

Electronic Thesis and Dissertation Repository

10-9-2019 1:30 PM

Automated Segmentation of Temporal Bone Structures

Daniel Allen, *The University of Western Ontario*

Supervisor: Ladak, Hanif M., *The University of Western Ontario*

Co-Supervisor: Agrawal, Sumit K., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Engineering
Science degree in Electrical and Computer Engineering

© Daniel Allen 2019

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biomedical Commons](#)

Recommended Citation

Allen, Daniel, "Automated Segmentation of Temporal Bone Structures" (2019). *Electronic Thesis and Dissertation Repository*. 6605.

<https://ir.lib.uwo.ca/etd/6605>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Mastoidectomy is a challenging surgical procedure that is difficult to perform and practice. As supplementation to current training techniques, surgical simulators have been developed with the ability to visualize and operate on temporal bone anatomy. Medical image segmentation is done to create three-dimensional models of anatomical structures for simulation. Manual segmentation is an accurate but time-consuming process that requires an expert to label each structure on images. An automatic method for segmentation would allow for more practical model creation. The objective of this work was to create an automated segmentation algorithm for structures of the temporal bone relevant to mastoidectomy. The first method explored was multi-atlas based segmentation of the sigmoid sinus which produced accurate and consistent results. In order to segment other structures and improve robustness and accuracy, two convolutional neural networks were compared. The convolutional neural network implementation produced results that were more accurate than previously published work.

Keywords

Automatic Image Segmentation, Deep Learning, Convolutional Neural Networks, Temporal Bone Anatomy, Atlas-Based Segmentation, Image Registration

Summary for Lay Audience

Surgeries in the area of the ear can be difficult to train and practice. There are many small important structures to be considered and there is a lot of variation between patients. If mistakes are made during the surgery, it can cause severe damage to the patient. By using a surgical simulator, surgical trainees can improve their skills before operating on real patients at a much lower cost than when using classical training methods. To create a surgical simulator, anatomical structures need to be labeled from images so that 3D models can be made. This is called image segmentation and can be done manually or automatically. Manual labelling is very accurate but takes a long time and requires an expert to do it. Automatic labelling is much easier and faster to do in a clinical setting. However, many parts of the anatomy that need to be labeled are small, variable in position and shape, and have low contrast edges (hard to distinguish from surrounding objects). These issues make automating the labelling of the structures very difficult. This work compares multiple methods for automatic image labelling. The first method tested and developed was done on the sigmoid sinus, a vein that passes near the ear. A set of high-resolution manually labelled examples of the vein were used and transformed to match the new lower resolution images to be labelled, and then these sets were combined. This method resulted in labels that were similar to the actual labels. The second method was done on several anatomic structures of the ear and used deep learning networks to learn patterns in the images and label them automatically. This method quickly and successfully created automatic labels from images that were also very close to the actual labels and showed better results than previous work on the same structures. This labelling method may be used to create 3D models for surgical simulators.

Co-Authorship Statement

This master's thesis is an integration of two articles, each integrated as a separate chapter. The first article, Chapter 2, has been published by the Biomedical Journal of Science & Technical Research (BJSTR) and the second article, Chapter 3, will be submitted to the International Journal of Computer Assisted Radiology and Surgery (IJCARS).

Chapter 2: D. G. Allen, K. Van Osch, B. Gare, S. K. Agrawal, H. M. Ladak, “Automated Segmentation of the Sigmoid Sinus using a Multi-Atlas Approach,” *Biomed. J. Sci. Tech. Res.*, vol. 20, no. 4, pp. 15313–15319, 2019.

The motivation for this study was presented by the primary supervisors S.K. Agrawal and H.M. Ladak. My contribution to this study was the development of the algorithm, collection and interpretation of the results, and the preparation of the manuscript, which was edited by S.K. Agrawal and H.M. Ladak and reviewed by all authors. K. Van Osch created the manual segmentations of the sigmoid sinus used to form the ground truth and atlases. B. Gare was consulted for data analysis.

Chapter 3: D. G. Allen, K. Van Osch, H. M. Ladak, S. K. Agrawal, “Automated Segmentation of Temporal Bone Structures using Convolutional Neural Networks”. *International Journal of Computer Assisted Radiology and Surgery*. Submitting November 2019.

The motivation for this study was presented by the primary supervisors H.M. Ladak and S.K. Agrawal. My role in this study was development of the segmentation algorithm, collection and interpretation of results, and the preparation of the manuscript. K. Van Osch and I created the manual segmentations from the images used in the development.

Acknowledgments

I would like to thank my supervisors without whom this work could not have happened. Prof. Hanif Ladak is everything and more than what you could ask for from a supervisor. He has kept not only me, but a very productive research group on track to success. He always keeps his students' best interests in mind and is always willing to provide advice. Dr. Sumit Agrawal shows that it is somehow possible to be a world-class otologist with a busy practice while maintaining hands-on knowledge of state-of-the-art technology and medical image processing techniques. He was instrumental in the inspiration for this work and in the entire development process start to finish.

I would also like to thank the past and present members of the Auditory Biophysics Lab. John Iyaniwura, Robert Koch, Carlos Salgado, Brad Gare and Dr. Mai Elfarnawany who were all former members of the lab during my time there who all made me feel welcome, provided advice and paved the way with their work for myself and the rest of the lab. The current members: Luke Helpard, Dr. Soodeh Nikan, and Dr. Alireza Rohani have been fantastic colleagues and friends who have given me guidance, assistance, and memorable experiences that will last a lifetime. Special thanks to Lauren Siegel for proofreading manuscripts.

Finally, I would like to thank all my friends and family. Thank you for being so patient and supportive throughout the years. Having a strong network of people around me has meant so much to the enjoyment of this chapter of my life and the completion of this work.

Table of Contents

Abstract.....	ii
Summary for Lay Audience.....	iii
Co-Authorship Statement.....	iv
Acknowledgments.....	v
Table of Contents.....	vi
List of Tables	ix
List of Figures.....	x
Chapter 1.....	1
1 Introduction.....	1
1.1 Anatomy.....	2
1.1.1 Sigmoid Sinus	4
1.2 Mastoidectomy.....	6
1.3 Simulation.....	7
1.4 Medical Image Segmentation	8
1.4.1 Image Registration	10
1.4.2 Atlas-based Methods.....	11
1.4.3 Convolutional Neural Networks	13
1.4.4 Objective Functions	16
1.4.5 Optimizers.....	19
1.4.6 Post-Processing Methods	20
1.4.7 Segmentation Evaluation	21
1.5 Objectives	24
1.6 Novelty.....	25
References.....	26

Chapter 2.....	31
2 Automated Segmentation of the Sigmoid Sinus using a Multi-Atlas Approach	31
2.1 Introduction.....	31
2.2 Materials and Methods.....	33
2.2.1 Data.....	33
2.2.2 Ground Truth Segmentations and Atlas Creation.....	34
2.2.3 Segmentation Algorithm.....	35
2.3 Evaluation and Metrics	40
2.4 Results.....	42
2.5 Discussion.....	45
2.6 Acknowledgments.....	47
References	48
Chapter 3.....	50
3 Automated Segmentation of Temporal Bones Structures using Convolutional Neural Networks	50
3.1 Introduction.....	50
3.2 Materials and Methods.....	51
3.2.1 Image Acquisition.....	51
3.2.2 Ground Truth Segmentations.....	52
3.2.3 Datasets for CNN Training, Validation, and Testing	52
3.2.4 Convolutional Neural Networks	53
3.2.5 Post-Processing	54
3.2.6 Evaluation and Metrics	55
3.3 Results.....	56
3.4 Discussion	59
3.5 Acknowledgements.....	61

References	62
Chapter 4.....	64
4 Conclusions and Future Directions	64
4.1 Conclusions.....	64
4.2 Future Directions	65
References	67
Curriculum Vitae	68

List of Tables

Table 2.1: Mean values of all metrics found from single atlas and multi-atlas approaches. Standard deviations given in parenthesis.....	42
Table 3.1: Quantitative metrics for DeepMedic segmentation results. Means are given with standard deviations in parentheses.....	56
Table 3.2: Quantitative metrics for DenseVNet segmentation results. Means are given with standard deviations in parentheses.....	58

List of Figures

Figure 1.1: Diagram of the parts of the temporal bone displaying the squamous (yellow), tympanic (green), and the combined petrous and mastoid (brown) parts. Image courtesy of Stanford Otolaryngology, Head and Neck Surgery [7].	2
Figure 1.2: Coronal (frontal) cross section view of the temporal bone and ear structures. Image courtesy of Stanford Otolaryngology, Head and Neck Surgery [7].....	4
Figure 1.3: Diagram of sigmoid sinus anatomy showing the sigmoid sinus and its relation to connected venous sinuses [9]. Copyright 2020, with permission from Elsevier.	5
Figure 1.4: Micro-CT scan of temporal bone showing the sigmoid sinus (indicated with arrow).....	6
Figure 1.5: Labelmap of temporal bone structures overlaid on a micro-CT image. The structures shown are the sigmoid sinus (blue), facial nerve (yellow), external auditory canal (green), tegmen (white), cochlea and semicircular canals (brown), malleus (magenta), incus (red), and the stapes (orange).	8
Figure 1.6: 3D rendering demonstrating an image transform being applied to an atlas. ..	11
Figure 1.7: DeepMedic 3D CNN architecture with two convolutional pathways followed by a fully connected conditional random field [39]. Licensed under CC BY 4.0.	15
Figure 1.8: DenseVNet network architecture [40]. © 2017 IEEE.	16
Figure 1.9: Venn diagram showing overlap between sets A and B.	22
Figure 2.1: Clinical-CT (left) compared to Micro-CT (right) with sigmoid sinus shown in blue.....	35
Figure 2.2: Block diagram displaying the multi-atlas segmentation algorithm. User specifies either left or right ear and appropriate atlases are chosen by algorithm.	36

Figure 2.3: Axial images of the temporal bone showing target clinical CT (magenta) and the μ CT atlas (green). The unregistered images are shown overlaid without registration (left) and the result of a complete registration (right). Observe the shift of the position and shape of the sigmoid sinus. 38

Figure 2.4: Screenshot from 3D Slicer showing an example automated segmentation in three slice views (a axial, b sagittal, c coronal) as well as in a 3D rendered temporal bone segment. The automated segmentation and 3D model of the sigmoid sinus are shown in blue against the grayscale clinical-CT volume from which it was segmented. 43

Figure 2.5: Colormaps projected on automated segmentation 3D models showing absolute distance in millimeters between the algorithm segmentations and the ground truth sigmoid sinus. 44

Figure 3.1: Screenshot from 3D Slicer of DeepMedic segmentation of a sample in three slice views (a axial, b sagittal, c coronal) and the structure segmentations rendered in 3D in the top right panel. 57

Figure 3.2: Screenshot from 3D Slicer of DenseVNet segmentation of a sample in three slice views (a axial, b sagittal, c coronal) and the structure segmentations rendered in 3D in the top right panel. 59

Chapter 1

1 Introduction

Hearing loss is a common and disabling health issue that can cause social-isolation, safety concerns, and many other potential consequences. The World Health Organization (WHO) estimates that there are approximately 466 million people worldwide with disabling hearing loss [1] and a Canadian survey on disability conducted in 2012 found that 3.2% of Canadians aged 15 years and older reported having a hearing disability [2], [3]. While hearing aids can treat a large portion of patients with hearing loss, they cannot be used in all patients. In these cases, other solutions such as cochlear implants or surgical interventions are needed to restore hearing.

Mastoidectomy is a challenging surgical procedure in which the air cells of the mastoid located in the temporal bone are removed using a high-speed surgical drill.

Mastoidectomy often precedes cochlear implantation or other procedures such as cholesteatoma removal [4]. Mastoidectomy is associated with many possible complications where permanent damage could occur such as facial nerve paralysis or balance problems [5], [6]. Therefore, it is important that surgical residents are well trained and able to sufficiently practice and be tested on their ability to perform the operation, as well as be knowledgeable of the temporal bone anatomy. Traditional training for mastoidectomy is done on cadavers, however this is expensive, and it is difficult to continually supply cadavers to labs. To address these problems, training may be supplemented through the use of surgical simulators. A mastoidectomy simulator allows for repeated practice on patient-specific models and a variety of cases. In order to

create visualizations of anatomical structures for a surgical simulator, it is necessary to delineate them in a process called medical image segmentation. However, it is impractical to create these segmentations manually for each case as it can take over an hour of expert time. Therefore, developing an accurate, automatic process for image segmentation is critical to being able to perform simulated mastoidectomies on patient specific anatomy in a clinical setting.

1.1 Anatomy

The temporal bones are located on either side of the head and contains the structures of the ear. The temporal bone consists of four major components shown from the interior of the skull in Figure 1.1: the petrous, squamous, tympanic, and mastoid parts.

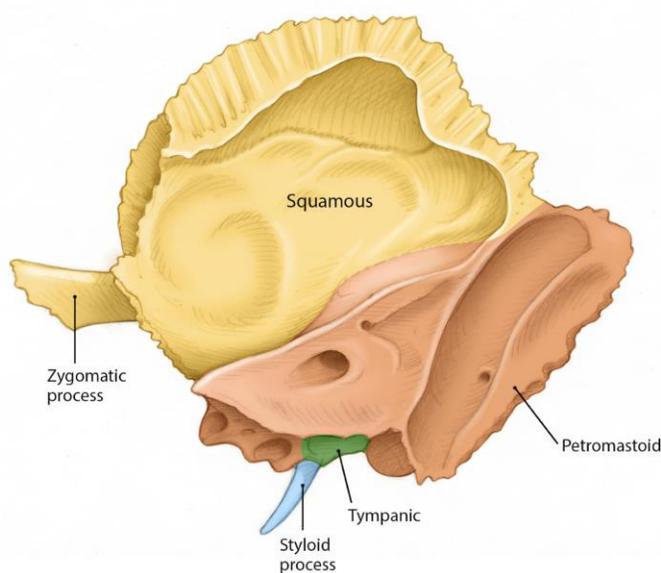


Figure 1.1: Diagram of the parts of the temporal bone displaying the squamous (yellow), tympanic (green), and the combined petrous and mastoid (brown) parts.

Image courtesy of Stanford Otolaryngology, Head and Neck Surgery [7].

The petrous part is pyramidal in shape and located inside the skull and contains the structures of the inner ear. The squamous part is the mostly thin, superior part of the bone which covers a large surface of the skull. The lower portion of the squamous contains the zygomatic process jutting laterally and connecting to the zygomatic bone. The tympanic part is found below the squamous and in front of the mastoid part. The tympanic part surrounds the external auditory canal. The mastoid part is found at the posterior end of the temporal bone and contains a multitude of air cells which are drilled out during mastoidectomy.

Figure 1.2 shows a coronal (frontal) view of the numerous structures within the temporal bone and the ear which must be considered during a mastoidectomy. Due to the close proximity, importance, and delicacy of the structures, an otologist (ear surgeon) performing a mastoidectomy must be aware of their precise locations and have a preoperative plan to avoid potentially severe complications.

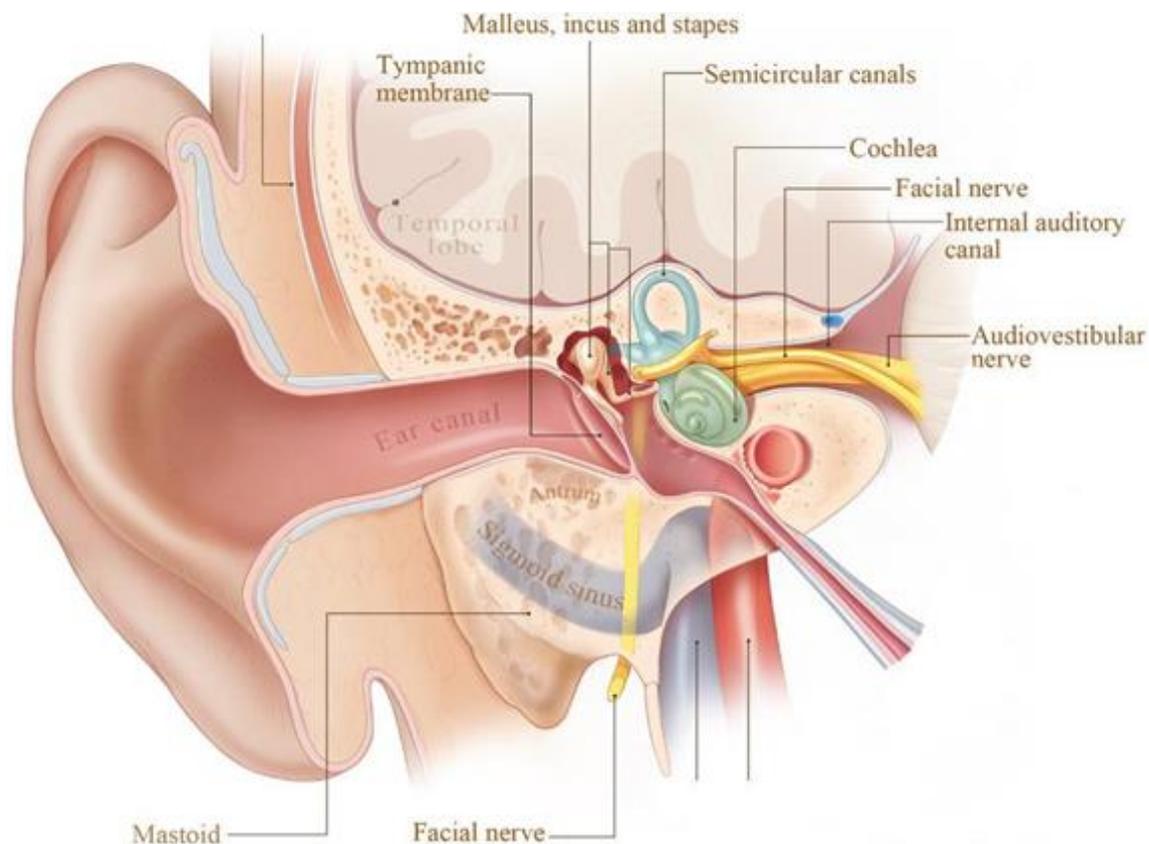


Figure 1.2: Coronal (frontal) cross section view of the temporal bone and ear structures. Image courtesy of Stanford Otolaryngology, Head and Neck Surgery [7].

1.1.1 Sigmoid Sinus

The sigmoid sinus (pictured in Figure 1.3) is a venous sinus that runs in an S-shaped groove through the temporal bone and is an important consideration in mastoidectomy. During the initial portion of the procedure the sigmoid sinus indicates the posterior boundary of where bone should be removed [4], [8]. An otologist must be able to locate and identify the sigmoid sinus without causing damage to it in order to avoid potential vascular complications and bleeding during the surgery.

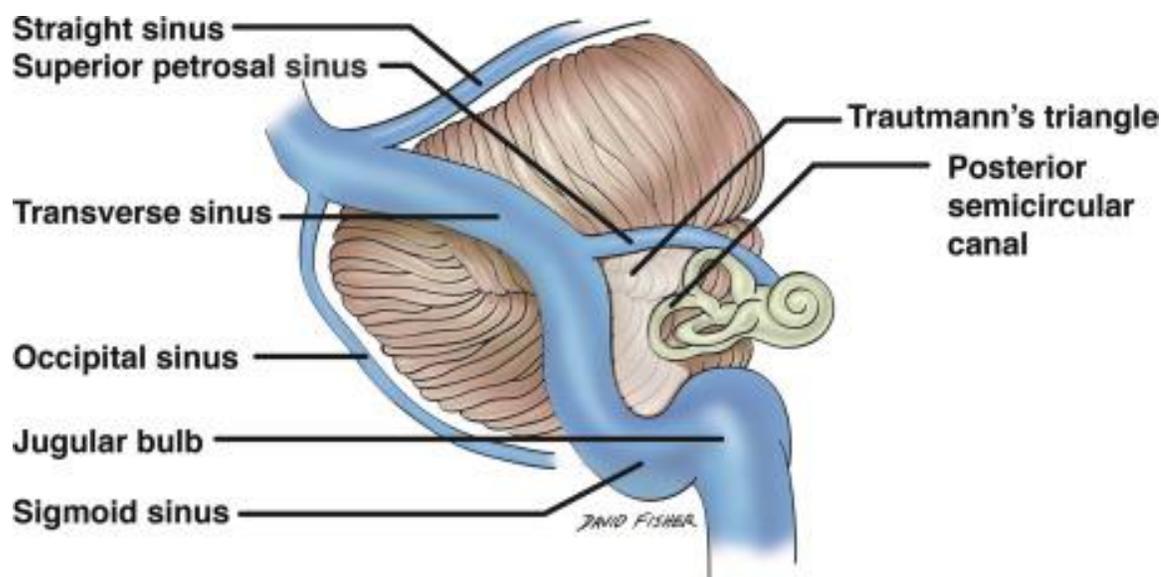


Figure 1.3: Diagram of sigmoid sinus anatomy showing the sigmoid sinus and its relation to connected venous sinuses [9]. Copyright 2020, with permission from Elsevier.

The sigmoid sinus is highly variable in terms of both location relative to other structures in the temporal bone as well as its shape, especially when considering that patients will have dominant and non-dominant (determined by relative size) sides [10], [11]. Most of the variation occurs in the transverse sinus where it connects to the sigmoid sinus [10] which can make this area especially difficult to segment. This means that any attempts at automating the segmentation of the sigmoid sinus should use development methods and datasets that cover as much of the variability as possible and include the ability to segment dominant and non-dominant sigmoid sinuses. Other than variability, the sigmoid sinus also presents challenges for automated segmentation in the form of poor contrast boundaries between the vessel and the surrounding areas on its medial side. This can be seen in the example CT scan of a temporal bone shown in Figure 1.4.

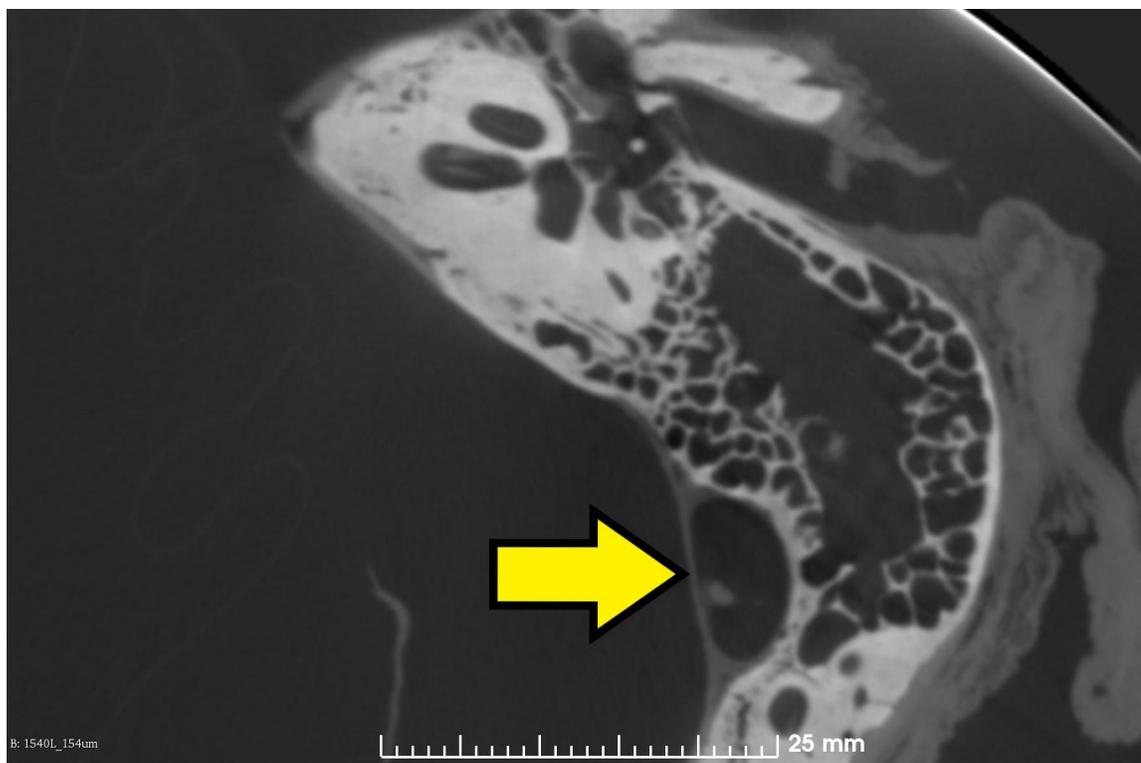


Figure 1.4: Micro-CT scan of temporal bone showing the sigmoid sinus (indicated with arrow).

1.2 Mastoidectomy

Mastoidectomy is a surgical procedure where the air cells in the mastoid are drilled away from the patient's temporal bone. The mastoid is found behind the ear where an incision can be made, and drilling is then performed. While drilling, it is important to correctly locate and identify the structural landmarks without damaging them to avoid complications. Mastoidectomy is performed for a variety of reasons including treatment of cholesteatoma, clearing out infected air cells, gaining access for the insertion of cochlear implants, among others. After a mastoidectomy is completed the facial recess can be drilled out to gain access to the round window for cochlear implant insertion.

Completed mastoidectomies and facial recesses can be evaluated by using the Welling scale which provides a score to assess completed temporal bone dissections [12], [13].

1.3 Simulation

Surgical simulation is a growing field in medical training. Many surgical simulators are being developed with 3D visualization and haptic (touch) feedback to improve the immersion, intuitiveness, and translatability of the experience to real-life situations [14]–[22]. While surgical simulators do not replace traditional training techniques, they do provide a helpful supplement to practice difficult procedures. Surgical simulation has the benefit of allowing trainees to practice on real anatomy while not introducing any risk to live patients or depleting expensive and difficult to procure cadavers.

Furthermore, with medical imaging technology it is possible to produce a scan of a patient's specific anatomy and import it for visualization in a surgical simulator. This provides the opportunity to practice on patient-specific cases repeated times for either training or preoperative planning, which is not possible with other methods. This also extends to the ability to simulate irregular temporal bone anatomy examples which may be difficult to obtain in cadaver labs.

Mastoidectomy simulators also provide the opportunity to assess a user's dissection with performance metrics and to provide the trainee with feedback. Applying the Welling scale to score a completed mastoidectomy is a quantifiable way of assessing a user's dissection and could be incorporated into a surgical simulator. However, medical image segmentation has to be done for each new case to import patient CT scans into a simulator and provide 3D models for visualization and assessment.

1.4 Medical Image Segmentation

Medical image segmentation is required to create models of anatomy for use in a surgical simulator. Image segmentation is the process of delineating objects or structures from an image [23]. When applied to medical image volumes this means creating labelmaps which paint the volumetric pixels (voxels) of each separate 3D structure with individual integer values. An example axial slice of a labelmap of several segmented temporal bone structures can be seen overlaid on a micro-CT image in Figure 1.5.

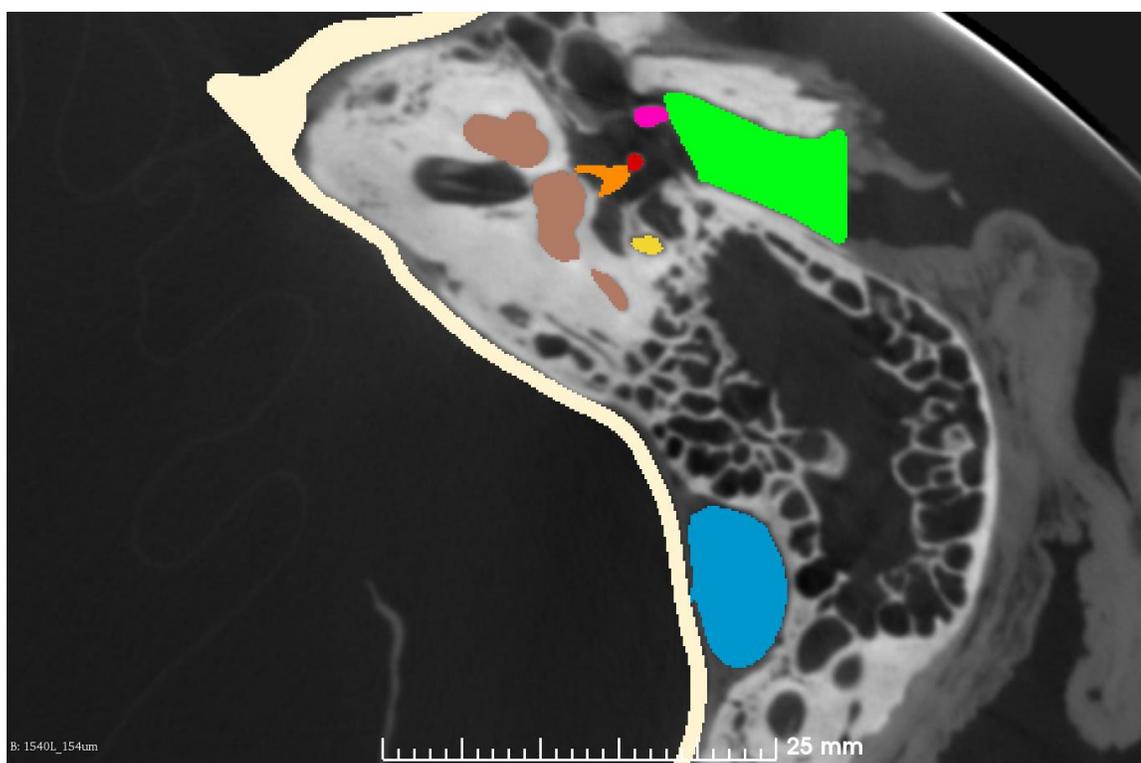


Figure 1.5: Labelmap of temporal bone structures overlaid on a micro-CT image. The structures shown are the sigmoid sinus (blue), facial nerve (yellow), external auditory canal (green), tegmen (white), cochlea and semicircular canals (brown), malleus (magenta), incus (red), and the stapes (orange).

There are three general approaches to the process of segmenting an image: manual, semi-automatic, and automatic. Manual segmentation is the process of painting the voxel of each structure in every slice it appears in from the medical image volume. This is a very time-consuming process but produces the most accurate resulting segmentations when performed by an expert. Semi-automated methods use reduced user inputs such as painting only intermittent sections of the target structure or placing identifier points on landmarks for the algorithm to use for generating a full segmentation. Examples of semi-automated methods include filling between slices [24] and fast grow cut [25]. Automatic methods seek to create segmentations with no user input. For the purpose of creating patient specific cases for surgical simulation, accurate automatic segmentations are optimal for ease of use and feasibility of implementation. However, many anatomical structures are challenging to segment automatically due to low contrast, patient variation, ambiguous boundaries, differing imaging protocols, etc. This means the process of selecting and optimizing an automatic method for segmentation must be done differently for each individual scenario and area of anatomy. While no semi-automated or automated methods for segmentation immediately produce segmentations of the same degree of accuracy as manual segmentation, they can be reviewed and edited after the algorithm is finished to reach the same level of accuracy. This is required for applications with very low error tolerance such as robotic surgery.

Given the challenges of segmentation described above, two knowledge-based approaches appear to dominate recent literature and are selected for further description here: atlas-based segmentation and convolutional neural networks (CNNs) [26], [27]. Both are

described next after a brief description of image registration which is particularly important for the former.

1.4.1 Image Registration

When working with medical images it is often necessary to align image volumes together. In order to perform atlas-based segmentation it is necessary to register the reference atlas image to the target image to be segmented. The transforms that result from these registrations are then applied to the reference atlas labelmap to transfer the labels to the target image and thus segment its components. The two categories of transforms that registrations can produce are linear transforms, including rigid and affine, which preserve points, straight lines, and planes; and non-rigid which can deform images.

1.4.1.1 Linear Registration

Rigid registration attempts to align two images by applying only translation and rotation operations. This is useful for aligning two identical images that are offset from each other. Affine registration expands on rigid registration by adding scaling and shear operations to the transformation. Linear registrations can act as an initial step to closely align two different images before non-rigid registration.

1.4.1.2 Non-rigid Registration

When aligning two different images from different samples such as in the atlas-based segmentation method it is common to follow a linear registration with a non-rigid registration. Non-rigid registrations are able to deform images locally to fit a target image's geometry. One effective example of non-rigid registration for medical images is deformation with cubic B-splines [28], [29]. Cubic B-spline deformation is applied by

optimizing the positions of a 3D grid of control points which are applied to the image being registered. Non-rigid registrations require much more computational resources than linear registrations and can create poor quality registrations when too much deformation is applied. A method for managing these issues in a cubic B-spline registration is by first applying an affine linear registration and then using a pyramidal approach of progressively smaller grid spacing between control points [28].

1.4.2 Atlas-based Methods

Atlas-based methods for medical image segmentation use labelled reference images to provide prior knowledge to make new segmentations. One or more known examples of the target anatomy are manually labelled and used as an atlas set which is registered to a new target medical image. Once the registration step is complete the image transform that corresponds to the movement of the registered atlas image is applied to the atlas labelmap. This transformed labelmap becomes the atlas segmentation for the target image. An example of this being applied to the 3D rendering of an atlas labelmap is shown in Figure 1.6.



Figure 1.6: 3D rendering demonstrating an image transform being applied to an atlas.

Often the target structure has variable anatomy and one atlas does not provide the ability to cover the full scope of variation. In this case, multiple atlases can be used in conjunction to create a segmentation for the target image. When using a set comprised of multiple atlases for segmentation there are two general methods to use the information from them for the final segmentation. The first method is atlas selection where different approaches are used to select the most appropriate atlas for the target image [30]–[33]. The second approach is to use the segmentations from each atlas and combine them using label fusion methods [26], [34], [35]. Both of these multi-atlas methods allow for more consistent and accurate final segmentations than are usually possible with just a single atlas.

Due to the registration transform being applied to the atlas labelmap, it is possible to use higher resolution atlases than the target image resolution. In the context of CT images this means that it is possible to use high-resolution micro-CT atlases for creating high-resolution segmentations of low-resolution clinical-CT scans like those used on patients.

1.4.2.1 Label Fusion

Label fusion is an operation which takes multiple segmentations (labelmaps) as inputs and uses them to calculate an output segmentation. The simplest version of label fusion is majority voting in which each voxel output is decided by the value that holds the majority among the input segmentations. Simultaneous truth and performance level estimation (STAPLE) improves on majority voting by creating a probabilistic estimate of the true segmentation by estimating the performance of each individual atlas segmentation and using those performance estimates to determine the weighting of each segmentation for combination [35]. Joint label fusion (JLF) uses intensity values from each registered

image and the target image and also takes into account not only the similarity between the atlases and the target but between the atlases themselves in order to reduce similar label errors [34]. Using multiple atlases with label fusion can increase segmentation consistency and accuracy especially in cases with a lot of structural variability. However, more complex methods such as STAPLE and JLF can be computationally expensive and can increase the time an algorithm takes to complete a segmentation.

1.4.3 Convolutional Neural Networks

CNNs, as applied to 3D medical image volumes, are fully connected deep neural networks which extract features and use them to learn patterns. Receptive areas of voxels of an image are taken as inputs, which then are fed through a hierarchy system of convolutional layers which extract features and pooling layers which reduce spatial size. The way CNN kernels sample images preserve some spatial information which is difficult to achieve in many other types of architectures. The hierarchy system of convolutional layers allows for extraction of low-level features such as edges and intensity values at the first layers and higher-level features in deeper layers. Due to these properties, CNNs are very suited to the task of medical image segmentation.

In order to provide more data to a CNN during training as well as improving robustness when presented with a variety of inputs, image augmentation is often done. Some common and easy-to-implement methods of augmentation for CNN medical image segmentation are based on transforms: rotation, axis flipping, scaling, and elastic deformations. These augmentations can be tuned to the level of generalization required as well as the expected input images that inference will be run on.

When using CNNs for medical image segmentation it is necessary to sample both the images and labelmaps. However, with the large size of medical images, memory constraints are an important consideration when loading image data onto a graphic processing unit (GPU). Therefore, it is often necessary when working with larger datasets to sample only parts of images. Samples can be drawn uniformly across the images or weights can be applied such as in a balanced sampling method, where each label has the same probability of being sampled [36], [37]. The order these image samples are processed can then be mixed randomly and trained in batches to update the optimizer function.

A major problem that often occurs in the development of a CNN is overfitting. Overfitting is when the model produced by training the network becomes too specific to the training dataset and fails to make optimal predictions on data that is unseen in the training process. One method to prevent this when training a CNN for image segmentation is to portion part of the dataset to be used as validation images. The validation images are evaluated with the loss function at set intervals in the training to determine if the CNN has the ability to make generalizations outside the training set. Validation can be an indicator of when to perform early stopping of the training if the validation loss increases. Another method for preventing overfitting is regularization. Regularization attempts to bias the training towards developing simpler models which are more likely to be better at generalization. L_2 regularization is a common method of regularization that adds to the value of the loss function based on the complexity represented by the sum of the squares of the current network weights. The amount of

regularization applied can be tuned by changing a lambda weight variable. This can be a small value in medical image segmentation applications [38].

1.4.3.1 DeepMedic

DeepMedic is a 3D CNN combined with a conditional random field that acts as a filter for incorrect results and was originally developed for brain lesion segmentation [39]. The DeepMedic network architecture shown in Figure 1.7 uses two different resolution scale pathways through eight convolutional layers each to capture local and wider regional information about the image. This makes it possible to learn the features such as edges and intensities locally while simultaneously learning relative positions of structures. The results of the two pathways are then fed through a fully connected conditional random field that attempts to remove any false positive segmentations. After passing through the conditional random field, the output is fed through a classification layer to generate the final segmentation.

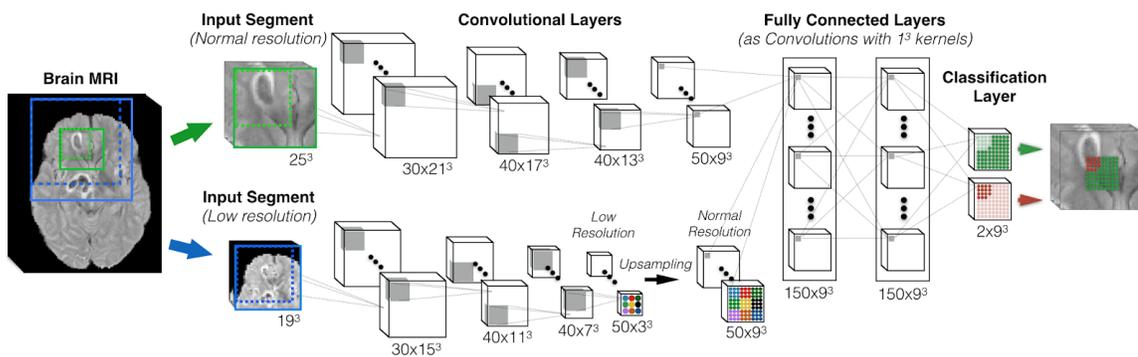


Figure 1.7: DeepMedic 3D CNN architecture with two convolutional pathways followed by a fully connected conditional random field [39]. Licensed under CC BY

1.4.3.2 DenseVNet

DenseVNet is another CNN originally developed for multi-class abdominal structure segmentation [40]. The DenseVNet architecture is shown in Figure 1.8. The image is first downsampled to a given size and then fed through dense feature stacks which each output a skip as well as a downsampled output to go into the next dense feature stack. This is done three times. All outputs and skips are then upsampled back to the original size. If there is a spatial prior map which provides the log-probability of a certain label being at given spatial coordinates [40], [41], it is added to the output prediction. DenseVNet has been shown to produce high-accuracy empirical results when compared to other methods on the multi-class abdominal segmentation problem, while being memory-efficient [40].

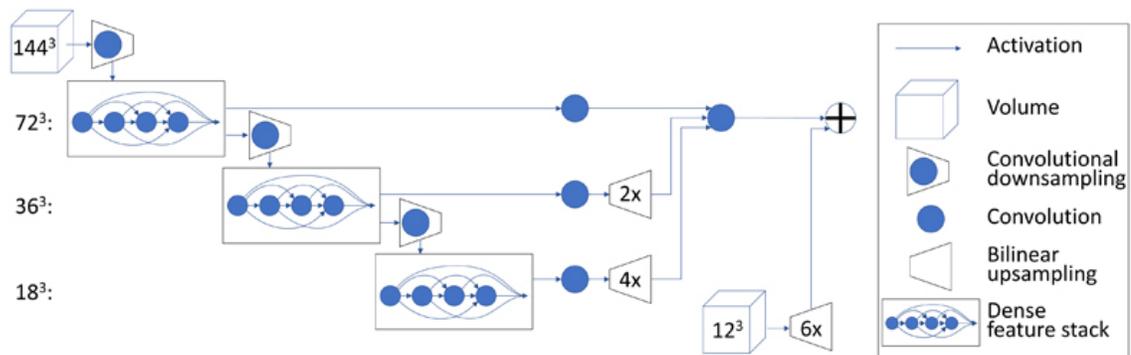


Figure 1.8: DenseVNet network architecture [40]. © 2017 IEEE.

1.4.4 Objective Functions

In order to guide an optimization problem such as image registration or CNN training it is necessary to provide an objective function (also known as a loss function when minimized). A loss function returns a numerical value, which when minimized will approach the optimum solution.

In the case of image registration there are several loss functions available to describe how similar images are. Normalized cross-correlation (NCC) and normalized mutual information (NMI) are examples of objective functions which use different approaches to evaluate the similarity between two overlapped images [42]. NCC uses the intensity of overlapping regions to perform a comparison. The method for calculating the NCC value from images A and B is shown in Equation 1.1.

$$NCC(A, B) = \frac{1}{N} \sum_x \sum_y \sum_z \frac{(A(x, y, z) - \bar{A})(B(x, y, z) - \bar{B})}{\sigma_A \sigma_B} \quad \text{Equation 1.1}$$

where (x, y, z) are the voxel coordinates, and σ_A and σ_B denote the variances for A and B .

While NCC is accurate and fast, it is only usable when there is a linear dependency between the intensity values in the two images [43]. Since a value of 1.0 indicates perfect image correlation, an optimizer for NCC should attempt to maximize the value.

Unlike NCC, NMI does not require intensity values to be linearly dependent between images. NMI uses the difference in information between the two images. The NMI of images A and B is given in Equation 1.2 [43], [44].

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)} \quad \text{Equation 1.2}$$

where $H(A)$ and $H(B)$ are the marginal entropies of images A and B and $H(A, B)$ is the joint entropy of the images. The entropies act as information measurements which estimate the probability of values occurring at each point and then averages the information across the image(s). The marginal entropy of an image A is defined in

Equation 1.3 and following a similar method except for the overlap of two images, the joint entropy of images A and B is shown in Equation 1.4 [44].

$$H(A) = -\sum_{a \in A} p\{a\} \log(p\{a\}) \quad \text{Equation 1.3}$$

$$H(A, B) = -\sum_{b \in B} \sum_{a \in A} p\{a, b\} \log(p\{a, b\}) \quad \text{Equation 1.4}$$

The NMI value when perfect overlap occurs is 2.0; meaning $H(A) = H(B) = H(A, B)$ and the lowest value possible is 1.0, indicating no overlap. NMI presents a robust and accurate similarity metric for image registration even when done across different imaging modalities such as CT and magnetic resonance imaging (MRI) [43].

While the objective functions for image registration focus on determining the amount of difference between images, objective functions for CNNs focus on the difference between the ground truth segmentations and the predicted results. Two common loss functions used for CNNs are Dice loss and cross entropy. Dice loss for a binary case between a ground truth segmentation G , and predicted segmentation P , is shown in Equation 1.5 where DL denotes Dice loss. Dice loss is based on the Dice coefficient evaluation metric used to determine the overlap between segmentations [45], [46].

$$DL = 1 - \frac{\sum_{i=1}^N p_i g_i + \varepsilon}{\sum_{i=1}^N p_i + g_i + \varepsilon} - \frac{\sum_{i=1}^N (1 - p_i)(1 - g_i) + \varepsilon}{\sum_{i=1}^N 2 - p_i - g_i + \varepsilon} \quad \text{Equation 1.5}$$

In Equation 1.5 p is the predicted probability voxel values, g is the ground truth voxel values, N is the number of image elements, and ε is a term used to avoid dividing by 0 if P and G are empty.

When training for segmenting multiple structures at once, changes can be made by manually adding weights to the Dice loss function; or when dealing with highly imbalanced classes (i.e. some structures have much larger volumes than others), weights can be determined based on the inverse of their volume [45]. Dice loss is often used because minimizing it is directly increasing the value of the Dice coefficient evaluation metric for determining segmentation overlap accuracy, and it naturally performs well at class imbalanced problems.

Cross entropy is a common alternative loss function for CNN image segmentation training [27], [47], [48]. It compares the predicted probability to the actual label and follows a logarithmic scale which rapidly reaches high levels of penalization (i.e. high cost value) as probability increases for incorrect labels. The cross entropy for a binary segmentation problem is denoted in Equation 1.6 as CE .

$$CE = -(g \log(p) + (1 - g) \log(1 - p)) \quad \text{Equation 1.6}$$

where g is the ground truth voxel values and p is the predicted probability voxel values.

Cross entropy has the advantage of having generally better gradients than Dice loss which is what must be traversed by an optimizer when attempting to minimize the loss function. Loss functions for segmentation should be selected on a case by case basis depending on the structures being segmented by the CNN and through iterative testing.

1.4.5 Optimizers

While traversing an optimization problem with a loss function, an optimizer is used to move towards a local minimum solution. Gradient descent attempts to move down the

steepest path of a gradient iteratively in steps. The Newton method also uses a gradient but also computes the Hessian matrix at every step. This can be extremely computationally intensive and is often not practical, especially in the context of large CNNs. Quasi-Newton methods are similar to Newton's method but use processes such as approximating the Hessian matrix by updating it instead of fully recalculating it in order to reduce the computational power needed [49].

A very efficient and effective implementation of gradient descent often recommended for large CNN applications is Adam [50], [51]. Adam stores past squared gradients and also exponentially decaying averages of past gradients. This results in a method of traversing the gradient that is adaptive during training while requiring minimal computational resources.

When using an optimizer, it is necessary to set a learning rate which determines the amount the optimization problem will update on each step of the function. If this is set too high the training can become unstable and not converge; if set too low the training will take excessively long and may get stuck at a small minimum. Therefore, it is important to tune the learning rate of an optimizer depending on the progress of training runs and the application of the network.

1.4.6 Post-Processing Methods

After the main portion of an automated segmentation is run it is often useful to run post-processing operations to clean and smooth the segmentation. Island removal is a common post-processing method which is used to delete separate components from the segmentation. Since the structures in the temporal bone are all each separate single

connected components, the most useful form of island removal is largest connected component analysis. Largest connected component analysis only keeps the largest component of the segmentation for each structure measured by volume and can be applied automatically without user intervention.

Another useful post-processing operation for automated segmentations is smoothing.

Smoothing can improve the aesthetics and clean the contour of the segmentation.

Morphological hole filling is a form of smoothing which uses surrounding information selected by an adjustable size and shape kernel to smooth and fill holes in a segmentation.

This is especially useful for post-processing thin structures that are more likely to have holes in the surface, such as the tegmen.

1.4.7 Segmentation Evaluation

While developing an algorithm for medical image segmentation it is necessary to determine its accuracy. Aside from manual visual inspection, the results have to be quantitatively compared to ground truth manual segmentations. There are a variety of quantitative metrics that can be used.

One of the most common metrics for medical image segmentation evaluation is the Dice coefficient [30], [31], [41], [52]–[55]. Dice values range between 0 and 1, with 1 being perfect overlap and 0 indicating no overlap. An important property to note is the correlation between region size and the Dice coefficient value [45]. Figure 1.9 shows the overlap between two segmentations *A* and *B* which is used to calculate the Dice coefficient.

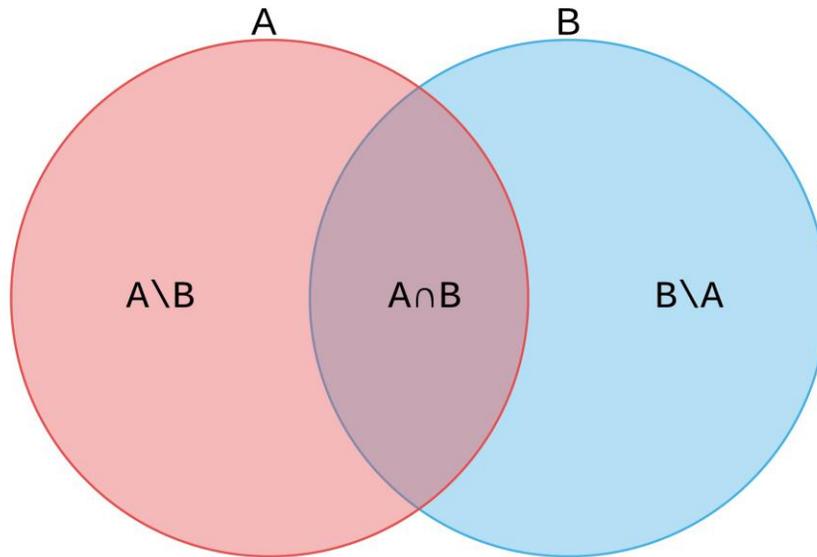


Figure 1.9: Venn diagram showing overlap between sets A and B.

Equation 1.7 shows the method for calculating the Dice value between segmentations A and B .

$$DICE = \frac{2|A \cap B|}{|A| + |B|} \quad \text{Equation 1.7}$$

Volumetric similarity (VS) is a measure that compares the volumes of the segmentations without comparing the overlap or their positions [52]. VS has been used often in medical image segmentation literature [52], [56], [57] and in two studies by Powell et al. [55], [58] on temporal bone structures. Similar to the Dice coefficient, VS ranges between values of 0 and 1. Equation 1.8 shows the calculation for the VS between segmentations using the true positives (TP), false negatives (FN), and false positives (FP).

$$VS = 1 - \frac{|FN - FP|}{2TP + FN + FP} \quad \text{Equation 1.8}$$

While comparing segmentations it is also necessary to use metrics for the distance between them. This can be done comprehensively using two metrics, the Hausdorff distance (HD) and the average Hausdorff distance (AHD). HD gives a value for the maximum distance between two segmentations [52]. The Hausdorff distance between segmentations A and B is shown in Equation 1.9.

$$HD(A, B) = \max(h(A, B), h(B, A)) \quad \text{Equation 1.9}$$

where $h(A, B)$ is the directed HD given in Equation 1.10.

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad \text{Equation 1.10}$$

HD gives an understanding of the maximum distance error between segmentations, but it is very sensitive to outliers and noise. It is best used on segmentations of single component structures that have had noise removed and are cropped to a region of interest (ROI) that matches the ground truth.

While HD is useful for determining the maximum errors in distance, AHD gives a value for the overall distance between the segmentations. AHD is the HD averaged over all points in the segmentation [52]. The AHD between segmentations A and B is given in Equation 1.11.

$$AHD(A, B) = \max(d(A, B), d(B, A)) \quad \text{Equation 1.11}$$

where $d(A, B)$ is the directed AHD described in Equation 1.12.

$$d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \|a - b\| \quad \text{Equation 1.12}$$

The AHD is much less sensitive to noise and outliers and is useful for describing the overall distance between the segmentations.

When used in comparison to a ground truth, the set of quantitative metrics described above provide an understanding of the accuracy of an automatically generated segmentation through overlap, similarity of volumes, and distances between corresponding points.

1.5 Objectives

The main objective of this thesis was (1) to develop an automated segmentation algorithm for accurate delineation of temporal bone structures relevant to the creation of 3D models for use in a mastoidectomy surgical simulator. Secondary objectives of this work were (2) to determine the best label fusion method for multi-atlas based segmentation of the sigmoid sinus and (3) to explore the capabilities of CNNs for simultaneous segmentation of temporal bone structures relevant to mastoidectomy.

Initially, Chapter 2 explores the use of multi-atlas based methods to segment one temporal bone structure, the sigmoid sinus; however, as it was realized that developing an atlas-based approach to segment all structures would be difficult because spatial inter-relationships between structures need to be modeled and the portability of the solution is low. To address these issues, CNNs were adapted to the problem of segmenting multiple temporal bone structures and are presented in Chapter 3.

1.6 Novelty

A CNN based automated segmentation algorithm was developed that quickly produces segmentations with minimal user intervention. The novel algorithm is more accurate than those developed in previous works while also being the first to include segmentations of the digastric ridge. This work is also the first to use multiple micro-CT atlases with label fusion to generate high quality segmentations of the sigmoid sinus.

References

- [1] “Deafness and hearing loss.” [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>. [Accessed: 20-Aug-2019].
- [2] “Hearing disabilities among Canadians aged 15 years and older, 2012.” [Online]. Available: <https://www150.statcan.gc.ca/n1/pub/89-654-x/89-654-x2016002-eng.htm>. [Accessed: 17-Aug-2019].
- [3] “The Daily — Canadian Survey on Disability: Seeing and hearing disabilities, 2012.” [Online]. Available: <https://www150.statcan.gc.ca/n1/daily-quotidien/160229/dq160229c-eng.htm>. [Accessed: 17-Aug-2019].
- [4] J. C. Mason, “Mastoidectomy – Canal Wall-Up Technique, Posterior Tympanotomy (Facial Recess),” in *Encyclopedia of Otolaryngology, Head and Neck Surgery*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1581–1583.
- [5] B. Schick and J. Długaiczek, “Surgery of the ear and the lateral skull base: pitfalls and complications.,” *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.*, vol. 12, p. Doc05, Dec. 2013.
- [6] N.-G. Ryu and J. Kim, “How to Avoid Facial Nerve Injury in Mastoidectomy?,” *J. Audiol. Otol.*, vol. 20, no. 2, pp. 68–72, Sep. 2016.
- [7] R. Jackler and C. Gralapp, “Ear Anatomy - Otologic Surgery Atlas - Otolaryngology Head and Neck Surgery (OHNS) - Stanford School of Medicine.” [Online]. Available: <http://med.stanford.edu/sm/ohns-otologic-surgery-atlas/ear-anatomy/>. [Accessed: 20-Aug-2019].
- [8] R. Jackler, *Atlas of skull base surgery and neurotology*. New York: Thieme, 2009.
- [9] I. Aly and R. S. Tubbs, “The Sigmoid Sinus,” in *Anatomy, Imaging and Surgery of the Intracranial Dural Venous Sinuses*, Elsevier, 2020, pp. 47–58.
- [10] K. Van Osch, D. Allen, B. Gare, T. J. Hudson, H. Ladak, and S. K. Agrawal, “Morphological analysis of sigmoid sinus anatomy: Clinical applications to neurotological surgery,” *J. Otolaryngol. - Head Neck Surg.*, vol. 48, no. 1, p. 2, Dec. 2019.
- [11] P. B. Sarmiento and F. G. Eslait, “Surgical Classification of Variations in the Anatomy of the Sigmoid Sinus,” *Otolaryngol. Neck Surg.*, vol. 131, no. 3, pp. 192–199, Sep. 2004.
- [12] N. N. Butler and G. J. Wiet, “Reliability of the Welling Scale (WS1) for Rating Temporal Bone Dissection Performance,” *Laryngoscope*, vol. 117, no. 10, pp. 1803–1808, Oct. 2007.
- [13] D. Wan, G. J. Wiet, D. B. Welling, T. Kerwin, and D. Stredney, “Creating a cross-institutional grading scale for temporal bone dissection.,” *Laryngoscope*, vol. 120, no. 7, pp. 1422–7, Jul. 2010.
- [14] A. Arora *et al.*, “Face and content validation of a virtual reality temporal bone simulator,” *Otolaryngol. - Head Neck Surg.*, vol. 146, no. 3, pp. 497–503, Mar.

2012.

- [15] D. Morris, C. Sewell, F. Barbagli, K. Salisbury, N. H. Blevins, and S. Girod, "Visuohaptic simulation of bone surgery for training and evaluation," *IEEE Comput. Graph. Appl.*, vol. 26, no. 6, pp. 48–57, Nov. 2006.
- [16] C. Sewell *et al.*, "Providing metrics and performance feedback in a surgical simulator," *Computer Aided Surgery*, vol. 13, no. 2, pp. 63–81, 06-Jan-2008.
- [17] G. J. Wiet *et al.*, "Virtual temporal bone dissection system: OSU virtual temporal bone system: Development and Testing," *Laryngoscope*, vol. 122, no. SUPPL. 1, pp. S1–S12, Mar. 2012.
- [18] R. Varshney *et al.*, "The McGill simulator for endoscopic sinus surgery (MSESS): a validation study," *J. Otolaryngol. Head Neck Surg.*, vol. 43, no. 1, p. 40, Dec. 2014.
- [19] B. Tolsdorff *et al.*, "Virtual reality: A new paranasal sinus surgery simulator," *Laryngoscope*, vol. 120, no. 2, pp. 420–426, Feb. 2010.
- [20] S. M. Anil, Y. Kato, M. Hayakawa, K. Yoshida, S. Nagahisha, and T. Kanno, "Virtual 3-dimensional preoperative planning with the dextroscope for excision of a 4th ventricular ependymoma," *Minim. Invasive Neurosurg.*, vol. 50, no. 2, pp. 65–70, Apr. 2007.
- [21] S. Weghorst *et al.*, "Validation of the Madigan ESS simulator.," *Stud. Health Technol. Inform.*, vol. 50, pp. 399–405, 1998.
- [22] S. Chan, "CardinalSim." [Online]. Available: <https://cardinalsim.stanford.edu/>. [Accessed: 19-Feb-2019].
- [23] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, no. 1, pp. 315–337, Aug. 2000.
- [24] D. Zukić, J. Vicory, M. McCormick, L. Wisse, G. Gerig, and P. Yushkevich, "ND morphological contour interpolation," *Insight J.*, 2016.
- [25] L. Zhu, I. Kolesov, Y. Gao, R. Kikinis, and A. Tannenbaum, "An Effective Interactive Medical Image Segmentation Method Using Fast GrowCut," *MICCAI Work. Interact. Med. Image Comput.*, 2014.
- [26] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Med. Image Anal.*, vol. 24, no. 1, pp. 205–219, Aug. 2015.
- [27] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [28] M. Modat *et al.*, "Fast free-form deformation using graphics processing units," *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 278–284, Jun. 2010.
- [29] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [30] B. M. Gare, "Multi-Atlas Segmentation of the Facial Nerve," Western University, 2019.

- [31] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, Jul. 2009.
- [32] B. Schipaanboord *et al.*, "An evaluation of atlas selection methods for atlas-based automatic segmentation in radiotherapy treatment planning," *IEEE Trans. Med. Imaging*, pp. 1–1, 2019.
- [33] G. Sanroma, G. Wu, Y. Gao, and D. Shen, "Learning-Based Atlas Selection for Multiple-Atlas Segmentation.," *Conf. Comput. Vis. Pattern Recognit. Work. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. Work.*, vol. 2014, pp. 3111–3117, Jun. 2014.
- [34] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Mar. 2013.
- [35] S. K. Warfield *et al.*, "Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation," *IEEE Trans Med Imaging*, vol. 23, no. 7, pp. 903–921, Jul. 2004.
- [36] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.
- [37] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [38] M. Avetisian, "Volumetric medical image segmentation with deep convolutional neural networks," *CEUR Workshop Proc.*, vol. 2022, pp. 5–9, 2017.
- [39] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Mar. 2017.
- [40] E. Gibson *et al.*, "Automatic Multi-Organ Segmentation on Abdominal CT with Dense V-Networks," *IEEE Trans. Med. Imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.
- [41] E. Gibson *et al.*, "Towards Image-Guided Pancreas and Biliary Endoscopy: Automatic Multi-organ Segmentation on Abdominal CT with Dense Dilated Networks," Springer, Cham, 2017, pp. 728–736.
- [42] J. Wu, M. Kim, J. Peters, H. Chung, and S. S. Samant, "Evaluation of similarity measures for use in the intensity-based rigid 2D-3D registration for patient positioning in radiotherapy," *Med. Phys.*, vol. 36, no. 12, pp. 5391–5403, Nov. 2009.
- [43] F. E.-Z. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egypt. Informatics J.*, vol. 17, no. 1, pp. 99–124, Mar. 2016.
- [44] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy

- measure of 3D medical image alignment,” *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.
- [45] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.
- [46] F. Milletari, N. Navab, and S. A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, Jun. 2016.
- [47] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology.,” *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018.
- [48] N. Ibtehaz and M. S. Rahman, “MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation,” Feb. 2019.
- [49] R. Battiti, “First- and Second-Order Methods for Learning: Between Steepest Descent and Newton’s Method,” *Neural Comput.*, vol. 4, no. 2, pp. 141–166, Mar. 1992.
- [50] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014.
- [51] S. Ruder, “An overview of gradient descent optimization algorithms,” Sep. 2016.
- [52] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.,” *BMC Med. Imaging*, vol. 15, no. 1, p. 29, Dec. 2015.
- [53] P. Lu *et al.*, “Highly Accurate Facial Nerve Segmentation Refinement From CBCT/CT Imaging Using a Super-Resolution Classification Approach,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 178–188, Jan. 2018.
- [54] V. Dill, P. C. Klein, A. R. Franco, and M. S. Pinho, “Atlas selection for hippocampus segmentation: Relevance evaluation of three meta-information parameters,” *Comput. Biol. Med.*, vol. 95, pp. 90–98, Apr. 2018.
- [55] K. A. Powell, T. Liang, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet, “Atlas-Based Segmentation of Temporal Bone Anatomy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 11, pp. 1937–1944, Nov. 2017.
- [56] R. Cárdenes, R. de Luis-García, and M. Bach-Cuadra, “A multidimensional segmentation evaluation for medical image data,” *Comput. Methods Programs Biomed.*, vol. 96, no. 2, pp. 108–124, Nov. 2009.
- [57] A. RamaswamyReddy, E. V. Prasad, and L. S. S. Reddy, “Comparative Analysis of Brain Tumor Detection using Different Segmentation Techniques,” *Int. J. Comput. Appl.*, vol. 82, no. 14, pp. 14–28, 2013.
- [58] K. A. Powell, T. Kashikar, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet, “Atlas-based segmentation of temporal bone surface structures,” *Int. J. Comput.*

Assist. Radiol. Surg., vol. 14, no. 8, pp. 1267–1273, Aug. 2019.

Chapter 2

2 Automated Segmentation of the Sigmoid Sinus using a Multi-Atlas Approach

2.1 Introduction

Mastoidectomy is a complex surgical procedure that requires extensive knowledge of the anatomy of the temporal bone and is often required in cochlear implantation surgery.

However, due to its complexity, mastoidectomy is a difficult procedure for trainees to master. Traditional training methods for surgical residents utilize cadavers, which can be expensive and difficult to access. To provide more consistent and accessible training for surgery involving the temporal bone, surgical simulators have been developed that provide haptic (touch) feedback and three-dimensional (3D) visualization [1]–[10].

Surgical simulation is becoming a widely accepted tool in Otolaryngology since it offers the ability to model difficult and varied cases and allows trainees to practice on patient-specific models.

Simulators such as CardinalSim [10] are able to import patient images and can be used for both training and in pre-operative planning. To maximize the variety and relevance of the anatomical cases available in the simulator, many clinical scans need to first be segmented (i.e., the anatomy needs to be delineated) for use. A major drawback to VR simulators is the need for manual image segmentation, which can take hours per scan and is infeasible in a clinical setting. Automated segmentation methods are preferred to be able to rapidly and automatically produce a variety of 3D digital models of the temporal bone.

One challenge associated with the development of simulators in Otolaryngology is the complex anatomy of the temporal bone. For example, in mastoidectomy one of the vital anatomical structures is the sigmoid sinus. The sigmoid sinus is a venous sinus that travels down an S-shaped groove in the temporal bone. During the initial portion of the procedure the sigmoid sinus represents the posterior boundary of bone removal [11]. It is critical for surgeons and trainees to be able to identify the sigmoid sinus to avoid catastrophic vascular complications during surgery.

Currently, creating 3D models of the sigmoid sinus requires manual delineation of structural boundaries by an individual with expertise in both the anatomy and the software tools used to segment medical image volumes. In addition, the sigmoid sinus is a highly variable structure in both shape and relative position to other structures in the temporal bone, and previous work has focused on evaluating its variability through statistical shape analysis [12]. Due to this variability, it is very time-consuming for an expert to manually perform segmentation (often taking up to 45 minutes). Therefore, an automated algorithm is required to produce sigmoid sinus segmentations accurately and quickly with minimal expert intervention. However, due to the vast anatomic variability and low contrast at the medial wall, purely intensity-based methods such as thresholding are inconsistent, making development of an automated segmentation method of the sigmoid sinus uniquely challenging.

Several approaches to automated segmentation of anatomical structures have been described in the literature. One of the simplest approaches is thresholding. Thresholding is a fast and effective method for delineating structures that have high contrast from the surrounding objects, but as noted above this is generally ineffective for many anatomical

structures such as the sigmoid sinus which has similar intensity values to its surrounding areas on its medial side. Atlas-based methods are far more promising for capturing variability in difficult to delineate anatomical structures due to their ability to capture the anatomical information and their relative robustness to poor contrast. One effective application of atlas-based segmentation on various structures of the temporal bone that excludes the sigmoid sinus has been presented by Powell. et al [13].

The present work describes the development and evaluation of a multi-atlas based segmentation algorithm that compares a variety of label fusion methods on clinical CT scans of cadaveric temporal bones with the goal of accurately segmenting the sigmoid sinus. The multi-atlas method presented herein has been applied previously to segment medical images in a variety of fields [14]–[16], and has been adapted in this work to allow for the usage of highly detailed micro-CT (μ CT) atlases to segment low resolution and poorly delineated clinical CT scans. This method was able to capture the high variability of the anatomy of the sigmoid sinus.

2.2 Materials and Methods

2.2.1 Data

Thirty-eight anonymized adult cadaveric temporal bones with normal anatomy were used. Samples had not been operated on in any previous surgeries and were scanned using a General Electric (GE) Healthcare eXplore Locus μ CT scanner at a resolution of $154\mu\text{m} \times 154\mu\text{m} \times 154\mu\text{m}$ and a voltage of 80kV. Clinical-CT scans of the same thirty-eight samples were also collected using a resolution of $234\mu\text{m} \times 234\mu\text{m} \times 625\mu\text{m}$ and a

voltage of 120kV on a Discovery CT750 HD Clinical Scanner with GE's Gemstone CT detector.

All cadaveric specimens were obtained with permission from the body bequeathal program at Western University, London, Ontario, Canada in accordance with the Anatomy Act of Ontario and Western's Committee for Cadaveric Use in Research (approval number: #19062014).

2.2.2 Ground Truth Segmentations and Atlas Creation

The sigmoid sinus was manually segmented from the μ CT images by an expert anatomist (KV) using a combination of semi-automated and manual tools in 3D Slicer, an open-source software for medical image processing and visualization [17]. Consensus interpretation of the segmentations was achieved by an experienced surgeon (SKA), the anatomist (KV), and the lead author (DA). Details of the segmentations and anatomic analysis have been previously described in [12].

Twelve segmentations were used to define atlases to drive the segmentation algorithm and the remainder formed the ground truth and were used to evaluate the segmentation algorithm. Two sets of atlases were defined from the 12 segmentations: 6 atlases for left temporal bones and 6 for right temporal bones. This number of atlases managed to capture the variability of the sigmoid sinus while producing results comparable to using 36 atlases. The use of μ CT images resulted in higher resolution and detail for both the ground truth and the algorithm-generated segmentations. The difference in resolution can be seen in Figure 2.1.

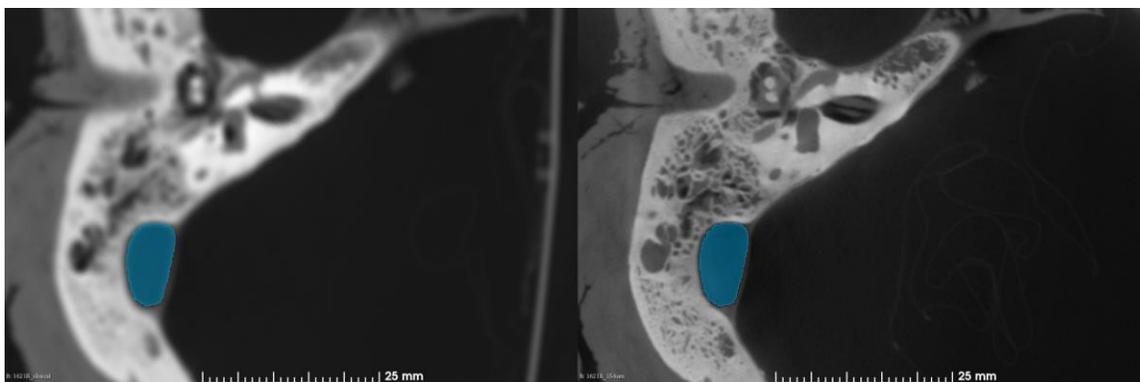


Figure 2.1: Clinical-CT (left) compared to Micro-CT (right) with sigmoid sinus shown in blue.

As the algorithm is applied to clinical-CT volumes, the μ CT segmentations were registered using a combination of rigid and affine techniques to their corresponding clinical-CT volumes and then reviewed by the research team to be used as a high-resolution ground truth.

2.2.3 Segmentation Algorithm

Figure 2.2 depicts the operation of the segmentation algorithm. All steps are completely automatic with the exception of basic rough cropping of the image. The major steps of the algorithm are described next in further detail. The algorithm was implemented as one single script written in Bash shell script.

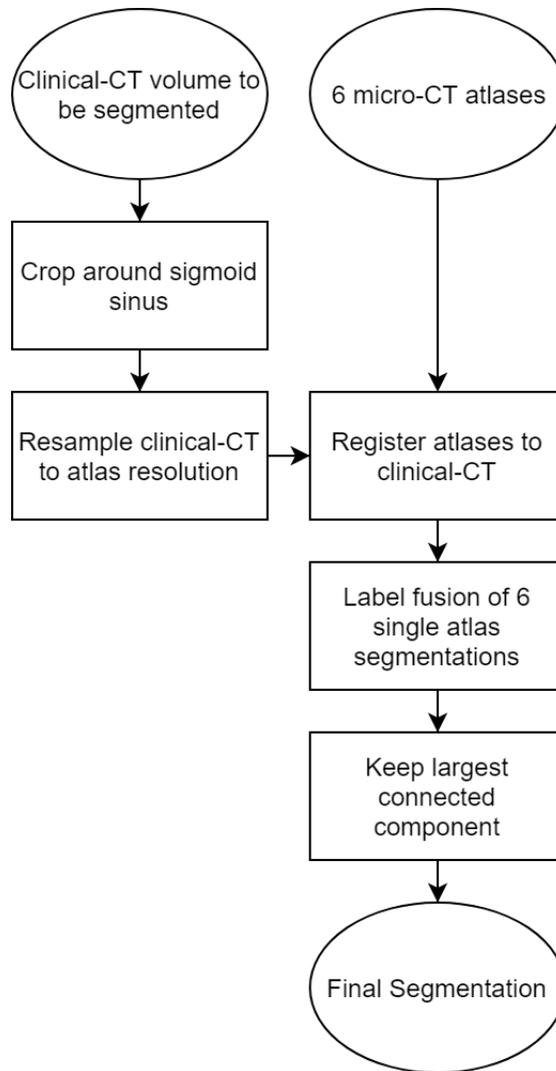


Figure 2.2: Block diagram displaying the multi-atlas segmentation algorithm. User specifies either left or right ear and appropriate atlases are chosen by algorithm.

2.2.3.1 Cropping and Resampling

The clinical-CT volumes were cropped around the area of the sigmoid sinus and resampled using linear interpolation from 0.234 mm x 0.234 mm x 0.625 mm to 0.154 mm x 0.154 mm x 0.154 mm isotropic to be approximately the same resolution as the μ CT volumes. This was done to reduce the loss of detail from the μ CT atlas

segmentations during registration and resampling to the resolution of the target clinical-CT volumes.

2.2.3.2 Rigid and Affine Registration

Prior to non-rigid registration, a rigid step and affine step were used to approximately align each atlas to the target clinical-CT volume. This two-step approach was taken to improve accuracy. By applying the rigid registration first, the time required for the affine transformation is reduced, which is needed to scale and skew the atlas to approximate the target image before non-rigid registration. This two-step registration approach was performed using the NiftyReg implementation of a symmetric (source to target and target to source, simultaneously) block-matching registration, applied in three pyramidal levels from coarse to fine, doubling the resolution on each step up to the original 0.154 mm isotropic voxel size [18].

2.2.3.3 Non-Rigid Registration

Non-rigid registration for the μ CT atlases was also accomplished using the NiftyReg implementation of a B-spline pyramidal approach in three progressively finer control point grids using 12 mm x 12 mm x 12 mm, 6 mm x 6 mm x 6 mm, and 3 mm x 3 mm x 3 mm grid spacings, respectively [19]. The loss function used by NiftyReg for the non-rigid registration was a combination of normalized mutual information (NMI) and bending-energy (BE), which was optimized using a conjugate gradient scheme. An example of the process of volume registration from rigid to the final non-rigid B-spline registration is given in Figure 2.3.

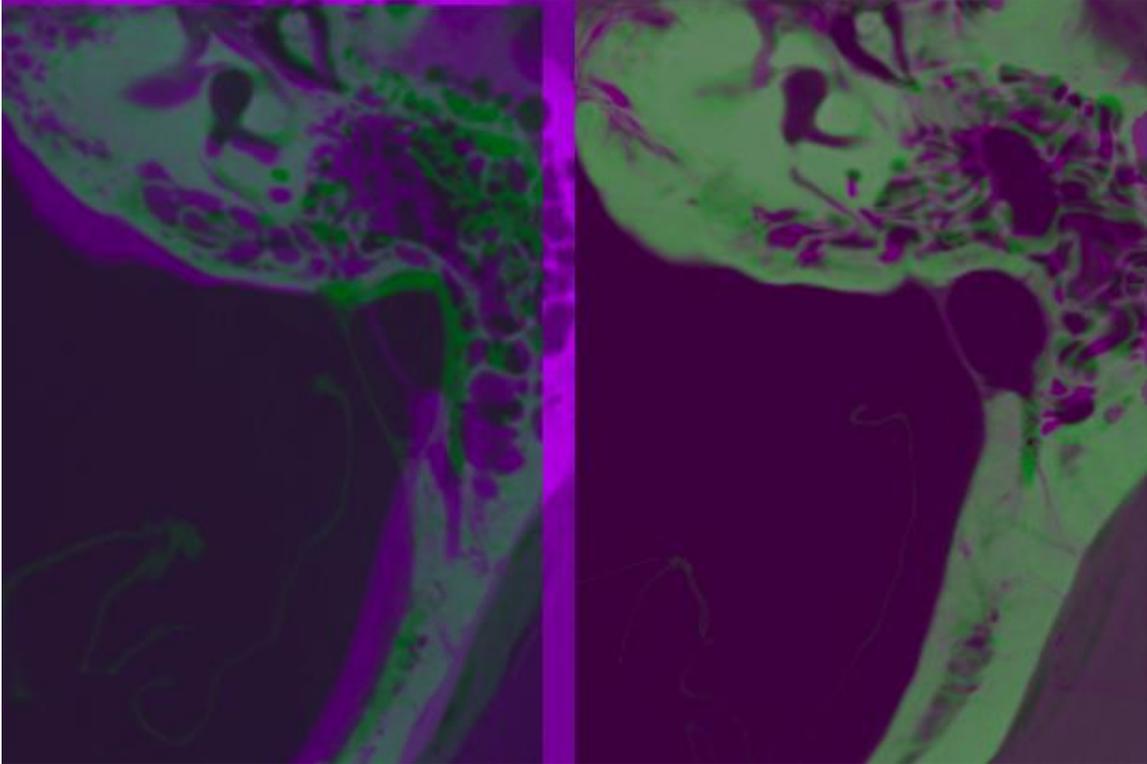


Figure 2.3: Axial images of the temporal bone showing target clinical CT (magenta) and the μ CT atlas (green). The unregistered images are shown overlaid without registration (left) and the result of a complete registration (right). Observe the shift of the position and shape of the sigmoid sinus.

2.2.3.4 Label Fusion

Individually, the accuracy when comparing Dice and Hausdorff distance values of the results of the single-atlas segmentations varied, but consistency and overall accuracy were greatly improved by combining the information from each of the six single-atlas segmentations for a given clinical-CT volume using label fusion methods.

Three established and widely available label fusion methods were applied to the completed registrations: majority voting, STAPLE [20], and joint label fusion [14]. The

registrations were then compared to determine the differences between the methods as they related to the sigmoid sinus.

2.2.3.4.1 Majority Voting

The first label fusion method applied was majority voting. Majority voting counts the value (either 0 or 1) of each individual binary segmentation at each voxel in the image volume and takes the majority decision as the result. This method is fast and simple to apply and improves the consistency of the segmentation quality when compared to single-atlas segmentations. A disadvantage of this method is that inaccurate and outlying segmentations are given the same weight as the more accurate segmentations.

2.2.3.4.2 STAPLE

STAPLE is an expectation maximization algorithm for evaluating the performance of multiple separate segmentations and produces a final probabilistic segmentation. As opposed to majority voting, STAPLE aims to use the data from all the individual segmentations to determine performance levels of each individual segmentation and then uses that information to find a final segmentation deemed closest to the true segmentation by the algorithm [20]. On average, STAPLE produces much better results compared to majority voting; however, it takes more time especially when evaluating multiple atlases.

2.2.3.4.3 Joint Label Fusion

The third and final label fusion method used was the Advanced Normalization Tools (ANTs) implementation of joint label fusion. Joint label fusion adopts a similar statistical approach to STAPLE, but also uses information from each registration image volume

result along with the generated label map. Joint label fusion applies the probability that multiple atlases would make the same error at a particular voxel [14].

2.2.3.5 Largest Connected Component

After the multi-atlas procedure was completed with label fusion, island removal was performed to remove noise and disconnected voxels from the segmentation usually caused by the label fusion methods. Largest connected component island removal is quick to apply and has been applied in previous segmentation applications [13]. Since the sigmoid sinus is one connected blood vessel, the approach used for island removal was to discard all but the largest connected component of the segmentation. This resulted in a clean single label with no noise that was ready for use in a surgical simulator.

2.3 Evaluation and Metrics

The automated segmentations generated by the multi-atlas based method were evaluated by comparing to the ground truth, manual segmentations completed by the anatomist using a variety of metrics. The segmentation algorithm was applied to clinical-CT volumes, but the assessment was done with comparison to ground truth μ CT which were registered to their corresponding clinical-CT since label maps are higher resolution in μ CT as well as boundaries being more visible in μ CT.

The first metric used was the Dice coefficient, which determines the overlap between the automated and manual segmentations. The second metric used was the Hausdorff distance, which measures the maximum distance from one segmentation to another. The Hausdorff distance is extremely sensitive to noise, such that algorithms which segment a

smaller or larger portion of the structure than the ground truth will return larger values while segmenting the correct areas. Island removal completed in post-processing is reasonably effective at negating this sensitivity to noise as it removes the unconnected components of the segmentation. The final metric used was the average Hausdorff distance (AHD), which considers the mean of all the Hausdorff distances between the two segmentations. The AHD metric is less sensitive to outliers than the Hausdorff distance and provides an understanding of the magnitude of the distance between the segmentations that cannot be seen in the Dice coefficient.

Using the Dice coefficient, Hausdorff distance, and AHD in conjunction with visual comparison provided an overall understanding of the differences in shape, size, outliers, and distance between the two compared segmentations (ground truth and algorithmic) while providing values that could easily be compared to previous segmentation projects in the literature [21].

2.4 Results

Results of the majority voting, STAPLE and joint label fusion metrics for segmentations created using just a single atlas (only one atlas is registered to the target image) as well as multi-atlas are shown in Table 2.1.

Table 2.1: Mean values of all metrics found from single atlas and multi-atlas approaches. Standard deviations given in parenthesis.

	Mean Dice	Mean Hausdorff Distance [mm]	Mean AHD [mm]
Single-Atlas	0.62 (0.17)	13.64 (7.02)	1.00 (1.03)
Majority Voting	0.75 (0.12)	11.26 (7.55)	0.48 (0.61)
STAPLE	0.76 (0.11)	10.36 (5.71)	0.46 (0.41)
Joint Label Fusion	0.77 (0.10)	10.39 (7.31)	0.30 (0.32)

For visual inspection, an example of an automated segmentation result in both two-dimensions (2D) and 3D is shown in Figure 2.4.

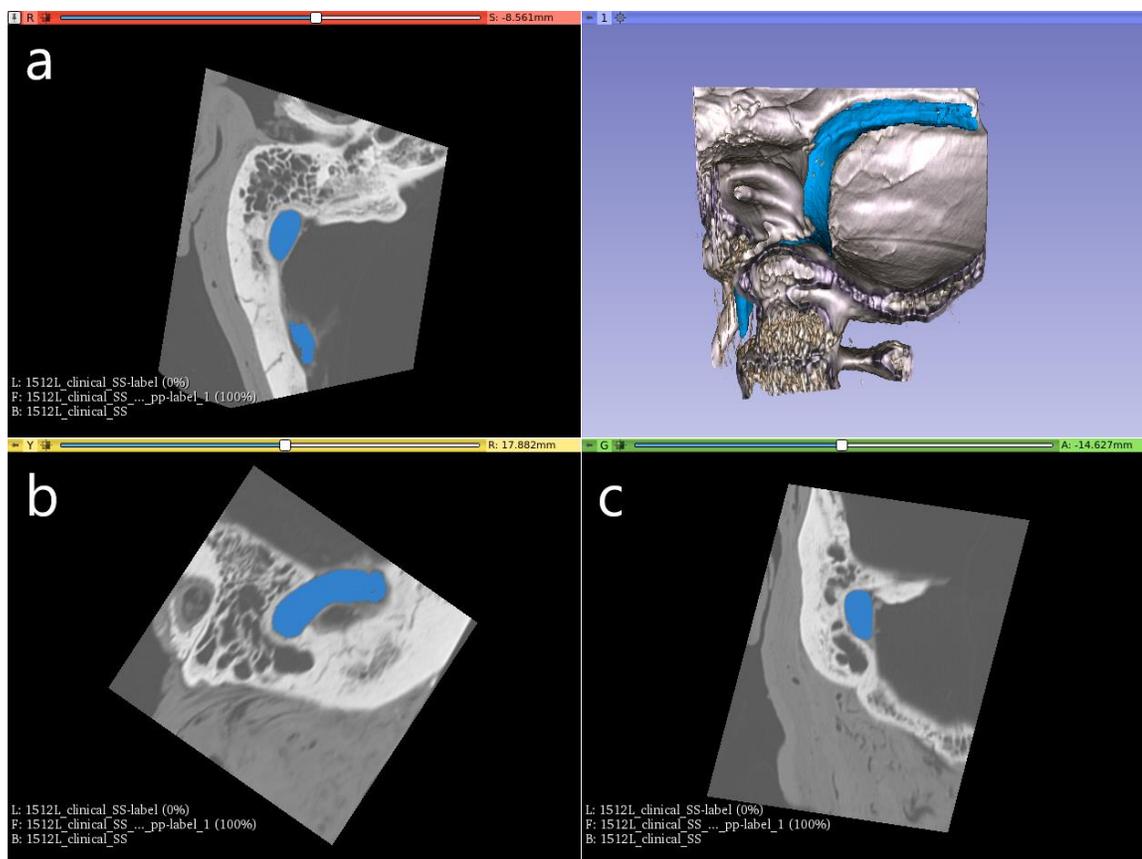


Figure 2.4: Screenshot from 3D Slicer showing an example automated segmentation in three slice views (a axial, b sagittal, c coronal) as well as in a 3D rendered temporal bone segment. The automated segmentation and 3D model of the sigmoid sinus are shown in blue against the grayscale clinical-CT volume from which it was segmented.

All of the multi-atlas methods outperformed the single-atlas segmentations, which only produced a mean Dice of 0.62, a mean Hausdorff distance of 13.64 mm, and a mean AHD of 1.00 mm. Majority voting improved on single-atlas with a mean Dice of 0.75 with a mean AHD of 0.48 mm. STAPLE was found to give a mean Dice score of 0.76 with a standard deviation of 0.11. Joint label fusion's resulting mean Dice score was slightly higher than STAPLE and provided a mean Dice of 0.77. Joint label fusion showed improvements in the average distance of 0.30 mm compared to STAPLE with an

average distance of 0.46. As expected when the AHD decreased the Dice coefficient increased.

A comparison of the distances between the nearest points of the automated label fusion method results and the ground truth segmentation using absolute distance color maps can be seen in Figure 2.5. The colormaps revealed that the largest distance differences occurred at the inferior end near the jugular bulb within the jugular foramen where it connects to the jugular vein in the neck and the posterior extreme of the transverse portion of the sigmoid sinus of the segmented area of the sigmoid sinus. These areas are outside of the clinically relevant portion for surgical simulation. The distances were likely caused by differences in the size of the portion of the sigmoid sinus segmented between each ground truth model and the amount of the sigmoid sinus segmented by the algorithm.

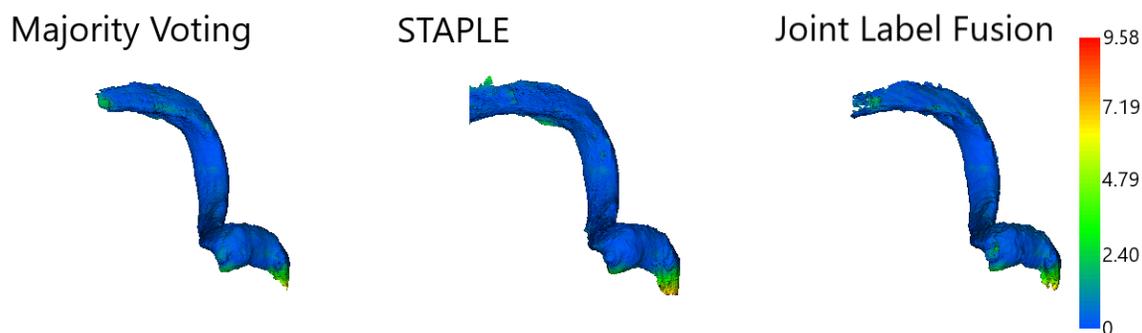


Figure 2.5: Colormaps projected on automated segmentation 3D models showing absolute distance in millimeters between the algorithm segmentations and the ground truth sigmoid sinus.

2.5 Discussion

While the single-atlas approach to automated segmentation can occasionally provide comparable metric results to individual multi-atlas segmentations, the average performance of the single-atlas applied across the dataset of temporal bone images is much lower and the results are less consistent. This is likely due to the difficulty of capturing the high degree of variability of the sigmoid sinus anatomy in one example. This performance increase from single-atlas to multi-atlas has been seen in previous works that segmented other anatomical structures, such as the brain from MRI scans [16], [22].

The current methods presented for multi-atlas based segmentation of the sigmoid sinus provided accurate segmentations from clinical-CT scans which may be used in future surgical simulation. The use of joint label fusion, the most successful label fusion approach, resulted in a mean Dice coefficient score of 0.77, a mean Hausdorff distance of 10.39 mm, and a mean AHD of 0.30 mm, along with reasonable visual results. The seemingly large mean Hausdorff distance may be attributed to the difference in the amount of the extremities segmented by the algorithm from the ground truth segmentations, since there were no large portions segmented outside the sigmoid sinus in the visual inspection.

One drawback to joint label fusion is that the method requires much more computation time and storage space for registered images than STAPLE or majority voting, which performed almost as well as joint label fusion in the current work. STAPLE and majority voting are therefore attractive options for segmenting the sigmoid sinus when time and storage space is at a premium.

By using μ CT atlases, detailed, high-resolution models were created that had comparable metric scores to other temporal bone structures segmented by previous groups using different methods. Other atlas-based approaches for segmenting the structures of the temporal bone differ from the one presented herein, as they do not use multiple atlases for one segmentation, do not use label fusion methods to better capture variability, do not use μ CT atlases, and do not target the sigmoid sinus [13], [23].

Despite the accurate results produced by the present approach, it is important to note that since these are automatically generated models there is risk of error. If used clinically, automatically generated segmentations should be reviewed and revised as needed by an expert. Even in cases where the automated segmentation requires revision, automation significantly reduces the time and labor associated with manual segmentation.

Finally, while the present study focused on the sigmoid sinus only, other temporal bone structures also need to be segmented for simulation purposes. Bony structures such as the middle-ear ossicles and cochlear shell can often be segmented using simple methods such as thresholding. Thresholding was not a viable option here due to the low contrast of the sigmoid sinus on the medial side. Nor is thresholding an option for other structures of the temporal bone that exhibit similar segmentation challenges to the sigmoid sinus such as the carotid artery, external auditory canal, tegmen, and the digastric ridge. Since capturing spatial inter-relationships between temporal bone structures using a few atlases is difficult because of the degree of variability, CNNs are utilized in Chapter 3 to segment all of these structures simultaneously while improving the portability of the solution.

2.6 Acknowledgments

This work was supported by the Otolaryngology Graduate Research Stipend from the Department of Otolaryngology - Head and Neck Surgery at Western University and by a Collaborative Health Research Project grant from the Canadian Institutes of Health Research and from the Natural Sciences and Engineering Research Council of Canada. Edits and review of the manuscript were performed by Lauren Siegel.

References

- [1] A. Arora et al., “Face and content validation of a virtual reality temporal bone simulator,” *Otolaryngol. - Head Neck Surg.*, vol. 146, no. 3, pp. 497–503, Mar. 2012.
- [2] C. Sewell et al., “Providing metrics and performance feedback in a surgical simulator,” *Computer Aided Surgery*, vol. 13, no. 2, pp. 63–81, 06-Jan-2008.
- [3] D. Morris, C. Sewell, F. Barbagli, K. Salisbury, N. H. Blevins, and S. Girod, “Visuohaptic simulation of bone surgery for training and evaluation,” *IEEE Comput. Graph. Appl.*, vol. 26, no. 6, pp. 48–57, Nov. 2006.
- [4] G. J. Wiet et al., “Virtual temporal bone dissection system: OSU virtual temporal bone system: Development and Testing,” *Laryngoscope*, vol. 122, no. SUPPL. 1, pp. S1–S12, Mar. 2012.
- [5] R. Varshney et al., “The McGill simulator for endoscopic sinus surgery (MSESS): a validation study,” *J. Otolaryngol. Head Neck Surg.*, vol. 43, no. 1, p. 40, Dec. 2014.
- [6] B. Tolsdorff et al., “Virtual reality: A new paranasal sinus surgery simulator,” *Laryngoscope*, vol. 120, no. 2, pp. 420–426, Feb. 2010.
- [7] M. A. Audette, H. Delingette, A. Fuchs, O. Burgert, and K. Chinzei, “A topologically faithful, tissue-guided, spatially varying meshing strategy for computing patient-specific head models for endoscopic pituitary surgery simulation,” in *Computer Aided Surgery, 2007*, vol. 12, no. 1, pp. 43–52.
- [8] S. M. Anil, Y. Kato, M. Hayakawa, K. Yoshida, S. Nagahisha, and T. Kanno, “Virtual 3-dimensional preoperative planning with the dextroscope for excision of a 4th ventricular ependymoma,” *Minim. Invasive Neurosurg.*, vol. 50, no. 2, pp. 65–70, Apr. 2007.
- [9] S. Weghorst et al., “Validation of the Madigan ESS simulator,” *Stud. Health Technol. Inform.*, vol. 50, pp. 399–405, 1998.
- [10] S. Chan, “CardinalSim.” [Online]. Available: <https://cardinalsim.stanford.edu/>. [Accessed: 19-Feb-2019].
- [11] R. Jackler, *Atlas of skull base surgery and neurotology*. New York: Thieme, 2009.
- [12] K. Van Osch, D. Allen, B. Gare, T. J. Hudson, H. Ladak, and S. K. Agrawal, “Morphological analysis of sigmoid sinus anatomy: Clinical applications to neurotological surgery,” *J. Otolaryngol. - Head Neck Surg.*, vol. 48, no. 1, p. 2, Dec. 2019.
- [13] K. A. Powell, T. Liang, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet, “Atlas-Based Segmentation of Temporal Bone Anatomy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 11, pp. 1937–1944, Nov. 2017.
- [14] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, “Multi-atlas segmentation with joint label fusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Mar. 2013.

- [15] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, Jul. 2009.
- [16] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Med. Image Anal.*, vol. 24, no. 1, pp. 205–219, Aug. 2015.
- [17] A. Fedorov et al., "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012.
- [18] M. Modat, D. M. Cash, P. Daga, G. P. Winston, J. S. Duncan, and S. Ourselin, "Global image registration using a symmetric block-matching approach," *J. Med. imaging* (Bellingham, Wash.), vol. 1, no. 2, p. 024003, Jul. 2014.
- [19] M. Modat et al., "Fast free-form deformation using graphics processing units," *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 278–284, Jun. 2010.
- [20] S. K. Warfield, K. H. Zou, W. M. Wells, S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation," *IEEE Trans Med Imaging*, vol. 23, no. 7, pp. 903–921, 2004.
- [21] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imaging*, vol. 15, no. 1, p. 29, Aug. 2015.
- [22] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *Neuroimage*, vol. 33, no. 1, pp. 115–126, Oct. 2006.
- [23] J. H. Noble, B. M. Dawant, F. M. Warren, and R. F. Labadie, "Automatic identification and 3D rendering of temporal bone anatomy," *Otol. Neurotol.*, vol. 30, no. 4, pp. 436–442, Jun. 2009.

Chapter 3

3 Automated Segmentation of Temporal Bones Structures using Convolutional Neural Networks

3.1 Introduction

Segmentation of the following temporal bone structures is important for mastoidectomy simulator training: the sigmoid sinus, carotid artery, external auditory canal, tegmen outlining the border of the dura mater, and digastric ridge. The sigmoid sinus, external auditory canal, and tegmen, along with other temporal bone structures have been segmented in previous works [1]-[5]; however, the results could be improved upon in terms of speed and accuracy. Accurate segmentations are not only important to improve mastoidectomy simulation, but also to use metrics for surgical assessment such as the Welling scale, which requires these structures to determine a score [6], [7]. Each of these structures present unique challenges for segmentation including dehiscence, differing scales, complex boundaries, and patient variation [8].

Automated image segmentation has been previously accomplished through a variety of methods, the simplest of which is thresholding. While thresholding is effective and quick when used for structures with high contrast and sharp edges from their surroundings, it has difficulty with more complex structures with poor intensity delineation as are common in the temporal bone. A more advanced and effective method that has been used in the context of temporal bone structures is atlas-based methods [1]-[3]. While atlas-based methods are far superior to thresholding for structures with less contrast, they require lengthy and difficult image registrations as well as large atlas databases on a user's computer to capture the anatomic variability that can be found in the temporal

bone. A proposed method to address these problems is with a convolutional neural network (CNN). CNNs are deep neural networks which use hierarchies to combine learned simple patterns to predict larger more complex ones. This ability to infer complex patterns as well as their ability to retain some spatial relations from the image makes CNNs suited to the task of image classification and segmentation. Multiple CNNs have previously been developed for the purpose of segmenting various parts of human anatomy from 3D medical image volumes [9]–[13].

The present work details the adaptation, evaluation, and comparison of two deep CNN architectures for the segmentation of the sigmoid sinus, carotid artery, external auditory canal, tegmen, and the digastric ridge from clinical computed tomography (CT) scans of cadaveric temporal bones. These structures were selected for their importance to the assessment of a finished mastoidectomy with the Welling scale. The networks used, DeepMedic and DenseVNet, have been previously shown to be effective for segmenting structures of other anatomical regions [12], [13].

3.2 Materials and Methods

3.2.1 Image Acquisition

Image acquisition was completed with the same method as Chapter 2. Thirty-eight anonymized adult cadaveric temporal bones with normal anatomy that had not been previously operated on were used to develop the algorithm. Clinical-CT scans of the samples were performed using a Discovery CT750 HD Clinical Scanner with GE's Gemstone CT detector at a resolution of $234\mu\text{m} \times 234\mu\text{m} \times 625\mu\text{m}$ and a voltage of 120kV. Micro-CT (μCT) images used for ground truth creation were obtained from the

same samples and scanned with a General Electric (GE) Healthcare eXplore Locus μ CT scanner at a resolution of $154\mu\text{m} \times 154\mu\text{m} \times 154\mu\text{m}$ and a voltage of 80kV.

The cadaveric specimens were obtained with permission from the body bequeathal program at Western University, London, Ontario, Canada in accordance with the Anatomy Act of Ontario and Western's Committee for Cadaveric Use in Research (approval number: #19062014).

3.2.2 Ground Truth Segmentations

The sigmoid sinus, carotid artery, external auditory canal, tegmen, and digastric ridge were segmented from micro-CT images by the lead author (DA) and an expert anatomist (KV) using a combination of semi-automated and manual tools in 3D Slicer, an open-source medical image processing software [14], [15]. The μ CT segmentations were rigidly registered with their corresponding clinical-CTs of the same sample to transfer the high-resolution ground truth segmentations. Consensus interpretation was provided by an experienced surgeon (SKA), and the lead author (DA).

3.2.3 Datasets for CNN Training, Validation, and Testing

The clinical-CT images were all normalized to intensity values between 0 and 1 prior to use in the training, validation, and testing of the networks. To improve accuracy the networks were trained, validated, and tested for left and right temporal bone samples separately. Both networks used the same samples for training, validation, and testing; approximately 70% were used for training, 10% for validation, and 20% for testing. This split allowed for enough data to test on despite the small total dataset. The validation set was used to monitor for potential overfitting and determine if early stopping was required

during the training of the networks. The left temporal bone training set consisted of 14 of the clinical-CT images, while the right temporal bone training set contained 12. The validation set consisted of two images per side, and the testing was done on four images per side.

3.2.4 Convolutional Neural Networks

Two deep convolutional neural networks, DeepMedic and DenseVNet, were optimized and trained. These networks were selected due to their significantly different design approaches as well as being originally developed for different anatomical regions. Both networks were implemented using NiftyNet, an open-source CNN platform for medical images built on TensorFlow [16].

3.2.4.1 DeepMedic

The first network used for comparison was DeepMedic [13]. DeepMedic is a 3D deep CNN originally designed for brain lesion segmentation that uses two resolution pathways followed by a conditional random field. During training, rotational augmentation of ± 10 degrees and scaling of $\pm 5\%$ was randomly applied to improve the robustness of the network. Window sampling was done with a spatial window size of $57 \times 57 \times 57$ voxels for the image and a spatial window size of $9 \times 9 \times 9$ voxels for the labels using a balanced method where each label had the same probability of being sampled. The downscaling factor for DeepMedic was set to 3. Dice loss was used as the loss function and optimized using the Adam method with a learning rate of 0.001 [11], [17]. The training was done on clinical-CT images and was run for 10000 iterations. L_2 regularization with a decay parameter of 0.00001 was applied to mitigate overfitting.

3.2.4.2 DenseVNet

The second network used for comparison was DenseVNet [12]. DenseVNet is a deep, fully convolutional neural network initially used for abdominal organ segmentation. Rotational and scaling augmentations of ± 10 degrees and ± 5 % were applied throughout the training. Window sampling was done with a spatial window size of $128 \times 128 \times 128$ voxels for the image and $128 \times 128 \times 128$ voxels for the labels and used the balanced sampling method. Dice loss was also used as the loss function and was optimized again using Adam. The training was run on the clinical-CT for 10000 iterations. L_2 regularization with a decay of 0.00001 was used for the same purpose described above.

3.2.5 Post-Processing

The output from the two networks was automatically post-processed using largest connected component island removal for all structures and morphological closing for the tegmen. Both operations were implemented using the Insight Toolkit (ITK) [18].

3.2.5.1 Largest Connected Component

After the networks produced the segmentations, island removal was performed to eliminate noise and disconnected portions. Noise is often created in the output of the networks due to misclassifications and presents as small clusters of incorrectly marked voxels. Since each structure is a single connected object, each label value had every component discarded except the largest. The largest connected component operation is very quick to apply and has been used as a final step in previous medical image segmentation applications [19]. Applying the largest connected component operation resulted in labels with no scattered noise.

3.2.5.2 Morphological Closing

Since the tegmen is less than 2 mm thick yet covers a relatively large surface area, it is common to get small holes in the tegmen's surface. To counteract this, after the largest connected component is applied the tegmen segmentation is finalized with a morphological closing operation with a spherical kernel with a diameter of approximately 5 mm (10 x 10 x 4 voxels at the clinical-CT resolution).

3.2.6 Evaluation and Metrics

Evaluation was completed with the use of quantitative metrics as well as visual inspection. Metrics were obtained by comparing the automated segmentation results to the ground truth manual segmentations created by the authors. The time required to produce a segmentation from a clinical-CT (time to segment) was also recorded.

The quantitative metrics used were the Dice coefficient, volumetric similarity, Hausdorff distance, and average Hausdorff distance (AHD). The Dice coefficient determines the overlap between two segmentations and is commonly used to evaluate medical image segmentations. The volumetric similarity metric provides a non-overlap-based comparison of the absolute volumes of two segmentations. The Hausdorff distance gives a value for the maximum distance from the automated to the ground truth segmentation, however it is extremely sensitive to both noise and cases where an algorithm segments more of the structure than was performed manually. While the largest connected component operation can mostly address the former, the latter would still present in the Hausdorff distance metric. Finally, the AHD metric provides a less sensitive approach to distance measurement and gives a mean of the Hausdorff distances for each point in the segmentation [20].

3.3 Results

The results of the DeepMedic network when run on the test set are shown in Table 3.1.

The trained DeepMedic network took approximately 1-2 minutes to create a segmentation from a test set clinical-CT when using a modern GPU for inference (estimated using a GTX1080Ti).

Table 3.1: Quantitative metrics for DeepMedic segmentation results. Means are given with standard deviations in parentheses.

	Dice	Volumetric Similarity	Hausdorff Distance [mm]	AHD [mm]
Sigmoid Sinus	0.72 (0.02)	0.91 (0.12)	18.81 (7.38)	0.74 (0.38)
Carotid Artery	0.20 (0.21)	0.22 (0.23)	40.90 (11.74)	7.76 (4.18)
External Auditory Canal	0.60 (0.02)	0.80 (0.16)	8.32 (1.96)	1.12 (0.57)
Tegmen	0.41 (0.12)	0.66 (0.21)	42.92 (18.60)	4.63 (2.19)
Digastric Ridge	0.02 (0.03)	0.12 (0.09)	47.00 (2.36)	10.79 (5.12)

All the metrics from the comparison of the DeepMedic segmentations to ground truth show reasonable values for the overlap, volume similarity, and distance metrics of the sigmoid sinus and external auditory canal but poor quantitative results were found for the carotid artery, tegmen, and digastric ridge. A visual example of the results from DeepMedic with 2D slices and a 3D rendering is shown in Figure 3.1.

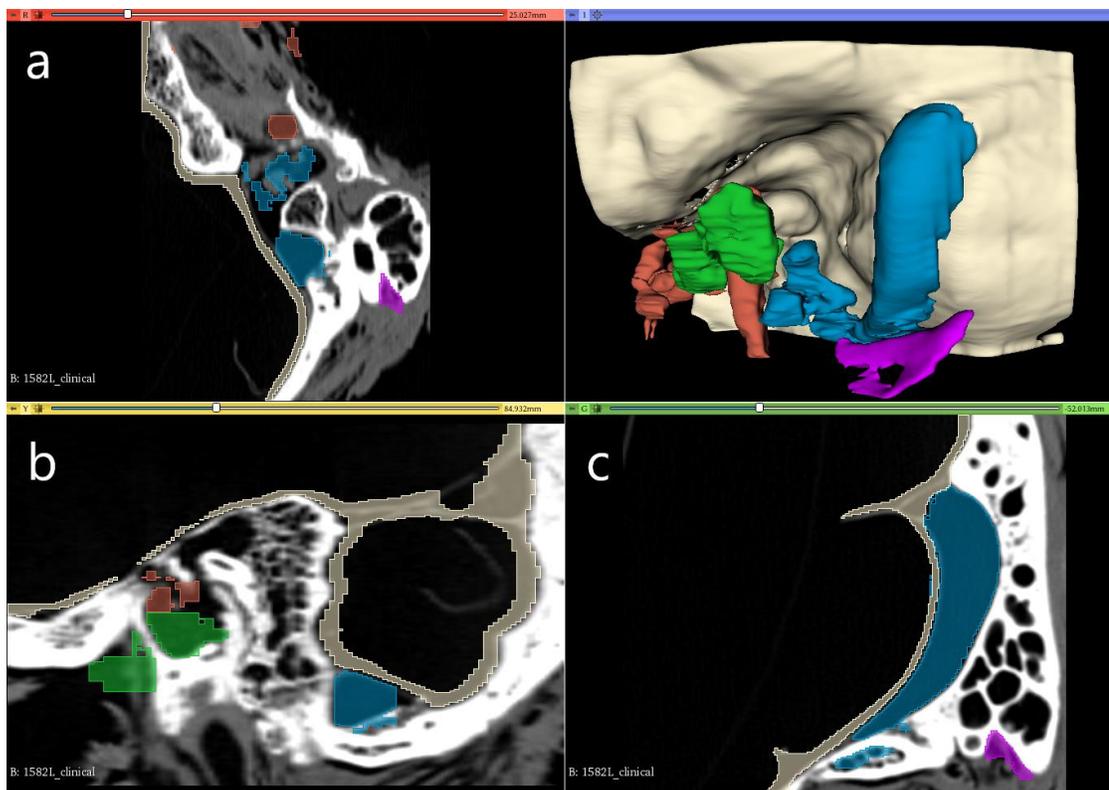


Figure 3.1: Screenshot from 3D Slicer of DeepMedic segmentation of a sample in three slice views (a axial, b sagittal, c coronal) and the structure segmentations rendered in 3D in the top right panel.

Visual inspection revealed that the DeepMedic implementation had issues segmenting many of the structures. The segmentation of the carotid artery had mistakenly filled many of the mastoid air cells; the tegmen had several holes in the surface; and the digastric ridge had been over segmented past its boundaries. Results for the sigmoid sinus and external auditory canal were better but still had issues in areas of low contrast with surrounding areas.

The metrics computed for the DenseVNet network are shown in Table 3.2. The trained DenseVNet also generally took approximately 1-2 minutes to generate a segmentation from a testing set clinical-CT.

Table 3.2: Quantitative metrics for DenseVNet segmentation results. Means are given with standard deviations in parentheses.

	Dice	Volumetric Similarity	Hausdorff distance [mm]	AHD [mm]
Sigmoid Sinus	0.85 (0.04)	0.96 (0.03)	11.41 (2.92)	0.37 (0.33)
Carotid Artery	0.77 (0.05)	0.94 (0.04)	3.97 (1.52)	0.19 (0.03)
External Auditory Canal	0.79 (0.06)	0.95 (0.04)	4.29 (0.37)	0.20 (0.11)
Tegmen	0.53 (0.14)	0.92 (0.07)	20.74 (6.19)	1.45 (0.15)
Digastric Ridge	0.51 (0.09)	0.56 (0.19)	3.76 (1.33)	1.67 (2.28)

The quantitative results of DenseVNet show high accuracy in overlap, similarity, and distance for all structures except the digastric ridge which had worse metrics than the other structures.

A visual example of the results from the DenseVNet with 2D slices as well and a 3D rendering is shown in Figure 3.2.

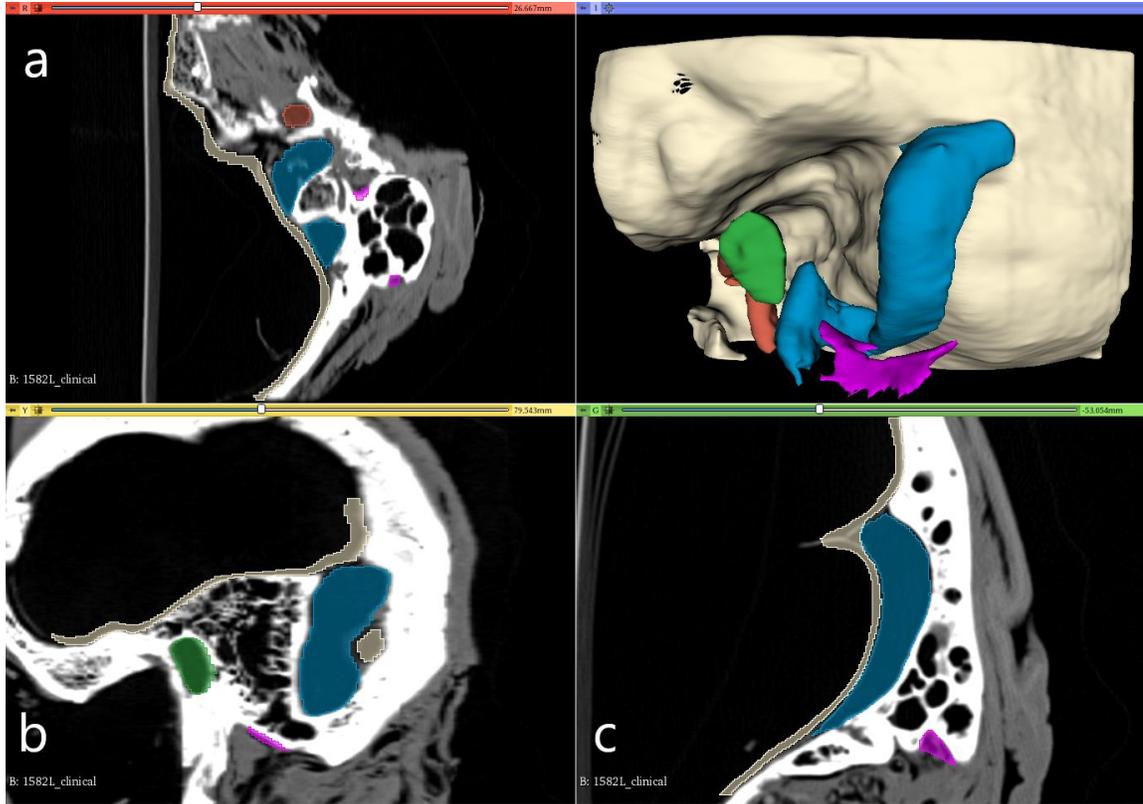


Figure 3.2: Screenshot from 3D Slicer of DenseVNet segmentation of a sample in three slice views (a axial, b sagittal, c coronal) and the structure segmentations rendered in 3D in the top right panel.

Visual inspection for the DenseVNet outputs revealed mostly high accuracy, clean segmentations except for the presence of occasional holes in the tegmen and inconsistent quality for the digastric ridge.

3.4 Discussion

The DenseVNet CNN produced far more accurate and consistent results across all structures in the test set overall when compared to the DeepMedic implementation. While the outputs from DenseVNet produced clean segmentations, DeepMedic had problems differentiating air cells in the mastoid from structures especially in the case of the carotid

artery. DeepMedic also had greater trouble with areas of low contrast. These issues resulted in significant noise in heavily pneumatized areas as well as overflowing segmentations in areas with poor contrast borders, making many of the test result segmentations unusable. DeepMedic was able to produce reasonably accurate segmentations of the sigmoid sinus and external auditory canal but was far outperformed by DenseVNet on all structures including the two mentioned. DeepMedic had an average Dice score of 0.39 across all structures compared to 0.69 with DenseVNet.

All temporal bone structures except for the digastric ridge had high accuracy results when using the DenseVNet. The digastric ridge is a uniquely challenging structure to segment due to its small size, lack of intensity delineation at its ends, proximity to the highly pneumatized portion of the mastoid, and its variability between samples. Furthermore, its small volume relative to other structures makes even small visual differences result in much lower Dice and volumetric similarity scores [20]. Improvements could be made with more data for training as well as potentially using a separate specific network with a region of interest cropped around the digastric ridge area, however this would increase the amount of manual intervention required.

Compared to previous works in the field of temporal bone image segmentation, the DenseVNet implementation was found to be an improvement over atlas and multi-atlas based methods for segmentation of the sigmoid sinus and external auditory canal. The multi-atlas based approach presented in Allen et. al [1] for segmenting the sigmoid sinus produced a mean Dice of 0.77 compared to a mean Dice of 0.85 in the present study. Powell et. al [3] reported a volumetric similarity of 0.75 for the sigmoid sulcus and 0.83 for the external auditory canal, and 0.82 for the tegmen when lefts and rights were

averaged together. The DenseVNet produced a mean volumetric similarity of 0.96 for the sigmoid sinus, 0.95 for the external auditory canal, and 0.92 for the tegmen.

While the majority of the segmentation results presented in the current paper were accurate, there is always a risk of error when using algorithmically generated models. Depending on the application, segmentations should be reviewed and potentially edited by an expert. However, while automated segmentations may occasionally require revisions, they still improve the process of segmentation considerably by reducing the amount of expert time and labor required.

The dataset used in this study was limited in size and because all data came from the same clinical-CT scanner. Future work may benefit from collecting data from different scanners and institutions which may improve the robustness, generalization, and potential accuracy of the automated networks.

3.5 Acknowledgements

This work was supported by the Otolaryngology Graduate Research Stipend from the Department of Otolaryngology - Head and Neck Surgery at Western University and by a Collaborative Health Research Project grant from the Canadian Institutes of Health Research and from the Natural Sciences and Engineering Research Council of Canada. Editing and review of the manuscript were performed by Lauren Siegel.

References

- [1] D. G. Allen, K. Van Osch, B. Gare, S. K. Agrawal, H. M. Ladak, “Automated Segmentation of the Sigmoid Sinus using a Multi-Atlas Approach,” *Biomed. J. Sci. Tech. Res.*, vol. 20, no. 4, pp. 15313–15319, 2019.
- [2] K. A. Powell, T. Liang, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet, “Atlas-Based Segmentation of Temporal Bone Anatomy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 11, pp. 1937–1944, Nov. 2017.
- [3] K. A. Powell, T. Kashikar, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet, “Atlas-based segmentation of temporal bone surface structures,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 8, pp. 1267–1273, Aug. 2019.
- [4] J. H. Noble, B. M. Dawant, F. M. Warren, and R. F. Labadie, “Automatic identification and 3D rendering of temporal bone anatomy,” *Otol. Neurotol.*, vol. 30, no. 4, pp. 436–442, Jun. 2009.
- [5] J. H. Noble, F. M. Warren, R. F. Labadie, and B. M. Dawant, “Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values,” *Med. Phys.*, vol. 35, no. 12, pp. 5375–5384, Nov. 2008.
- [6] D. Wan, G. J. Wiet, D. B. Welling, T. Kerwin, and D. Stredney, “Creating a cross-institutional grading scale for temporal bone dissection,” *Laryngoscope*, vol. 120, no. 7, pp. 1422–7, Jul. 2010.
- [7] N. N. Butler and G. J. Wiet, “Reliability of the Welling Scale (WS1) for Rating Temporal Bone Dissection Performance,” *Laryngoscope*, vol. 117, no. 10, pp. 1803–1808, Oct. 2007.
- [8] K. Van Osch, D. Allen, B. Gare, T. J. Hudson, H. Ladak, and S. K. Agrawal, “Morphological analysis of sigmoid sinus anatomy: Clinical applications to neurotological surgery,” *J. Otolaryngol. - Head Neck Surg.*, vol. 48, no. 1, p. 2, Dec. 2019.
- [9] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015.
- [10] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” Jun. 2016.
- [11] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” Jun. 2016.
- [12] E. Gibson et al., “Automatic Multi-Organ Segmentation on Abdominal CT With Dense V-Networks,” *IEEE Trans. Med. Imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.
- [13] K. Kamnitsas et al., “Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation,” Mar. 2016.

- [14] “3D Slicer.” [Online]. Available: <https://www.slicer.org/>. [Accessed: 15-Aug-2019].
- [15] A. Fedorov et al., “3D Slicer as an image computing platform for the Quantitative Imaging Network,” *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012.
- [16] E. Gibson et al., “NiftyNet: a deep-learning platform for medical imaging,” *Comput. Methods Programs Biomed.*, 2018.
- [17] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014.
- [18] “ITK - Segmentation & Registration Toolkit.” [Online]. Available: <https://itk.org/>. [Accessed: 15-Aug-2019].
- [19] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert, “Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm,” *Med. Image Anal.*, vol. 8, no. 3, pp. 255–265, Sep. 2004.
- [20] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool,” *BMC Med. Imaging*, vol. 15, no. 1, p. 29, Dec. 2015.

Chapter 4

4 Conclusions and Future Directions

4.1 Conclusions

This work detailed the development and testing of methods for automatically segmenting structures of the temporal bone for the purpose of generating accurate, patient-specific models from medical images for mastoidectomy simulation.

In Chapter 2, a novel multi-atlas based approach for segmenting the sigmoid sinus from clinical computed tomography (CT) images was described where multiple micro-CT atlases were registered to target clinical-CT images and then combined with label fusion methods. Developing an automated segmentation of the sigmoid sinus was a difficult task due to its variability [1], [2] and low contrast on the medial side. Three label fusion methods were compared to single-atlas techniques and it was found that the label fusion methods greatly improved quantitative and qualitative accuracy and consistency. Joint label fusion produced the best results which when compared to ground truth manual segmentations, with a Dice coefficient value of 0.77, a Hausdorff distance of 10.39 mm, and an average Hausdorff distance of 0.30 mm which is less than two voxels. Visual inspection revealed that the segmentations were accurate and high-resolution.

In Chapter 3, an accurate CNN-based approach for segmenting five of the temporal bone structures required for mastoidectomy simulation was developed. The five structures were the sigmoid sinus, carotid artery, external auditory canal, tegmen, and the digastric ridge, which had not been segmented automatically in previous works. The DenseVNet implementation produced superior results when compared to the implementation of

DeepMedic. While DeepMedic struggled to segment many of the structures mostly due to false positives, DenseVNet produced accurate quantitative metric results when compared to the ground truth segmentations. DenseVNet had mean values averaged across all test samples and structures of 0.69 for Dice, 0.87 volumetric similarity, 8.83 mm Hausdorff distance, and 0.77 mm average Hausdorff distance. The segmentation algorithm had a time to segmentation of approximately one to two minutes for a temporal bone image. When previously segmented structures were compared to previous works, the DenseVNet implementation had more accurate results.

4.2 Future Directions

Automated segmentation algorithms for medical imaging are constantly improving. As new and improved methods and architectures for learning how to segment anatomy are released, they should be assessed for their effectiveness in the temporal bone.

The automated algorithm described here would benefit from the addition of more relevant structures in the temporal bone such as the ossicles, cochlea, internal auditory canal, etc. This would further improve the 3D models available in a simulator for visualization and trainee assessment.

Increasing the amount and variety of temporal bone imaging data would improve the robustness and accuracy of the software when used at other institutions with different imaging protocols. Furthermore, adding diseased or abnormal anatomy to the dataset used for development would allow for automated segmentation of these special cases.

Many of the tools specific to the automated segmentation of temporal bone structures previously described in the literature are not available to the public either for download or

in open-source format. Making the trained models and code open-source and accessible to other groups would help to translate the research findings into real-life clinical use. In addition to making the algorithm code open-source, implementing an easy-to-use graphical user interface (GUI) for all major platforms would increase accessibility among users with less technological experience. This could be done either as a standalone program or as an extension within a freely available medical image processing platform such as 3D Slicer [3], [4]. Finally, integrating this segmentation software directly into a mastoidectomy simulator such as CardinalSim [5] along with automated assessment using the Welling scale [6] would allow for surgical simulation with assessment all within one application.

References

- [1] K. Van Osch, D. Allen, B. Gare, T. J. Hudson, H. Ladak, and S. K. Agrawal, “Morphological analysis of sigmoid sinus anatomy: Clinical applications to neurotological surgery,” *J. Otolaryngol. - Head Neck Surg.*, vol. 48, no. 1, p. 2, Dec. 2019.
- [2] P. B. Sarmiento and F. G. Eslait, “Surgical Classification of Variations in the Anatomy of the Sigmoid Sinus,” *Otolaryngol. Neck Surg.*, vol. 131, no. 3, pp. 192–199, Sep. 2004.
- [3] “3D Slicer.” [Online]. Available: <https://www.slicer.org/>. [Accessed: 15-Aug-2019].
- [4] A. Fedorov et al., “3D Slicer as an image computing platform for the Quantitative Imaging Network,” *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012.
- [5] S. Chan, “CardinalSim.” [Online]. Available: <https://cardinalsim.stanford.edu/>. [Accessed: 19-Feb-2019].
- [6] N. N. Butler and G. J. Wiet, “Reliability of the Welling Scale (WS1) for Rating Temporal Bone Dissection Performance,” *Laryngoscope*, vol. 117, no. 10, pp. 1803–1808, Oct. 2007.

Curriculum Vitae

Name: Daniel Gregory Allen, MEng.

Education:

2017 – Present M.Eng (Electrical & Computer Engineering)
Western University
London, Ontario, CA

2013 – 2017 B.Eng (Computer Engineering)
Western University
London, Ontario, CA

Honours and Awards:

2017 - 2018 Otolaryngology Graduate Research Scholarship

2017 University of Western Ontario Gold Medal in
Computer Engineering

2013 – 2017 Dean's Honour List

2014 NSERC Undergraduate Student Research Award

Related Work Experience:

2018 - 2019 Graduate Teaching Assistant (Introduction to
Electrical Engineering – ECE2238)

2017 - 2018 Graduate Teaching Assistant (Programming
Fundamentals for Engineers – ES 1036)

Publications:

1. M. Elfarnawany, A. Rohani, S. Ghomashchi, **D. Allen**, N. Zhu, S. K. Agrawal, and H. M. Ladak, "Improved middle-ear soft-tissue visualization using synchrotron radiation phase-contrast imaging," *Hear. Res.*, vol. 354, pp. 1-8, Oct. 2017.
2. K. Van Osch, **D. Allen**, B. Gare, T. J. Hudson, H. Ladak, and S. K. Agrawal, "Morphological analysis of sigmoid sinus anatomy: Clinical applications to

- neurotological surgery,” *J. Otolaryngol. - Head Neck Surg.*, vol. 48, no. 1, p. 2, Dec. 2019.
3. T. Hudson, B. Gare, **D. Allen**, H. Ladak, and S. Agrawal, “Intrinsic Measures and Shape Analysis of the Intratemporal Facial Nerve,” *J. Otolaryngol. – Head Neck Surg.*, Submitted Feb 25, 2019.
 4. A. Rohani, **D. Allen**, B. Gare, N. Zhu, H. Ladak, S. Agrawal, “High-Resolution Imaging of the Human Incudostapedial Joint Using Synchrotron-Radiation Phase-Contrast Imaging,” *Journal of Microscopy*, Submitted March 2019.
 5. B. Gare, T. Hudson, A. Rohani, **D. Allen**, S. Agrawal, H. Ladak, “Multi-Atlas Segmentation of the Facial Nerve from Clinical CT,” *International Journal of Computer Assisted Radiology and Surgery*. Submitted April 29, 2019.
 6. **D. G. Allen**, K. Van Osch, B. Gare, S. K. Agrawal, H. M. Ladak, “Automated Segmentation of the Sigmoid Sinus using a Multi-Atlas Approach,” *Biomed. J. Sci. Tech. Res.*, vol. 20, no. 4, pp. 15313–15319, 2019.
 7. **D. G. Allen**, K. Van Osch, H. M. Ladak, S. K. Agrawal, “Automated Segmentation of Temporal Bone Structures using Convolutional Neural Networks,” *International Journal of Computer Assisted Radiology and Surgery*. Submitting November 2019.

Conference Presentations:

* denotes presenting author

1. **D. Allen***, K. Van Osch, B. Gare, S. Agrawal and H. Ladak, "Automated Digital Modeling of the Sigmoid Sinus from 3D Micro-CT Images for Surgical Training," in *41st Midwinter Research Meeting of the Association for Research in Otolaryngology*, San Diego, CA, February 2018.
2. K. Van Osch*, **D. Allen**, B. Gare, T. Hudson, H. Ladak, S. Agrawal, “Morphological Analysis of Sigmoid Sinus Anatomy: Clinical Applications to Neurotological Surgery,” in *Canadian Society of Otolaryngology – Head and Neck Surgery Annual Meeting*, Quebec, QC, June 2018.
3. T. Hudson*, B. Gare, **D. Allen**, H. Ladak, S. Agrawal, “Segmentation of the Facial Nerve and other Temporal Bone Structures for Patient-Specific Simulation in Surgical Otology,” in *Canadian Society of Otolaryngology – Head and Neck Surgery Annual Meeting*, Quebec, QC, June 2018.