surrounding anatomical structures. This has set the foundations for developing a more thorough understanding of the eustachian tube and has the potential to transform the clinical care for patients with eustachian tube dysfunction. Furthermore, this framework may be integrated into a CT-based registration system for performing eustachian tube dilation and ultimately reducing potential associated complications.

6. CONCLUSION

We have developed the first deep learning platform for automated segmentation of the eustachian tube via CT images using the nnU-Net configuration. This framework avoids the need for manual annotation of the eustachian tube and its surrounding structures. Although the torus tubarius and internal carotid artery demonstrated high performance metrics, there is room for improvement for the eustachian tube. This is largely due to the difficulty in visualizing the entire tract of the ET on CT images speaking to the limitation of this imaging modality. Future work will integrate MRI and label propagation techniques to enhance the reliability of the ground truth prior to training on nnUNet. Furthermore, we will explore active shape modeling methodologies to improve the segmentation of the middle portion of the eustachian tube.

7. MANAGEMENT SUMMARY

The team met with Dr. Manish Sahu on a weekly basis on Monday afternoons to review project updates and devise future tasks. Additionally, we participated in the weekly LCSR meetings on Wednesday mornings with Dr. Russell Taylor, Dr. Francis Creighton, and Dr. Mathias Unberath to present our weekly progress.

7.1 Team member contribution

During the project, all team members were involved with all aspects of the project. However, the tasks were divided according to the following to meet the deliverables by the end of the semester:

Tasks	Assigned Member
Obtain publicly available images until IRB access granted	Ameen
Manual segmentation of the eustachian tube and surrounding structures	Ameen

Registration script via ANTs	Yuliang and Ameen
nnUNet model	Yuliang, Chanha, and Ameen
Validation scripts (DSC + HD95)	Yuliang
Trained VoxelMorph Model	Chanha
Ongoing documentation	Ameen, Chanha, and Yuliang

7.2 Deliverables

Minimum	**Planned Date**	**Completion Date**
CT co-registration script	February 25	February 25
Trained nnUNet model	March 11	March 11
Dataset containing ground truth segmentations	March 25	April 4
Documentation	May 17	May 17
Final report	May 1	May 1
Expected		
Validation script computing dice score and Hausdorff distance on predicted labels	March 25	March 25
Heat map script	March 21	March 21
Maximum		
Weighted Hausdorff Distance script	April 30	Ongoing
Trained VoxelMorph image registration models	April 15	Ongoing
Validation script comparing nnUNet model with VoxelMorph image registration models	April 30	Ongoing
Conference presentation and manuscript draft	July 1	Ongoing

To date, we have met the minimum and expected deliverables pertaining to the project. There was a delay with obtaining our imaging dataset for which we resolved by developing a de-facing script. With regards to the maximum deliverables, we have begun with the unsupervised registration script with VoxelMorph; however, it needs to be further optimized by modifying hyperparameters. Therefore, we are going to continue to work on the maximum deliverables during the summer. The weighted hausdorff distance script has been completed; however, we need to define the probability map which will be done in consultation with Dr. Creighton.

7.3 Lessons Learned

The eustachian tube is a complex anatomical structure that has both a cartilaginous and bony component. The former part appears to be routinely collapsed and its path is difficult to decipher on CT images. Therefore, this presented a challenge when performing manual annotation. We have learned several takeaways from this. First, we would like to obtain a set of images where patients have been asked to perform a Valsalva maneuver which naturally opens the eustachian tube and allows the cartilaginous segment to be visualized. Second, soft tissue structures are better captured on MRI images. Therefore, we are currently working to obtain a set of MRIs for patients with corresponding CTs; we will then manually annotate the MRI images and using registration-label propagation techniques, transfer the eustachian tube label from the MRI to the CT. This will ultimately enhance the reliability of the ground truth and may result in improved performance from the nnUNet predictions.

7.4 Future Steps

Future work will incorporate MRI images and utilize registration-label propagation techniques to further enhance the reliability of the ground truth segmentations. We are currently exploring learning-based registration frameworks such as VoxelMorph for completion of this task. Additionally, we have also curated an imaging database containing segmentations of clinical structures which can be used for education purposes or integrated into image-guided procedures such as for eustachian tube dilation. Finally, given that one of the risks of dilation procedures includes injury to the ICA, we plan to develop an automated measurement extraction pipeline that will delineate the distance between the ICA and ET. Ultimately, this can be used in the preoperative evaluation of patients with ETD.

8. ACKNOWLEDGEMENTS

We would like to thank the mentors of this project Dr. Francis Creighton, Dr. Manish Sahu, Dr. Mathias Unberath, and Dr. Russell Taylor for their invaluable guidance throughout the course of this project. We would also like to thank the LCSR and ARCADE lab for providing remote GPU access for the deep learning training.

9. REFERENCES

[1] I. Magro, D. Pastel, J. Hilton, M. Miller, J. Saunders, and K. Noonan, "Developmental Anatomy of the Eustachian Tube: Implications for Balloon Dilation," *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.*, vol. 165, no. 6, pp. 862–867, Dec. 2021, doi: 10.1177/0194599821994817.

[2] M. H. Froehlich, P. T. Le, S. A. Nguyen, T. R. McRackan, H. G. Rizk, and T. A. Meyer, "Eustachian Tube Balloon Dilation: A Systematic Review and Meta-analysis of Treatment Outcomes," *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.*, vol. 163, no. 5, pp. 870–882, Nov. 2020, doi: 10.1177/0194599820924322.

[3] D. Angeletti *et al.*, "Chronic obstructive Eustachian tube dysfunction: CT assessment with Valsalva maneuver and ETS-7 score," *PloS One*, vol. 16, no. 3, p. e0247708, 2021, doi: 10.1371/journal.pone.0247708.

[4] N. J. Tustison *et al.*, "Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements," *NeuroImage*, vol. 99, pp. 166–179, Oct. 2014, doi: 10.1016/j.neuroimage.2014.05.044.

[5] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019, doi: 10.1109/TMI.2019.2897538.

[7] K. H. Zou *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004, doi: 10.1016/s1076-6332(03)00671-8.

[8] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proceedings of 12th International Conference on Pattern Recognition*, Oct. 1994, vol. 1, pp. 566–568 vol.1. doi: 10.1109/ICPR.1994.576361.

[9] J. Ribera, D. Güera, Y. Chen, and E. J. Delp, "Locating Objects Without Bounding Boxes," *ArXiv180607564 Cs*, Apr. 2019, Accessed: Apr. 30, 2022. [Online]. Available: http://arxiv.org/abs/1806.07564

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.

10. TECHNICAL APPENDICES

Code - The Python code with respective documentation can be referenced in the GitHub Repository which has been made available to the mentors and members of this project via https://github.com/mikami520/CIS2-EustachianTube.git.

Documentation - The documentation pertaining to the project goals, milestones, and deliverables can be referenced in the Course Wiki via https://ciis.lcsr.jhu.edu/doku.php?id=courses:456:2022:projects:456-2022-03:project-03.