AN ACTIVE SHAPE MODEL FRAMEWORK FOR SEGMENTING COCHLEAR ANATOMY AND OTHER SMALL SHAPE LIBRARIES

By

Rueben Ansel Banalagay

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

December 17, 2022

Nashville, Tennessee

Approved:

Jack H. Noble, Ph.D.

Benoit M. Dawant, Ph.D.

Robert F. Labadie, M.D, Ph.D.

Ipek Oguz, Ph.D.

Don M. Wilkes, Ph.D.

TABLE OF CONTENTS

Page

LIST OF TABLES . . . . . **iv**

LIST OF FIGURES . . . . . **vi**

1 Introduction **1**

1.1 Overview 1
1.2 Generic Markers for CI Insertion 1
1.3 Segmentation of the Chorda Tympani 2
1.4 Validation of the Active Shape Model for Intra-Cochlear Segmentation 3
1.5 Limitations of the Active Shape Model 3
1.6 Organization of Work 4

2 Related Work **5**

2.1 Determining CI Placement from Cochlear Measurements 5
2.2 Segmentation of the Chorda Tympani 5
2.3 Segmentation of the Intra-Cochlear Anatomy 5
2.4 Variations on Statistical Shape Models 6
2.5 Modeling Shape 7
2.6 Modeling Shape Spaces 8
2.7 Shape Modeling and Small Sample Sizes 9

3 Insertion Depth for Optimized Positioning of Pre-Curved Cochlear Implant Electrodes . . **10**

3.1 Abstract 10
3.2 Introduction 10
3.3 Materials and Methods 13
3.3.1 Measuring Ideal Marker Depths Relative to Anatomical Landmarks 14
3.4 Results 15
3.5 Discussion 16
3.6 Appendix 18
3.6.1 Measuring distances from landmarks to markers 18

4 Segmentation of the Chorda Tympani Using a Weakly Supervised GAN **20**

4.1 Introduction 20
4.2 Related Work 21
4.3 Methods 22
4.3.1 Dataset 22
4.3.2 Network 22
4.3.3 Training 22
4.3.4 Testing 25
4.3.5 Implementation Details 25
4.3.6 Evaluation 25
4.4 Results 26
4.5 Discussion 26

5 Validation of Active Shape Model Techniques for Intra-Cochlear Anatomy Segmentation in CT Images . . . 29

5.1 Abstract . . . 29
5.2 Introduction . . . 29
5.2.1 Related Work . . . 31
5.3 Approach . . . 33
5.3.1 Overview . . . 33
5.3.2 ASM training . . . 33
5.3.3 ASM estimation . . . 34
5.3.4 Image segmentation with the ASM . . . 34
5.3.5 Evaluation metrics . . . 35
5.3.6 Parameter sensitivity study . . . 35
5.3.7 Library size study . . . 36
5.3.8 Robustness study . . . 36
5.4 Results . . . 36
5.4.1 Parameter sensitivity study . . . 36
5.4.2 Library size study . . . 37
5.4.3 Robustness study . . . 38
5.5 Discussion . . . 39

6 An Active Shape Model with Locally Weighted Elements . . . 47

6.1 Abstract . . . 47
6.2 Introduction . . . 47
6.2.1 Related Work . . . 49
6.3 Multi-Element ASM . . . 50
6.3.1 Overview . . . 50
6.3.2 ASM Element Creation . . . 51
6.3.3 Shape Estimation with the Multi-Element ASM . . . 52
6.3.4 Training the Full Multi-Element ASM . . . 53
6.3.5 Segmentation with the Multi-Element ASM . . . 54
6.3.6 Extension to Multi-Structure Settings . . . 54
6.4 Evaluation . . . 55
6.4.1 Synthetic Box Dataset . . . 55
6.4.2 Labyrinth Dataset . . . 56
6.4.3 Full Ear Dataset . . . 56
6.4.4 Chest Dataset . . . 57
6.4.5 Implementation Details . . . 58
6.5 Results . . . 58
6.5.1 Synthetic Box Dataset . . . 58
6.5.2 Labyrinth Dataset . . . 59
6.5.3 Full Ear Dataset . . . 59
6.5.4 Chest Dataset . . . 60
6.5.5 Model Training . . . 62
6.6 Discussion . . . 63

7 Conclusion . . . 78

References . . . 79

LIST OF TABLES

Table Page

2.1 Comparison of the different methods for segmenting intra-cochlear anatomy in pre-operative CT images, along with our optimized ASM. Methods are compared based on the number of training samples, surface errors, and the Dice score for each ST/SV. Mean surface errors are calculated from the aggregated ST/SV. (N/R) indicates not reported, while * indicates that the mean Dice score was reported instead of for individual ST/SV. 6

3.1 Percentage of failure cases (N=131) that would lead to insertions that are too deep (≤ -0.5 mm) or too shallow (≥ 0.5 mm) when aligning the secondary marker located 9.5mm from the E1 base medially $\overrightarrow{t}^{M}_{FN}$, flush $\overrightarrow{t}^{F}_{FN}$, or laterally $\overrightarrow{t}^{L}_{FN}$ to the FR. $\overline{\mathbf{p}}_{FN}$ shows failure rates if the secondary marker was instead placed at the mean FN distance and aligned flush with each patient's FN. 16

3.2 Table 2: Percentage of failure cases (N=131) that would lead to insertions that are too deep (≤ -0.5 mm) or too shallow (≥ 0.5 mm) when aligning the patient RW with the three depth markers $\overrightarrow{p}^{M}_{rw}$, $\overrightarrow{p}^{F}_{rw}$, and $\overrightarrow{p}^{L}_{rw}$ located at 2mm, 3mm, and 4mm from the E1 base. $\bar{t}^{F}_{RW}$ shows failure rates for a depth marker located at the average RW location and aligned with each patient's RW. 17

5.1 Tested and optimized parameter settings for the ASM 37

5.2 One-sided Wilcoxon signed rank p-values of the paired samples at $w = 1$ comparing ($\xi = rigid$, $\sigma = 2$) performing better than the other combinations of ξ and σ. Significant p-values ($p < .017$) are shown in bold 38

5.3 Results of the 11 cases where the optimized ASM produced a segmentation that differed from the original ASM segmentation by 0.5mm point-to-point when performing our robustness study over 137 cases. In bold are the 4 cases where the clinician preferred the result of the original ASM segmentation over the result produced by the new model. . . . 45

5.4 Comparison of the different methods for segmenting intra-cochlear anatomy in pre-operative CT images, along with our optimized ASM. Methods are compared based on the number of training samples, surface errors, and the Dice score for each ST/SV. Mean surface errors are calculated from the aggregated ST/SV. (N/R) indicates not reported, while * indicates that the mean Dice score was reported instead of for individual ST/SV. 46

6.1 ASM settings used for segmentation in our experiments. Numbers marked with * are multiplied by the pixel size per resolution level. 58

6.2 Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the synthetic boxes dataset. 0* indicates that the largest eigenmode already captures >95% of the variation in the model. 63

6.3 Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the labyrinth dataset 64

6.4 Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the full ear dataset in a "structure-agnostic" scenario 65

6.5 Comparision of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and an individual structure specific (SS) ASM on the full ear dataset in a "structure-specific" scenario 66

6.6 Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the chest dataset in a "structure-agnostic" scenario 66

6.7 Comparision of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and an individual structure specific (SS) ASM on the chest dataset in a ”structure-specific” scenario 66

LIST OF FIGURES

Figure Page

1.1 Example of electrode insertion depths that are too deep (left) and too shallow (right) for two patients in our database. The black electrode arrays are the actual electrode positions for these patients. The green electrode arrays show the positioning of the same array type that is optimized to well match the shape of the modiolar wall (indicated by blue curve) using the optimal insertion vector (orange). The actual array positions (black) are sub-optimal due to incorrect insertion depths. 2

1.2 3-D rendering of the facial recess with chorda and facial nerve labelled 2

1.3 Example of two independently rotating boxes (Top) A traditional ASM, due to the single linear vector space will produce complex eigenmodes and thus an incorrect fit when given a new example. (Bottom) A model involving two ASMs that utilize importance weightings on a different box will produce estimates that, when combined, will produce the desired fit. 4

3.1 Example of electrode insertion depths that are too deep (left) and too shallow (right) for two patients in our database. The black electrode arrays are the actual electrode positions for these patients. The green electrode arrays show the positioning of the same array type that is optimized to well match the shape of the modiolar wall (indicated by blue curve) using the optimal insertion vector (orange). The actual array positions (black) are sub-optimal due to incorrect insertion depths. 12

3.2 Example of differing optimal insertion depths for different insertion plans into the same cochlea. The light grey electrode array has an optimal insertion depth aligned with the central RW depth marker. However, a different insertion plan (darker grey) will have an optimal depth aligned with the distal depth marker 13

3.3 Facial nerve, scala tympani, scala vestibuli, and an electrode array placed in the ideal location. Relevant locations and coordinate frames for analysis are indicated in the figure. When calculating the position of $\overrightarrow{t}^{F}_{FR}$, the region of the FN searched as possible candidates are between the thicker dotted lines on the FN. Actual marker distances on E1 with respect to the array base are marked at the bottom. 14

3.4 Boxplots of distances from the most basal electrode to the anatomical locations of the FR and RW. For reference, the generic markers of $\overrightarrow{p}^{M}_{rw}$, and $\overrightarrow{p}_{FR}$ are marked. 16

4.1 Example 3-D rendering of facial recess, with chorda and facial nerve labelled. 20

4.2 Example image of a chorda in CT images (outlined in green). Note the faint appearance of the structure along with similar image features occurring from pneumatized bone . . . 21

4.3 Illustrations of the networks used in our GAN setup. Numbers above blocks indicate number of channels in that layer 23

4.4 Example images with corresponding probability maps for each channel. For visualization purposes, colored outlines in each image represent an isosurface of 0.5 from the probability maps. Note also how the contours in the weak labels are close, but not quite accurate to the true chorda position. 24

4.5 Boxplot of mean surface segmentation errors for the original automatic method (Original) and our proposed GAN method (Proposed) 26

4.6 Boxplot of paired differences (Original - Proposed) in surface segmentation error between the original automatic method and our proposed GAN method 27

4.7 Mean surface error for each vertex in our proposed segmentation method. Larger vertex indices indicate more inferior positions along the chorda. Bars represent +/- 1 standard deviation 28

4.8 Example segmentations of the best (top) and worst (bottom) cases from our proposed method. Green is ground truth, Red is the result from the original segmentation method, and blue is the result from our proposed method. For visualization purposes, we show each contour as a tube with radius of 0.1mm 28

5.1 Example μCT (left) and registered CT slice (right) showing ground truth segmentations of the scala tympani (red), scala vestibuli (blue) and modiolus (green). Labeled also is the Osseous Spiral Lamina (OSL) . . . 30
5.2 Example ASM segmentation with labeling of exterior (blue) vs.interior (red) points of the cochlea. Note that the exterior points generally correspond to stronger image gradients in the corresponding CT. . . . 31
5.3 Graphical overview of the experiments conducted in this study . . . 33
5.4 Boxplots of segmentation point-to-point errors for the parameter tuning study. Means are marked with green triangles. . . . 38
5.5 Boxplots of segmentation surface errors for the parameter tuning study. Means are marked with green triangles. . . . 39
5.6 Boxplots of segmentation Dice scores for the parameter tuning study. Means are marked with green triangles. . . . 40
5.7 Comparison of initialization positions for a typical case (left) and the observed outlier case in our parameter sensitivity study (right). Ground truth is shown in green while the initialization position is shown in red. . . . 41
5.8 Mean segmentation performance for the library size study at different percentage thresholds of model variation. (Clockwise from top left): Point-to-point errors, surface errors, Dice scores . . . 42
5.9 Mean fitting errors for the library size study at different percentage thresholds of model variation. (Left) Point-to-point errors (Right) Surface errors . . . 43
5.10 Mean total variance explained by an ASM for different training sizes. Total variance for an ASM is obtained by taking the sum of the eigenvalues in that model. Vertical bars represent 1 standard deviation. . . . 43
5.11 Mean number of eigenmodes present in an ASM when at different percentage thresholds of model variation. Vertical bars represent 1 standard deviation . . . 44
5.12 Heatmap of mean surface (top) and point-to-point errors (bottom) when validating 15 sample ASMs with our optimized settings from Table 5.1 . . . 44
5.13 Example segmentation slices from our robustness study where the new ASM segmentation differed by >0.5mm from the original ASM result. Segmentations shown in green were produced by the original automatic method and the segmentations produced by our optimized ASM are shown in red. (Left) A result where the clinician selected the new ASM segmentation. (Right) One of the 4 results where the clinician selected the original ASM segmentation. . . . 45

6.1 Example of two independently rotating boxes (Top) A traditional ASM, due to the single linear vector space will produce complex eigenmodes and thus an incorrect fit when given a new example. (Bottom) A model involving two ASMs that utilize importance weightings on a different box will produce estimates that, when combined, will produce the desired fit. 67
6.2 Mask utilized in the full ear dataset. As the lateral region of the ear canal is an arbitrary boundary, the points labeled in blue are not used in the training and evaluation stages of any of our models. . . . 68
6.3 Boxplot of fitting errors comparing a traditional vs. multi-element ASM for our boxes dataset. . . . 68
6.4 Boxplot of mean surface segmentation error for the labyrinth dataset . . . 68
6.5 Boxplot of surface segmentation error paired differences for the labyrinth dataset when comparing a multi-element ASM to a traditional ASM. More positive values indicate better performance for the multi-element model. . . . 69
6.6 Heatmap of mean paired surface errors comparing the 16-element ASM vs a traditional ASM for a training set size of 6 (top row) and 9 (bottom row). Each row shows two views of the labyrinth. Negative values (red hues) indicate better performance of a traditional ASM while positive values (blue hues) indicate better performance for the multi-element model. . . . 69

6.7 Boxplot of overall mean segmentation error when not using structure-specific ASMs (No) or using structure-specific ASMs (Yes) 70
6.8 Paired differences for the two structure-specific scenarios with the ear dataset. For the 'No' case, paired differences are compared with a traditional ASM. For the 'Yes' case, paired differences are calculated against a traditional ASM created for each individual structure. For both scenarios, more positive numbers indicate better performance of the multi-element ASM. Red asterisk indicates statistical significance ($p < .05$) 70
6.9 Boxplot of mean surface segmentation error for each structure in the ear when structure groupings of points are known 71
6.10 Boxplot of surface segmentation error paired differences between different structure-specific multi-element ASMs compared with a traditional ASM for each structure. More positive values indicate a better performance for the multi-element ASM. Red asterisk indicates significance at Bonferoni corrected level of ($p < 0.013$) 71
6.11 Example segmentation of the ossicle outlier case in our structure-specific scenario. Green is ground truth. Blue is the segmentation result of the multi-element ASM. Red is the result of the traditional structure-specific ASM for the ossicles. 72
6.12 Heatmap of mean paired surface errors comparing the 8-element structure-specific multi-element ASM to a traditional ASM targeting each individual structure. Negative values (red hues) indicate better performance of the individual structure specific ASM while positive values (blue hues) indicate better performance for the multi-element model. Regions of interest are circled in green. (Region A) Contact point of ossicles with labyrinth. (Region B) Contact point of ossicles with ear canal. (Region C) Body of facial nerve. (Region D) Inferior region of the ear canal. 72
6.13 Boxplot of surface segmentation differences between our 8-element structure-specific ASMs and the original automatic segmentations of our clinical dataset of 132 patient CTs. . . . 73
6.14 Example sagittal slices of the largest discrepancies in surface error between the original segmentations (Blue) and the 8-element structure specific segmentations (Orange) in our robustness test. 73
6.15 Mean surface segmentation errors for the parameter pilot study in our chest dataset when determining the optimal number of elements to use in a multi-element ASM for a structure agnostic case. 74
6.16 Overall mean surface segmentation errors for the parameter pilot study in the chest dataset when determining the optimal number of elements to use for the multi-element ASM in the structure specific case. 74
6.17 Boxplot of overall surface error comparing the multi-element ASM against its corresponding traditional ASM counterpart. (Left) A structure-agnostic scenario where the traditional ASM is compared against a multi-element ASM with the number of elements determined from the parameter pilot study. (Right) Structure specific scenario comparing traditional structure-specific ASMs against a 5-element multi-element ASM. 75
6.18 Boxplot of paired differences comparing the multi-element ASM against its corresponding traditional ASM counterpart. (Left) A structure-agnostic scenario where the traditional ASM is compared against a multi-element ASM with the number of elements determined from the parameter pilot study. (Right) Structure specific scenario comparing traditional structure-specific ASMs against a 5-element multi-element ASM. More positive values indicate better performance for the multi-element ASM. Red asterisk indicates statistical significance ($p < 0.05$) 75
6.19 Boxplot of mean surface error for each of the chest structures when comparing a 5-element structure specific ASM against a structure-specific traditional ASM for each structure. . . 76
6.20 Boxplot of paired differences comparing surface segmentation error of a 5-element structure specific ASM against a structure-specific traditional ASM for each structure. More positive values indicate better performance for the multi-element ASM. Red asterisk indicates significance under Bonferroni corrected ($p < .01$) 76

6.21 Heatmap of mean paired surface errors comparing the 5-element structure specific ASM and traditional structure-specific ASMs. Redder hues indicate better performance of the structure-specific traditional ASMs and bluer hues indicate better performance of the 5-element model. (Left) comparison of both models when using a training set size of 5 samples. (Right) comparison of both models when using a training set size of 10 samples. 77
6.22 Examples of training for each of the 4 datasets in this study. Left column is the initialization of the gaussian elements. Middle column is the result of the gaussian elements after the initial training stage. Colors shown in left/middle columns are the mixing weight of the global element in the model. Right column is the mixing weights for each of the models after the final stage of training. 77

CHAPTER 1

Introduction

1.1 Overview

The cochlear implant (CI) is a widely successful neural prosthetic device and is the preferred method of treatment for severe to profound hearing loss. Implants available today produce remarkable results for the vast majority of recipients with average postoperative word and sentence recognition approximating 60% and 70% correct, respectively, for unilaterally implanted recipients and 70% and 80% correct for bilateral recipients(1; 2; 3; 4; 5). However, patient outcomes remain highly variable, and there is a substantial fraction of individuals who experience poor speech recognition outcomes with CIs. CIs use an array of electrodes surgically implanted in the cochlea to directly stimulate the auditory nerve, inducing the sensation of hearing.

When electrodes are close to their neural stimulation sites, spread of excitation is relatively small, and the electrodes stimulate more localized regions of the auditory nerve. (6; 7) This is important because previous research (8; 9; 10; 11; 12; 13)has shown positioning the electrodes close to the nerves is one of several factors associated with better speech recognition with CIs. Other factors include demographics, such as younger age and shorter duration of hearing loss, as well as scalar location of the electrode array in the cochlea, where perimodiolar positioning within the scala tympani is associated with better outcomes for pre-curved arrays.

1.2 Generic Markers for CI Insertion

With the traditional surgical approach, the array is threaded into a small opening into the cochlea with most of its intra-cochlear path as well as its ultimate position in the cochlea blind to the surgeon. A system of markers on the array lead proximal to the electrodes is used to visually indicate when the generically recommended overall insertion depth of the array has been realized once the markers reach the cochlear entry site. However, these generic guidelines do not account for high variability in cochlear size and shape (8; 9). The result is that most pre-curved arrays are not well positioned.(13; 14) Techniques have been proposed for determining an optimal, patient-customized insertion vector and depth.(14; 15) The same studies also showed that using the patient-specific insertion depth was needed to avoid over- or under-insertion of the array.(14; 15) Under-inserted arrays result in shallow depth of the tip of the array and sub-optimal cochlear coverage, and over-inserted arrays have electrodes lifted away from the modiolar surface where they would be more effective (see Figure 1.1). Such sub-optimal positioning has been shown to be associated with poorer hearing outcomes.(12; 13)

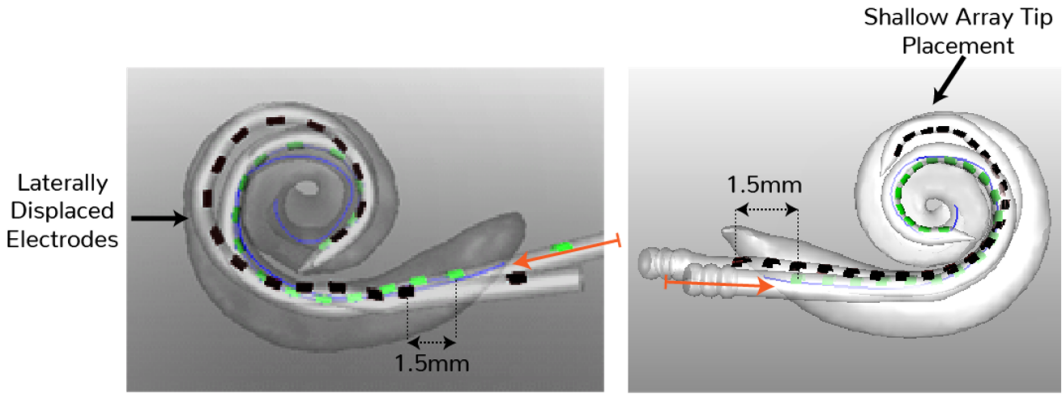

Figure 1.1: Example of electrode insertion depths that are too deep (left) and too shallow (right) for two patients in our database. The black electrode arrays are the actual electrode positions for these patients. The green electrode arrays show the positioning of the same array type that is optimized to well match the shape of the modiolar wall (indicated by blue curve) using the optimal insertion vector (orange). The actual array positions (black) are sub-optimal due to incorrect insertion depths.

1.3 Segmentation of the Chorda Tympani

Furthermore, when performing percutaneous cochlear access, a single hole through the skull surface is drilled to provide direct access to the cochlea where the CI can be threaded. The trajectory of this insertion typically involves passing through the facial recess, a region approximately 1.0–3.5 mm in width bounded posteriorly by the facial nerve and anteriorly by the chorda tympani (Figure 1.2). The determination of a safe drilling trajectory is highly important, as damage to these structures during surgery may result in a loss of taste (chorda) or facial paralysis (facial nerve). The use of image-guided techniques to determine safe trajectories therefore relies heavily on accurate segmentations of the chorda to operate effectively(15). Additionally, the understanding of inter-subject variability of spatial relationships between ear structures, such as the chorda, is helpful for the design of electrode array insertion guidelines for more optimal CI placement (16).

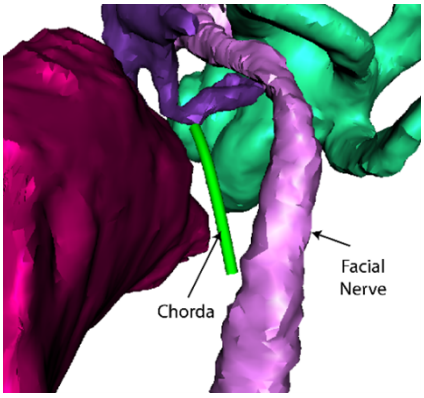

Figure 1.2: 3-D rendering of the facial recess with chorda and facial nerve labelled

1.4 Validation of the Active Shape Model for Intra-Cochlear Segmentation

Ultimately, the CI is placed within the cochlea with preference to the scala tympani as the location of the electrodes over the scala vestibuli. As such, proper segmentation of both scala is useful for creating insertion plans that achieve this desired result. Furthermore, understanding the shape of the intracochlear anatomy with respect to the locations of electrode positions is useful for determining on a patient specific basis the proper activation of electrodes in techniques such as image guided cochlear implant programming. To this end, active shape models (ASM) have been demonstrated by (17) to provide such intra-cochlear segmentations, even though clinical CT images normally lack the resolution to observe delineating features such as the osseous spiral lamina. The method has been shown to be quite robust, additionally being used in large scale studies to investigate the relationship between intra-cochlear electrode position and hearing outcomes (9) and for use in image guided cochlear implant planning and programming techniques (15; 18; 19). However, the ASM's ability to capture population variability, its quantitative accuracy, and the effectiveness of its parameters have not been fully explored and were evaluated on a limited dataset of only 5 μCT/CT pairs due to the limited availability of specimens.

1.5 Limitations of the Active Shape Model

Active shape models (20) (ASMs) themselves have been an effective technique for medical image segmentation, particularly in these situations with limited availability of training samples. To accomplish this goal, the basic framework of the ASM as described by Cootes (20) involves representing shape variations across all the global-pose-normalized training exemplars in a single, linear vector space. This resulting 'shape space' is then used in an iterative process that alternates between 1) searching for a noisy segmentation from an image search and 2) refining the segmentation to the closest 'plausible' shape as defined by the learned shape space.

While effective, complexities in shape variation that are not well suited to the described model creation process, such as non-linear shape variations or localized rotations, will be difficult to capture by the ASM, particularly in low-sample-size situations. Similar to (21), we show a hypothetical example of a shape comprised of two fixed location rectangles. Each rectangle contains a single mode of variation, but each box is allowed to rotate independently. The resulting shape variation modeled by the single linear vector space of a traditional ASM will be rather complex and have difficulty in producing an accurate fit to a new exemplar from the population. Meanwhile, a model composed of two weighted-ASMs could each place their importance weighting on a different rectangle in the shape. The resulting shape variations modeled by each ASM would be correct for the rectangle with the higher importance weighting, but still skewed for the lower importance weighted rectangle. However, when the estimates from each weighted-ASM are aggregated, the correct overall shape is produced by the combined model. (Figure 1.3)

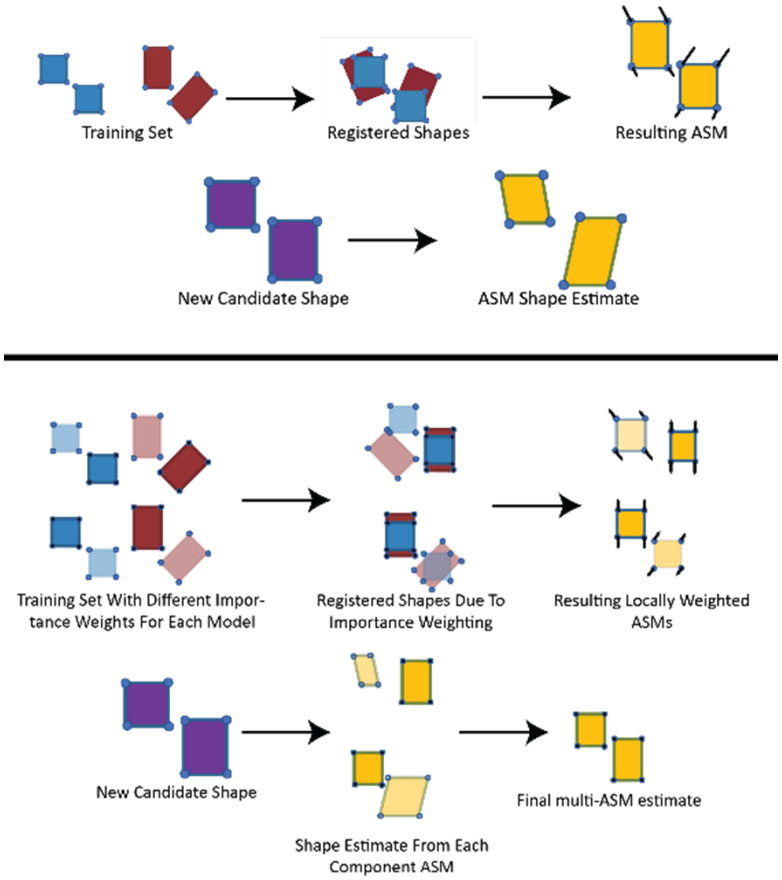

Figure 1.3: Example of two independently rotating boxes (Top) A traditional ASM, due to the single linear vector space will produce complex eigenmodes and thus an incorrect fit when given a new example. (Bottom) A model involving two ASMs that utilize importance weightings on a different box will produce estimates that, when combined, will produce the desired fit.

1.6 Organization of Work

The work presented in this dissertation addresses these stated issues regarding clinical outcomes of CI recipients. The organization of this work is presented as follows: Chapter 2 is an overview of related work regarding these topics. Chapter 3 aims to study the variability in the position of landmarks that can be used to determine insertion depth relative to the location of the modiolar wall which the array wraps around. Chapter 4 shows a generative adversarial network and weak supervision approach to chorda segmentation in clinical CT images. Chapter 5 is a validation study of the intra-cochlear ASM by leveraging an expanded dataset of 16 specimens. Chapter 6 shows a novel extension to the traditional ASM framework called the "multi-element ASM" for use in high-dimension low-sample-size scenarios. Finally, Chapter 7 presents the conclusions of this work.

CHAPTER 2

Related Work

2.1 Determining CI Placement from Cochlear Measurements

Several studies from multiple groups have studied the geometry of the various CI devices available and discuss implications for electrode positioning (22; 23). Particular interest has been placed on understanding correlations with angular insertion depth, with studies describing variation in angular depth of lateral wall and precurved electrodes (24) and others describing the insertion depth where surgeons encounter resistance. (25) Other studies quantify the effects of cochlear size on final insertion depth angle. (26; 27; 28; 29; 30; 31; 32; 17) Other studies propose methods to estimate cochlear duct length from the diameter of the basal turn in order to select the most appropriate CI array length (31).

2.2 Segmentation of the Chorda Tympani

Little related work exists for automatic segmentation of the chorda tympani. A semi-automatic method exists in (32) that utilizes probabilistic active shape models (PASM), but requires manual initialization by the user to select a few points on the chorda. (17) is the first fully automatic method for adults with (33) extending this work for paediatric cases. These utilize a minimum cost path to extract the medial-axis of the chorda based on trained models of intensity appearance. Finally, (34) uses a U-net architecture to initialize segmentation of the chorda and multiple other ear structures, and then refine those results with a PASM. The authors report low dice scores of <0.5 for the chorda, however a direct comparison with the results here are difficult, as we are more concerned with the location of the chorda rather than its very thin width, thus treating the chorda as a contour rather than a volume.

2.3 Segmentation of the Intra-Cochlear Anatomy

Many different approaches exist in the literature regarding image segmentation of the cochlea; see (35) for an extensive survey. Traditional image processing algorithms have been described for this purpose, with automatic region growing and level set methods in (36), and semi-automatic active contours and level sets utilized in (37) and (38; 39) respectively. Atlas based methods also exist in (40; 41) to predict the shape of the cochlea in the target image by projecting a known atlas segmentation through the registration transformation. As mentioned earlier, subtle details of the cochlea required for accurate segmentation are difficult to differentiate in clinical images, thus leading to the necessity of training a model to constrain the segmentation to be a valid shape. Using a shape prior, i.e. learning “plausible” shapes from training data, one can regularize

segmentations obtained from noisy images; the first formal application of this concept in cochlear segmentation utilized active shape models in (42) for pre-implantation CT's, with expansions to post-implantation CT's in (43) and a multi-region approach in (44). An evaluation of the active shape model's candidate search strategy was performed in (45). In addition, (46; 47) also use the idea of shape priors to constrain the solutions from their level set algorithm and a random walk model respectively, though their segmentation task is for the entire inner-ear rather than intra-cochlear anatomy. (48), on the other hand, uses statistical modeling to represent both shape and image appearance for use in intra-cochlear segmentation. Alternatively, instead of learning valid shapes, statistical approaches can learn valid deformations of a reference shape. Using a statistical deformation model with a dataset of 18 samples, (49) reports the best Dice scores and max surface errors, but their samples are comprised of much higher resolution and contrast CT images than our study. Finally, deep learning has been a popular approach in other image processing applications, but its requirement of large training sizes poses a difficulty in this domain; cochlear specimens, particularly those containing a high resolution μCT, remain difficult to obtain and manually segment. Nonetheless, (50) shows a successful application of deep learning in cochlear segmentation, producing the best mean surface error, even when utilizing a limited training set. However, the failure cases displayed in that approach produce shapes that are clearly impossible, whereas statistical models will, by construction, always produce a valid shape even in the worst case. A comparison of the various statistical modeling approaches for segmenting intra-cochlear anatomy in pre-operative CTs can be seen in Table 2.1.

2.4 Variations on Statistical Shape Models

Different approaches to statistical shape modeling exist in the literature, with numerous survey papers (51; 52; 53; 54; 55; 56; 57; 58; 59; 60; 61; 62) covering statistical shape models and the many variations arising out of each aspect of the shape modeling pipeline. At its core, the active shape model (ASM) proposed (20) and applied to cochlear segmentation in (42) rely on creating a global shape model with linear shape variations.

Method	Number of Samples	ST Dice Score	SV Dice Score	Mean Surface Error (mm)
(42)	6	0.77	0.72	0.21
(48)	9	0.77	0.73	0.12
(49)	18	0.88*	0.88*	0.11
(50)	11	0.87	0.86	0.08

Table 2.1: Comparison of the different methods for segmenting intra-cochlear anatomy in pre-operative CT images, along with our optimized ASM. Methods are compared based on the number of training samples, surface errors, and the Dice score for each ST/SV. Mean surface errors are calculated from the aggregated ST/SV. (N/R) indicates not reported, while * indicates that the mean Dice score was reported instead of for individual ST/SV.

Subsequent variations on shape modeling, such as the one proposed in this work, focus on capturing non-linear shape variations with the goal of increasing model flexibility. We highlight the different methods under their mathematical representations of shapes and shape spaces.

2.5 Modeling Shape

While some methods deal with shape through direct interaction with the images and deformation fields (63; 64; 65; 66; 67; 68; 69), most variations on shape representation still consider using original surface representing the modeled object. A natural extension to a simple point based model would be to use different families of functions to approximate the object's surface (70; 71; 72; 73; 74; 75; 76; 77). Furthermore, domain constraints can be integrated through the particular choice of shape representation; (78) explicitly considers the selection of their PDM landmarks to account for the modeling of tubular structures, while derived features from the original vertices provide anatomical constraints (79) or particular classes of transformation invariance (80; 81). Geometric constraints are implicitly modeled in (82; 83; 84) by considering the landmark connectivity of the original shape as the vertices and edges in a probabilistic graphical model.

Though the current work uses the original 3-D landmarks as shape representation, we note the use of alternate shape encoding methods attempt to also model local and global shape variations; (75) and (76) capture local and global structure by respectively utilizing polynomials and localized kernels in their surface parameterization. Other approaches to simultaneous global/local shape modeling utilize a coarse-to-fine strategy; wavelet methods (68; 85; 86; 87; 88; 89; 90; 91; 92; 93; 94; 95) are a common example, as localized decompositions similar to Fourier analysis are used. Alternatively, medial-representations (m-reps) (96; 97; 98; 99; 100; 101; 102; 103) represent shape as a skeleton of points with linkages that traverse the interior of the object; depending on the scale and type of structure under analysis, different numbers of linkages can be used to capture different levels of variation. (100; 102) also extend this m-rep idea to multi-organ structures.

Regardless of the shape encoding method used, a local/global understanding of shape is captured by some methods with a hierarchical approach. Models that utilize a multi-resolution approach (88; 90; 100; 104; 105; 106; 107; 108; 109; 110; 111; 112) exemplify this philosophy; local variations are captured by models at finer resolutions while global shape dependencies are captured at coarser resolutions.

Within these hierarchical approaches, methods that are particularly relevant to the current work explicitly try to model localized 3-D regions of shape anatomy. (111) utilize a piecewise approach that groups together regions with similar variation, while (104; 112), use polyaffine trees to break the anatomy into different hierarchical clusters based on the singular value decomposition of the 3-D spatial arrangement of the landmark points. Furthermore, hierarchical clustering approaches have also focused on creating clusters of salient landmarks (110) or shape points (105; 90).

2.6 Modeling Shape Spaces

As mentioned earlier, in the original active shape model, shapes were assumed to be drawn from a single gaussian in a linear subspace. Relaxing either of these constraints leads to another avenue through which statistical shape models achieve additional flexibility. Spaces modeled using polynomials (113; 114), Hilbert spaces (115), or different non-linear manifolds (116; 117; 118; 119; 120; 121; 122; 123; 124) have been proposed to relax the constraint of the linear subspace. Other approaches focus on altering the distribution of shapes by creating alternate probability models using gaussian mixtures (115; 125; 126), Boltzman machines (127; 128; 129), other Bayesian-style approaches (116; 69; 77), or the integration of prior knowledge (130; 131). Deep learning methods to shape modeling and image segmentation (129; 132; 133; 134; 135; 136; 137; 138) have also been applied, but without special treatment, generally still require many samples to be effective.

An indirect approach to modeling the underlying shape space, as is taken in this work, is to modify the importance weighting of landmarks during the modeling and segmentation phases of the active shape model. A mass-spring system (139) uses weighting to model geometric relationships, while local image appearance (140; 141; 142) drives the weights to mitigate outliers in the segmentation process. (143) shares the most similarity with our approach of using importance weighting to achieve spatial locality; however, their approach considers localized weighting at each landmark in the shape and only during the segmentation process. Similarly, (144) applies localized weighting for both training and segmentation in a multi-organ framework, but considers only considers local regions as defined by individual organs instead of arbitrarily sized and located regions in 3-D space.

Our method uses the original decomposition of the shape space using principal components analysis (PCA). We further mention other variations on shape space decomposition utilized by other methods. As the general goal of shape space decomposition is dimensionality reduction, sparsity techniques (145; 146; 147; 148; 149; 150; 151; 152; 153; 154; 155; 156) are a natural approach. Additionally, PCA itself has been generalized to tensors (157; 68), kernels (158; 159), or geodesics (64; 96; 98; 101; 102; 103; 118; 160), as well as a localized PCA model (161), robust variants (162; 163), or decompositions involving eigenfunctions (63). Related decompositions of maximum autocorrelation analysis (164) and canonical correlation analysis has also been used on its own to model unobserved landmarks (165) or with a generalized PCA to model multi-organ structures (166). (126) also uses a mixture of PCA models but still performs the analysis within the context of a hierarchical framework.

2.7 Shape Modeling and Small Sample Sizes

Because some medical applications have difficulty obtaining samples for analysis, an additional complication is introduced of trying to model high dimensional complex structures with very limited data. While our method indirectly addresses this problem by finding compact representations of local regions with relatively simple radial basis functions, other methods address this problem from a data augmentation approach; (167) compares some of these techniques with respect to cardiac data. An example approach to data augmentation involves generating additional examples by creating perturbations in the original training set mesh (168; 169; 170). In contrast, (171) shuffles local patches from different training samples to create new examples. Other methods of data augmentation try to combine datasets with no point correspondence (172) or from different modalities (173) to increase the dataset size. Meanwhile, the method in (174) mitigates the problem of high dimensionality by modeling shape with as few manual landmarks as possible.

CHAPTER 3

Insertion Depth for Optimized Positioning of Pre-Curved Cochlear Implant Electrodes

This work was published in: Banalagay RA, Labadie RF, Chakravorti S, Noble JH. Insertion Depth for Optimized Positioning of Precurved Cochlear Implant Electrodes. Otol Neurotol. 2020 Sep;41(8):1066-1071. doi: 10.1097/MAO.0000000000002726. PMID: 32569133; PMCID: PMC8054969.

3.1 Abstract

Hypothesis: Generic guidelines for insertion depth of pre-curved electrodes are sub-optimal for many individuals.

Background: Insertion depths that are too shallow result in decreased cochlear coverage, and ones that are too deep lift electrodes away from the modiolus and degrade the electro-neural interface. Guidelines for insertion depth are generically applied to all individuals using insertion depth markers on the array that can be referenced against anatomical landmarks.

Methods: To normalize our measurements, we determined the optimal position and insertion vector where a pre-curved array best fits the cochlea for each patient in an IRB approved, N=131 subject CT database. The distances from the most basal electrode on an optimally placed array to anatomical landmarks, including the round window (RW) and facial recess (FR), was measured for all patients.

Results: The standard deviations of the distance from the most basal electrode to the FR and RW are 0.65mm and 0.26 mm, respectively. Due to the high variability in FR distance, using the FR as a landmark to determine insertion depth results in >0.5mm difference with ideal depth in 44% of cases. Alignment of either of the two most proximal RW markers with the RW would result in over-insertion failures for >80% of cases, whereas the use of the third, most medial marker would result in under-insertion in only 19% of cases.

Conclusions: Normalized measurements using the optimized insertion vector show low variance in distance from the basal electrode position to the RW, thereby suggesting it as a better landmark for determining insertion depth than the FR.

3.2 Introduction

The cochlear implant (CI) is a widely successful neural prosthetic device and is the preferred method of treatment for severe to profound hearing loss. Implants available today produce remarkable results for the vast majority of recipients with average postoperative word and sentence recognition approximating 60% and 70% correct, respectively, for unilaterally implanted recipients and 70% and 80% correct for bilateral

recipients (1; 3; 4; 5). However, patient outcomes remain highly variable, and there is a substantial fraction of individuals who experience poor speech recognition outcomes with CIs. CIs use an array of electrodes surgically implanted in the cochlea to directly stimulate the auditory nerve, inducing the sensation of hearing. The types of electrode arrays available today can be broadly divided into two classes – straight (aka lateral wall) and precurved (aka perimodiolar). Straight arrays position the electrodes along the outer ('lateral') wall of the cochlea, whereas precurved arrays are advanced into the cochlea off of a straightening stylet, or out of a straightening sheath, and coil to attempt to match the shape of the inner ('modiolar') wall of the cochlea. The coiled shape of precurved arrays is designed to position the electrodes against the modiolus because this is where the auditory nerves are housed. When electrodes are close to their neural stimulation sites, spread of excitation is relatively small, and the electrodes stimulate more localized regions of the auditory nerve. (6; 7) This is important because previous research (8; 9; 10; 11; 12; 13) has shown positioning the electrodes close to the nerves is one of several factors associated with better speech recognition with CIs. Other factors include demographics, such as younger age and shorter duration of hearing loss, as well as scalar location of the electrode array in the cochlea, where perimodiolar positioning within the scala tympani is associated with better outcomes for pre-curved arrays.

With the traditional surgical approach, the array is threaded into a small opening into the cochlea with most of its intra-cochlear path as well as its ultimate position in the cochlea blind to the surgeon. A system of markers on the array lead proximal to the electrodes is used to visually indicate when the generically recommended overall insertion depth of the array has been realized once the markers reach the cochlear entry site. However, these generic guidelines do not account for high variability in cochlear size and shape (8; 9). The result is that most pre-curved arrays are not well positioned. (13; 14) Techniques have been proposed for determining an optimal, patient-customized insertion vector and depth.(14; 15) The approach is to align a geometric model of an electrode array with the patient's scala tympani, estimated using the pre-operative CT, such that the array positioning best agrees with the inner wall of the scala tympani and uses a round window or extended round window entry site. Example results of this process are shown in Figure 4. The aligned model shows the position where the natural shape of the array is in best agreement with the shape of the inner wall of the patient's ST. The depth of the aligned model thus implies the insertion depth that should be used for this patient's unique cochlea shape when the array insertion vector is collinear with the base of the aligned array model. The same studies also showed that using the patient-specific insertion depth was needed to avoid over- or under-insertion of the array. (14; 15) Under-inserted arrays result in shallow depth of the tip of the array and sub-optimal cochlear coverage, and over-inserted arrays have electrodes lifted away from the modiolar surface where they would be more effective (see Figure 3.1). Such sub-optimal positioning has been shown to be associated with poorer hearing outcomes. (12; 13)

In this work, we aim to study the variability in the position of landmarks that can be used to determine insertion depth relative to the location of the modiolar wall which the array wraps around. This relationship is important because if, for example, the distance from the modiolar wall to a depth landmark is highly variable across patients, then inserting the precurved array until a generic depth marker reaches the landmark will lead to over- or under-insertion for many individuals. Conventionally, the round window is used as a landmark, which is an obvious choice as this is where the array enters the cochlea. As such, all currently available electrode arrays are equipped with visual markers that the surgeon can reference against the round window landmark for insertion depth information. A recently marketed array, the Cochlear ™ (Sydney, Australia) 532/632 electrode array (referred to as "E1" in the remainder of the manuscript), is equipped with both a round window marker as well as another more proximal marker used to deploy the electrode from the straightening sheath, which often aligns with the facial recess when the array is fully inserted. We have learned through discussions among surgeons from several institutions that some are relying on the alignment of this proximal marker with the facial recess, or using a notch in the facial recess narrower than the protruding marker to secure the marker just medial or lateral to the facial recess, in order to determine final insertion depth. Thus, in this study, we investigate the reliability of using the facial recess, as well as positions lateral and medial to the facial recess, as a depth landmark in comparison with the round window. Inter-patient variability in the relative locations of these landmarks could motivate the use of patient-customized insertion depths. Conversely, landmarks found to have low variability with the optimal depth would suggest that the same insertion depth can be generically applied to all patients.

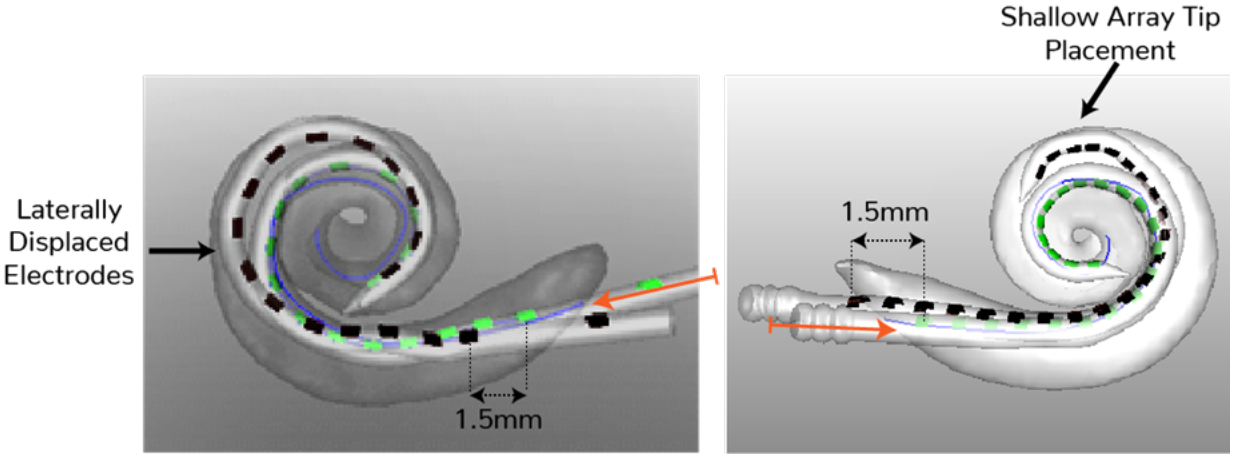

Figure 3.1: Example of electrode insertion depths that are too deep (left) and too shallow (right) for two patients in our database. The black electrode arrays are the actual electrode positions for these patients. The green electrode arrays show the positioning of the same array type that is optimized to well match the shape of the modiolar wall (indicated by blue curve) using the optimal insertion vector (orange). The actual array positions (black) are sub-optimal due to incorrect insertion depths.

3.3 Materials and Methods

In this study, we rely on a retrospective dataset of CT scans of 131 cochlear implant patients obtained with local IRB approval to study variability in optimal insertion depth of pre-curved arrays. The CTs were pre-implantation CTs obtained with standard multi-slice temporal bone protocol acquisitions with voxel size around 0.25 x 0.25 x 0.4 mm^3. For each CT scan, we used previously published "active shape model" (20) methods to accurately localize the scala tympani (ST), scala vestibuli, and modiolus (42); and other model-based techniques (17) to localize the FN for each of the patient CTs. These methods have been validated extensively and shown to be highly accurate.

Since the modiolar wall has complex 3D geometry, it is not possible to directly measure variability in distance between it and the round window or facial recess in a meaningful way. Thus, in this work, we leverage the previously proposed approach to align a geometric model of an electrode array with the patient's scala tympani. (15) This approach defines the optimal depth as the depth of the aligned model. We can then measure variability of the position of depth landmarks to this optimized electrode position for individual patients. It is of note that optimal electrode depth can differ for a specific cochlea when varying entry site or insertion angle. For example, in Figure 3.2 the central marker would represent the preferred depth to the round window when inserting the array through the center of the round window approaching from a superior angle, whereas the distal marker would be preferred for an extended round window insertion from an inferior angle. Herein, we define the optimal depth as the one when using the optimized insertion vector as defined above.

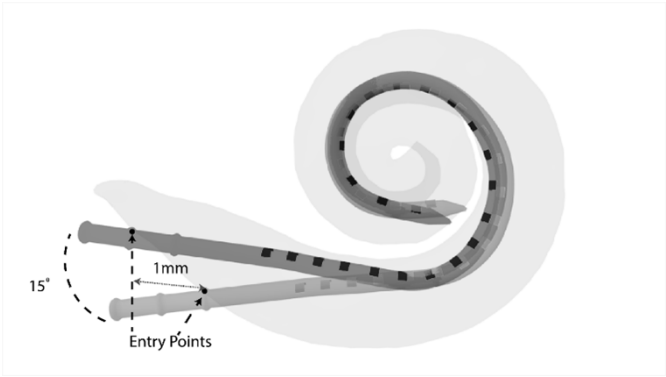

Figure 3.2: Example of differing optimal insertion depths for different insertion plans into the same cochlea. The light grey electrode array has an optimal insertion depth aligned with the central RW depth marker. However, a different insertion plan (darker grey) will have an optimal depth aligned with the distal depth marker

We align a shape model of array E1 with the cochlear modiolus in each of the 131 CT scans in our dataset. E1 is outfitted with three insertion depth markers that the surgeon can choose to align with the round window (RW) cochlear entry site once the generically recommended insertion depth is reached (see Figure

3.3). These markers are white in color and do not protrude from the lead on E1, but we indicate their position using protruding ribs in the 3D surface models shown in the figures in this paper. E1 is also equipped with a secondary marker that we hypothesize might well align with the facial recess (FR) when the array has reached the generically recommended depth of insertion (Figure 3.3).

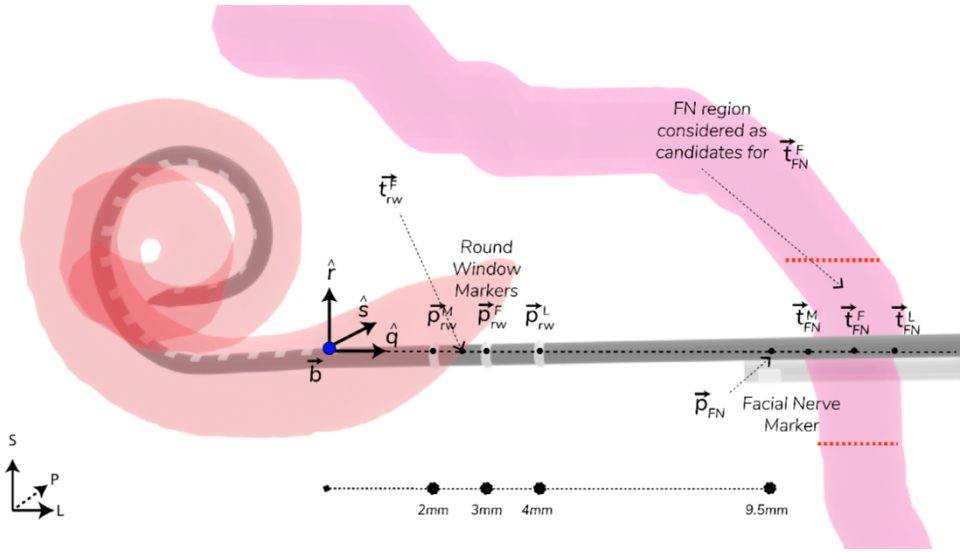

Figure 3.3: Facial nerve, scala tympani, scala vestibuli, and an electrode array placed in the ideal location. Relevant locations and coordinate frames for analysis are indicated in the figure. When calculating the position of $\vec{t}^{F}_{FR}$, the region of the FN searched as possible candidates are between the thicker dotted lines on the FN. Actual marker distances on E1 with respect to the array base are marked at the bottom.

3.3.1 Measuring Ideal Marker Depths Relative to Anatomical Landmarks

A detailed diagram of all relevant locations and measurements can be found in Figure 3.3. To compare the ideal depth of insertion to the generic depth recommended using the RW markers, we simply measure the distance from each of the medial, $\vec{p}^{M}_{rw}$, central, $\vec{p}^{F}_{rw}$, and lateral, $\vec{p}^{L}_{rw}$, RW markers on the registered E1 model to the point on the electrode array axis $\hat{q}$ closest to the center of the RW membrane of the ST, $\vec{t}_{RW}$. To compare ideal depth of insertion to the depth indicated by aligning the FR marker, $\vec{p}_{FR}$, with the center of the FR, or 1mm medial or 1 mm lateral to this position, we measure the distance from $\vec{p}_{FR}$ to the point on the electrode array axis $\hat{q}$ closest to the center of the FR, $\vec{t}^{F}_{FR}$, as well as 1mm medial, $\vec{t}^{M}_{FR}$, or 1 mm lateral, $\vec{t}^{L}_{FR}$, to this position along $\hat{q}$. While the FR is shaped by surgical resection and thus is subject to inter-surgeon variability, we assume that positioning the notched marker medial or lateral to the facial recess will need to fall at least 1mm medial or lateral to the facial nerve. Mathematical details about how these landmarks and marker positions are found are included in the Appendix. Because the distances are measured relative to positions along $\hat{q}$, we can measure signed distances, where positive distances indicate the generically recommended depth is lateral to ideal (which would lead to an insertion that is too shallow),

and negative distances indicate the generically recommended depth is medial to ideal (which would lead to an insertion that is too deep).

We then measure the mean and standard deviation of these signed distances across our entire dataset. As such, a mean close to zero for a given marker would indicate that it is, on average, very close to a landmark when optimally placed. More importantly, however, a low standard deviation for a given marker would indicate that its position relative to a landmark is highly consistent when optimally placed. A marker with a consistent position relative to a landmark can serve as a reliable depth indicator even if its mean distance to the landmark is non-zero.

3.4 Results

Mean +/- standard deviation of the distances from the most basal electrode on the ideally positioned E1 to the FR and RW sites were 9.49mm +/- 0.65mm and 2.28 +/- 0.26mm respectively. Given the FR marker is 9.5 mm and the RW markers are 2, 3, and 4 mm from the most basal electrode on E1, the FR and RW markers are, on average, very well located to align with both anatomical landmarks. However, it is of note that the standard deviation is much larger for FR distance than RW distance. Thus, the RW is a more reliable anatomical landmark than the FR for determining insertion depth.

In previous temporal bone studies where a custom depth relative to the RW was planned, it was found that an experienced surgeon could achieve a preplanned depth accurate to within 0.5mm on average (15). Thus, we consider a failure case to be one in which the use of a generic depth marker would lead to absolute depth difference from ideal exceeding 0.5mm, and thus could be improved with a custom depth. A count of failure cases when using each of the marker options is presented in Tables 3.1 and 3.2. As the mean distance $\overline{\left\|\overrightarrow{t}^{F}_{FR} - \overrightarrow{b}\right\|}$ from the base electrode to the FR across all cases is in almost exact agreement with the generic marker distance (9.5 mm), the number of over and under-inserted failure cases are approximately the same. However, the effects of a larger variance in $\overrightarrow{t}^{F}_{FR}$ are apparent, as even when simulating a marker placed at the average FR distance, using the FR as a depth landmark would result in failures in 44% of cases. This percentage of failure cases further increases to 79% when lateral $\overrightarrow{t}^{L}_{FR}$ or medial $\overrightarrow{t}^{M}_{FR}$ positioning of the FR marker is used. On the other hand, the lower variance of RW distances allow many fewer failures, as only 5% of cases would fail when using a simulated marker placed at the average RW distance $\overline{\left\|\overrightarrow{t}_{RW} - \overrightarrow{b}\right\|}$. If using the generic markers, $\overrightarrow{p}^{F}_{rw}$ or $\overrightarrow{p}^{L}_{rw}$, insertion depths tend to be too deep, with over-insertion occurring in 80% and 100% of cases, respectively. Meanwhile, using $\overrightarrow{p}^{M}_{rw}$ results in a 19% under-insertion failure rate.

In 3.4, we show boxplots of the distances of the most basal electrode to the FR, $\overrightarrow{t}^{F}_{FR}$, and RW, $\overrightarrow{t}_{RW}$, for each patient. For both RW and FR, the distributions are highly symmetric with very few outliers. Also shown for reference are the distances of both the medial RW $\overrightarrow{p}^{F}_{rw}$ and secondary $\overrightarrow{p}_{FR}$ markers of E1. As can be

seen in the figure and discussed in the last paragraph, positioning the marker flush with the FR is already in good agreement with the average ideal position. Therefore, positioning the current secondary marker medial or lateral to the facial nerve (with no slack in the array lead) would produce less than ideal results. Lastly, the generic location of $\overrightarrow{p}^F_{rw}$, is not well positioned with respect to the average RW depth. An alignment of the RW between $\overrightarrow{p}^F_{rw}$ and $\overrightarrow{p}^M_{rw}$ at 2.28mm from the array base electrode would agree more with the dataset average.

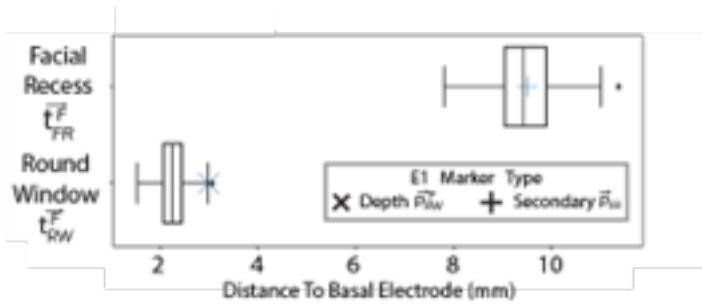

Figure 3.4: Boxplots of distances from the most basal electrode to the anatomical locations of the FR and RW. For reference, the generic markers of $\overrightarrow{p}^M_{rw}$, and $\overrightarrow{p}_{FR}$ are marked.

3.5 Discussion

Several studies from multiple groups have studied the geometry of the various CI devices available and discuss implications for electrode positioning (22; 23). Particular interest has been placed on understanding correlations with angular insertion depth, with studies describing variation in angular depth of straight and precurved electrodes (24) and others describing the insertion depth where surgeons encounter resistance. (25) Other studies quantify the effects of cochlear size on final insertion depth angle. (26; 27; 28; 29; 30) Other studies propose methods to estimate cochlear duct length from the diameter of the basal turn in order to select the most appropriate CI array length (31).

In contrast with prior work, in this work, we have investigated variance in ideal electrode insertion depth of pre-curved arrays when using optimized electrode entry site and insertion vectors. We have found that overall, the RW is a more reliable positioning landmark than the FN because variability in the position of the

Failure Type	$\vec{t}^M_{FN}$	$\vec{t}^F_{FN}$	$\vec{t}^L_{FN}$	$\bar{p}_{FN}$
Deep	*78%*	*21%*	*1%*	*21%*
Shallow	*1%*	*22%*	*78%*	*23%*

Table 3.1: Percentage of failure cases (N=131) that would lead to insertions that are too deep (≤ -0.5 mm) or too shallow (≥ 0.5 mm) when aligning the secondary marker located 9.5mm from the E1 base medially $\overrightarrow{t}^M_{FN}$, flush $\overrightarrow{t}^F_{FN}$, or laterally $\overrightarrow{t}^L_{FN}$ to the FR. $\bar{\mathbf{p}}_{FN}$ shows failure rates if the secondary marker was instead placed at the mean FN distance and aligned flush with each patient's FN.

RW relative to the ideal electrode position is much smaller than variability in the position of the FN (Figure 3.4). Due to this contrast in variance, a 5% failure rate is estimated to occur if using a RW marker placed at the ideal generic position as opposed to a 44% failure rate when using a FR marker at the ideal generic position (Table 2). While our analyses rely on one particular array (E1) as a reference, these data are indicative of the effect of general variability of available anatomical landmarks (the round window and the facial recess) relative to the modiolar wall, and thus are applicable for other precurved electrode arrays in general. While the result that the RW is more reliable a landmark than the FR is intuitive, this is the first study to the best of our knowledge to confirm this intuition, and to provide data informing surgeons who may be using the FR as a depth landmark.

When using an ideal electrode insertion trajectory, which is associated with reduced rates of scalar translocation with other precurved arrays (15), the population mean ideal depth to the RW for the array is ˜2.28mm from the most basal electrode, which corresponds to aligning the RW closer to the distal than the middle marker. For the FR marker, positioning it laterally to the facial nerve would lead to few overinsertion errors but would often lead to underinsertion and reduced cochlear coverage. However, positioning the FR marker at the mid-point of the FN is often difficult as it's notched nature (being used to deploy off of the straightening stylet), causes it to rest either medial or lateral to the nerve. One limitation of this work is that we assume a medial or lateral positioning of the FR marker to be +/- 1mm to the facial nerve. In reality this is likely an underestimation in the variability of these points since the boundaries of the FR are created by surgical resection. As such, the reported variability in depth when placing the marker lateral or medial to the facial recess is likely underestimated.

In a number of cases, the optimal insertion vector was very close to, and in some cases slightly intersected, the facial nerve. The closest realizable trajectory can be found by adjusting the insertion vector angle in the anterior direction towards the facial recess. Fortunately, optimal depth is likely highly insensitive to slight adjustment of this vector in the anterior direction when using the pre-planned optimal entry site because this will cause the array to deflect slightly on the floor of the scala tympani upon entering the cochlea. As the scala tympani is narrow in the anterior-posterior direction, small deflections from the floor of the scala

Failure Type	$\vec{t}^{M}_{RW}$	$\vec{t}^{F}_{RW}$	$\vec{t}^{L}_{RW}$	$\bar{t}_{RW}$
Deep	*0%*	*80%*	*100%*	*2%*
Shallow	*19%*	*0%*	*0%*	*3%*

Table 3.2: Table 2: Percentage of failure cases (N=131) that would lead to insertions that are too deep (≤ -0.5 mm) or too shallow (≥ 0.5 mm) when aligning the patient RW with the three depth markers $\overrightarrow{p}^{M}_{rw}$, $\overrightarrow{p}^{F}_{rw}$, and $\overrightarrow{p}^{L}_{rw}$ located at 2mm, 3mm, and 4mm from the E1 base. $\bar{t}^{F}_{RW}$ shows failure rates for a depth marker located at the average RW location and aligned with each patient's RW.

tympani likely do not significantly impact the intra-cochlear path of the array.

While information presented in this work is directly relevant with the use of an ideal insertion trajectory, ideal depth is likely highly sensitive to the entry site and angle of the insertion vector in the superior-inferior direction. As shown in Figure 3.2, with slight adjustment of the entry site and superior-inferior insertion angle, optimal insertion depth changes dramatically. Thus, while ideal depth is relatively consistent when implementing an optimized, pre-planned insertion vector, this property may not hold when not using pre-planned trajectories. In separate work we are investigating preplanning of the optimal insertion vector. The results of this work imply that when using the optimized vector, in 95% of cases a generic insertion depth relative to the RW would be successful. This represents an important step towards optimizing placement of the electrodes. It is possible that the same insertion depth is also optimal when using non-optimized insertion trajectories, but this will need to be investigated in future work.

3.6 Appendix

3.6.1 Measuring distances from landmarks to markers

To determine coordinate locations for $\overrightarrow{t}^F_{FR}$, $\overrightarrow{t}^L_{FR}$, and $\overrightarrow{t}^M_{FR}$, we define a coordinate system using three orthonormal vectors $\widehat{q}$, $\widehat{r}$, $\widehat{s}$ originating at the location of the most basal electrode $\overrightarrow{b}$ of the registered E1 model. $\widehat{q}$ is defined as a unit vector pointing along the axis of the electrode array laterally towards the facial recess. $\widehat{r}$ is defined as a vector orthonormal to $\widehat{q}$ pointing in the superior direction. $\widehat{s}$ is defined as a vector orthonormal to $\widehat{q}$ and $\widehat{r}$ pointing in the posterior direction.

$\overrightarrow{t}^F_{FR}$, the location along $\widehat{q}$ that is even with the middle of the FR in the medial-to-lateral direction is found by finding the closest point along $\widehat{q}$ to the facial nerve by solving:

$$\min_{a,\overrightarrow{f_i}} \left| \overrightarrow{b} + a\widehat{q} - \overrightarrow{f_i} \right|^2$$

where $\overrightarrow{f_i}$ is a vertex on the mesh of the FN. To consider only the portion of the FN adjacent to the facial recess when finding $\overrightarrow{t}^F_{FR}$, the elements $\overrightarrow{f_i}$ are constrained to be those vertices in the FN that fall within 20° of the plane normal to $\widehat{r}$ and containing $\overrightarrow{b}$. The closed form solution for a is given by

$$a = \left(\overrightarrow{f_i} - \overrightarrow{b} \right) \cdot \widehat{q}$$

Finally, the coordinate location for $\overrightarrow{t}^F_{FR}$ is defined as
$\overrightarrow{t}^F_{FR} = \overrightarrow{b} + a\widehat{q}$.

In cases where the $\widehat{q}$ axis intersects the facial nerve, a is chosen to correspond to the midpoint between the entry and exit intersection points of the $\widehat{q}$ axis with the facial nerve.

The distances from the array base location, $\overrightarrow{b}$, to the $\overrightarrow{p}_{FR}$ FR marker location is 9.5 mm and to the $\overrightarrow{p}^{M}_{rw}$, $\overrightarrow{p}^{F}_{rw}$, and $\overrightarrow{p}^{L}_{rw}$ RW marker locations are 2, 3, and 4 mm along $\widehat{q}$, the ideal insertion vector. Thus, the signed distance between a generic RW marker location, $\overrightarrow{b}+s^{\alpha}\widehat{q}$, and the ideal location $\overrightarrow{t}_{RW}$ is $\left(\overrightarrow{t}_{RW}-\overrightarrow{b}\right)\cdot\widehat{q}-s^{\alpha}$, where $\alpha\in[M,F,L]$, $s^{[M,F,L]}=[2,\ 3,\ 4]$, and $\cdot$ indicates the vector dot product. Similarly, the signed distance between the ideal FR marker locations and the generic recommendations are $\left(\overrightarrow{t}^{\alpha}_{FR}-\overrightarrow{b}\right)\cdot\widehat{q}-9.5$.

CHAPTER 4

Segmentation of the Chorda Tympani Using a Weakly Supervised GAN

4.1 Introduction

Cochlear implants (CI) are a widely successful neural-prosthetic device for improving quality of life for individuals experiencing severe to profound hearing loss. These devices operate by directly stimulating the nerve fibres inside the cochlea to create the sensation of hearing. As such, the placement of these devices within the cochlea requires surgical procedures. One such minimally invasive technique, percutaneous cochlear access, involves drilling a single hole through the skull surface to provide direct access to the cochlea where the CI can be threaded. The trajectory of this insertion typically involves passing through the facial recess, a region approximately 1.0–3.5 mm in width bounded posteriorly by the facial nerve and anteriorly by the chorda tympani (Figure 4.1). The determination of a safe drilling trajectory is highly important, as damage to these structures during surgery may result in a loss of taste (chorda) or facial paralysis (facial nerve). The use of image-guided techniques to determine safe trajectories therefore relies heavily on accurate segmentations of the chorda to operate effectively(15). Furthermore, the understanding of inter-subject variability of spatial relationships between ear structures, such as the chorda, is helpful for the design of electrode array insertion guidelines for more optimal CI placement (16).

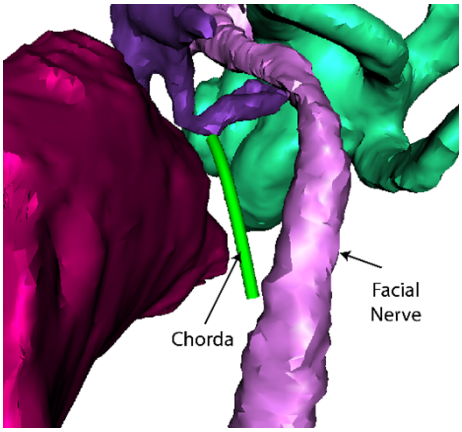

Figure 4.1: Example 3-D rendering of facial recess, with chorda and facial nerve labelled.

Automated methods have been developed in previous work (17; 33) for the segmentation of these structures. These techniques have been integrated into our current clinical workflow which, while helpful, still typically requires some manual intervention by the user to correct some inaccuracies from the segmentation result. In particular, the segmentation of the chorda tympani is a difficult problem for traditional image pro-

cessing techniques, as its thin nature leads to a very faint appearance in CT scans and can easily be confused with air pockets from the surrounding pneumatized bone. (Figure 4.2)

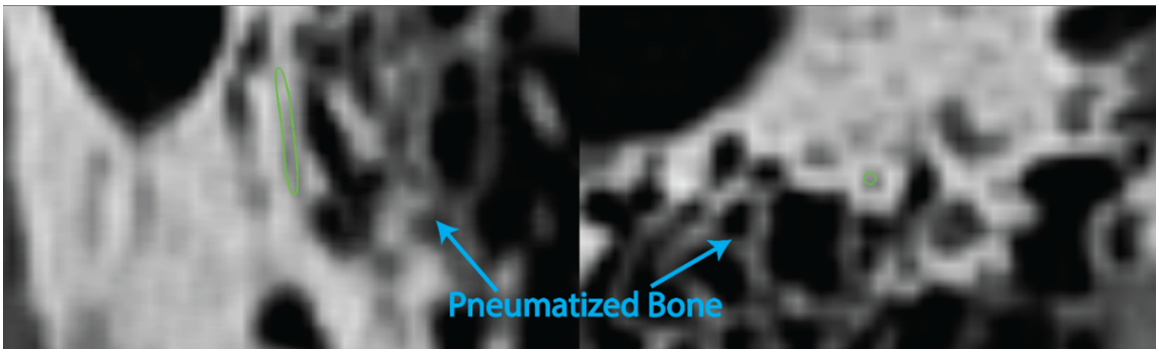

Figure 4.2: Example image of a chorda in CT images (outlined in green). Note the faint appearance of the structure along with similar image features occurring from pneumatized bone

The use of deep learning techniques, particularly generative adversarial networks (GAN), may provide some benefit in this area, as they have been successfully employed to a wide range of segmentation tasks in other medical image domains. However, the typical issue with employing a deep learning solution is the requirement of sufficiently large quantities of labelled training data. This can be problematic, as the creation of high-quality manual segmentations of this structure quickly becomes tedious and time-consuming. However, the acquisition of weakly-labelled data that is only of moderate quality, can easily be attained in large quantities from the use of other automated techniques such as atlas-based methods. Therefore, we propose using weak supervision to augment a relatively small dataset of manually segmented images with automatically segmented images from our large patient CT database. The goal of this work is to leverage this idea, along with a GAN architecture to create an improved automatic segmentation method for the chorda tympani.

4.2 Related Work

Little related work exists for automatic segmentation of the chorda tympani. A semi-automatic method exists in (32) that utilizes probabilistic active shape models (PASM), but requires manual initialization by the user to select a few points on the chorda. (17) is the first fully automatic method for adults with (33) extending this work for paediatric cases. These utilize a minimum cost path to extract the medial-axis of the chorda based on trained models of intensity appearance. Finally, (34) uses a U-net architecture to initialize segmentation of the chorda and multiple other ear structures, and then refine those results with a PASM. The authors report low dice scores of <0.5 for the chorda, however a direct comparison with the results here are difficult, as we are more concerned with the location of the chorda rather than its very thin width, thus treating the chorda as a contour rather than a volume.

4.3 Methods

4.3.1 Dataset

Our overall dataset consists of 146 patient CT scans (anisotropic voxel size ˜0.3mm). To create our strongly labelled dataset we create a contour representing the medial-axis of the chorda for 16 of these images. We put these segmentations in point correspondence by evenly resampling 20 points on a spline interpolated representation of these contours. These curves then serve as the ground-truth segmentations for our dataset. One of these samples is reserved as an atlas image that is utilized in the training for all methods but excluded from testing.

The remaining 130 images are then used as our weakly labeled dataset. To create these weak segmentations, we take the image from the atlas sample and automatically register it to each of these CT scans using an existing mutual-information based approach (175; 176). The ground-truth contour from the atlas sample is then projected through the found transformation to create our weak segmentation labels.

4.3.2 Network

Our network architecture remains consistent with the traditional GAN approach of training two competing networks: a generator and discriminator. The generator network uses a U-net variant that employs residual blocks as described in (177). For the discriminator network, we employ an architecture that is the same as the encoder portion of our U-net. Detailed illustrations of each of the networks can be found in Figure 4.3.

4.3.3 Training

Regardless of whether the sample is weak or strongly labeled, the input image $I \in \mathbf{R}^{1x32x32x32}$ to the generator is a single channel 32x32x32 region of interest (ROI) around the sample's chorda segmentation with anisotropic voxel size of .3125x.3125x.625. Data augmentation is achieved by random translation and rotation of this ROI with respect to the original image. The output of the generator $I^G \in \mathbf{R}^{3x32x32x32}$ is a 3-channel probability map each with dimension 32x32x32. The first channel in I^G is a probability map of the chorda body while the second and third channels are probability maps of the superior and inferior ends of the chorda respectively. Example images and corresponding probability map for each of the channels can be found in Figure 4.4. While ultimately, we will only use the body channel in the creation of our final segmentation contours, we found during our testing that adding the extra tasks of localizing the endpoints of the chorda was helpful for the overall chorda segmentation process.

The target of the generator $I^{target} \in \mathbf{R}^{3x32x32x32}$ is created from the image sample via distance maps. Let $\overrightarrow{x}_{i,j,k} \in R^3$ be the 3-D world coordinates of the $(i,j,k)^{th}$ voxel in I. The probability value for voxel (i,j,k) in channel 0 is defined as:

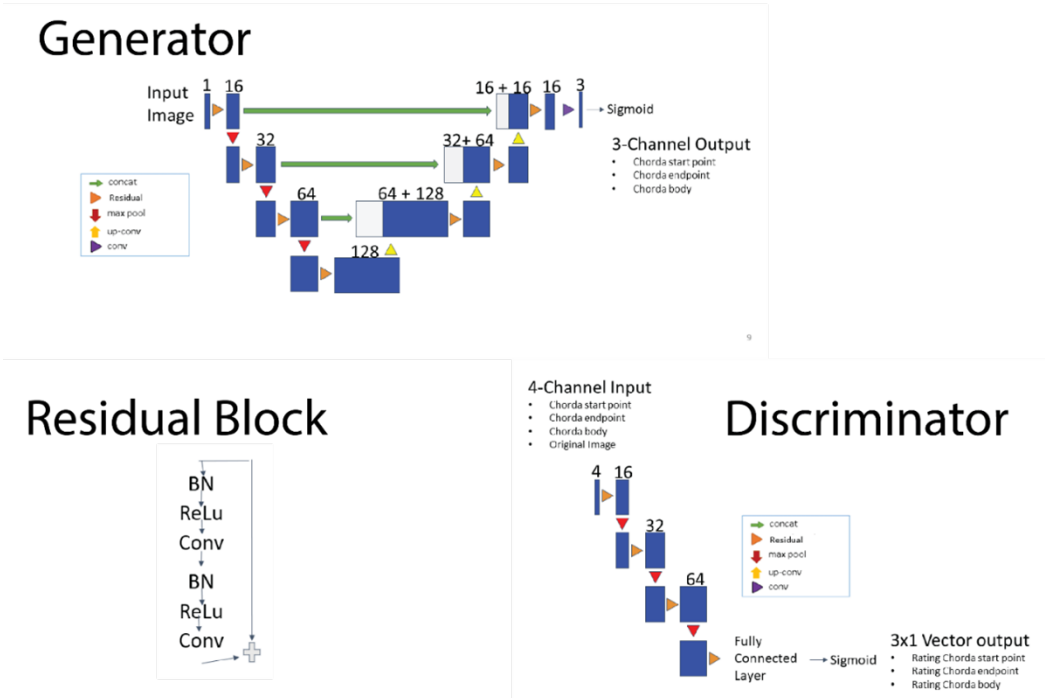

Figure 4.3: Illustrations of the networks used in our GAN setup. Numbers above blocks indicate number of channels in that layer

$$I_0^{target}(i,j,k) = \exp\left(-0.5 * —\overrightarrow{x}_{i,j,k} - \overrightarrow{s}_{close}—^2\right)$$

Where $\overrightarrow{s}_{close}$ is the closest location on the contour to $\overrightarrow{x}_{i,j,k}$. The values the other channels, I_1^{target} and I_2^{target} are defined similarly, with $\overrightarrow{s}_{close}$ being replaced with the fixed 3-D coordinates of the most superior and inferior vertices on the contour respectively. The final target I^{target} of the generator network is then defined as the concatenation of I_0^{target}, I_1^{target}, and I_2^{target} along the channel dimension.

The loss function for the generator is then defined as

$$L^G = k * BCE_{weak}\left(I^G, I^{target}\right) + BCE_{strong}\left(I^G, I^{target}\right) - D\left(\left[I, I^G\right]\right)$$

Where $BCE_{weak,strong}(\cdot,\cdot)$ are the binary cross entropy loss functions depending on whether I was taken from the weak or strongly labelled dataset. $[\cdot,\cdot]$ represents the concatenation operator along the channel dimension, and therefore $D\left(\left[I, I^G\right]\right) \in \mathbf{R}$ is the output of the Discriminator network when given the concatenation of I and I^G as input. Since we know that the probability maps that are created from the weakly labelled dataset are not entirely accurate, we do not wish to incentivize the generator to completely match those probability maps. Thus, we down weight the value of BCE_{weak} in the loss function with a user defined

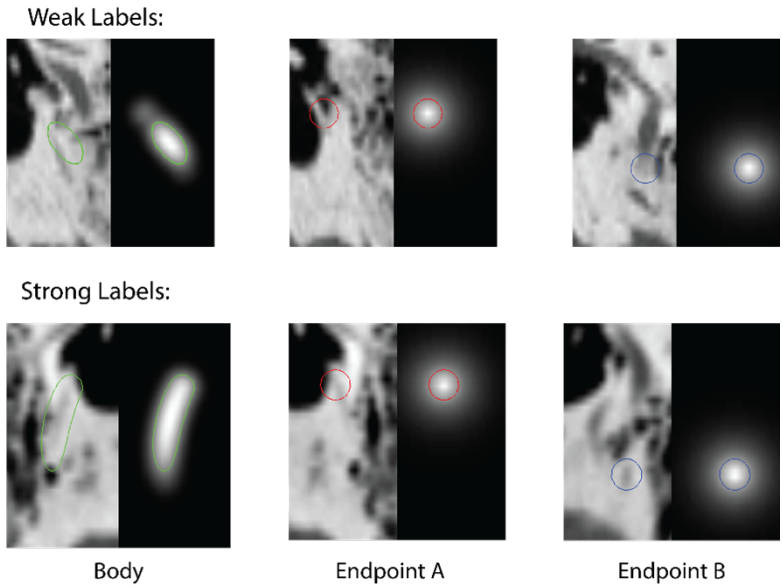

Figure 4.4: Example images with corresponding probability maps for each channel. For visualization purposes, colored outlines in each image represent an isosurface of 0.5 from the probability maps. Note also how the contours in the weak labels are close, but not quite accurate to the true chorda position.

value k.

The discriminator network takes a 4-channel image where the probability maps of the chorda body and endpoints are concatenated with the original image along the channel dimension. The output of the discriminator $D^{out} \in \mathbf{R}^3$ is a vector indicating the probability that each channel in input tensor was created by the generator network or was taken from the strongly labelled dataset. The idea is that we would like the network to produce high quality maps of all three channels due to the features learned from the image rather than producing a result that would 'pass' the discriminator purely because the network learned the correct spatial relationship between the body and endpoints. Furthermore, setting up the discriminator output in this manner allows for the future possibility of mixing real and fake images for each of the channel to further increase network robustness. However, for this work, the target value D^{target} is defined as $[0,0,0]^T$ if taken from the generator, or $[1,1,1]^T$ if taken from the probability maps created from the strongly labelled dataset. The loss function of the discriminator is then defined as

$$L^D = BCE\left(D^{out}, D^{target}\right)$$

With $BCE(\cdot,\cdot)$ again representing the binary cross entropy function.

With these two loss functions, the adversarial learning objective of each network is then set up as the discriminator network trying to distinguish if an image and its corresponding probability map segmentation

come from the generator or the strongly labeled dataset. The competing goal of the generator is then to 'fake' segmentations from the input images that 'convince' the discriminator into classifying it as coming from the strongly labelled dataset.

4.3.4 Testing

During testing, we use only the generator network to create the probability maps from the input image. We then threshold the produced probability map for the chorda body and apply medial-axis extraction via voxel thinning (178). The resulting skeleton is then converted to a contour via spline interpolation between the locations of the remaining voxels and then placed in point correspondence with the rest of the labelled dataset by resampling the curve to 20 equally placed positions along the contour. With this point-corresponding contour, we can now make comparisons with the ground truth contour from our dataset.

4.3.5 Implementation Details

For the generator we use a batch size of 17. It should be helpful for training to ensure at least one strongly labeled image existed in each batch when taking an optimization step. Therefore, for each batch, 16 images are taken from the weakly labelled dataset and the remaining image is taken from the strongly labelled dataset. Additionally, we use $k = .01$. as the down weighting value for the weak label loss. Due to the small sizes of the images and discriminator network, we were able to fit the entire dataset into GPU memory and not need to use batches for training. Furthermore, since the discriminator network will tend to learn much faster than the generator, we set the discriminator to take only one optimization step for every 10 epochs stepped by the generator. Both networks are trained using the Adam optimizer (179) with a learning rate of 1e-4 for the generator and 1e-5 for the discriminator. Lastly, as recommended in (180) weights for batch norm and convolutional layers are initialized from a normal distribution of mean 0 and standard deviation of 0.02.

4.3.6 Evaluation

To evaluate our method, we take the 16 images from our strongly labelled dataset and perform 2-fold cross validation. Eight samples are taken for training and tested on the remaining samples and vice versa. As mentioned previously, the atlas sample is used during training for its corresponding fold but excluded from testing. As a point of comparison, we further create automatic segmentations using the existing method (17) for the 15 non-atlas images. We then report mean symmetric surface distance between the segmentations produced by each method and the ground truth contours for the evaluation.

4.4 Results

Boxplots of overall and paired surface segmentation errors can be seen in Figures 4.5 and 4.6. Our proposed method produces a mean surface segmentation error of 0.49mm +/- 0.18mm, which is lower than the original method of 0.96mm +/- 0.55mm. The mean paired difference of 0.47mm translates to a 49% improvement in mean segmentation performance of our method over the original method. A Wilcoxon signed-rank test further shows this to be statistically significant. Only in 3 of the 15 cases tested did the proposed method perform slightly worse than the original. However, we note that the original method performed unusually well for these cases, and the segmentation error from our method was still quite low at 0.83mm, 0.59mm, and 0.39mm. Mean surface error for each vertex using our proposed method show the largest errors to occur at the endpoints. However, even in the worst case, mean surface error for a vertex is still sub-millimeter at 0.82mm. (Figure 4.7) Visual examples of the best and worst performing cases for our method can be found in Figure 4.8 .

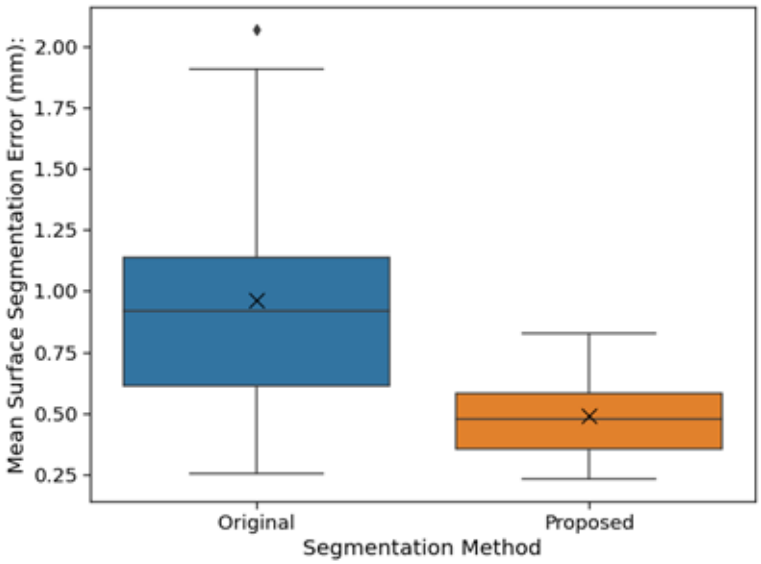

Figure 4.5: Boxplot of mean surface segmentation errors for the original automatic method (Original) and our proposed GAN method (Proposed)

4.5 Discussion

In this work we have proposed a GAN based method for improving the automatic segmentation of the chorda tympani. We have described a weakly supervised training process for our method that can leverage automatic segmentations from a large dataset to create an effective network even when only few manual segmentations exist. On our dataset of 130 weakly labelled images and 16 manually labelled images, we have shown that

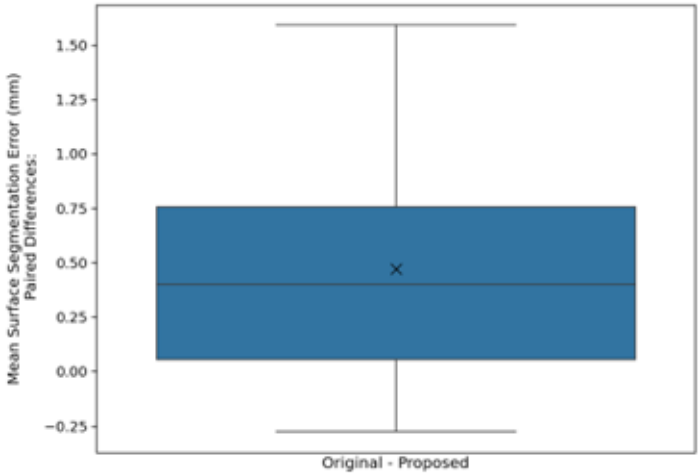

Figure 4.6: Boxplot of paired differences (Original - Proposed) in surface segmentation error between the original automatic method and our proposed GAN method

networks utilizing only 8 manual segmentations can produce mean surface segmentation errors of 0.49mm and significantly outperforms the existing technique. As mentioned in (181; 182) for 1mm drill bits, safe trajectories must lie at least 1mm away from the facial nerve and chorda, with (183) using a 2.5mm facial recess size for their 1.8mm robotic procedure. Therefore, since our mean segmentation errors are sub-millimeter even in the worst case, such results are encouraging for use in trajectory planning in clinical settings. Since these worst-case errors occur at the endpoints of the chorda, immediate future work should focus on integrating the information learned by the network for the localization of these endpoints into the final chorda segmentation. That, along with an extension of this method for use with the facial nerve will help increase the flexibility of choosing robust and safe drilling trajectories for cochlear implant surgery.

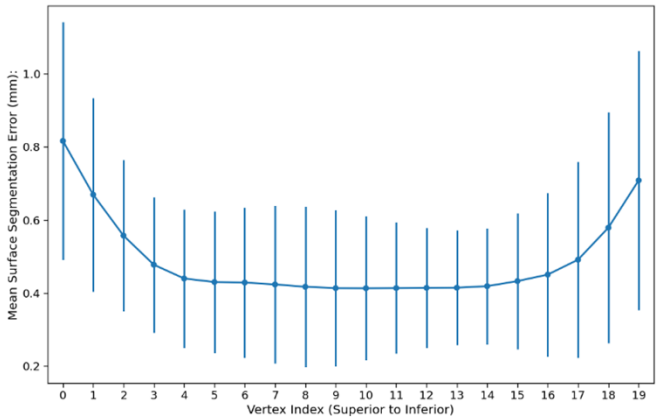

Figure 4.7: Mean surface error for each vertex in our proposed segmentation method. Larger vertex indices indicate more inferior positions along the chorda. Bars represent +/- 1 standard deviation

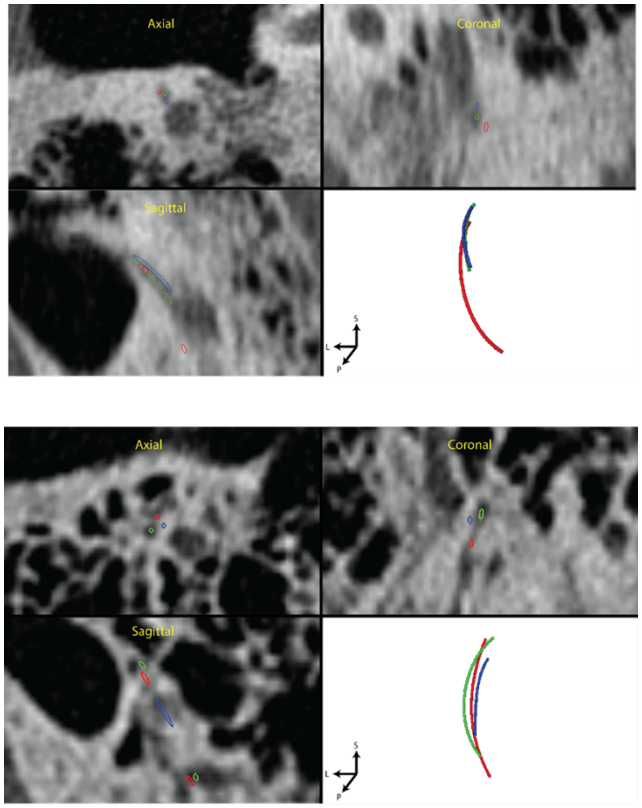

Figure 4.8: Example segmentations of the best (top) and worst (bottom) cases from our proposed method. Green is ground truth, Red is the result from the original segmentation method, and blue is the result from our proposed method. For visualization purposes, we show each contour as a tube with radius of 0.1mm

CHAPTER 5

Validation of Active Shape Model Techniques for Intra-Cochlear Anatomy Segmentation in CT Images

5.1 Abstract

Purpose: Cochlear implants (CIs) have been shown to be highly effective restorative devices for patients suffering from severe-to-profound hearing loss. Hearing outcomes with CIs depend on electrode positions with respect to intra-cochlear anatomy. Intra-cochlear anatomy can only be directly visualized using high resolution modalities such as μCT, which cannot be used in vivo. However, active shape models (ASM) have been shown to be robust and effective for segmenting intra-cochlear anatomy in large scale datasets of patient CTs. In this study, we present an extended dataset of μCT specimens and aim to evaluate the ASM's performance more comprehensively than has been previously possible.

Approach: Using a dataset of 16 manually segmented cochlea specimens on μCTs we found parameters that optimize mean CT segmentation performance and then evaluate the effect of library size on the ASM. The optimized ASM was further evaluated on a clinical dataset of 134 CT images to assess method reliability

Results: Optimized parameters lead to mean CT segmentation performance to 0.36mm point-to-point error, 0.10mm surface error, and 0.83 Dice score. Larger library sizes provide diminishing returns on segmentation performance and total variance captured by the ASM. We found our method to be clinically reliable with the main performance limitation was found to be the candidate search process rather than model representation.

Conclusions: We have presented a validation of the ASM for use in intra-cochlear anatomy segmentation. These results are critical to understand the limitations of the method for clinical use and for future development.

5.2 Introduction

Cochlear implants (CIs) have been shown to be highly effective restorative devices for patients suffering from severe-to-profound hearing loss. Hearing outcomes with CIs are dependent on many factors, including the positions of the electrodes with respect to intra-cochlear anatomy (6; 2) . In recent studies, we have shown that when the audiologist, who programs the patient's CI by selecting a number of stimulation parameters, is provided an estimation of the intra-cochlear positioning of the electrodes, patient-customized programs can be created that significantly improve hearing outcomes (184; 185; 18; 19). We call this Image-Guided CI Programming (IGCIP). To enable IGCIP, we have developed an active-shape model (ASM) (20) based

solution for localizing the intra-cochlear anatomy, including the scala tympani (ST), scala vestibuli (SV), and modiolus in patient CT images (42).The method has been shown to be robust as it has been used in large scale studies to investigate the relationship between intra-cochlear electrode position and hearing outcomes (186; 187; 188; 189; 13; 9), despite the fact that micron-level intra-cochlear structures being localized by the ASM, such as the Osseous Spiral Lamina (OSL), which separates the ST and SV, are so small that they are not visible in CT images (Figure 5.1). Since clinical MRI of the temporal bone at resolutions obtainable by CT remains unsolved, CT is currently the best in vivo modality for imaging of the cochlea. However, it is inadequate for construction of an intra-cochlear ASM because fine-scale structures such as the OSL are not visible.

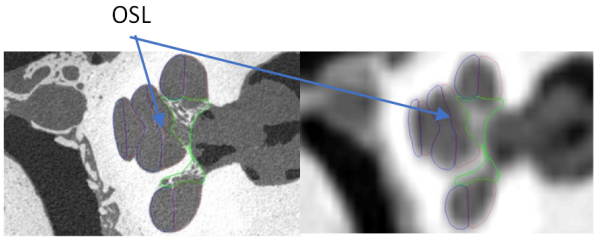

Figure 5.1: Example μCT (left) and registered CT slice (right) showing ground truth segmentations of the scala tympani (red), scala vestibuli (blue) and modiolus (green). Labeled also is the Osseous Spiral Lamina (OSL)

The approach originally proposed in (42) was to construct the ASM using a dataset of μCTs of 6 cochleae specimens. The μCT modality cannot be used in vivo but has high enough resolution to visualize fine-scale intra-cochlear anatomy in specimens. The ASM constructed using specimens could then be registered to new patient CTs. Nonrigid image registration-based methods were proposed to automatically initialize the ASM. While highly reliable for this task, atlas image registration alone was not accurate enough to directly provide an accurate segmentation of the cochlea. Thus, after initialization, the ASM was optimized using the standard method of iteratively identifying candidate positions for model points and fitting the ASM to those candidates until convergence (20). In the fitting procedure, two types of model points were weighted with different importance. The points in the ASM that correspond to the external walls of the cochlea, termed "edge" points, have strong gradients in the image as shown in Figure 5.2. For these points, a line search was performed along the surface normal to find a candidate with the strongest image gradient. The remaining points, termed "non-edge" points, correspond to locations that have no salient image features. In the original implementation, the ASM was found to be less likely to converge to errant local minima of the edge features if the non-edge points were fitted to their initialization positions provided by nonrigid registration with the atlas instead of searching for new candidates for these points. The fit of the edge points, which are driven

by image features, was weighted with very high importance relative to non-edge points, which were only weighted with enough importance to ensure the ASM does not diverge to solutions far from the initialization position.

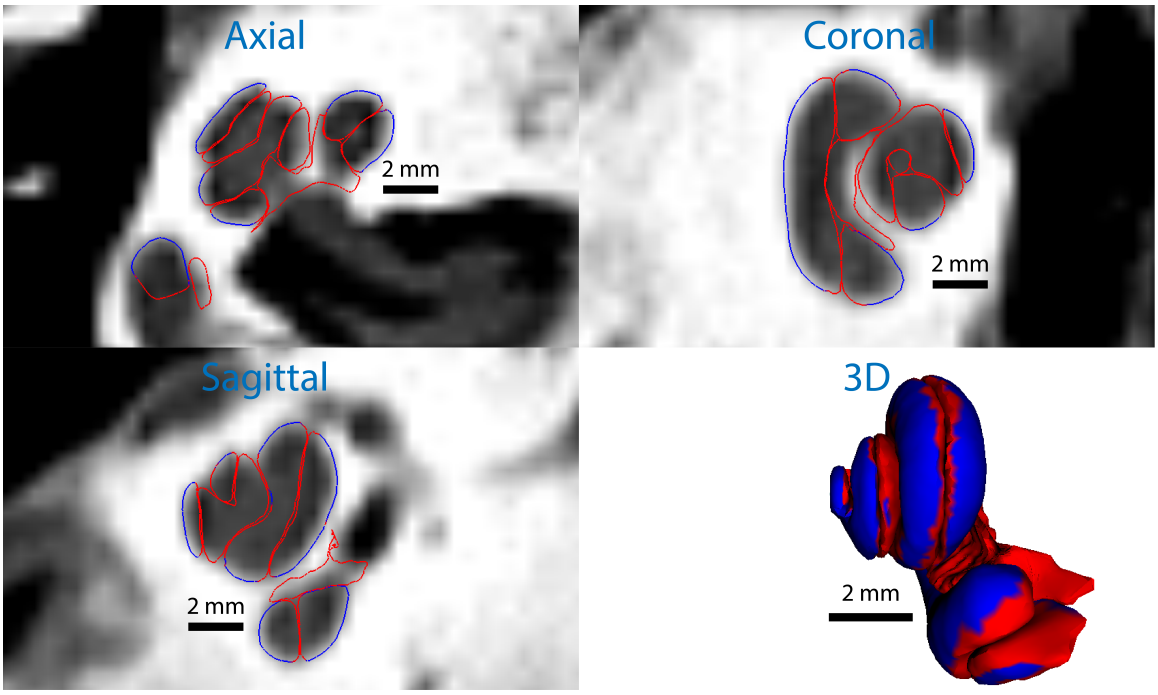

Figure 5.2: Example ASM segmentation with labeling of exterior (blue) vs.interior (red) points of the cochlea. Note that the exterior points generally correspond to stronger image gradients in the corresponding CT.

While the method has been robust when applied to large scale datasets, its ability to capture population variability, its quantitative accuracy, and the effectiveness of its parameters were evaluated on a limited dataset of only 5 μCT/CT pairs due to the limited availability of manually segmented μCTs of specimens in the original study (42). In this study, we expand the dataset to 16 specimens and aim to evaluate the method's performance more comprehensively. More specifically, our goals are to determine: 1) optimized ASM parameters, 2) the effect of training size on performance, and 3) the viability of using the ASM with our determined parameters on a diverse population of clinical CT images. These issues are critical to understand the limitations of the method for use in IGCIP and other studies evaluating the relationship between intra-cochlear electrode position and outcomes.

5.2.1 Related Work

Different approaches exist in the literature regarding image segmentation of the cochlea; see (35) for an extensive survey. Traditional image processing algorithms have been described for this purpose, with automatic region growing and level set methods in (36), and semi-automatic active contours and level sets utilized in

(37) and (38; 39), respectively. Atlas-based methods are also proposed in (40; 41; 190) to predict the shape of the cochlea in the target image by projecting a known atlas segmentation through a registration transformation. Each of those methods, while effective at segmenting the high contrast bones of the inner ear, do not deal specifically with the segmentation problem of the intra-cochlear anatomy. As mentioned earlier, the fine-scale structures of the intra-cochlear anatomy required to accurately differentiate between the ST/SV are difficult if not impossible to see in clinical images, thus leading to the necessity of training a model to estimate the location of invisible structures from the location of visible landmarks.

Using a shape prior, i.e. learning "plausible" shapes from training data, enables regularizing segmentations obtained from noisy or partially occluded images. The first formal application of this concept in intra-cochlear segmentation utilized active shape models by Noble et al. (42) for pre-implantation CT's, with extensions to post-implantation CT's in Reda et al. (43) and a multi-region approach by Romera et al. (44). An evaluation of the active shape model's candidate search strategy was performed Gaa et al. (45). A shape prior approach used to constrain the solution of a level set of the entire cochlear labyrinth, but not the intra-cochlear anatomy, was developed by Zhu et al. (46), while Pujadas et al. (65) developed a solution for intra-cochlear anatomy but only for segmenting μCT images. Demarcy et al. (48) propose using statistical modeling to represent a shape parameterization and image appearance for use in intra-cochlear segmentation. Further, Kjer et al. (49) propose a statistical deformation model with a dataset of 18 samples that results in strong Dice scores and max surface errors in their validation study, but a direct comparison to our method is difficult as the much higher quality and resolution of the CTs used in their study may also be a factor enabling the relatively high performance of their method. Finally, deep learning has been a popular approach in other image processing applications, but its requirement of large training sizes poses a difficulty in this domain. High resolution μCTs of cochlear specimens remain difficult to obtain and manually segment. Nonetheless, Zhang et al. (50) propose a successful deep learning cochlear segmentation architecture that produces results with impressive mean surface errors, even when utilizing a limited training set. However, one limitation of the approach is that in the small number of failure cases that occur, the approach tends to fail dramatically and produce shapes that are clearly impossible. In contrast, the ASM approach leverages statistical regularization to ensure highly robust performance with far fewer failures on large datasets. A performance comparison of the various statistical modeling approaches for segmenting intra-cochlear anatomy in pre-operative CTs, along with our current model, will be presented in the results section.

5.3 Approach

5.3.1 Overview

Our approach for validation of the ASM-based method for segmentation of intra-cochlear anatomy consists of three separate studies. First, we first conduct a parameter sensitivity study to determine how the various user-defined parameters in the ASM affect segmentation performance. Using the parameters that optimize mean segmentation errors, second, we then perform a study on the training library size, with the goal of determining how model performance changes with the size of the training set. Finally, we perform a robustness study, in which we use an ASM with optimized parameters to create segmentations on a large dataset of patient CTs and compare the results with expert verified segmentations. A visual outline of our study is shown in Figure 5.3. The following sections detail our approach.

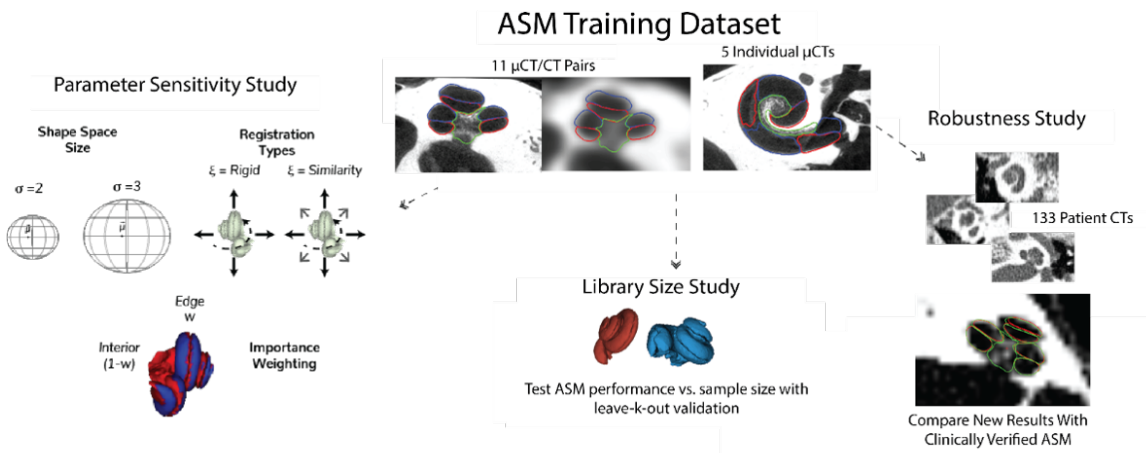

Figure 5.3: Graphical overview of the experiments conducted in this study

5.3.2 ASM training

Given a set of T training shapes, we define a shape S_i as the collection of the N 3-D points $\{\overrightarrow{p}_{ij}\}_{j=1}^{N}$ such that the jth point in the ith shape, $\overrightarrow{p}_{ij}$, is in point correspondence with the jth point in the kth shape, $\overrightarrow{p}_{kj}$, for all $i,k \in \{1 \ldots T\}$. Our use of the ASM in cochlear segmentation then generally follows the procedure described by [7]. We first choose a single shape from the training set S_0 and perform a point registration between the reference and remaining $T-1$ training shapes. The function $\Gamma_{S_i,S_k}(\cdot)$ that registers shape S_i to shape S_k is defined by the rotation R, scale s, and translation $\overrightarrow{t}$ that solves

$$\Gamma_{S_i,S_k} = argmin_{R,s,\overrightarrow{t}} \sum_{j=1}^{N} w_j^2 \parallel sR\overrightarrow{p}_{ij} + \overrightarrow{t} - \overrightarrow{p}_{kj} \parallel^2$$

where w_j is used to define the importance weighting for the jth point. In the context of intracochlear

segmentation, w_j is defined as w for point indices labeled edge points or $(1-w)$ for those that are labeled non-edge points. We further consider the parameter $\xi = \{similarity, \ rigid\}$ to determine the form of the transformations used when computing Γ_{S_i,S_k}. The scale factor s is optimized for the case of $\xi = similarity$ but constrained to $s = 1$ when $\xi = rigid$.

An initial mean shape is computed as $\overline{S}' = \frac{1}{T}\sum_{i=1}^{T}\Gamma_{S_i,S_0}(S_i)$ followed by further improvement by iterating $\overline{S}' = \frac{1}{T}\sum_{i=1}^{T}\Gamma_{S_i,\overline{S}'}(S_i)$ until convergence to the final mean shape $\overline{Y}$. We then register each shape in the training set to $\overline{Y}$ to create a registered training set of shapes $\{Y_i\}_{i=1}^{T}$. The corresponding shape vector representations for the mean shape $\overline{y} \in \mathbf{R}^{3N}$ and registered training shapes $\overrightarrow{y}_{i\in 1\ldots T} \in \mathbf{R}^{3N}$ are defined as the concatenation of the N points of the respective shape into a single vector. The resulting covariance matrix for the training set is then calculated as $C = \frac{1}{T}\sum_{i=1}^{T}(\overrightarrow{y}_i - \overline{y})(\overrightarrow{y}_i - \overline{y})^T$. The ASM is then fully defined by $\overline{y}$ and the $T-1$ non-zero eigenvalues λ_i and corresponding eigenvectors $\overrightarrow{e}_i$ of C, respectively. For our implementation of the ASM, to ensure numerical stability, we discard any eigenmodes with eigenvalues $\lambda_i \leq 10^{-8}$

5.3.3 ASM estimation

Using the shape vector representation of a candidate shape $\overrightarrow{\kappa} \in \mathbf{R}^{3N}$ we first register $\overrightarrow{\kappa}$ from the CT image space to the ASM mean shape $\overline{y}$ to create a registered candidate shape $\widehat{\kappa}$. The form of the registration function matches the value of ξ used in the training process. The ASM's estimate of the shape space coordinates $\widehat{b} \in \mathbf{R}^{T-1}$ is calculated using weighted least squares.

$$\widehat{b} = (P^T W^T W P)^{-1} P^T W^T W(\widehat{\kappa} - \overline{y})$$

Where $P = \begin{bmatrix} \overrightarrow{e}_1 & \overrightarrow{e}_2 & \ldots & \overrightarrow{e}_{T-1} \end{bmatrix}$ is the concatenation of the eigenvectors of the ASM into a matrix and $W = diag([w_1 w_1 w_1 \ldots w_j w_j w_j \ldots w_N w_N w_N])$ is the diagonal matrix $W \in \mathbf{R}^{3Nx3N}$ of importance weightings w_j defined as before. To enforce the user-selected constraint that valid shapes have a Mahalanobis distance that falls within σ standard deviations of the mean shape, the coefficients of $\widehat{b}$ are scaled to satisfy $\sum_{j=1}^{T-1}\frac{b_j^2}{|\lambda_j|} < \sigma^2$. The final fitted shape can then be reconstructed as $\overline{y} + P\widehat{b}$ and projected back into the target image space as the segmentation estimate.

5.3.4 Image segmentation with the ASM

The segmentation is initialized using atlas-based segmentation techniques that rely on affine and nonlinear image registration between the patient CT and an atlas CT. These registration methods are described in (175; 176). The pre-defined cochlea surface of the atlas is first projected through the resulting non-rigid transformation from the atlas to the patient CT. This atlas-based segmentation is then fitted to the ASM, and the resulting estimate serves as the initial segmentation of the image.

To determine new candidate update positions for each vertex and from a candidate shape $\overrightarrow{\kappa}$, vertex locations are moved according to their classification as edge or non-edge points. If the point is classified as an edge point, we use the location of the strongest image gradient observed in a line search along the surface normal over a range of ±1 mm as the new candidate position. The gradients are computed using 0.1 mm wide central differences on linearly interpolated image intensities. Alternatively, vertices classified as non-edge points are not searched but are always set to their original positions determined from the initial atlas registration. The candidate vertex locations for the mesh are subsequently fit to the ASM to create a new shape estimate. This segmentation is then iteratively improved by repeating this process of alternating line searches and ASM fits until convergence or a maximum of 100 iterations is completed.

5.3.5 Evaluation metrics

For our tests, we use three different error criteria: point-to-point error, symmetric surface error(191), and Dice score (192). The point-to-point error e_j^{point} is defined as the Euclidean distance $e_j^{point} = \left\| \overrightarrow{p}_j^{est} - \overrightarrow{p}_j^{gt} \right\|$ of the corresponding j^{th} 3-D vertex locations in both meshes. Furthermore, the surface error e_j^{surf} of the jth vertex is the Euclidean distance $e_j^{surf} = \left\| \overrightarrow{p}_j^{est} - \overrightarrow{s}_j^{gt} \right\|$, where $\overrightarrow{s}_j^{gt}$ is the closest location on the surface of the ground truth mesh to the jth estimated vertex $\overrightarrow{v}_j^{est}$. As the surface error is asymmetric from estimate to ground-truth and vice-versa, we also include backwards distance $e_j^{gt} = \left\| \overrightarrow{p}_j^{gt} - \overrightarrow{s}_j^{est} \right\|$ of the ground-truth to estimate as part of our surface error calculations. Thus, the reported mesh-based errors for our evaluation are the mean point-to-point error $\frac{1}{N}\sum_{j=1}^{N} e_j^{point}$ and mean symmetric surface error $\frac{1}{2N}\sum_{j=1}^{N}\left(e_j^{surf} + e_j^{gt}\right)$. For the volume-based Dice score calculations we convert the ground truth and estimated meshes into binarized volumes I^{est} and I^{gt} respectively. Dice scores for each estimated structure was calculated as $\mathscr{D} = \frac{2\cdot\#(I^{est}\cap I^{gt})}{\#(I^{est})+\#(I^{gt})}$ where $\#(\cdot)$ is the number of foreground voxels in the volume contained in the argument.

5.3.6 Parameter sensitivity study

The goal of this experiment is to find the optimal user-selected parameters for use in intra-cochlear segmentation. We investigate the effect of the following parameters on ASM performance: the edge point weighting factor w, the Mahalanobis distance constraint σ, and whether a rigid or similarity transformations is used, ξ. A range of values for $\{w,\ \sigma,\ \xi\}$ are exhaustively evaluated using a leave-one-out validation approach. (see Table 5.1) ASMs were trained with 15 of the samples to segment the remaining sample's CT. Since only 11 of the samples had a clinical CT available, we note that only 11 train/test splits could be properly evaluated for each parameter combination.

5.3.7 Library size study

Since μCT images of cochlear specimens are costly to acquire, this library size experiment is designed to demonstrate the number of samples needed for the ASM to achieve different levels of performance. We evaluate the effect of the training library size on ASM segmentation performance through a sweep of all ASM training sizes $t \in \{1 \ldots 15\}$. At each training size, we create an ASM with t samples and segment the CT image of one of the remaining $16 - t$ samples. Furthermore, we evaluate the segmentation performance of each model when using only the eigenmodes that correspond to the largest 80%, 90%, 95% and 100% of the total variation in the dataset. This training size and eigenvalue sweep is performed exhaustively for $t \in \{1,\ 2,\ 14,\ 15\}$. To limit computation for the other training set sizes, we randomly sampled 100 train/test pairs for $t \in \{3 \ldots 13\}$. Furthermore, since no eigenmodes exist for $t = 1$ we use the single training sample as a "mean-shape" and directly register this mean-shape to the candidate CT points at each segmentation iteration. As will be elaborated on in the results section, we conducted all tests using the best results from the parameter optimization study, which were registration type $\xi = rigid$, distance constraint $\sigma = 2$, and edge point weighting $w = 1$.

5.3.8 Robustness study

We validated the robustness of our ASM when trained on the full dataset of 16 samples and using the optimal parameters found in the parameter sensitivity study discussed above. A previous ASM generated from (42) had been used to create automatic segmentations of the intracochlear anatomy on a clinical CT database of 137 ear volumes. No high-resolution ground truth segmentations exist for these volumes, but the automatic segmentations were all visually verified to be clinically usable. We used our new model to re-segment these images and compared the mean point-to-point distance between the two meshes. Any segmentations produced by our new ASM that differ by a mean point-to-point error greater than 0.5mm were selected for further visual inspection by an expert Otologist. For each case, the expert was presented with both our optimized ASM segmentation and the segmentation created using the original, unoptimized ASM, masked to their identity. The expert was asked to assess the clinical viability of both segmentations and asked to assess which one was more accurate in a forced choice evaluation. A high rate of viability and quality assessment over the unoptimized ASM would indicate high robustness of our optimized ASM.

5.4 Results

5.4.1 Parameter sensitivity study

Boxplots for our parameter study can be seen in Figures 5.4 - 5.6. The effect of w on the mean segmentation accuracy appears to be consistent across different combinations of ξ and σ. Furthermore, when $w > .9$

the plots show a consistent outlier in the surface error and Dice score metrics as seen in Figure 5.5 and 5.6. This outlier corresponds to one case that had a poorer initialization than the other cases in our dataset (Figure 5.7). Nonetheless, there is generally an improvement to mean segmentation performance across all metrics when using larger w values. Given these observations on (w, ξ, σ) we decided on parameters that yielded the smallest mean point-to-point segmentation error over our dataset as shown in Table 5.1. While we noted that the improvements from ξ and σ in terms of absolute segmentation accuracy are small, a Wilcoxcon signed rank test (193) at $w = 1$ does show a statistically significant improvement in point-to-point accuracy of $(\xi, \sigma) = (rigid, 2)$ over $(rigid, 3)$ and $(similarity, 3)$ when using the Bonferroni corrected (194) $\alpha = .05/3 = .017$. A full listing of the calculated p-values for the comparisons of results between $(\xi, \sigma) = (rigid, 2)$ and results using the different parameter combinations of (ξ, σ) at $w = 1$ is shown in Table 5.2.

5.4.2 Library size study

Segmentation and fitting performance for our library size study can be seen in Figure 5.8 and Figure 5.9 respectively. Surface errors are generally small, as the segmentation surface errors of ˜0.10mm of 15 sample ASMs approaches the corresponding fitting errors of ˜0.08mm. Point-to-point errors, however, have a much larger discrepancy, since 15 sample ASMs have a segmentation error of ˜0.40mm with a corresponding fitting error of ˜0.20mm. Overall, there appears to be a general trend of diminishing returns on both the segmentation and fitting performance as training size increases. Such decreasing benefits to the ASM from more samples can further be seen in plots of the mean total variation (Figure 5.10) and in the mean number of eigenmodes (Figure 5.11). Indeed, the addition of more samples does little to increase the mean number of eigenmodes in the model since 15 sample ASMs only have 9 effective eigenmodes to represent the intracochlear anatomy (Figure 5.11). While such models would be expected to have $15 - 1 = 14$ eigenmodes, 5 eigenmodes on average are discarded since they have eigenvalues below our threshold of $\lambda_i \leq 10^{-8}$.

Despite the diminishing returns from more samples, we do find that ASM segmentation performance is

	Tested Values	Optimal Value
Edge point weighting w	0.0, 0.0001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, 0.999, 1.0	1.0
Registration type ξ	Rigid, Similarity	Rigid
Mahalanobis Distance Constraint σ	2,3	2
% Variation Cutoff in ASM	80%, 90%, 95%, 100%,	90%

Table 5.1: Tested and optimized parameter settings for the ASM

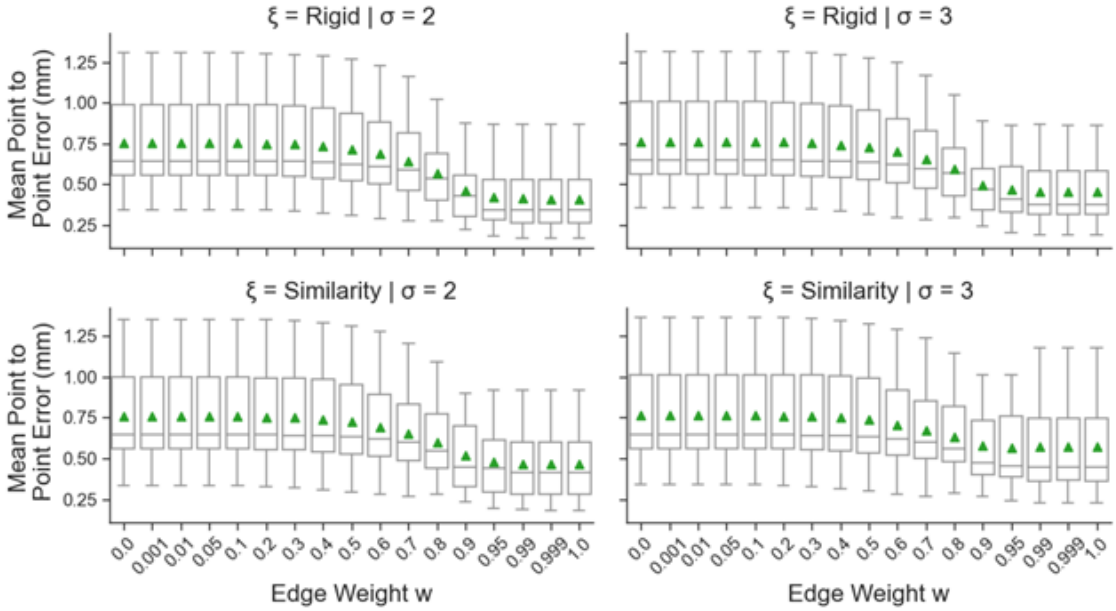

Figure 5.4: Boxplots of segmentation point-to-point errors for the parameter tuning study. Means are marked with green triangles.

increased when using the largest 90% of the eigenmodes in a 15 sample ASM. This leads to an optimized ASM performance of 0.36mm point-to-point error, 0.10mm surface error, and 0.83 Dice score. In Figure 5.12, we show a heatmap of mean performance errors of our optimized 15-sample ASM on our ground truth dataset. We find that larger surface errors tend to localize at the basal end of the cochlea, while the large point-to-point errors are distributed across the basal end and the first turn of the cochlea (Figure 5.12). Observing point-to-point error in the basal turn where surface errors are low indicates point-to-point errors tend to be tangential to the surface in this region.

5.4.3 Robustness study

Out of the 137 ear volumes tested, we found 11 cases that diverged in mean surface distance from the original automatic solution by 0.5mm. In our masked qualitative evaluation experiment with these 11 cases, the clin-

	ξ = rigid, $\sigma = 3$	ξ = similarity, $\sigma = 2$	ξ = similarity, $\sigma = 3$
Mean point-to-point	**0.001**	0.034	**0.001**
Mean Surface	0.034	0.483	0.139
Mean Dice	0.027	0.650	0.207

Table 5.2: One-sided Wilcoxon signed rank p-values of the paired samples at $w = 1$ comparing ($\xi = rigid$, $\sigma = 2$) performing better than the other combinations of ξ and σ. Significant p-values ($p < .017$) are shown in bold

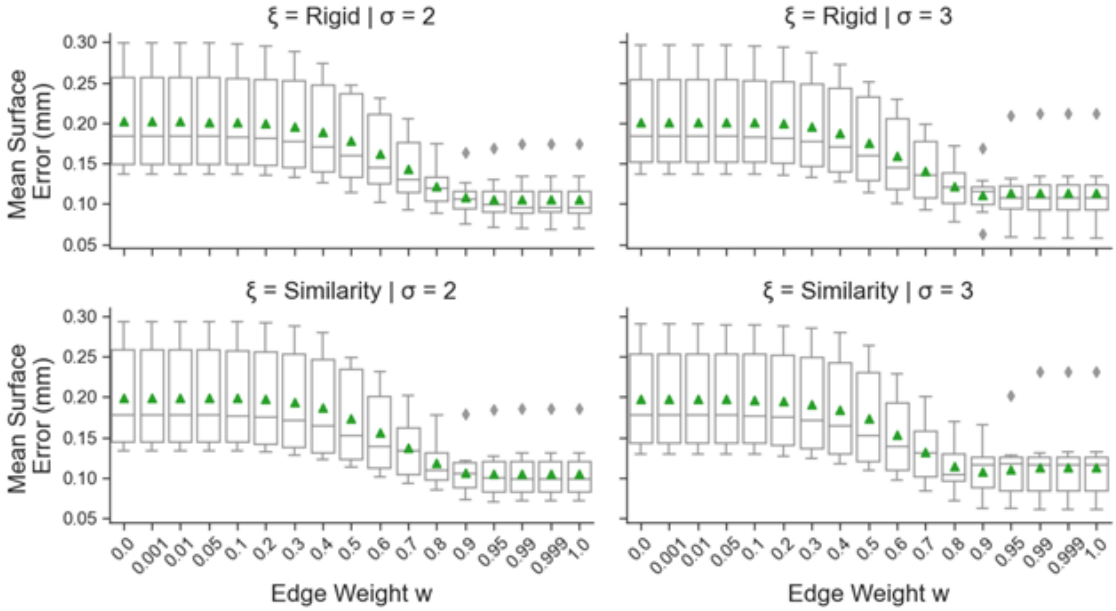

Figure 5.5: Boxplots of segmentation surface errors for the parameter tuning study. Means are marked with green triangles.

ician found all segmentations produced by our optimized ASM to be clinically viable. Further, the clinician preferred the segmentations produced by our proposed optimized ASM over the original ASM for 7 of the 11 cases (Table 5.3). While the experiment was performed with forced choice, the expert mentioned that for many cases the two segmentations were virtually equivalent. In the 4 cases where the optimized ASM result was not selected (shown as bold rows in the table), mean-surface differences between the original and optimized ASM segmentations were among the smallest of all 11 cases, implying that for these cases the two results were likely nearly equally accurate. Representative cases that visually demonstrate this are shown in Figure 5.13.

5.5 Discussion

In this work, we have used our dataset of 16 specimens to comprehensively validate the ASM's use in intra-cochlear anatomy segmentation. We found optimal parameter settings of $w = 1$, $\xi = rigid$, and $\sigma = 2$ as our optimal parameter settings for the ASM. We further found that segmentation performance is most optimized when using the eigenmodes representing largest 90% of the variation. These settings lead to an optimized ASM performance of 0.36mm point-to-point error, 0.10mm surface error, and 0.83 Dice score. We tested this optimized ASM on our clinical dataset of 137 volumes and found that it was quite robust by producing clinically viable results in all cases. Though 4 cases were not preferred by a clinician over

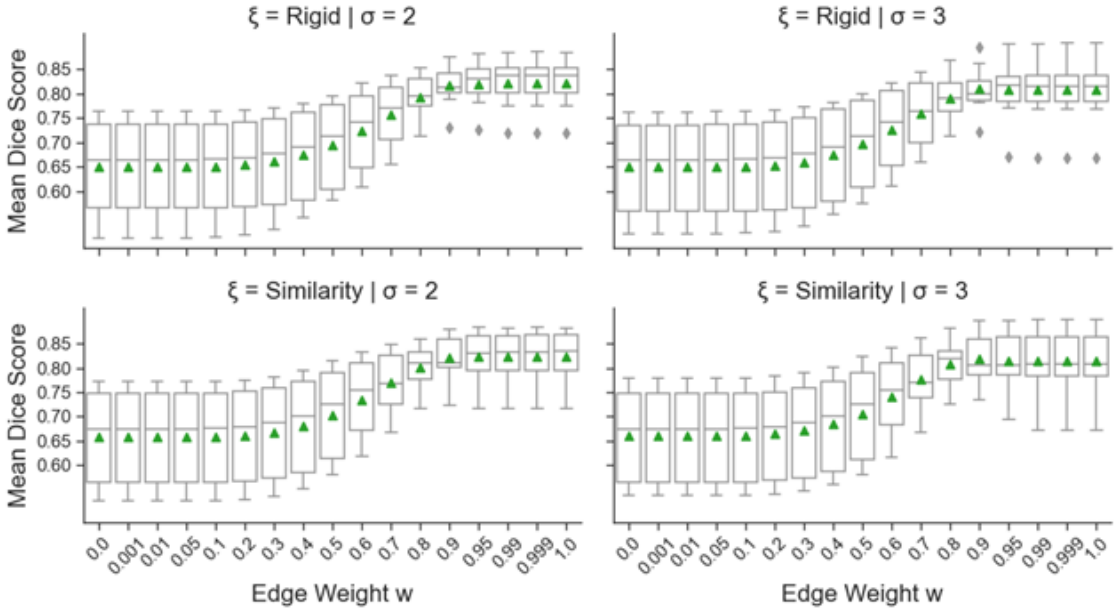

Figure 5.6: Boxplots of segmentation Dice scores for the parameter tuning study. Means are marked with green triangles.

the original ASM segmentation, it should be noted that those cases had the lowest mean surface difference between both segmentations. In these cases, the shapes produced by both methods were visually similar where one segmentation would be only slightly preferred over the other. Meanwhile, for those cases that had the largest discrepancies, the clinician preferred the result of our parameter-tuned model, thereby suggesting those differences as clear visual improvements to the segmentation result over the original ASM.

While the acquisition of more samples would generally lead to better performance and more captured shape variation, our results from our library size study suggest the onset of diminishing returns for larger training sizes. Indeed, the total number of eigenmodes in our models still resembles a linearly increasing trend with more samples, but we note that the effective number of eigenmodes utilized by ASMs trained with T samples is much lower than the theoretical $T-1$ eigenmodes that should exist in those models. For example, a 15 sample ASM has, on average, only 9 eigenmodes with eigenvalues greater than our threshold of 1e-8. Furthermore, the converging trend of fitting errors and mean total variation suggests progressively less shape variation is added to the model with the addition of more samples. Finally, when considering segmentation performance across all our metrics, the observed converging trends suggests progressively fewer benefits provided by larger training sizes to the accuracy of the segmentation result.

However, our results do show some limitations to our current implementation of the ASM. From our outlier case found in our parameter study and depicted in Figure 5.7, we find that segmentation performance

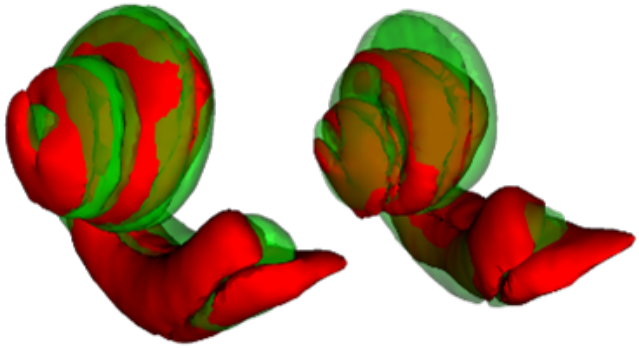

Figure 5.7: Comparison of initialization positions for a typical case (left) and the observed outlier case in our parameter sensitivity study (right). Ground truth is shown in green while the initialization position is shown in red.

can still be adversely affected from bad atlas initializations to the ASM. While the ASM is designed to offer a level of robustness to some noisy candidate locations, a highly incorrect initialization will be less likely to lead to an accurate ASM segmentation.

Another limitation to our approach stems from the rotational symmetry of the cochlea's shape. We found that segmentation surface errors are generally quite small, even achieving rough parity with the fitting errors. However, point-to-point errors are much larger and have a higher discrepancy between measured segmentation and fitting performance. These results suggest that the overall surface shape of the intra-cochlear anatomy is accurate, but the vertex locations are shifted tangentially on the surface. Such behavior shows the limitation of our simple line search method since candidate locations are chosen without considering information in the tangent plane of the surface. Instead, candidates are chosen with line searches and intensity gradients measured normal to the surface. Therefore, for a very rotationally symmetric shape such as the cochlea, these edge gradients would indeed be on the surface, but would be ambiguous as to where on the surface the correct location should be. The estimate surface produced by the ASM would then match very closely to the true cochlea surface but would be rotated about the axis of symmetry. This effect can be clearly seen by the localization of larger surface errors predominantly in the basal end of the cochlea.

A comparison of our optimized ASM with the other results from the literature can be seen in Table 4. Overall, the mean performance numbers compare favorably to the other methods. The Dice score improves upon the scores of (42; 48) and the mean surface errors are slightly better to that of (48; 49). As the Dice score reported in (49) was calculated for a model with a combined ST/SV volume, we cannot directly compare our

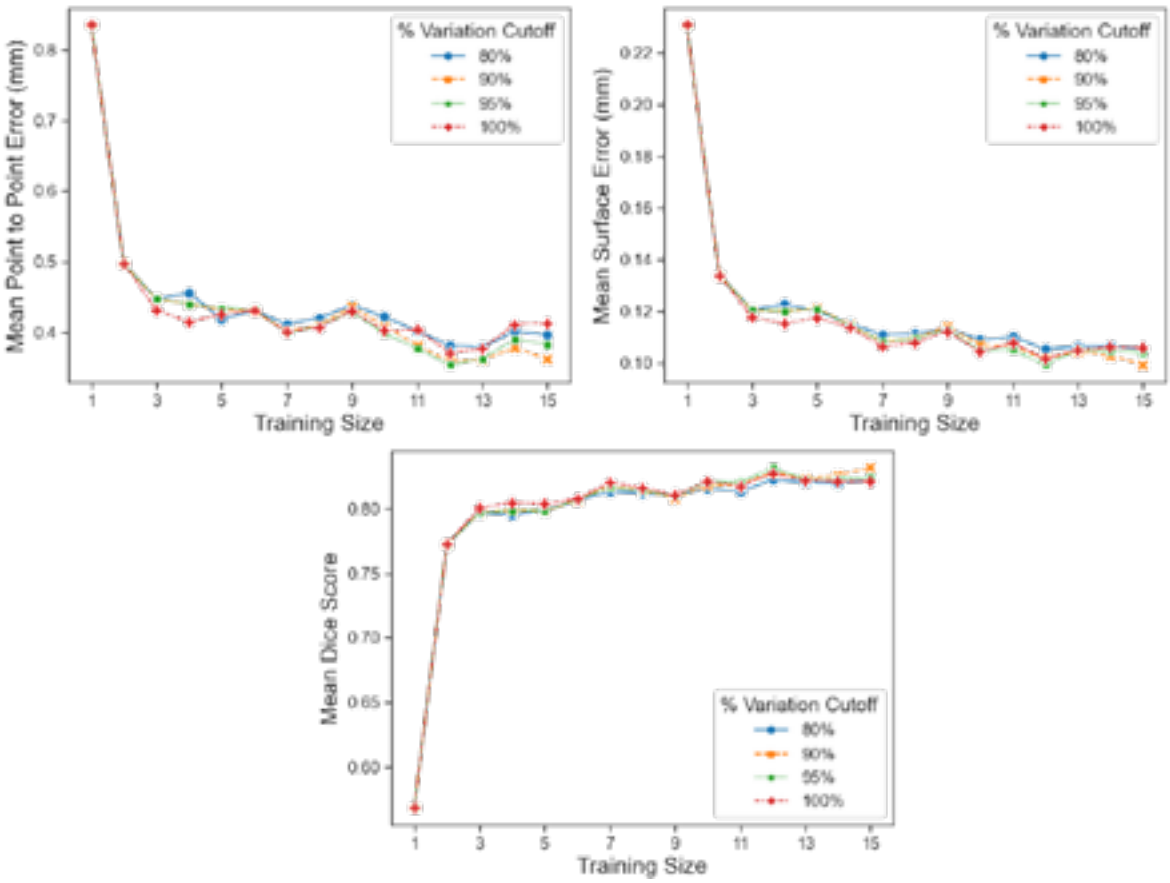

Figure 5.8: Mean segmentation performance for the library size study at different percentage thresholds of model variation. (Clockwise from top left): Point-to-point errors, surface errors, Dice scores

Dice results which were calculated for each individual scala. However, we do note that our method was able to slightly outperform that method in terms of surface errors despite our method using fewer samples and segmenting lower resolution CT images. Finally, while the results from (50) still provides the best overall performance across all metrics, we note that the method sometimes fails under difficult image conditions. In contrast, our optimized ASM can produce clinically viable segmentations for images with a range of different image qualities.

Overall, the level of performance achieved with our optimized model achieves similar performance to the rest of the literature and is robust enough for use in clinical applications such as IGCIP. However, due to the rotational errors resulting from our line search method, care should be taken in applications that require high localization accuracy of landmarks in the basal end of the cochlea. While the acquisition of more samples and larger library sizes will generally lead to better performance, our results suggest that future work in improving ASM segmentation performance should be more focused on improvements to the candidate searching strategy

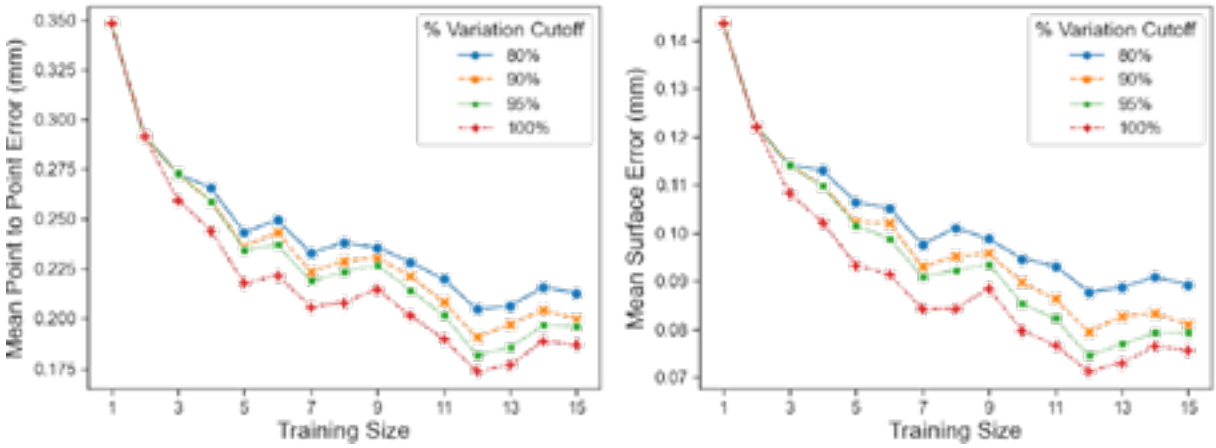

Figure 5.9: Mean fitting errors for the library size study at different percentage thresholds of model variation. (Left) Point-to-point errors (Right) Surface errors

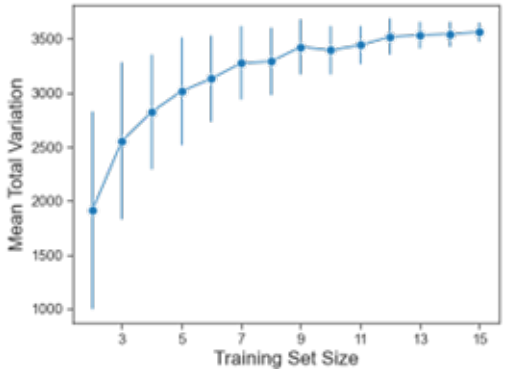

Figure 5.10: Mean total variance explained by an ASM for different training sizes. Total variance for an ASM is obtained by taking the sum of the eigenvalues in that model. Vertical bars represent 1 standard deviation.

and atlas initialization method.

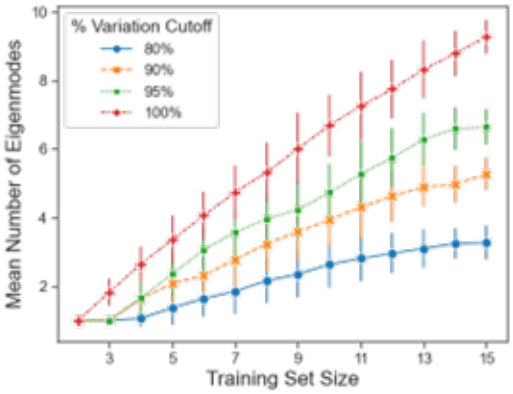

Figure 5.11: Mean number of eigenmodes present in an ASM when at different percentage thresholds of model variation. Vertical bars represent 1 standard deviation

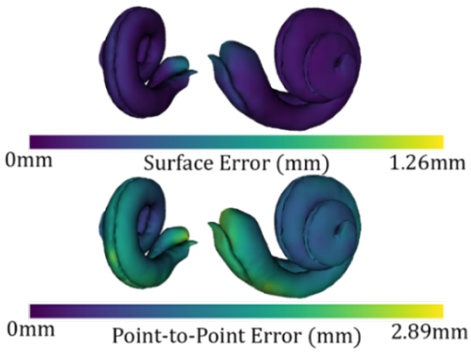

Figure 5.12: Heatmap of mean surface (top) and point-to-point errors (bottom) when validating 15 sample ASMs with our optimized settings from Table 5.1

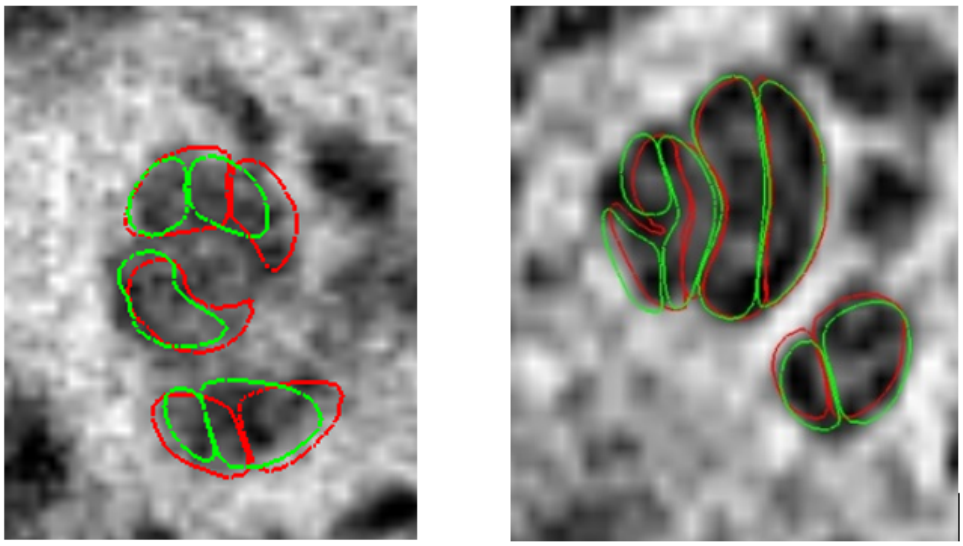

Figure 5.13: Example segmentation slices from our robustness study where the new ASM segmentation differed by >0.5mm from the original ASM result. Segmentations shown in green were produced by the original automatic method and the segmentations produced by our optimized ASM are shown in red. (Left) A result where the clinician selected the new ASM segmentation. (Right) One of the 4 results where the clinician selected the original ASM segmentation.

Case Number	Mean Point-to-Point Difference (mm)	Mean Surface Distance (mm)	Selected By Clinician	Clinically Viable
1	**0.522**	**0.102**	**N**	**Y**
2	0.517	0.173	Y	Y
3	**0.544**	**0.110**	**N**	**Y**
4	0.600	0.195	Y	Y
5	**0.522**	**0.102**	**N**	**Y**
6	0.668	0.144	Y	Y
7	**0.544**	**0.139**	**N**	**Y**
8	0.711	0.170	Y	Y
9	0.576	0.158	Y	Y
10	0.883	0.224	Y	Y
11	0.513	0.139	Y	Y

Table 5.3: Results of the 11 cases where the optimized ASM produced a segmentation that differed from the original ASM segmentation by 0.5mm point-to-point when performing our robustness study over 137 cases. In bold are the 4 cases where the clinician preferred the result of the original ASM segmentation over the result produced by the new model.

Method	Number of Samples	ST Dice Score	SV Dice Score	Mean Surface Error (mm)
(42)	6	0.77	0.72	0.21
(48)	9	0.77	0.73	0.12
(49)	18	0.88*	0.88*	0.11
(50)	11	0.87	0.86	0.08
Optimized ASM	16	0.83	0.83	0.10

Table 5.4: Comparison of the different methods for segmenting intra-cochlear anatomy in pre-operative CT images, along with our optimized ASM. Methods are compared based on the number of training samples, surface errors, and the Dice score for each ST/SV. Mean surface errors are calculated from the aggregated ST/SV. (N/R) indicates not reported, while * indicates that the mean Dice score was reported instead of for individual ST/SV.

CHAPTER 6

An Active Shape Model with Locally Weighted Elements

6.1 Abstract

While state-of-the-art deep learning methods consistently provide the most accurate image segmentation results when sufficient training datasets are available, in some applications large datasets are difficult to acquire. Active shape models (ASM) have been a successful technique in medical image segmentation and require less extensive datasets for training. However, the ability of the ASM framework to capture complex shape variation is limited by representing variations across all T global-pose-normalized training exemplars in a single, linear vector space. In this work, we describe a novel non-linear extension to the ASM in the form of a mixture of landmark weighted ASMs. We describe the training procedure and demonstrate the advantage of this approach with a synthetic box dataset and 3 real-world datasets of the cochlear labyrinth, full ear, and chest. The trained models appear to use the limited training data efficiently by capturing simple intuitive regions of the overall shape. We find that our proposed model can outperform a traditional ASM with similar sample sizes in a variety of scenarios and can leverage information from surrounding structures to display robust performance against anatomical outliers. These results show the method's potential advantages in applications that are limited by small shape libraries.

6.2 Introduction

While the current state-of-the art trend for achieving high-accuracy segmentations typically involves deep learning approaches with sufficiently sized datasets, achieving good accuracy with these techniques on smaller sample size applications is not particularly straightforward (195). These cases of limited training sample availability are still relevant, however. For example, a model for inner ear structures can only be constructed by ex vivo specimen imaging modalities such as μCT (42), and manual segmentations on a large scale can make dataset acquisition costly and time consuming.

Active shape models (20) (ASMs) have been an effective technique for medical image segmentation, particularly in these situations with limited availability of training samples. To accomplish this goal, the basic framework of the ASM as described by (20) involves representing shape variations across all the global-pose-normalized training exemplars in a single, linear vector space. This resulting 'shape space' is then used in an iterative process that alternates between 1) searching for a noisy segmentation from an image search and 2) refining the segmentation to the closest 'plausible' shape as defined by the learned shape space.

While effective, complexities in shape variation that are not well suited to the described model creation

process, such as non-linear shape variations or localized rotations, will be difficult to capture by the ASM, particularly in low-sample-size situations. Similar to (21), we show a hypothetical example of a shape comprised of two fixed location rectangles. Each rectangle contains a single mode of variation, but each box is allowed to rotate independently. The resulting shape variation modeled by the single linear vector space of a traditional ASM will be rather complex and have difficulty in producing an accurate fit to a new exemplar from the population. Meanwhile, a model composed of two weighted-ASMs could each place their importance weighting on a different rectangle in the shape. The resulting shape variations modeled by each ASM would be correct for the rectangle with the higher importance weighting, but still skewed for the lower importance weighted rectangle. However, when the estimates from each weighted-ASM are aggregated, the correct overall shape is produced by the combined model. (Figure 6.1)

The goal of this work is to this multiple weighted-ASM idea to address the limitations of the traditional ASM described above. We propose the 'multi-element ASM' as an extension to the ASM framework. In our model, we describe a way of increasing the degrees of freedom in the model while determining those degrees of freedom using anatomical variations that exist within the dataset, rather than, e.g., adding random variance. This is accomplished by determining regions of the shape that can be best described with specialized weighted-ASMs which we label as "elements" and combining the estimates of these elements into a global shape. In doing so, we aim to use our limited training samples to represent shape variation as efficiently as possible.

When given a high dimension, low sample size scenario involving a complex shape with many landmarks and a small training library, there exists the possibility of determining spurious correlations between distant unrelated regions due to random chance. A traditional ASM will be more likely in this scenario to capture such relationships between uncorrelated parts of the shape as all landmark variation will be modelled by a single linear vector space. As shown by our box example, the traditional ASM modeled a linear correlation between both rotating boxes since there were insufficient samples, and therefore degrees of freedom in the linear vector space, for the ASM to fully 'understand' the independent rotations of each constituent box. Our multi-element model can reduce this type of error as we 1) increase the model degrees of freedom with multiple elements and 2) use the intuition that points in closer proximity will tend to have more meaningful correlations with each other than more distant points. The use of the locally weighted ASMs in our box example shows how the added degrees of freedom introduced by each element allowed for the flexibility to each capture the true local correlations of each sub region and combine their estimates together to produce the correct result. Furthermore, the ability for an element to still place non-zero weight to distant points provides the capacity to model correlations between distant points as needed. An advantage of our approach is that we determine these importance weightings for each element in a data-driven way through the minimization of

fitting error on a left-out sample in the training set.

There is a tradeoff, though, between model flexibility and robustness. The addition of more degrees of freedom is useful for capturing more shape variation present in complex shapes, thereby providing more accurate segmentation estimates when given good candidate points. However, too many degrees of freedom run the risk of falling into poor optima during segmentation, as the model not only fits to the shape described by the candidate points, but also the noise present in those points from the image. This can be particularly true in cases where there is little extra shape variation to be gained from additional degrees of freedom. For example, if we were to have a shape comprising of just one of our boxes, a traditional ASM would perform well here by capturing the single mode of variation. A multi-element ASM would add more degrees of freedom than the traditional ASM but would be more likely to produce worse estimates during segmentation time since those degrees of freedom would be used to fit to any noise present in the candidate points.

With our multi-element ASM, we propose a method that balances this flexibility/robustness tradeoff. We use a data-driven approach in our specialized training process to optimize elements that capture robust regions described from localized rotations. We further test our model against a traditional ASM on a variety of dataset scenarios to better understand instances where our model would provide the most benefit. Taken together, this gives the framework for using the multi-element ASM as a tool in analyzing and segmenting small shape libraries.

6.2.1 Related Work

Different approaches to statistical shape modeling exist in the literature, with numerous survey papers (57; 53; 54; 55; 56; 52; 58; 59; 60; 61; 62; 51) highlighting statistical shape models and the many variations arising out of each aspect of the shape modeling pipeline.

The concept of using gaussian mixture models within the ASM has been proposed in (125) to use as a way of modelling the space of valid shapes in shape space. While our work does use gaussians to draw the importance weighting for different elements, the shape space for each element still utilizes a single linear vector space. Furthermore, weighted ASMs have been proposed in various forms such as (140), (196). This concept is used heavily in our method by allowing for the weighting of landmark points so that weighted least squares can be used rather than regular least squares fit to the shape space as in the traditional ASM. However, on its own, the weighted-ASM still uses a single linear vector space to describe shape variation. Partitioning approaches are taken in (21), (196) to separate the overall shape into distinct sub-regions for modelling by different ASMs. Like our work, these partitions are determined through optimization of an objective function. However, interactions between different ASMs are limited, effectively capturing the shape variability of different regions independently from one another. Hierarchical models such as (196; 88; 110)

utilize multiple ASMs to represent shape variation in a coarse to fine manner. These increasing degrees of locality do impose relationships between models, but this is not the approach that we use in this work. In our approach, multiple element-ASMs can have importance weighting that overlap across landmarks, thereby allowing for interactions between different element-ASMs in predicting landmark location. However, we do not impose any hierarchical relationships between these models, instead allowing each element to determine its own level of locality needed to produce the best estimate in concert with the other elements in the model.

As opposed to the hierarchical or partitioning approaches described above, our model uses weighted elements to determine regions that can be modeled with localized rotations. We hypothesize our model performing well for bony structures where shape variation of different regions can be intuitively seen in this localized manner rather than as acting part of an overarching hierarchical structure. Furthermore, the use of weighted elements throughout all points does allow for the interaction between different regions, thereby providing robustness through interactions that would otherwise not be present when using multiple independently working models.

6.3 Multi-Element ASM

6.3.1 Overview

Our proposed muti-element ASM framework uses multiple, weighted-ASMs to capture regional shape variation. For clarity, we will refer to an individual weighted-ASM in our model as an "element". Each element utilizes its own unique set of optimized importance weights for all N 3-D points defining the shape. Therefore, each point will have a different importance weighting, and thus estimated position, based on the element. By combining these different estimates for the point to create the final location, we allow for the modeling of a point's location from the interaction of multiple ASM elements rather than the responsibility of any single ASM. Finally, our model can be easily implemented in the readily available deep-learning frameworks such as Pytorch (197). Doing so allows us to leverage the automatic differentiation features already available in these frameworks to easily optimize the weights of each of these elements.

A straightforward method to determine the N importance weights for each of the M elements in our model would be an optimization method where all $N * M$ values are selected to minimize the overall fitting error on a training set. However, our initial experiments showed that such a formulation is a difficult optimization that is too prone to poor-performing local minima to be a viable approach. Therefore, we use a gradual method that first performs a simpler optimization requiring fewer parameters to select the importance weights for the M elements. Those determined importance weights are then used as the initialization point of the full optimization method involving all $N * M$ importance weights as free parameters.

For the initial optimization, we simplify the number of free parameters by introducing the concept of

a gaussian-shaped element. For these elements, we constrain the distribution of its importance weighting by modeling it with a full-covariance gaussian placed in 3-D space with respect to an atlas shape. As a result, the importance weighting of a single gaussian-shaped element is fully defined by the optimization of a mean vector $\overrightarrow{\mu} \in \mathbf{R}^3$ and covariance parameters $\overrightarrow{\mathbf{c}} \in \mathbf{R}^6$. Therefore, a multi-element ASM utilizing multiple gaussian-shaped elements can capture the shape variability of different contiguous regions while keeping the optimization manageable. We additionally add a single global element with N free parameters at this stage to help model any correlated disjoint regions or portions of the shape which are not given high importance weighting from any of the gaussian-shaped elements.

In the following, we will first describe how individual ASM elements are defined, and then we show how a final shape estimate is created from a multi-element ASM comprised of M elements. We then describe the full optimization process of the multi-element ASM for single and multi-structure settings. Finally, we describe our segmentation and evaluation methods for our different datasets: simulated boxes, cochlear labyrinth, multiple ear structures, and multiple chest structures.

6.3.2 ASM Element Creation

For a multi-element ASM comprised of M elements, if the ith element is a gaussian-shaped element, we define a 3-D gaussian with mean vector $\overrightarrow{\mu_\mathbf{i}} \in \mathbf{R}^3$ and covariance matrix $\Gamma_i \in \mathbf{R}^{3x3}$ with respect to an atlas shape. Therefore, the importance weighting $w_{i,j}$ of the jth point in this i^{th} element is calculated as

$$w_{i,j} = exp\left(-\left(\overrightarrow{\mu_\mathbf{i}} - \overrightarrow{\mathbf{a_j}}\right)^T \Gamma_i^{-1} \left(\overrightarrow{\mu_\mathbf{i}} - \overrightarrow{\mathbf{a_j}}\right)\right)$$

Where $\overrightarrow{\mathbf{a}}_{\mathbf{j}\in\mathbf{1}\ldots\mathbf{N}}$ are the 3-D coordinates of the jth atlas point. Since this calculation will be part of an optimization process, we utilize a log-cholesky parameterization for Γ_i^{-1}. Using a parameter vector $\overrightarrow{\gamma_\mathbf{i}} = [\gamma_{i,1}\gamma_{i,2}\ldots\gamma_{i,6}]$, we create the lower triangular matrix L_i as

$$L_i = \begin{bmatrix} e^{\gamma_{i,1}} & 0 & 0 \\ \gamma_{i,2} & e^{\gamma_{i,4}} & 0 \\ \gamma_{i,3} & \gamma_{i,5} & e^{\gamma_{i,6}} \end{bmatrix}$$

The resulting inverse covariance matrix $\Gamma_i^{-1} = L_i L_i^T$ used in the calculation of $w_{i,j}$ will have the required property of being symmetric positive semi-definite while remaining uniquely defined by the unconstrained parameter vector $\overrightarrow{\gamma_\mathbf{i}}$

Global elements, in contrast, have N free parameters $\phi_{i,j} \in \mathbf{R}$ which we constrain with a differentiable mapping for optimization purposes to importance weights in the range of $(0,\ 1)$ via the sigmoid function.

$$w_{i,j} = \frac{1}{1+e^{-\phi_{i,j}}}$$

Regardless of the type of element, the training of the ith element then proceeds similarly to the original ASM formulation as described by [3], but we use as the values of the calculated $w_{i,j}$ as the importance weighting during the Procrustes shape registration. We choose a shape from the T samples in the training set as reference and register the remaining $T-1$ training shapes to it in a weighted least squares sense using rigid transformations. The registered training shapes $\overrightarrow{\mathbf{y}}_{\mathbf{i},\ \mathbf{t}\in 1\ldots T}$ are then used to iteratively create a mean shape $\overline{\mathbf{y}}_{\mathbf{i}} = \frac{1}{T}\sum_{t=1}^{T}\overrightarrow{\mathbf{y}}_{\mathbf{i,t}}$ for this ASM element. We note that $\overrightarrow{\mathbf{y}}_{\mathbf{i,t}}$ and $\overline{\mathbf{y}}_{\mathbf{i}}$ are $3N$ shape vectors created from the concatenation of the N 3-D points in their corresponding shapes. Furthermore, unlike the covariance matrix used in the original ASM formulation, we calculate a weighted covariance matrix

$$C = \frac{1}{T}\sum_{t=1}^{T}\left(\overrightarrow{\mathbf{y}}_{\mathbf{i,t}} - \overline{\mathbf{y}}_{\mathbf{i}}\right) W_i W_i^T \left(\overrightarrow{\mathbf{y}}_{\mathbf{i,t}} - \overline{\mathbf{y}}_{\mathbf{i}}\right)^T$$

Where the diagonal matrix $W_i = diag\left([w_{i,1}w_{i,1}w_{i,1}\ldots w_{i,N}w_{i,N}w_{i,N}]\right)$ is formed from the learned importance weights $w_{i,j}$. By using this weighting of the covariance matrix, we focus the variation captured by the resulting eigenmodes on those regions of the overall shape that are more highly weighted for this ASM element. The ith element in our multi-element ASM is fully defined with the resulting $T-1$ eigenvectors $\overrightarrow{\mathbf{e}}_{\mathbf{i,t}}$ with non-zero corresponding eigenvalues $\lambda_{i,t}$ of C along with the mean shape $\overline{\mathbf{y}}_{\mathbf{i}}$.

6.3.3 Shape Estimation with the Multi-Element ASM

Given a candidate shape $\overrightarrow{\mathbf{c}} \in \mathbf{R}^{3N}$ we produce the ith shape estimate $\widehat{\mathbf{x}}_{\mathbf{i}} \in \mathbf{R}^{3N}$ from the ith element by first rigidly registering $\overrightarrow{\mathbf{c}}$ to $\overline{\mathbf{y}}_{\mathbf{i}}$ to create a registered candidate shape $\widehat{\mathbf{c}}_{\mathbf{i}}$. Since we used a weighted covariance matrix in the creation of the element, we calculate the shape space coordinates $\widehat{\mathbf{b}}_{\mathbf{i}} \in \mathbf{R}^{T-1}$ from this element using the weighted displacements from the mean shape $\widehat{\mathbf{dx}}_{\mathbf{i}} = W_i\left(\widehat{\mathbf{c}}_{\mathbf{i}} - \overline{\mathbf{y}}_{\mathbf{i}}\right)$ and using weighted least squares

$$\widehat{\mathbf{b}}_{\mathbf{i}} = \left(P_i^T W_i^T W_i P_i\right)^{-1} P_i^T W_i^T W_i(\widehat{\mathbf{dx}}_{\mathbf{i}})$$

Where $P_i = \left[\overrightarrow{\mathbf{e}}_{\mathbf{i,1}}\overrightarrow{\mathbf{e}}_{\mathbf{i,2}}\ldots\overrightarrow{\mathbf{e}}_{\mathbf{i,T-1}}\right]$ is the concatenation of the eigenvectors of the ith element, and W_i is the same diagonal importance weighting matrix used in the element's creation. Each of the coefficients of $\widehat{\mathbf{b}}_{\mathbf{i}}$ are then clamped to fall into the range of user selected number of standard deviations σ. The fitted shape is then created by $\overline{\mathbf{y}}_{\mathbf{i}} + P_i\widehat{\mathbf{b}}_{\mathbf{i}}$ and then projected back into the original space of $\overrightarrow{\mathbf{c}}$ to create the shape estimate $\widehat{\mathbf{x}}_{\mathbf{i}}$ for this element. The final position estimate $\widehat{\mathbf{x}}_{\mathbf{j}} \in \mathbf{R}^3$ of the jth point in the shape is achieved by

$$\widehat{\mathbf{x}}_{\mathbf{j}} = \sum_{i=1}^{M} \left(w'_{i,j} \cdot \widehat{\mathbf{x}}_{\mathbf{i,j}} \right)$$

Where $\widehat{\mathbf{x}}_{\mathbf{i,j}}$ is the ith element estimate of the jth shape point and $w'_{i,j}$ are the L1 normalized importance weights across all M elements for the jth shape point given by:

$$w'_{i,j} = \frac{w_{i,j}}{\sum_{k=1}^{M} \text{—} w_{k,j} \text{—}}$$

With these normalized weights, the final shape estimate $\widehat{\mathbf{x}}$ is therefore calculated as the weighted average of the shape estimates $\widehat{\mathbf{x}}_{\mathbf{i}}$ produced by each element in our multi-element ASM.

6.3.4 Training the Full Multi-Element ASM

As mentioned in the methods overview, the optimization of the M element ASM starts with a simpler model comprised of $M-1$ gaussian-shaped elements and a single global element. We will use the convention that the first $M-1$ elements are gaussian-shaped elements and the Mth element in the model is the global element.

The gaussian-shaped elements are initialized with a spherical k-means clustering with $M-1$ clusters on the N 3-D points in the atlas shape. The means of each of the found clusters are used as the initial mean parameters $\overrightarrow{\mu}_{\mathbf{i} \in 1 \ldots M-1}$. Since we will be optimizing the log-cholesky parameters of the inverse covariance matrix, the radius $r_{i \in 1 \ldots M-1}$ of the ith cluster results in an initial parameter $\overrightarrow{\gamma_{\mathbf{i}}}$ vector for the ith element as:

$$\overrightarrow{\gamma_{\mathbf{i}}} = \begin{bmatrix} \ln(1/r_i) & 0 & 0 & \ln(1/r_i) & 0 & \ln(1/r_i) \end{bmatrix}$$

The global component, conversely, is initialized with $\phi_{M,\ j \in \{1 \ldots N\}} = 0$. At each training epoch we use a leave-one-out strategy so that a sample $\overline{\mathbf{x_t}}$ serves as the validation shape and the remaining $T-1$ samples train the M components as described above. To mitigate overfitting and to simulate noisy candidate shapes that would be encountered during segmentation, instead of fitting the noiseless validation sample $\overline{\mathbf{x_t}}$ to the resulting model, it is helpful to create a noisy validation shape $\overline{\mathbf{x_t}} + \overrightarrow{\varepsilon}$ by adding a spherical, zero mean gaussian noise vector $\overrightarrow{\varepsilon} \in \mathbf{R}^{3N}$ with radius on the order of the search range bounds that will be used in the segmentation process.

The noisy validation shape $\overline{\mathbf{x_t}} + \overrightarrow{\varepsilon}$ is then fit to the overall model to produce a shape estimate $\widehat{\mathbf{x}}_{\mathbf{t}}$. The objective function is then the mean squared point-to-point fitting error over all leave-one-out splits

$$L = \frac{1}{TN} \sum_{t=1}^{T} \sum_{j=1}^{N} \text{—} \widehat{\mathbf{x}}_{\mathbf{t,j}} - \overline{\mathbf{x}}_{\mathbf{t,j}} \text{—}^2$$

The gradients with respect to the model parameters are computed using the automatic differentiation

mechanism in PyTorch [22] and then optimized with the Adam optimizer (179) until convergence.

Once converged, the resulting $w_{i,j}$ are used as the starting point for a final refinement optimization. However, in this process, the first $M-1$ elements are no longer treated as gaussian-shaped elements, and the optimization of the M element ASM proceeds as containing all global elements with all $N*M$ importance weights as free parameters. Since global elements use the sigmoid function to constrain elements in the range $[0,\ 1]$ the $\phi_{i,j}$ for this optimization stage should be initialized as

$$\phi_{i,j} = \log\left(\frac{w_{i,j}^{initial}}{1-w_{i,j}^{initial}}\right)$$

Where $w_{i,j}^{initial}$ is the importance weight learned from the initial optimization. The optimization process for this stage then continues as before, still using the mean squared point-to-point error over all leave-one-out splits as the objective. Once converged, the learned importance weightings are used as the final values for the $w_{i,j}$ used in our multi-element ASM.

6.3.5 Segmentation with the Multi-Element ASM

Segmentation of a new image with a trained multi-element ASM involves an iterative process that begins with an initial segmentation. The specifics of each dataset's initialization will be elaborated in the evaluation section when describing each of our datasets in greater depth. Regardless, the initial segmentation is first fit to the multi-element ASM. New candidate points are found with line searches within a specified range along the surface normal as in the traditional ASM. For our line searches, we use the normalized derivative intensity profiles method described in (198) to find improved candidate locations from the image. We then repeat the process of fitting the candidate shape to our multi-ASM mixture and finding new candidate points from the image until convergence.

6.3.6 Extension to Multi-Structure Settings

For shapes comprising of multiple known structures, such as the chest and full ear datasets that will be described later, we can extend our multi-element ASM framework to integrate the information of surrounding structures in the modeling of an individual structure. Suppose the shape is comprised of Ψ anatomical structures $\Lambda_{s\in\{1\ldots\Psi\}}$. If we wish to create a structure-specific multi-element ASM for the sth structure Λ_s, we start with the initial spherical k-means clustering only on the points contained in Λ_s to create the initial parameter vectors $\overrightarrow{\mu_i}$ and $\overrightarrow{\gamma_i}$ for each element. The optimization then continues for both stages as before, with weights $w_{i,j}$ being optimized regardless of whether the jth point belongs to Λ_s. However, the objective is modified to optimize only the mean squared point-to-point error for the points belonging to Λ_s

$$L = \frac{1}{T * \#(\Lambda_s)} \sum_{t=1}^{T} \sum_{j \in \Lambda_s} \left| \widehat{\mathbf{x}}_{\mathbf{t,j}} - \overline{\mathbf{x}}_{\mathbf{t,j}} \right|^2$$

Where $\#(\Lambda_s)$ is the number of points belonging to Λ_s. With this modified optimization, shape variation captured by our model is mainly focused on the individual structure. However, due to the optimization of $w_{i,j}$ regardless of structure, information provided by the surrounding structures that may be beneficial to the modeling of Λ_s can still be retained.

During segmentation, we still perform the iterative process of line-searching and fitting candidate points to create the final estimate of all structures in the shape. However, in this case, we can use all $\mathscr{M}_{s \in \{1 \ldots \Psi\}}$ multi-element models simultaneously, where $\mathscr{M}_s$ is the structure-specific multi-element ASM for Λ_s. At iteration τ the estimate of the full shape from the previous iteration $\widehat{\mathbf{x}^{\tau-1}} \in \mathbf{R^{3N}}$ is comprised of the N points $\widehat{\mathbf{x}_\mathbf{j}^{\tau-1}} \in \mathbf{R}^3$. A line-search in the image from those points then create updated positions $\overrightarrow{\mathbf{x}_\mathbf{j}^{\tau}} \in \mathbf{R}^3$. We then create Ψ estimates by fitting candidate shapes $\overrightarrow{\mathbf{c}_\mathbf{s}^{\tau}}$ to each corresponding model $\mathscr{M}_s$. For each candidate shape $\overrightarrow{\mathbf{c}_\mathbf{s}^{\tau}}$, the N points $\overrightarrow{\mathbf{p}_{\mathbf{s,j}}}$ comprising that shape are defined as:

$$\overrightarrow{\mathbf{p}_{\mathbf{s,j}}} = \begin{cases} \overrightarrow{\mathbf{x}_\mathbf{j}^{\tau}} \; if \; j \in \Lambda_s \\ \widehat{\mathbf{x}_\mathbf{j}^{\tau-1}} \; otherwise \end{cases}$$

The resulting Ψ number of fits $\widehat{\mathbf{x}_\mathbf{s}^{\tau}}$, are then combined to create the final estimate $\widehat{\mathbf{x}^{\tau}}$ for this iteration by taking only the points corresponding to Λ_s from each structure-specific estimate $\widehat{\mathbf{x}_\mathbf{s}^{\tau}}$. By performing segmentation in this round-robin manner we are leveraging the most recent values of surrounding structures to help in the estimation of each individual structure without having to consider to the order of the estimates performed.

6.4 Evaluation

6.4.1 Synthetic Box Dataset

We demonstrate the multi-element concept by creating a synthetic dataset that extends the rectangle example to 3-D. We generate 100 random shapes where each shape consists of two boxes whose centers are located at (40, 40, 40) and (60, 60, 60). Each box has a fixed width of 5 units and a normally distributed height with a mean of 15 units and standard deviation of 7 units. Each box is then given a random rotation about its own center. This setup imitates our rectangle scenario in that each box has a single mode of variation (i.e. the height) and an independent rotation in space. We then divide the dataset into a 50/50 train/test split. The training set is further divided into non-overlapping sets of 5 samples. Each smaller subset then trains a traditional ASM and a multi-element ASM initialized with 2 gaussian elements. Since a multi-element ASM would be initialized with a k-means clustering, this would create an initialization that would produce

the already correct result for this setup. As such, we randomly initialize the gaussians in a 10-unit standard deviation neighborhood around each box and initialize the covariance parameters with a normal distribution of mean 0.05 and standard deviation 0.1. We perform this random initialization over 15 tests and choose the model that has the lowest training error on that smaller subset. We then report fitting errors of the testing set on each of the models created by the training subsets.

6.4.2 Labyrinth Dataset

The labyrinth dataset consists of 18 patient CT (voxel size ˜.3mm anisotropic) scans of the right ear with associated hand segmentations of the labyrinth created in a prior study [25]. These point corresponding segmentations across all our samples serve as our ground truth. We evaluate the segmentation performance of multi-element ASMs with $M \in \{4,\ 8,\ 12,\ 16\}$ using 2- and 3-fold cross validation to correspond with respective training set sizes $T \in \{6,\ 9\}$. For each training subset we create the multi-element ASMs and a traditional ASM and test on the remaining $18 - T$ images.

6.4.3 Full Ear Dataset

Our multi-structure dataset of the ear consists of 4 structures: the labyrinth, facial nerve, ear canal, and the ossicles. Hand segmentations for each of these structures are created for 10 patient CT images (voxel size ˜.3mm anisotropic) to serve as the ground truth. Due to the arbitrary boundary that defines the lateral side of the ear canal, any variability captured by a shape model would reflect inconsistencies in boundary interpretation rather than true shape variation. To minimize this effect, we create a mask that forces the vertices in this region to always have an importance weighting of zero. Furthermore, these points do not factor into our error metrics for evaluation. A depiction of this mask can be seen in Figure 6.2.

One of the 10 samples is reserved as an atlas image. The remaining 9 samples are split into two groups where one group serves as the training set to test on the opposite group and vice versa. Since the segmentation process begins with a registration from the atlas, we only use the atlas sample as always added to the training set but omitted from the testing set.

We first evaluate the performance of our multi-element ASM in a scenario where we are given a complex shape but have no a priori knowledge of any delineations that may exist between points. For clarity, we will call this a "structure-agnostic" scenario. As such, we treat the vertices in our dataset as a single shape without any structure-specific knowledge and compare the segmentation performance of a multi-element ASM consisting of $M \in \{2,\ 4,\ 8,\ 12\}$ against a traditional ASM. In our second assessment, we look at performance when there are known groupings of vertices for each structure. Again, for clarity, we will call this a "structure-specific" scenario. For this scenario we compare structure-specific multi-element ASMs com-

posed of $M \in \{2,\ 4,\ 8,\ 12\}$ for each structure against traditional ASMs that target each individual structure separately.

Additionally, we have a database of 132 clinical CT images that have already been automatically segmented with previously developed algorithms. These results have all been visually verified to be acceptable for clinical use. As a test of our multi-element ASM's robustness we train 8-element ASMs for each structure using all 10 of our original hand-segmented samples. These models are used to create automatic segmentations of our 132 clinical CT images. We then visually examine for clinical viability the cases where the multi-element segmentation has the largest discrepancy with respect to surface distance from the original automatic segmentation for each structure.

6.4.4 Chest Dataset

The chest dataset is taken from the JSRT database [26] of 247 chest radiographs with a resolution of 2048x2048 and an isotropic pixel size of 0.175mm. Point corresponding hand segmentations of the lungs, clavicles, and heart for each of the images are taken from [27]. In keeping with the same segmentation approach taken in [27], our segmentation process will utilize a multi-resolution approach based on a standard image pyramid. We use 5 resolution levels starting from 32x32 at the coarsest resolution until 512x512 at the finest resolution. For our evaluation, we discard 132 of the images due to the structures being too close to the image border and thus adding complications to the segmentation method that are outside of the scope of this paper. With the remaining subset of 115 images, we evaluate the segmentation performance of multi-element ASMs at different training sizes $T \in \{5,\ 10\}$. We randomly divide our dataset into a train/test split of 40 training images and 75 testing images. The training set is then further split into 8 and 4 non-overlapping sets for $T = 5$ and $T = 10$ respectively. Our evaluation for this dataset again looks at two scenarios depending on whether anatomical structure information is present ("structure-specific") or absent ("structure-agnostic") to the user.

To simulate how model parameter selection would work in a practical small-sample size situation, for each T we select one of the smaller subsets and perform leave-one-out validation on that subset for both the "structure-agnostic" and "structure-specific" scenarios. For the "structure agnostic" scenario we test $M \in \{2,\ 4,\ 6,\ \ldots,\ 16\}$ and $M \in \{2,\ 3,\ 4,\ 5\}$ for the "structure-specific" case. We choose the number of elements used in our final evaluation based on the best performing models with respect to surface error from this smaller test. As we will show in the results section, for the "structure-agnostic" scenario we will choose 8 and 14 elements for $T = 5$ and $T = 10$ respectively. For the "structure-specific" scenario we choose $M = 5$ for both sample sizes.

Using each of the subsets of the full training set, we train a multi-element ASM with the found optimal parameters for that training set size, a traditional ASM for each individual anatomical structure, and a

traditional ASM for the entire combined shape. We then evaluate the segmentation performance of each of these models on the testing set. Each segmentation is initialized from the mean-shape of that respective training subset at our coarsest resolution of 32x32. ASM segmentation proceeds using only a traditional global ASM for progressively finer resolutions until our highest resolution of 512x512. We found that the increased degrees of freedom provided by the traditional structure-specific ASM and our multi-element ASMs easily fall into very poor local minima on difficult structures such as the clavicles and heart and thus need better initialization locations than the original mean-shape. Therefore, we only perform the segmentations for these models at the 512x512 resolution starting from the segmentation result of the traditional global ASM.

6.4.5 Implementation Details

For all tests, we use the Adam optimizer and set the learning rates to 0.1 for $\overrightarrow{\mu}_{\mathbf{i}\in 1\ldots M-1}$ and 1e-3 for $\overrightarrow{\gamma}_{\mathbf{i}\in 1\ldots M-1}$ in the gaussian-shaped elements, while for all global elements in both optimization rounds the learning rate is set to 1e-4. Furthermore, a summary of the specific ASM settings used for each dataset can be found in Table 6.1.

6.5 Results

6.5.1 Synthetic Box Dataset

Fitting errors for our box dataset can be found in Figure 6.3. The multi-element clearly outperforms the original ASM with a fitting error of 5.46e-3mm. As mentioned in the training results section, the correct importance weighting distributions have been found for each of the elements, thus producing the nearly perfect result. It should be mentioned that a perfect result is not actually possible with our current setup, as the use of exponentials to have smooth differentiable functions for optimization mean that a weight can never be equal to zero. Therefore, there will always exist a very small importance weighting placed on the box with low importance weighting and still exerts some influence on the result.

Dataset	Standard deviation of ε used in training (mm)	Shape space cutoff threshold σ	Size of intensity profile (mm)	Number of points in intensity profile	Length of line search in image (mm)	Number of locations evaluated within search range	Max number of segmentation iterations
Boxes	0.01	-	-	-	-	-	-
Labyrinth/Ear	0.5	3	2	21	2	21	100
Chest	1	2.5	10*	11	4*	5	20

Table 6.1: ASM settings used for segmentation in our experiments. Numbers marked with * are multiplied by the pixel size per resolution level.

6.5.2 Labyrinth Dataset

Boxplots of mean surface segmentation errors for our multi-element ASMs and a traditional ASM can be found in Figure 6.4. Our multi-element ASMs do improve segmentation performance over their regular ASM counterparts for both training sizes tested, with progressively better performance associated with the use of more elements in the model. Furthermore, the increased degrees of freedom provided by our multi-element ASM appears to utilize the limited training set more efficiently than the traditional ASM, as all multi-element ASMs trained on 6 samples outperform a traditional ASM trained with 9 samples. However, there does appear to be a converging trend in performance improvements with respect to the number of elements. This can be seen in both sample size cases, since the addition of 4 elements from 12 to 16 provides smaller mean benefits to performance than the change of 4 to 8 elements. Regardless, the best results of 0.104mm and 0.097mm occur when using a 16 element ASM for both training sizes of 6 and 9 respectively. These values both translate to an improvement of 26% over their traditional ASM counterparts. Paired difference comparison using the Wilcoxon signed-rank test further show these improvements to be statistically significant ($p < 0.05$). (Figure 6.5)

A heatmap showing the paired differences between a traditional ASM and our best performing 16 element ASMs for both sample sizes can be seen in Figure 6.6. The regions where one model outperforms the other model appear similar for both sample sizes. The multi-element ASM has better surface error for much of the labyrinth, particularly in the semi-circular canals, while the traditional ASM has better segmentation results in only a few isolated pockets in the structure.

6.5.3 Full Ear Dataset

Overall mean surface segmentation errors can be found in Figure 6.7. Generally, the "structure-specific" scenario produces better overall performance regardless of the type of ASM. However, for both scenarios, the multi-element ASM does outperform its traditional ASM counterpart, with a general trend of better mean performance when using more elements in the model. In the "structure-agnostic" scenario a multi-element ASM with 12 elements performs best, with a mean surface error of 0.246mm. For the "structure-specific" scenario, the best results of 0.214mm occur with structure-specific multi-element ASMs containing 8 elements. A paired difference analysis further shows that these multi-element ASMs are significantly better than their traditional ASM counterparts ($p < 0.05$), leading to a 14% mean improvement in overall segmentation performance for both scenarios. (Figure 6.8)

For the "structure-specific" scenario, we further show a breakdown of segmentation performance with regards to each of the ear structures. Boxplots of segmentation errors and paired differences compared to the use of traditional structure-specific ASMs can be found in Figure 6.9 and Figure 6.10. Using structure-

specific multi-element ASMs with 12 elements provide the best mean segmentation performance on the ossicles and labyrinth of 0.142mm and 0.145mm respectively. For the facial nerve and ear canal, the best respective mean results of 0.295mm and 0.261mm occur with multi-element ASMs utilizing 8 elements. These values correspond to performance improvements over traditional structure-specific ASMs of 0.103mm for the ossicles, 0.026mm for the labyrinth, 0.011mm for the facial nerve, and 0.048mm for the ear canal. While these values translate to respective paired difference improvements of 42%, 15%, 3%, and 15%, only the ossicles meet the Bonferroni corrected significance threshold of 0.05/4 = 0.013. We further note the relatively poor mean performance of a traditional ASM on the ossicles due to the large outlier at 0.900mm. Further investigation into this case points to the relatively abnormal ossicle anatomy of the subject, thus leading to a poor solution when using the traditional ASM method. However, all multi-element ASMs tested remained robust to this abnormality and still produced reasonable results for the ossicles. (Figure 6.11)

A heatmap of paired surface errors comparing traditional structure-specific ASMs with the overall best performing model, the 8-element structure-specific ASMs can be seen in Figure 6.12. Qualitatively, our model does provide benefits in regions where one would expect correlations between points belonging to nearby structures. For example, the contact point of ossicles and the labyrinth (Region A), or the contact point of the tympanic membrane with the ossicles (Region B). Particularly in Region A, we observe a large improvement over a traditional ASM due to the multi-element ASM's robustness against the outlier case mentioned previously. However, for regions that are not expected to correlate well with other structures, such as the body of the facial nerve (Region C), or the inferior part of the ear canal (Region D) we observe a worse result than the traditional ASM, as the added degrees of freedom provided by our model are likely susceptible to noise.

For our robustness test using 8-element structure specific models on our clinical database of 132 CT images, we show boxplots of mean surface differences between the original automatic segmentations and our proposed model in Figure 6.13. Mean discrepancies are 0.19mm for the ossicles, 0.11mm for the labyrinth, 0.70mm for the facial nerve, and 0.40mm for the ear canal. From previous experience using these original segmentations in clinical practice, these discrepancies are already clinically acceptable. However, when visually inspecting the cases of highest difference between the two automatic segmentation methods, we find that our multi-element model has a better segmentation overall. Visual examples of the largest discrepancies between both segmentation methods can be found in Figure 6.14

6.5.4 Chest Dataset

Boxplots of the results of our parameter pilot study can be found in Figure 6.15 and Figure 6.16. For the "structure-agnostic" scenario, we find that an 8-element model and a 14-element model produce the best

overall mean surface errors of 4.96mm and 3.90mm for the 4-sample and 9-sample case respectively. For the "structure-specific" scenario, the 5-element model performs the best for both 4 and 9-sample cases. This results in a mean surface error of 5.02mm for the 4-sample case and 3.96mm for the 9-sample case.

From the results of our pilot parameter study, we therefore choose an 8-element and 14-element model for the "structure-agnostic" scenario involving 5 and 10 samples respectively. For the "structure-specific" scenario, we choose a 5-element model for both sample sizes. Figure 6.17 shows the overall mean segmentation errors and paired differences (Figure 6.18) when using this setup for the full dataset. We find the multi-element ASM outperforms its traditional ASM counterparts for both scenarios and for all training sizes tested. When structure-specific information is not present, the multi-element ASMs produce a mean surface segmentation error of 4.83mm for the 5-sample training set and 3.33mm for the 10-sample training set. This improves over their respective traditional ASM counterparts by 0.05mm and 0.30mm. For the structure-specific scenario, the 5-element ASM produces a mean surface segmentation error of 4.83mm and 3.23mm or an improvement of 0.01mm and 0.05mm over the traditional structure-specific ASMs for a 5 and 10 training size case respectively. Though these are relatively small improvements in overall segmentation error, a Wilcoxon paired difference test shows them to be statistically significant ($p < 0.05$)

For the structure-specific scenario, the boxplots of surface segmentation errors and paired differences broken down by each structure can be found in Figure 6.19 and Figure 6.20. For a training set size of 5 samples, mean surface segmentation errors are 4.37mm and 4.99mm for the left and right lungs, 3.54mm and 3.58mm for the left and right clavicles, and 7.51mm for the heart. These result in 1-2% improvements for all structures except for the heart, where the multi-element ASM results in a 2% decrease in performance. However, paired difference testing shows statistical significance for both lungs and the left clavicle under the Bonferroni corrected value of $p<.05/5 = .01$. Meanwhile, for a training set size of 10, mean surface segmentation errors are 3.30mm and 2.82mm for the left and right lungs, 2.33mm and 2.28mm for the left and right clavicles, and 6.02mm for the heart. These result in 2-3% improvement over the structure specific traditional ASM for the lungs, $<1\%$ for the clavicles, and 1% for the heart. Only the right lung meets the significance threshold of $p < .01$.

A heatmap comparing the paired surface error distances between the 5-element structure specific ASMs and a traditional structure-specific ASM can be seen in Figure 6.21. Our multi-element ASM seems to perform well in regions that have clear boundaries such as the bottom and side portions of the lungs, and very poorly in the difficult to segment areas of the clavicles and heart. The latter regions are generally associated with very noisy features in the image, thus suggesting that the added degrees of freedom introduced by our model are prone to capturing poor local minima in these regions.

6.5.5 Model Training

Some visual examples of the training process can be found in 6.22. In the initial phase of training, the locations of the gaussian elements remain relatively close to their initialization positions and do not tend to migrate to distant regions of the shape from where they started. Instead, gaussian elements increase their regional influence via shaping of the covariance matrix. Interestingly, the resulting alignment of the principal axes for each of these elements seems to correlate with local region under the gaussian's influence. For example, in the labyrinth dataset there are gaussian elements that seem to align with the contours of the semi-circular canals. In the ear dataset, two of the elements align with the primary axis of the facial nerve body, while another two elements are aligned with the 'bend' in the inferior part of the ear canal. The same is true for the chest, where elements are aligned with the clavicles and the bottom parts of the lungs and heart. In the box dataset, the gaussians are generally aligned with each box, but since this would already produce a correct distribution of importance weighting for each of the boxes, seems to have converged on this local minimum.

For the final stage of training, there seems to be only subtle changes to the regions under each element in the model. The box dataset, as mentioned, has already converged to a local minimum that produces the correct importance weighting distribution for the boxes so little more optimizations can occur in the final training stage. For the remaining datasets, the regions with highest importance for each element appears to be consistent with their corresponding regions from the gaussian optimization stage which suggests that our training process has converged to a stable solution.

We further show the number of the largest eigenmodes needed to represent 95% of the variation for both traditional ASMs and each of the elements in our multi-element models. As mentioned in our 2-D rectangle example and shown by the multi-element models in the 3-D synthetic box dataset, simpler models using fewer eigenmodes would be more likely to capture more meaningful shape variation and thus use its limited training set much more efficiently. The full eigenmode comparison involving our best performing multi-element models for each dataset and scenario can be found in Tables 6.2-6.7. Though a direct statistical comparison of these quantities between the multi-element model and their traditional ASM counterparts is not possible due to the small number of trained models, there appears to be a general trend of multi-element models requiring less eigenmodes to represent 95% of the variation per element than their traditional ASM counterparts. Furthermore, chest dataset appears to be a much simpler dataset than the labyrinth or the full ear, since a traditional ASM already requires relatively few samples with respect to the training size (Tables 6.6 and 6.7).

6.6 Discussion

In this work, we have introduced the multi-element ASM as an extension to the traditional ASM, particularly for small sample size applications. We have described the optimization process for this model and how it can be adapted for analysing and segmenting complex single or multi-structure shapes. Furthermore, we have validated the model on 4 different datasets to show the broad applicability of this approach. We found that across all the datasets, our optimization method produces stable and intuitive regions of influence for each of the elements in the model. Particularly in our box dataset, we show that our training process can indeed find the correct distribution of importance weights for each model element. Furthermore, our multi-element model produces elements that capture simpler modes of variation due to the fewer eigenmodes needed to reach the 95% model variation threshold. Such training results do provide confidence that each element is able to model real variation that exists in the shape rather than spurious correlations that may exist between different landmarks.

For each of our datasets, we show our multi-element ASM can outperform the traditional ASM in both single-structure and multi-structure scenarios. For the labyrinth dataset, we found a 16-element ASM produces a mean surface segmentation error of ˜0.1mm for both sample sizes tested, thereby significantly improving segmentation performance by 26%. For our ear dataset, the comparison of structure agnostic traditional ASMs and multi-element ASMs lead to a 14% improvement in overall segmentation performance. In the structure-specific case, we get the best results from an 8 or 12 element ASM producing segmentation errors of ˜0.15mm for the labyrinth and ossicles, ˜0.25mm for the ear canal, and ˜0.3mm for the facial nerve. Furthermore, we found that our method's use of surrounding structures in a "structure-specific" scenario was beneficial in providing robustness in an abnormal ossicle segmentation. Overall, for the chest dataset, our model was confirmed to have robust performance in a larger clinical dataset of 132 patient CT images. Finally in our chest dataset, our model provided a 10% improvement in the "structure-agnostic" scenario when using a training set size of 10 samples, but generally provided small 1-3% improvements in the remaining cases.

For the labyrinth dataset, we see a clear advantage of our approach. In this dataset, salient boundary features exist for the landmarks, thereby making the added degrees of freedom provided by our model beneficial. A traditional ASM would fail to capitalize on the good candidate locations found by an image search

Training Set Size	Global ASM	2-Element Model
5	2.9	0*

Table 6.2: Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the synthetic boxes dataset. 0* indicates that the largest eigenmode already captures >95% of the variation in the model.

due to the limited representational power of a single linear vector space. On the other hand, a multi-element model can distribute different regions to each element, thereby increasing the overall degrees of freedom of the model. As shown in Figure 6.22, the presence of elements that capture shape variation in the semi-circular canals help model what would otherwise be complex geometry for a single ASM to capture at low sample sizes. Therefore, when given good candidate locations from the image search, the multi-element model can use the extra degrees of freedom to produce more accurate segmentations.

This reasoning can be extended to the rest of the datasets, particularly when observing the better performance of the multi-element ASM in structure-agnostic scenarios. For these cases, we treat all landmarks as coming from a single shape, thereby creating a highly complex geometry that a traditional ASM would have difficulty modelling with a single linear vector space. Though the image saliency of individual landmarks may vary across the shape, overall, the image search is able to find enough good candidates for the multi-element ASM to use its increased representational ability to provide better segmentation results.

Another benefit of our approach in structure-specific settings can be seen when segmenting the region around the ossicles near the contact point of the tympanic membrane and labyrinth (Regions A and B in Figure 6.12). Since these regions must be in contact as part of the function of the ear, we would expect the landmark variations of these regions to be highly correlated and predictive of each other. Indeed, the increased performance of our structure-specific multi-element ASMs in this region confirm this hypothesis, as elements can place some importance weighting to surrounding structures when creating the model. Therefore, the information provided by surrounding structures can help provide robustness during the segmentation process. The clear example of this interplay can be seen from our improved segmentation result for the ossicle outlier case in Figure 6.11.

However, some of the structure specific cases of the ear and chest dataset do highlight some limitations of our approach. Since we are operating on a particular structure, the modes of variation are simpler than what would otherwise be required for a full structure-agnostic case. Furthermore, in the case of the chest dataset, we see that a traditional ASM requires relatively few eigenmodes compared to the training size to represent 95% of the shape variation for both scenarios (Tables 6.6 and 6.7). Therefore, for these cases a traditional ASM would already be able to represent the shape variation on its own and any additional degrees of freedom

Training Set Size	Global ASM	16-Element Model
6	3.67	3.27
9	5.5	5.13

Table 6.3: Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the labyrinth dataset

introduced by our model would be prone to image noise. Additionally, for regions where we would not expect any correlation with surrounding structures, such as the base of the facial nerve (Region C) or the inferior region of the ear canal (Region D) there will be little information to constrain solutions when provided poor candidate positions. Therefore, the poorer performance of our model in these regions suggest that the degrees of freedom provided by our model are falling into poor local minima from image noise. Another example of poor results due to noise can be seen in the heart segmentation of our chest dataset (Figure 6.21). Since landmarks in this region have few visible features from the image, the noisy candidate locations found by the image search produce poorer results with our model.

Overall, our method demonstrates itself to be a viable extension to the traditional ASM framework, providing benefits to segmentation performance across a variety of different shape libraries, particularly when constrained by limited training data. Our results suggest that landmark saliency in the image is the main contributing factor in the quality of segmentations produced by our model. Therefore, future work will be focused on integrating image feature saliency as part of the optimization in our model. Since to our knowledge there are no publicly available implementations of the other mentioned extensions to the ASM, further future work will involve implementing and testing those methods for comparison with our framework. Doing so will further increase and understand scenarios where our approach can be best utilized.

Training Set Size	Global ASM	16-Element Model
5	3	2.62
6	3	3.37

Table 6.4: Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the full ear dataset in a ”structure-agnostic” scenario

Training Set Size	Structure	Traditional SS ASM	8-Element SS Models
5	Ossicles	3.00	2.87
	Facial Nerve	3.00	2.62
	Labyrinth	3.00	2.50
	Ear Canal	3.00	2.50
6	Ossicles	4.00	3.50
	Facial Nerve	3.00	3.37
	Labyrinth	4.00	2.62
	Ear Canal	3.00	3.00

Table 6.5: Comparision of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and an individual structure specific (SS) ASM on the full ear dataset in a "structure-specific" scenario

Training Set Size	Global ASM	8-Element ASM	14-Element ASM
5	2.25	1.64	-
10	4.75	-	2.23

Table 6.6: Comparison of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and a traditional global ASM on the chest dataset in a "structure-agnostic" scenario

Training Set Size	Structure	Traditional SS ASM	8-Element SS Models
5	Heart	1.75	1.53
	Left Clavicle	1.88	1.63
	Right Clavicle	1.75	1.48
	Left Lung	2.00	1.63
	Right Lung	1.88	1.45
10	Heart	3.00	2.75
	Left Clavicle	2.50	2.95
	Right Clavicle	2.50	2.90
	Left Lung	3.25	3.20
	Right Lung	2.75	2.65

Table 6.7: Comparision of the mean number of the largest eigenmodes needed to represent 95% of the total variance captured in a particular element and an individual structure specific (SS) ASM on the chest dataset in a "structure-specific" scenario

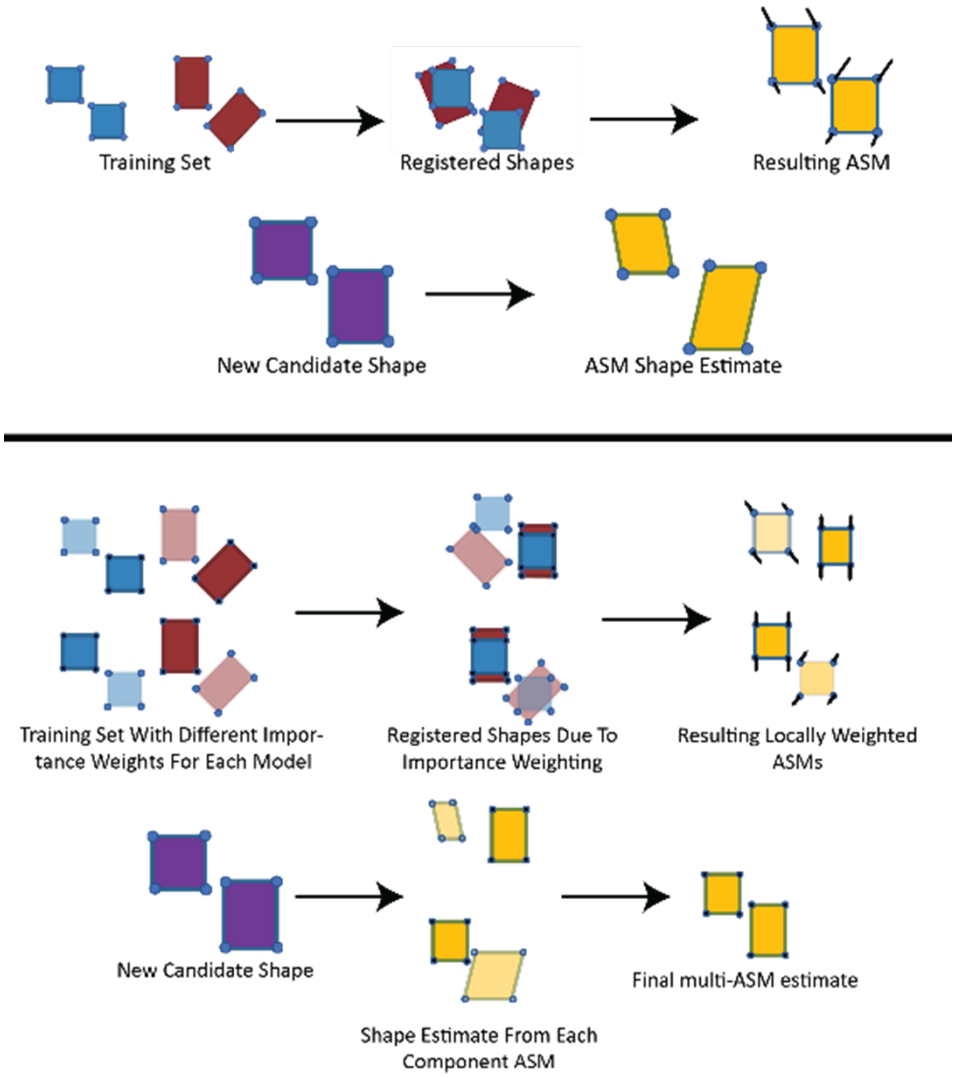

Figure 6.1: Example of two independently rotating boxes (Top) A traditional ASM, due to the single linear vector space will produce complex eigenmodes and thus an incorrect fit when given a new example. (Bottom) A model involving two ASMs that utilize importance weightings on a different box will produce estimates that, when combined, will produce the desired fit.

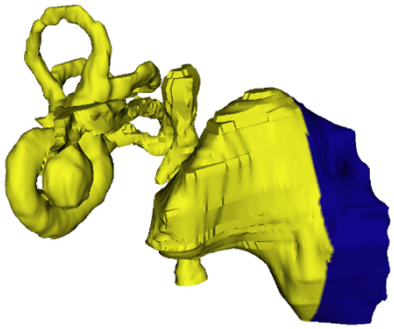

Figure 6.2: Mask utilized in the full ear dataset. As the lateral region of the ear canal is an arbitrary boundary, the points labeled in blue are not used in the training and evaluation stages of any of our models.

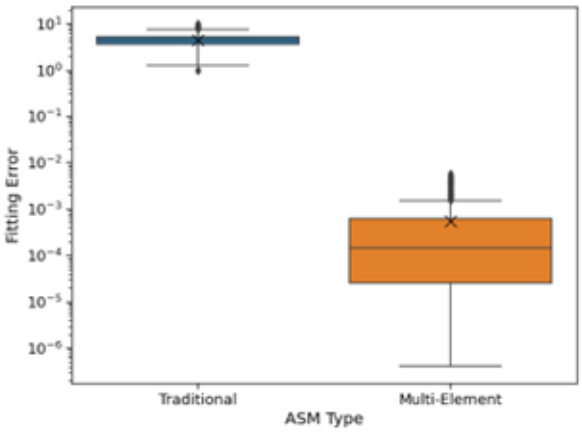

Figure 6.3: Boxplot of fitting errors comparing a traditional vs. multi-element ASM for our boxes dataset.

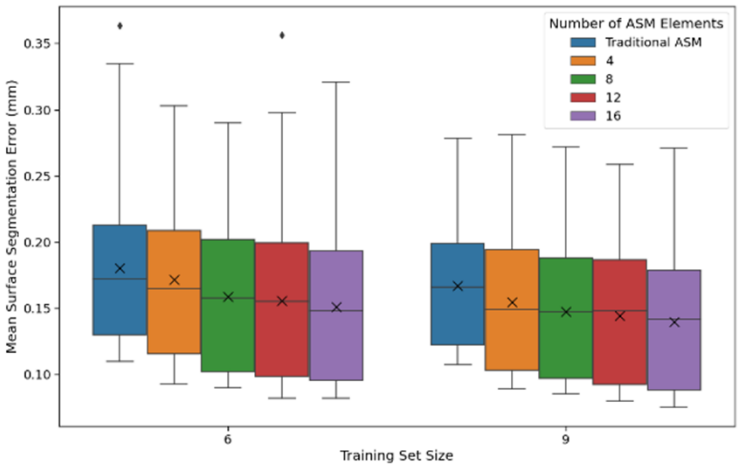

Figure 6.4: Boxplot of mean surface segmentation error for the labyrinth dataset

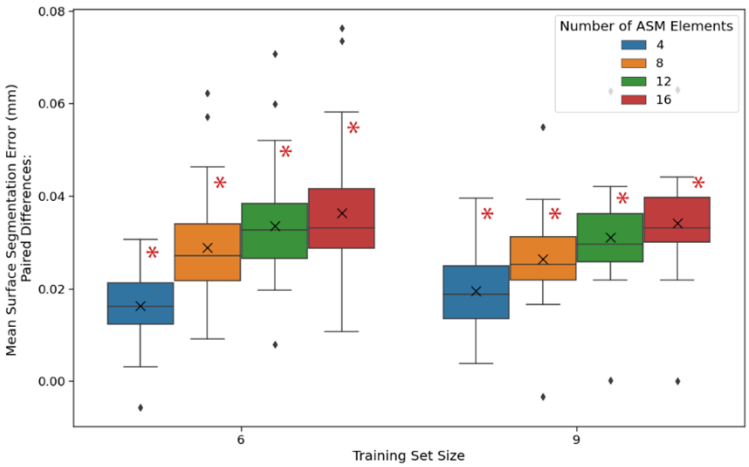

Figure 6.5: Boxplot of surface segmentation error paired differences for the labyrinth dataset when comparing a multi-element ASM to a traditional ASM. More positive values indicate better performance for the multi-element model.

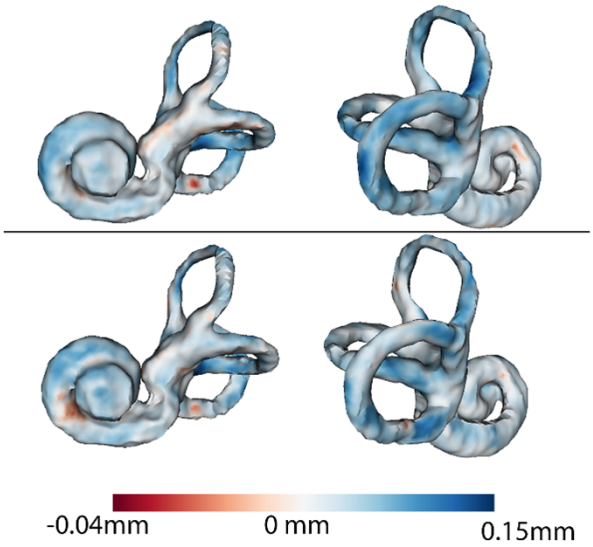

Figure 6.6: Heatmap of mean paired surface errors comparing the 16-element ASM vs a traditional ASM for a training set size of 6 (top row) and 9 (bottom row). Each row shows two views of the labyrinth. Negative values (red hues) indicate better performance of a traditional ASM while positive values (blue hues) indicate better performance for the multi-element model.

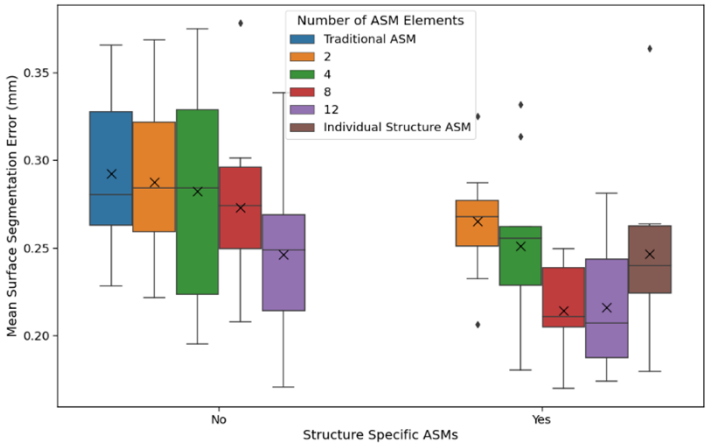

Figure 6.7: Boxplot of overall mean segmentation error when not using structure-specific ASMs (No) or using structure-specific ASMs (Yes)

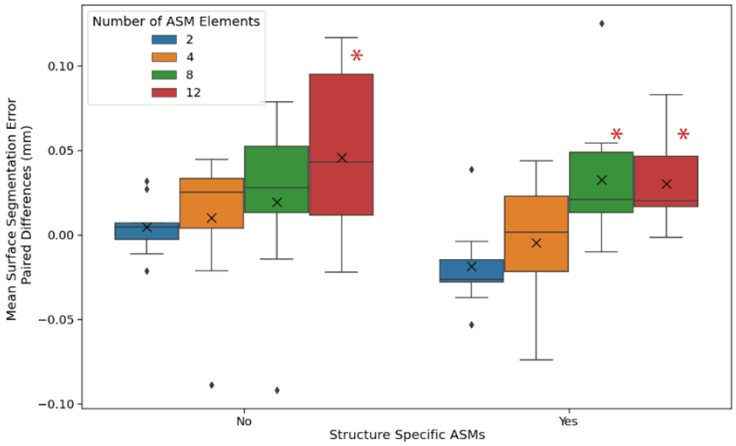

Figure 6.8: Paired differences for the two structure-specific scenarios with the ear dataset. For the 'No' case, paired differences are compared with a traditional ASM. For the 'Yes' case, paired differences are calculated against a traditional ASM created for each individual structure. For both scenarios, more positive numbers indicate better performance of the multi-element ASM. Red asterisk indicates statistical significance ($p < .05$)

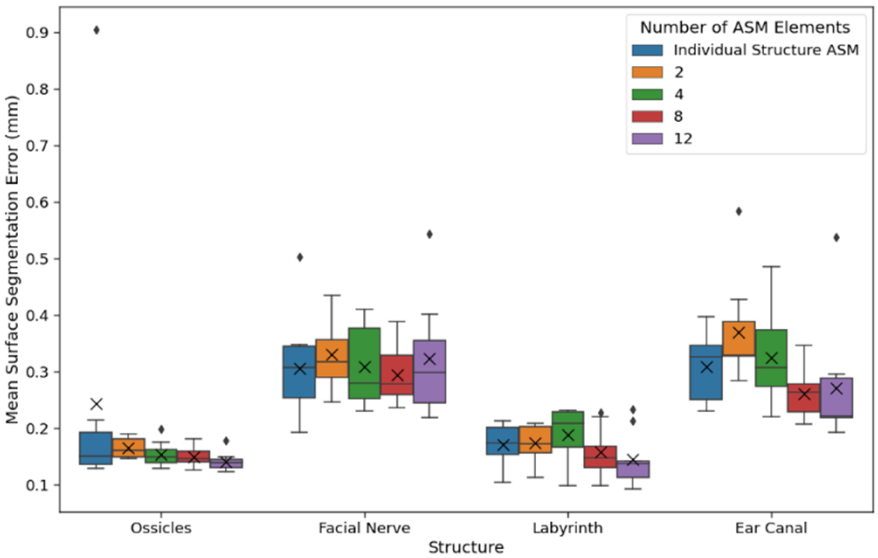

Figure 6.9: Boxplot of mean surface segmentation error for each structure in the ear when structure groupings of points are known

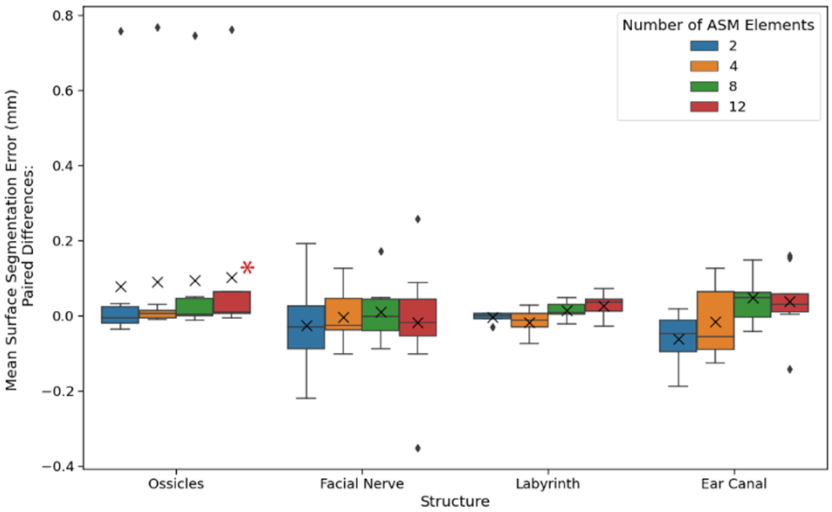

Figure 6.10: Boxplot of surface segmentation error paired differences between different structure-specific multi-element ASMs compared with a traditional ASM for each structure. More positive values indicate a better performance for the multi-element ASM. Red asterisk indicates significance at Bonferoni corrected level of ($p < 0.013$)

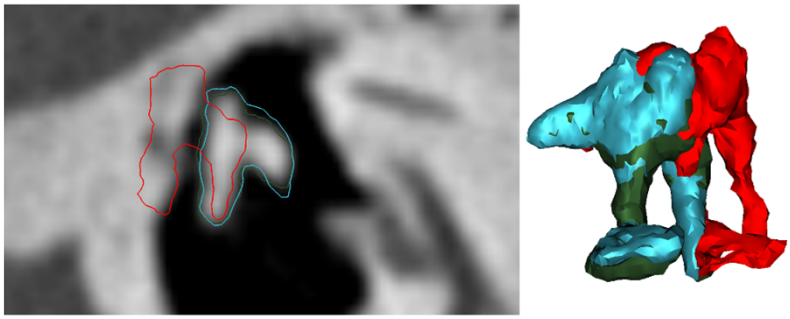

Figure 6.11: Example segmentation of the ossicle outlier case in our structure-specific scenario. Green is ground truth. Blue is the segmentation result of the multi-element ASM. Red is the result of the traditional structure-specific ASM for the ossicles.

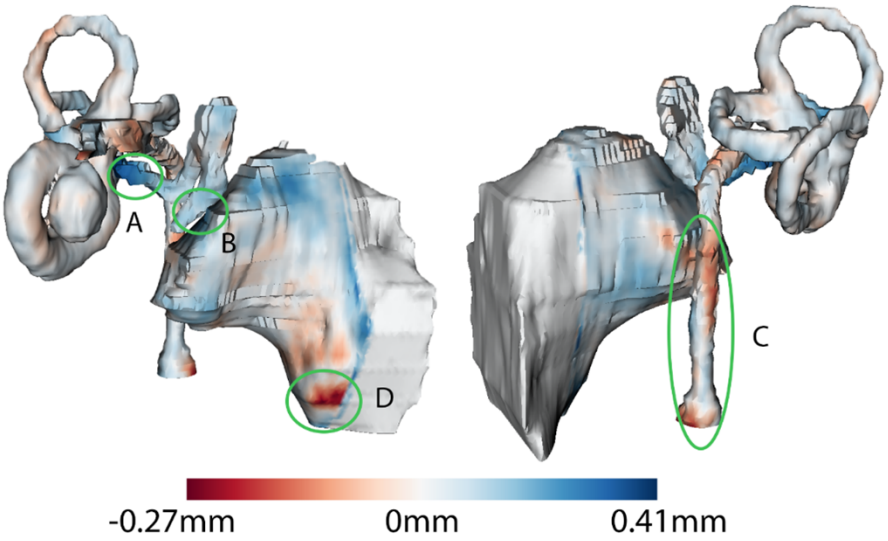

Figure 6.12: Heatmap of mean paired surface errors comparing the 8-element structure-specific multi-element ASM to a traditional ASM targeting each individual structure. Negative values (red hues) indicate better performance of the individual structure specific ASM while positive values (blue hues) indicate better performance for the multi-element model. Regions of interest are circled in green. (Region A) Contact point of ossicles with labyrinth. (Region B) Contact point of ossicles with ear canal. (Region C) Body of facial nerve. (Region D) Inferior region of the ear canal.

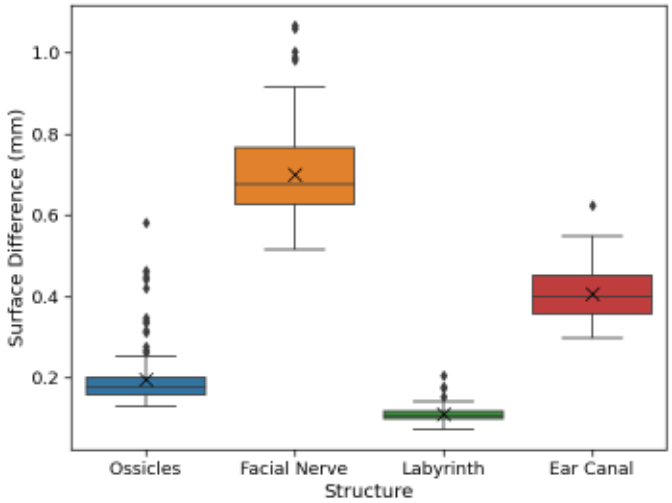

Figure 6.13: Boxplot of surface segmentation differences between our 8-element structure-specific ASMs and the original automatic segmentations of our clinical dataset of 132 patient CTs.

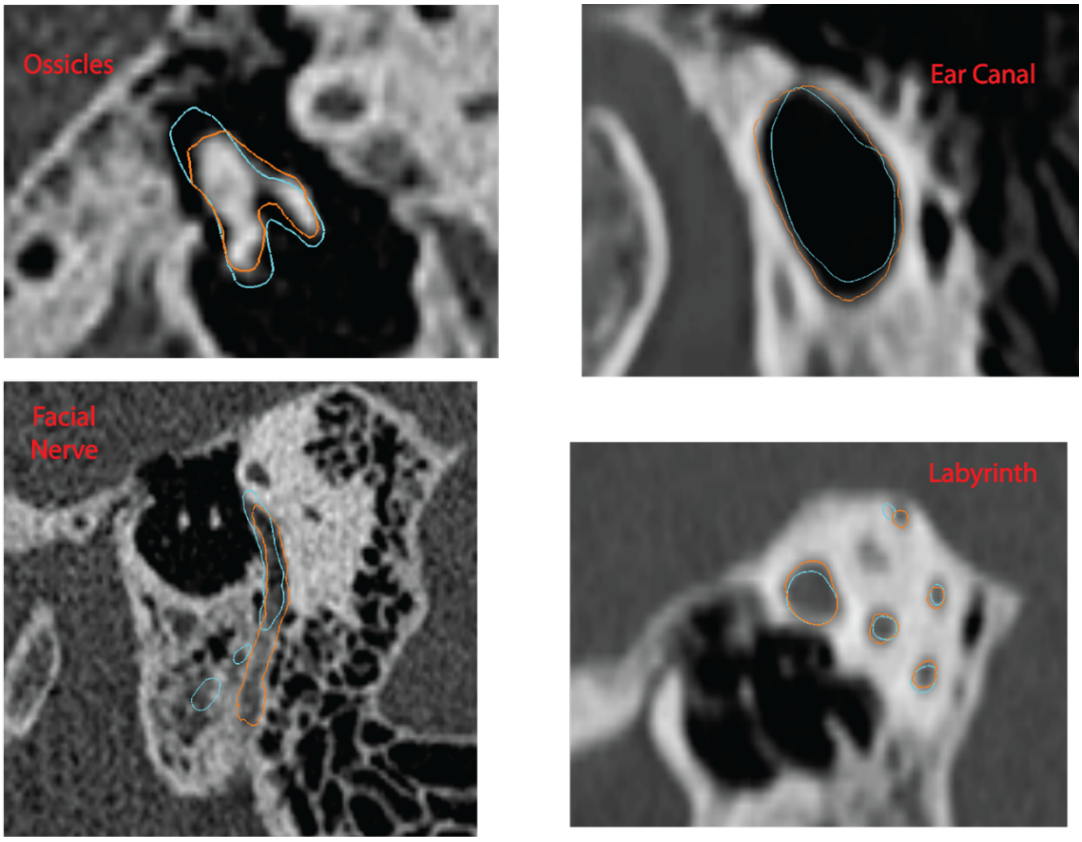

Figure 6.14: Example sagittal slices of the largest discrepancies in surface error between the original segmentations (Blue) and the 8-element structure specific segmentations (Orange) in our robustness test.

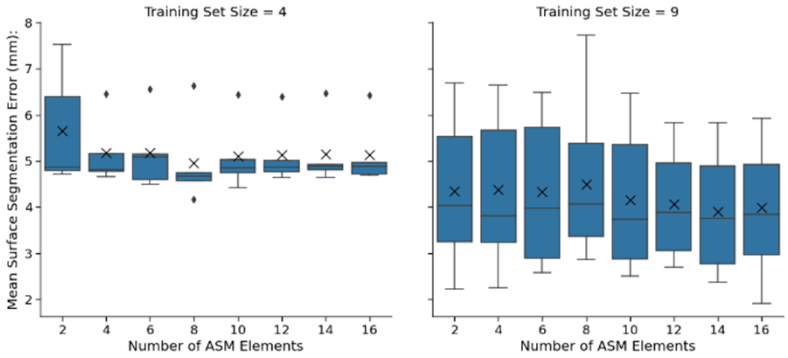

Figure 6.15: Mean surface segmentation errors for the parameter pilot study in our chest dataset when determining the optimal number of elements to use in a multi-element ASM for a structure agnostic case.

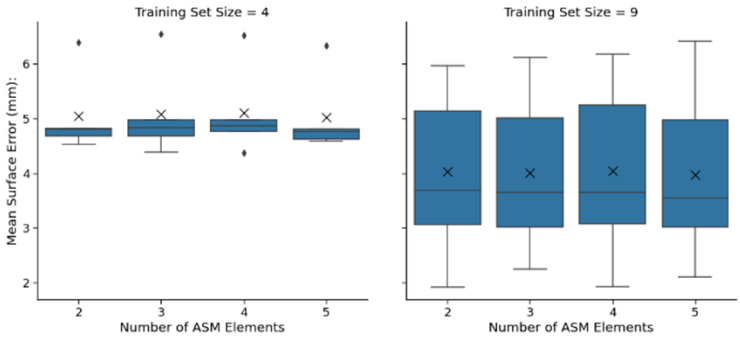

Figure 6.16: Overall mean surface segmentation errors for the parameter pilot study in the chest dataset when determining the optimal number of elements to use for the multi-element ASM in the structure specific case.

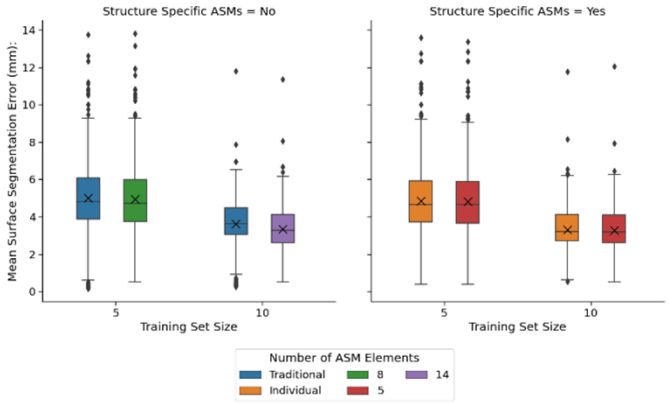

Figure 6.17: Boxplot of overall surface error comparing the multi-element ASM against its corresponding traditional ASM counterpart. (Left) A structure-agnostic scenario where the traditional ASM is compared against a multi-element ASM with the number of elements determined from the parameter pilot study. (Right) Structure specific scenario comparing traditional structure-specific ASMs against a 5-element multi-element ASM.

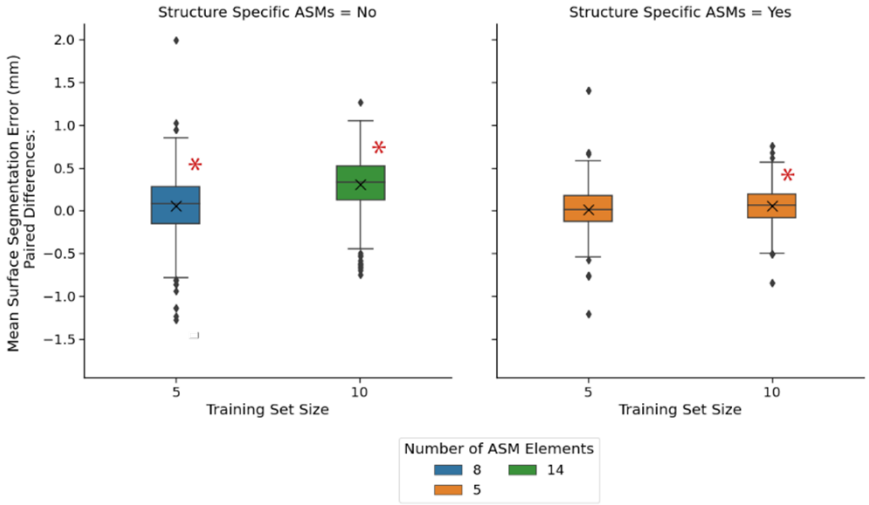

Figure 6.18: Boxplot of paired differences comparing the multi-element ASM against its corresponding traditional ASM counterpart. (Left) A structure-agnostic scenario where the traditional ASM is compared against a multi-element ASM with the number of elements determined from the parameter pilot study. (Right) Structure specific scenario comparing traditional structure-specific ASMs against a 5-element multi-element ASM. More positive values indicate better performance for the multi-element ASM. Red asterisk indicates statistical significance ($p < 0.05$)

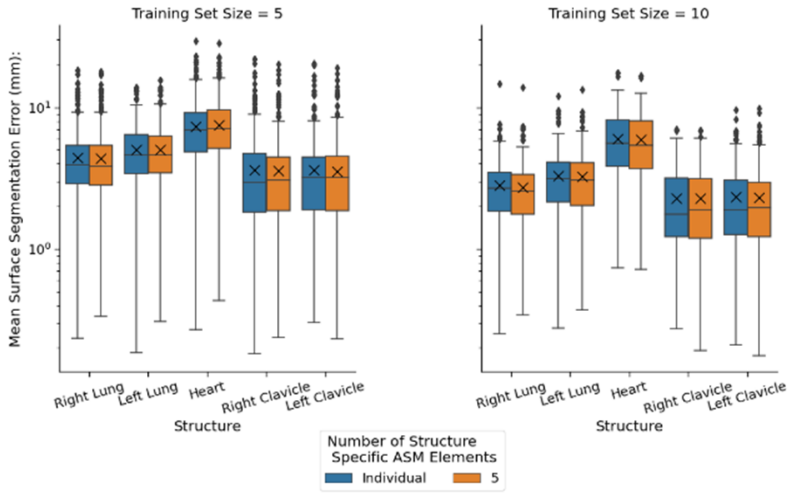

Figure 6.19: Boxplot of mean surface error for each of the chest structures when comparing a 5-element structure specific ASM against a structure-specific traditional ASM for each structure.

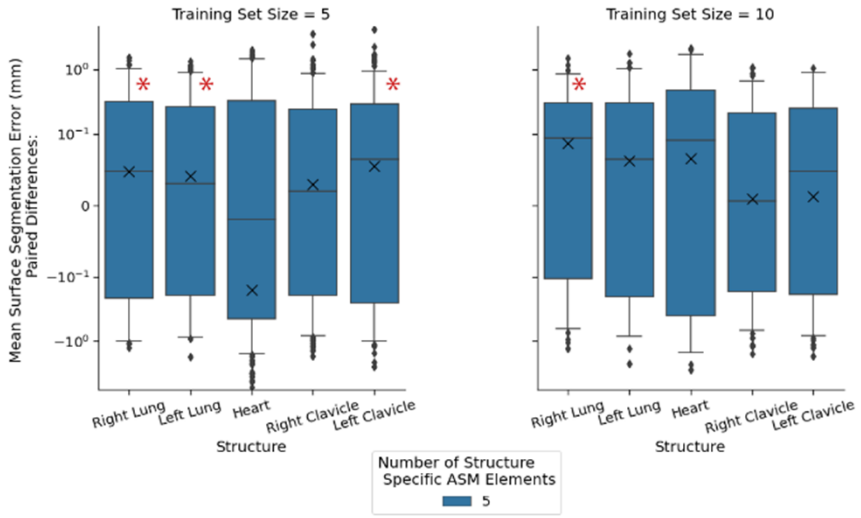

Figure 6.20: Boxplot of paired differences comparing surface segmentation error of a 5-element structure specific ASM against a structure-specific traditional ASM for each structure. More positive values indicate better performance for the multi-element ASM. Red asterisk indicates significance under Bonferroni corrected ($p < .01$)

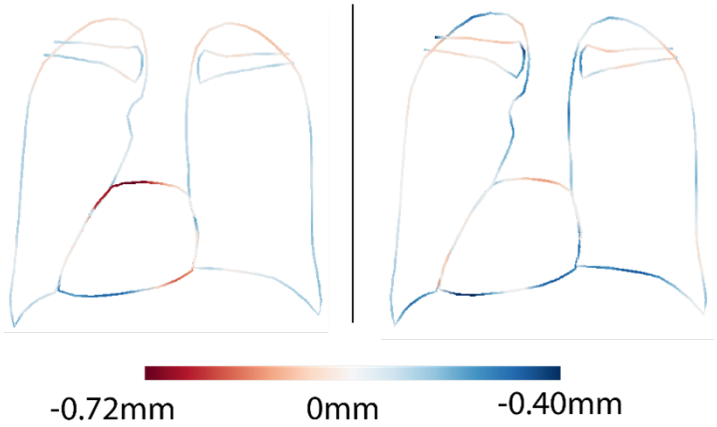

Figure 6.21: Heatmap of mean paired surface errors comparing the 5-element structure specific ASM and traditional structure-specific ASMs. Redder hues indicate better performance of the structure-specific traditional ASMs and bluer hues indicate better performance of the 5-element model. (Left) comparison of both models when using a training set size of 5 samples. (Right) comparison of both models when using a training set size of 10 samples.

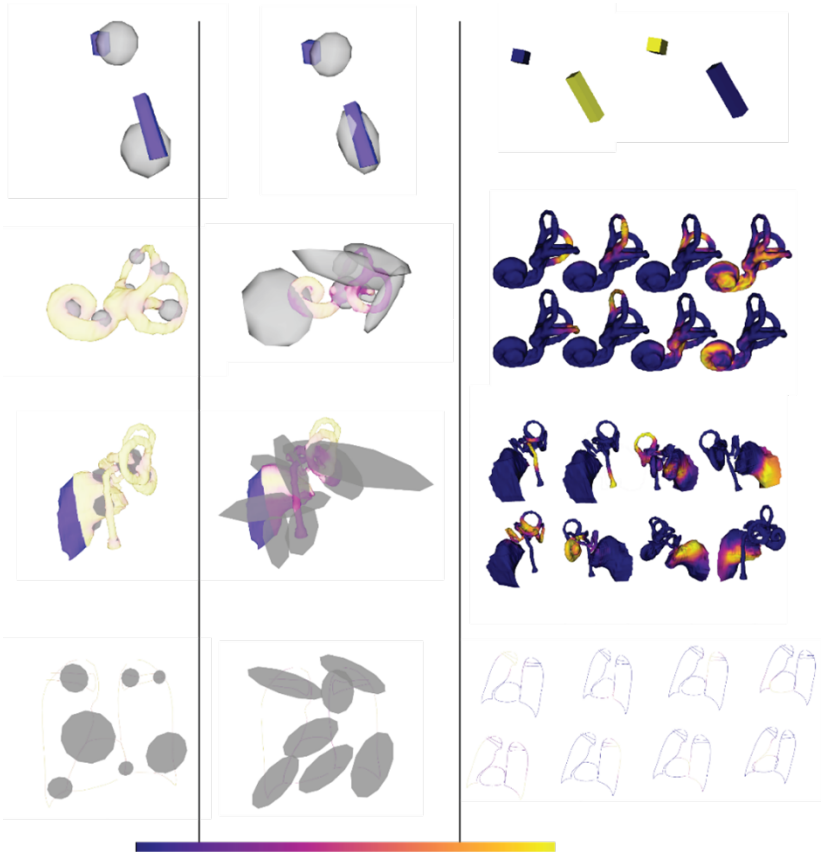

Figure 6.22: Examples of training for each of the 4 datasets in this study. Left column is the initialization of the gaussian elements. Middle column is the result of the gaussian elements after the initial training stage. Colors shown in left/middle columns are the mixing weight of the global element in the model. Right column is the mixing weights for each of the models after the final stage of training.

CHAPTER 7

Conclusion

In this dissertation we have addressed various issues for improving outcomes for CI recipients with a variety of image processing techniques. In Chapter 3 we investigated generic insertion depth guidelines for pre-curved arrays and found the round window to be a more reliable marker than the facial recess for achieving the correct insertion depth when using the optimal insertion trajectory. Furthermore, a round window marker placed at the population mean of 2.28 mm would produce fewer over or under insertions than the current round window markers when using the ideal trajectory. In Chapter 4 we demonstrated a weakly-supervised approach to automatic chorda segmentation using a GAN architecture. We showed that our proposed approach performs significantly better than the current automatic method in use for chorda segmentation. In Chapter 5 we used an expanded dataset of 16 samples to perform an extensive validation of the active shape model for use in segmenting the intra-cochlear anatomy. We found optimal parameters for the model and showed robust performance over a large database of clinical CT images. We found that additional training samples would lead to only diminishing returns on segmentation performance and future work in that area should be focused on the candidate search process in the image. Finally, in Chapter 6 we proposed the multi-element ASM as an extension to the ASM framework. We showed its efficacy on datasets of synthetic boxes, cochlear labyrinth, full ear, and chest. Our results suggest the model's viability for improving segmentation performance in scenarios involving small shape libraries. Taken together, this work represents a contribution to improving outcomes for CI recipients. The work in Chapter 3 now provides evidence for better CI insertion guidelines to minimize poor outcomes in patients. The improved chorda tympani segmentation provided by our GAN in Chapter 4 provides the possibility of safer and more robust insertion planning techniques. The validation of the ASM in Chapter 5 now gives optimal parameters for intra-cochlear segmentation and provides evidence for its robust performance in clinical settings. Finally, the multi-element ASM proposed in Chapter 7 not only provides better segmentation performance over traditional ASM techniques for relevant ear structures, but also gives a general method for the modeling and segmentation of structures when faced with a small shape library constraint.

References

[1] E. Buss, H. C. Pillsbury, C. A. Buchman, C. H. Pillsbury, M. S. Clark, D. S. Haynes, R. F. Labadie, S. Amberg, P. S. Roland, P. Kruger, M. A. Novak, J. A. Wirth, J. M. Black, R. Peters, J. Lake, P. A. Wackym, J. B. Firszt, B. S. Wilson, D. T. Lawson, R. Schatzer, P. S. C. D'Haese, and A. L. Barco, "Multicenter U.S. Bilateral MED-EL Cochlear Implantation Study: Speech Perception over the First Year of Use," *Ear and Hearing*, vol. 29, pp. 20–32, Jan. 2008.

[2] B. S. Wilson and M. F. Dorman, "Cochlear implants: current designs and future possibilities," *Journal of Rehabilitation Research and Development*, vol. 45, no. 5, pp. 695–730, 2008.

[3] R. H. Gifford, J. K. Shallop, and A. M. Peterson, "Speech Recognition Materials and Ceiling Effects: Considerations for Cochlear Implant Programs," *Audiology and Neurotology*, vol. 13, no. 3, pp. 193–205, 2008. Publisher: Karger Publishers.

[4] R. H. Gifford, M. F. Dorman, S. W. Sheffield, K. Teece, and A. P. Olund, "Availability of binaural cues for bilateral implant recipients and bimodal listeners with and without preserved hearing in the implanted ear," *Audiology & Neuro-Otology*, vol. 19, no. 1, pp. 57–71, 2014.

[5] R. Litovsky, A. Parkinson, J. Arcaroli, and C. Sammeth, "Simultaneous Bilateral Cochlear Implantation in Adults: A Multicenter Clinical Study," *Ear and hearing*, vol. 27, p. 714, Dec. 2006. ISBN: 9780000246813 Publisher: NIH Public Access.

[6] J. T. Rubinstein, "How cochlear implants encode speech," *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 12, pp. 444–448, Oct. 2004.

[7] R. H. Gifford, J. H. Noble, S. M. Camarata, L. W. Sunderhaus, R. T. Dwyer, B. M. Dawant, M. S. Dietrich, and R. F. Labadie, "The Relationship Between Spectral Modulation Detection and Speech Recognition: Adult Versus Pediatric Cochlear Implant Recipients," *Trends in Hearing*, vol. 22, p. 2331216518771176, 2018. Publisher: SAGE Publications Inc.

[8] L. K. Holden, C. C. Finley, J. B. Firszt, T. A. Holden, C. Brenner, L. G. Potts, B. D. Gotter, S. S. Vanderhoof, K. Mispagel, G. Heydebrand, and M. W. Skinner, "Factors Affecting Open-Set Word Recognition in Adults With Cochlear Implants," *Ear & Hearing*, vol. 34, pp. 342–360, May 2013.

[9] G. B. Wanna, J. H. Noble, M. L. Carlson, R. H. Gifford, M. S. Dietrich, D. S. Haynes, B. M. Dawant, and R. F. Labadie, "Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes," *The Laryngoscope*, vol. 124 Suppl 6, pp. S1–7, Nov. 2014.

[10] G. B. Wanna, J. H. Noble, R. H. Gifford, M. S. Dietrich, A. D. Sweeney, D. Zhang, B. M. Dawant, A. Rivas, and R. F. Labadie, "Impact of Intrascalar Electrode Location, Electrode Type, and Angular Insertion Depth on Residual Hearing in Cochlear Implant Patients: Preliminary Results," *Otology & Neurotology*, vol. 36, pp. 1343–1348, Sept. 2015.

[11] K. Yukawa, L. Cohen, P. Blamey, B. Pyman, V. Tungvachirakul, and S. O'Leary, "Effects of Insertion Depth of Cochlear Implant Electrodes upon Speech Perception," *Audiology and Neurotology*, vol. 9, no. 3, pp. 163–172, 2004.

[12] C. C. Finley, T. A. Holden, L. K. Holden, B. R. Whiting, R. A. Chole, G. J. Neely, T. E. Hullar, and M. W. Skinner, "Role of Electrode Placement as a Contributor to Variability in Cochlear Implant Outcomes," *Otology & Neurotology*, vol. 29, pp. 920–928, Oct. 2008.

[13] S. Chakravorti, J. H. Noble, R. H. Gifford, B. M. Dawant, B. P. O'Connell, J. Wang, and R. F. Labadie, "Further Evidence of the Relationship Between Cochlear Implant Electrode Positioning and Hearing Outcomes," *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 40, pp. 617–624, June 2019.

[14] J. Wang, B. M. Dawant, R. F. Labadie, and J. H. Noble, "Retrospective Evaluation of a Technique for Patient-Customized Placement of Precurved Cochlear Implant Electrode Arrays," *Otolaryngology–Head and Neck Surgery*, vol. 157, pp. 107–112, July 2017.

[15] R. F. Labadie and J. H. Noble, "Preliminary Results With Image-guided Cochlear Implant Insertion Techniques," *Otology & Neurotology*, vol. 39, pp. 922–928, Aug. 2018.

[16] S. Pelosi, J. H. Noble, B. M. Dawant, and R. F. Labadie, "Analysis of Intersubject Variations in Intracochlear and Middle Ear Surface Anatomy for Cochlear Implantation," *Otology & Neurotology*, vol. 34, pp. 1675–1680, Dec. 2013.

[17] 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," *Medical Physics*, vol. 35, pp. 5375–5384, Dec. 2008.

[18] J. H. Noble, A. J. Hedley-Williams, L. Sunderhaus, B. M. Dawant, R. F. Labadie, S. M. Camarata, and R. H. Gifford, "Initial Results With Image-guided Cochlear Implant Programming in Children," *Otology & Neurotology*, vol. 37, pp. e63–e69, Feb. 2016.

[19] J. H. Noble, R. H. Gifford, A. J. Hedley-Williams, B. M. Dawant, and R. F. Labadie, "Clinical evaluation of an image-guided cochlear implant programming strategy," *Audiology & neuro-otology*, vol. 19, no. 6, pp. 400–411, 2014.

[20] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, Jan. 1995.

[21] G. Langs, P. Peloschek, and H. Bischof, "Optimal sub-shape models by minimum description length," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 310–315 vol. 2, June 2005. ISSN: 1063-6919.

[22] A. Dhanasingh and C. Jolly, "An overview of cochlear implant electrode array designs," *Hearing Research*, vol. 356, pp. 93–103, Dec. 2017.

[23] P. Mistrík, C. Jolly, D. Sieber, and I. Hochmair, "Challenging aspects of contemporary cochlear implant electrode array design," *World Journal of Otorhinolaryngology - Head and Neck Surgery*, vol. 3, pp. 192–199, Dec. 2017.

[24] A. Radeloff, M. Mack, M. Baghi, W. Gstoettner, and O. Adunka, "Variance of Angular Insertion Depths in Free-Fitting and Perimodiolar Cochlear Implant Electrodes," *Otology & Neurotology*, vol. 29, pp. 131–136, Feb. 2008.

[25] W. Gstoettner, P. Franz, J. Hamzavi, H. Plenk, W. Baumgartner, and C. Czerny, "Intracochlear position of cochlear implant electrodes," *Acta Oto-Laryngologica*, vol. 119, pp. 229–233, Mar. 1999.

[26] A. Franke-Trieger and D. Mürbe, "Estimation of insertion depth angle based on cochlea diameter and linear insertion depth: a prediction tool for the CI422," *European Archives of Oto-Rhino-Laryngology*, vol. 272, pp. 3193–3199, Nov. 2015.

[27] D. R. Ketten, M. W. Skinner, G. Wang, M. W. Vannier, G. A. Gates, and J. G. Neely, "In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays," *The Annals of Otology, Rhinology & Laryngology. Supplement*, vol. 175, pp. 1–16, Nov. 1998.

[28] B. Escudé, C. James, O. Deguine, N. Cochard, E. Eter, and B. Fraysse, "The size of the cochlea and predictions of insertion depth angles for cochlear implant electrodes.," *Audiology & neuro-otology*, vol. 11 Suppl 1, pp. 27–33, 2006.

[29] A. Franke-Trieger, C. Jolly, A. Darbinjan, T. Zahnert, and D. Mürbe, "Insertion Depth Angles of Cochlear Implant Arrays With Varying Length: A Temporal Bone Study," *Otology & Neurotology*, vol. 35, p. 58, Jan. 2014.

[30] K. S. van der Marel, J. J. Briaire, R. Wolterbeek, J. Snel-Bongers, B. M. Verbist, and J. H. M. Frijns, "Diversity in Cochlear Morphology and Its Influence on Cochlear Implant Electrode Position:," *Ear and Hearing*, vol. 35, no. 1, pp. e9–e20, 2014.

[31] P. Mistrík and C. Jolly, "Optimal electrode length to match patient specific cochlear anatomy," *European Annals of Otorhinolaryngology, Head and Neck Diseases*, vol. 133, pp. S68–S71, June 2016.

[32] M. Becker, M. Kirschner, and G. Sakas, "Segmentation of risk structures for otologic surgery using the Probabilistic Active Shape Model (PASM)," in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036, pp. 172–178, SPIE, Mar. 2014.

[33] F. A. Reda, J. H. Noble, A. Rivas, T. R. McRackan, R. F. Labadie, and B. M. Dawant, "Automatic segmentation of the facial nerve and chorda tympani in pediatric CT scans," *Medical Physics*, vol. 38, pp. 5590–5600, Oct. 2011.

[34] J. Fauser, I. Stenin, M. Bauer, W.-H. Hsu, J. Kristin, T. Klenzner, J. Schipper, and A. Mukhopadhyay, "Toward an automatic preoperative pipeline for image-guided temporal bone surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 967–976, June 2019.

[35] A. Ferreira, F. Gentil, and J. M. R. S. Tavares, "Segmentation algorithms for ear image data towards biomechanical studies," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, no. 8, pp. 888–904, 2014.

[36] C. Todd, M. Kirillov, M. Tarabichi, F. Naghdy, and G. Naghdy, "An analysis of medical image processing methods for segmentation of the inner ear," *University of Wollongong in Dubai - Papers*, pp. 213–218, Jan. 2009.

[37] a. M. W. V. G. W. Sun K. Yoo, Jay T. Rubinstein, "Semiautomatic Segmentation of the Cochlea Using Real-Time Volume Rendering and Regional Adaptive Snake Modeling," *Journal of Digital Imaging*, vol. 14, pp. 173–181, Dec. 2001.

[38] Diao Xianfen, Chen Siping, Liang Changhong, and Wang Yuanmei, "3D Semi-automatic Segmentation of the Cochlea and Inner Ear," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 6285–6288, Jan. 2005.

[39] C. Bellos, G. Rigas, I. F. Spiridon, A. Bibas, D. Iliopoulou, F. Böhnke, D. Koutsouris, and D. I. Fotiadis, "Reconstruction of cochlea based on micro-CT and histological images of the human inner ear," *BioMed Research International*, vol. 2014, p. 485783, 2014.

[40] G. E. Christensen, J. He, J. A. Dill, J. T. Rubinstein, M. W. Vannier, and G. Wang, "Automatic measurement of the labyrinth using image registration and a deformable inner ear atlas," *Academic Radiology*, vol. 10, pp. 988–999, Sept. 2003.

[41] A. Ferreira, F. Oliveira, F. Gentil, and J. M. R. S. Tavares, "A method to automatically segment the internal ear structures in CT images," 2013.

[42] J. H. Noble, R. F. Labadie, O. Majdani, and B. M. Dawant, "Automatic Segmentation of Intracochlear Anatomy in Conventional CT," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 2625–2632, Sept. 2011.

[43] F. A. Reda, T. R. McRackan, R. F. Labadie, B. M. Dawant, and J. H. Noble, "Automatic segmentation of intra-cochlear anatomy in post-implantation CT of unilateral cochlear implant recipients," *Medical Image Analysis*, vol. 18, pp. 605–615, Apr. 2014.

[44] J. Romera, H. M. Kjer, G. Piella, M. Ceresa, and M. A. G. Ballester, "Multi-region statistical shape model for cochlear implantation," in *Medical Imaging 2016: Image Processing*, vol. 9784, p. 97840T, International Society for Optics and Photonics, Mar. 2016.

[45] J. Gaa, S. Müller, G. J. Lexow, O. Majdani, L. A. Kahrs, and T. Ortmaier, "Statistical Shape Model for Automated Cochlear Segmentation: A Comparison of Fitting Strategies.," in *CURAC*, pp. 137–142, 2016.

[46] S. Zhu, W. Gao, Y. Zhang, J. Zheng, Z. Liu, and G. Yuan, "3D automatic MRI level set segmentation of inner ear based on statistical shape models prior," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6, Oct. 2017.

[47] E. P. Ruiz, H. M. Kjer, G. Piella, M. Ceresa, and M. B. González, "Random walks with shape prior for cochlea segmentation in ex vivo CT.," *International journal of computer assisted radiology and surgery*, vol. 11, pp. 1647–1659, Sept. 2016.

[48] T. Demarcy, *Segmentation and study of anatomical variability of the cochlea from medical images*. PhD Thesis, Université Côte d'Azur, July 2017.

[49] H. M. Kjer, J. Fagertun, W. Wimmer, N. Gerber, S. Vera, L. Barazzetti, N. Mangado, M. Ceresa, G. Piella, T. Stark, M. Stauber, M. Reyes, S. Weber, M. Caversaccio, M. González Ballester, and R. R. Paulsen, "Patient-specific estimation of detailed cochlear shape from clinical CT images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, pp. 389–396, Mar. 2018.

[50] D. Zhang, R. Banalagay, J. Wang, Y. Zhao, J. H. Noble, and B. M. Dawant, "Two-level training of a 3D U-Net for accurate segmentation of the intra-cochlear anatomy in head CTs with limited ground truth training data," in *Medical Imaging 2019: Image Processing*, vol. 10949, p. 1094907, International Society for Optics and Photonics, Mar. 2019.

[51] P. Mesejo, O. Ibáñez, O. Cordón, and S. Cagnoni, "A Survey on Image Segmentation using Metaheuristic-based Deformable Models: State of the Art and Critical Analysis," *Applied Soft Computing*, vol. 44, pp. 1–29, July 2016.

[52] A. Brunton, A. Salazar, T. Bolkart, and S. Wuhrer, "Statistical shape spaces for 3d data: a review," in *Handbook of Pattern Recognition and Computer Vision*, pp. 217–238, WORLD SCIENTIFIC, Jan. 2015.

[53] F. Ambellan, H. Lamecker, C. von Tycowicz, and S. Zachow, "Statistical Shape Models: Understanding and Mastering Variation in Anatomy," in *Biomedical Visualisation : Volume 3* (P. M. Rea, ed.), Advances in Experimental Medicine and Biology, pp. 67–84, Cham: Springer International Publishing, 2019.

[54] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Medical Image Analysis*, vol. 13, pp. 543–563, Aug. 2009.

[55] M. Zhang and P. Golland, "Statistical shape analysis: From landmarks to diffeomorphisms," *Medical Image Analysis*, vol. 33, pp. 155–158, Oct. 2016.

[56] K. D. Toennies, "Shape, Appearance and Spatial Relationships," in *Guide to Medical Image Analysis: Methods and Algorithms* (K. D. Toennies, ed.), Advances in Computer Vision and Pattern Recognition, pp. 405–472, London: Springer London, 2017.

[57] A. Brunton, A. Salazar, T. Bolkart, and S. Wuhrer, "Review of statistical shape spaces for 3D data with comparative analysis for human faces," *Computer Vision and Image Understanding*, vol. 128, pp. 1–17, Nov. 2014.

[58] N. Regonda and M. Janga Reddy, "Review of Medical Image Segmentation with Statistical Approach -State of the Art and Analysis," pp. 36–55, June 2017.

[59] M. S. Nosrati and G. Hamarneh, "Incorporating prior knowledge in medical image segmentation: a survey," *arXiv:1607.01092 [cs]*, July 2016. arXiv: 1607.01092.

[60] S. Zhang, Y. Zhan, Y. Zhou, and D. N. Metaxas, "Efficient sparse shape composition with its applications in biomedical image analysis: An overview," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 976–979, May 2012.

[61] E. Oost, S. Tomoshige, and A. Shimizu, "Condition Relaxation in Conditional Statistical Shape Models," in *Subspace Methods for Pattern Recognition in Intelligent Environment* (Y.-W. Chen and L. C. Jain, eds.), Studies in Computational Intelligence, pp. 33–56, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[62] T. F. Cootes and C. J. Taylor, "Anatomical statistical models and their role in feature extraction," *The British Journal of Radiology*, vol. 77, pp. S133–S139, Dec. 2004.

[63] M. Lüthi, T. Gerig, C. Jud, and T. Vetter, "Gaussian Process Morphable Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 1860–1873, Aug. 2018.

[64] M. Zhang and P. Thomas Fletcher, "Bayesian principal geodesic analysis for estimating intrinsic diffeomorphic image variability," *Medical Image Analysis*, vol. 25, pp. 37–44, Oct. 2015.

[65] E. Ruiz Pujadas, H. M. Kjer, G. Piella, and M. A. G. Ballester, "Statistical Shape Model with Random Walks for Inner Ear Segmentation," in *Spectral and Shape Analysis in Medical Imaging* (M. Reuter, C. Wachinger, and H. Lombaert, eds.), Lecture Notes in Computer Science, pp. 92–102, Springer International Publishing, 2016.

[66] M. Rumpf and B. Wirth, "A Nonlinear Elastic Shape Averaging Approach," *SIAM Journal on Imaging Sciences*, vol. 2, pp. 800–833, Jan. 2009.

[67] A. Y. Gao and B. S. Bouix, "Signed poisson map for shape analysis," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 389–392, Apr. 2014.

[68] Z. Xue, D. Shen, B. Karacali, and C. Davatzikos, "Statistical Representation and Simulation of High-Dimensional Deformations: Application to Synthesizing Brain Deformations," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005* (J. S. Duncan and G. Gerig, eds.), Lecture Notes in Computer Science, pp. 500–508, Springer Berlin Heidelberg, 2005.

[69] S. Elhabian and R. Whitaker, "ShapeOdds: Variational Bayesian Learning of Generative Shape Models," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2185–2196, July 2017.

[70] P. Manousopoulos, V. Drakopoulos, and T. Theoharis, "Fractal Active Shape Models," in *Computer Analysis of Images and Patterns* (W. G. Kropatsch, M. Kampel, and A. Hanbury, eds.), Lecture Notes in Computer Science, pp. 645–652, Springer Berlin Heidelberg, 2007.

[71] Sang-Mook Lee, A. L. Abbott, N. A. Clark, and P. A. Araman, "A Shape Representation for Planar Curves by Shape Signature Harmonic Embedding," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1940–1947, June 2006.

[72] W. Xu, H. Wang, F. Qi, and C. Lu, "Explicit Shape Encoding for Real-Time Instance Segmentation," *arXiv:1908.04067 [cs]*, Aug. 2019. arXiv: 1908.04067.

[73] M. K. Chung, K. J. Worsley, B. M. Nacewicz, K. M. Dalton, and R. J. Davidson, "General multivariate linear modeling of surface shapes using SurfStat," *NeuroImage*, vol. 53, pp. 491–505, Nov. 2010.

[74] J. H. Tan and U. Rajendra Acharya, "Active spline model: A shape based model—interactive segmentation," *Digital Signal Processing*, vol. 35, pp. 64–74, Dec. 2014.

[75] A. Bhalerao and R. Wilson, "Local Shape Modelling Using Warplets," in *Image Analysis* (H. Kalviainen, J. Parkkinen, and A. Kaarna, eds.), Lecture Notes in Computer Science, pp. 439–448, Springer Berlin Heidelberg, 2005.

[76] B.-W. Hong, S. Soatto, and L. A. Vese, "Enforcing local context into shape statistics," *Advances in Computational Mathematics*, vol. 31, pp. 185–213, Oct. 2009.

[77] G. McNeill and S. Vijayakumar, "Linear and nonlinear generative probabilistic class models for shape contours," pp. 617–624, Jan. 2007.

[78] M. d. Bruijne, B. v. Ginneken, M. A. Viergever, and W. J. Niessen, "Adapting Active Shape Models for 3D Segmentation of Tubular Structures in Medical Images," in *Information Processing in Medical Imaging*, pp. 136–147, Springer, Berlin, Heidelberg, July 2003.

[79] R. Blanc, M. M. Reyes, C. Seiler, and G. Székely, "Conditional Variability of Statistical Shape Models Based on Surrogate Variables," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12, no. Pt, pp. 84–91, 2009.

[80] K. Lekadir, R. Merrifield, and G. Yang, "Outlier Detection and Handling for Robust 3-D Active Shape Models Search," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 212–222, Feb. 2007.

[81] C. Zhang, J. Liang, J. Zhang, and H. Zhao, "A new shape prior model with rotation invariance," *Pattern Recognition Letters*, vol. 54, pp. 82–88, Mar. 2015.

[82] B. Xiang, "Knowledge-based image segmentation using sparse shape priors and high-order MRFs," 2013.

[83] M. Yamada, H. Hontani, and H. Matsuzoe, "A Study on Model Selection from the q-Exponential Distribution for Constructing an Organ Point Distribution Model," in *Image and Video Technology – PSIVT 2015 Workshops* (F. Huang and A. Sugimoto, eds.), Lecture Notes in Computer Science, pp. 258–269, Springer International Publishing, 2016.

[84] S. R. Arashloo, "Incorporating higher-order point distribution model priors into MRFs using convex quadratic programming," *Machine Vision and Applications*, vol. 27, pp. 821–832, Aug. 2016.

[85] J. J. Cerrolaza, M. Reyes, R. M. Summers, M. González-Ballester, and M. G. Linguraru, "Automatic multi-resolution shape modeling of multi-organ structures," *Medical Image Analysis*, vol. 25, pp. 11–21, Oct. 2015.

[86] J. J. Cerrolaza, A. Villanueva, F. M. Sukno, C. Butakoff, A. F. Frangi, and R. Cabeza, "Full Multiresolution Active Shape Models," *Journal of Mathematical Imaging and Vision*, vol. 44, pp. 463–479, Nov. 2012.

[87] S. Essafi, G. Langs, and N. Paragios, "Hierarchical 3D diffusion wavelet shape priors," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 1717–1724, Sept. 2009.

[88] C. Davatzikos, Xiaodong Tao, and Dinggang Shen, "Hierarchical active shape models, using the wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 414–423, Mar. 2003.

[89] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Hierarchical multi-resolution decomposition of statistical shape models," *Signal, Image and Video Processing*, vol. 9, pp. 1473–1490, Sept. 2015.

[90] J. Cerrolaza, S. Vera, A. Bagué, M. Ceresa, P. Migliorelli, M. G. Linguraru, and M. G. Ballester, "Hierarchical Shape Modeling of the Cochlea and Surrounding Risk Structures for Minimally Invasive Cochlear Implant Surgery," in *Clinical Image-Based Procedures. Translational Research in Medical Imaging* (M. G. Linguraru, C. Oyarzun Laura, R. Shekhar, S. Wesarg, M. González Ballester, K. Drechsler, Y. Sato, and M. Erdt, eds.), Lecture Notes in Computer Science, pp. 59–67, Springer International Publishing, 2014.

[91] C.-J. Du, J. G. Ferguson, P. T. Hawkins, L. R. Stephens, and T. Bretschneider, "Local Shape Representation in 3D: from Weighted Spherical Harmonics to Spherical Wavelet," in *BMVC*, 2012.

[92] L. YANG, *Medical image analysis using statistical shape model based on subdivision surface wavelet*. Thesis, June 2008.

[93] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum, "Multiscale 3-D Shape Representation and Segmentation Using Spherical Wavelets," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 598–618, Apr. 2007.

[94] A. Mohamed and C. Davatzikos, "Shape Representation via Best Orthogonal Basis Selection," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004* (C. Barillot, D. R. Haynor, and P. Hellier, eds.), Lecture Notes in Computer Science, pp. 225–233, Springer Berlin Heidelberg, 2004.

[95] B. Dong, Y. Mao, I. D. Dinov, Z. Tu, Y. Shi, Y. Wang, and A. W. Toga, "Wavelet-Based Representation of Biological Shapes," in *Advances in Visual Computing* (G. Bebis, R. Boyle, B. Parvin, D. Koracin, Y. Kuno, J. Wang, J.-X. Wang, J. Wang, R. Pajarola, P. Lindstrom, A. Hinkenjann, M. L. Encarnação, C. T. Silva, and D. Coming, eds.), Lecture Notes in Computer Science, pp. 955–964, Springer Berlin Heidelberg, 2009.

[96] Xiaoxiao Liu, Ja-Yeon Jeong, J. H. Levy, R. R. Saboo, E. L. Chaney, and S. M. Pizer, "A large-to-fine-scale shape prior for probabilistic segmentations using a deformable m-rep," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–8, June 2008.

[97] C. McIntosh and G. Hamarneh, "Medial-Based Deformable Models in Nonconvex Shape-Spaces for Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 33–50, Jan. 2012.

[98] P. T. Fletcher, Conglin Lu, S. M. Pizer, and Sarang Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 995–1005, Aug. 2004.

[99] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 851–865, Oct. 1999.

[100] C. Lu, S. M. Pizer, S. Joshi, and J.-Y. Jeong, "Statistical Multi-Object Shape Models," *International Journal of Computer Vision*, vol. 75, pp. 387–404, Dec. 2007.

[101] P. T. Fletcher, Conglin Lu, and S. Joshi, "Statistics of shape via principal geodesic analysis on Lie groups," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1, pp. I–I, June 2003.

[102] J.-Y. Jeong, S. M. Pizer, and S. Ray, "Statistics on Anatomic Objects Reflecting Inter-Object Relations," Oct. 2006.

[103] D. Merck, G. Tracton, R. Saboo, J. Levy, E. Chaney, S. Pizer, and S. Joshi, "Training models of anatomic shape variability," *Medical Physics*, vol. 35, no. 8, pp. 3584–3596, 2008.

[104] C. Seiler, X. Pennec, and M. Reyes, "Capturing the multiscale anatomical shape variability with polyaffine transformation trees," *Medical Image Analysis*, vol. 16, pp. 1371–1384, Oct. 2012.

[105] J. J. Cerrolaza, A. Villanueva, M. Reyes, R. Cabeza, M. A. González Ballester, and M. G. Linguraru, "Generalized Multiresolution Hierarchical Shape Models via Automatic Landmark Clusterization," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014* (P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, eds.), Lecture Notes in Computer Science, pp. 1–8, Springer International Publishing, 2014.

[106] D. Valdés-Amaro and A. Bhalerao, "Hierarchical Contour Shape Analysis," *Computación y Sistemas*, vol. 19, pp. 233–242, June 2015.

[107] A. Baumberg, "Hierarchical shape fitting using an iterated linear filter," *Image and Vision Computing*, vol. 16, pp. 329–335, Apr. 1998.

[108] M. Wilms, H. Handels, and J. Ehrhardt, "Multi-resolution multi-object statistical shape models based on the locality assumption," *Medical Image Analysis*, vol. 38, pp. 17–29, May 2017.

[109] F. Lecron, J. Boisvert, M. Benjelloun, H. Labelle, and S. Mahmoudi, "Multilevel statistical shape models: A new framework for modeling hierarchical structures," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1284–1287, May 2012.

[110] Y. H. Le, U. Kurkure, and I. A. Kakadiaris, "PDM-ENLOR: Learning Ensemble of Local PDM-Based Regressions," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1878–1885, June 2013.

[111] P. Du, H. S. I. Horace, J. Feng, and B. Hua, "Segmenting deformable soft-body meshes based on statistical variation information for piecewise Active Shape Model," in *2009 11th IEEE International Conference on Computer-Aided Design and Computer Graphics*, pp. 223–226, Aug. 2009.

[112] C. Seiler, "Trees on Geometrical Deformations to Model the Statistical Variability of Organs in Medical Images," 2012.

[113] G. Sanroma, "Modelling Shapes with Intermittently Present Landmarks,"

[114] P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. C. Di-mauro, "A non-linear generalisation of PDMs using polynomial regression," in *In Proc. BMVC, volume II*, pp. 397–406, BMVA Press, 1994.

[115] L. Le Folgoc, A. V. Nori, and A. Criminisi, "Spectral Kernels for Probabilistic Analysis and Clustering of Shapes," in *Information Processing in Medical Imaging* (M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, and D. Shen, eds.), Lecture Notes in Computer Science, pp. 67–79, Springer International Publishing, 2017.

[116] J. Ma, M. I. Miller, and L. Younes, "A Bayesian Generative Model for Surface Template Estimation," *International Journal of Biomedical Imaging*, vol. 2010, 2010.

[117] B. Gris, S. Durrleman, and A. Trouvé, "A sub-riemannian modular framework for diffeomorphism-based analysis of shape ensembles," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 802–833, 2018.

[118] C. von Tycowicz, F. Ambellan, A. Mukhopadhyay, and S. Zachow, "An efficient Riemannian statistical shape model using differential coordinates: With application to the classification of data from the Osteoarthritis Initiative," *Medical Image Analysis*, vol. 43, pp. 1–9, Jan. 2018.

[119] M. S. Hefny, "Analysis of Discrete Shapes Using Lie Groups," 2014.

[120] S. Sommer, A. Tatu, Chen Chen, D. R. Jurgensen, M. d. Bruijne, M. Loog, M. Nielsen, and F. Lauze, "Bicycle chain shape models," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 157–163, June 2009.

[121] A. Patel and W. A. P. Smith, "Exploring the Identity Manifold: Constrained Operations in Face Space," in *Computer Vision – ECCV 2010* (K. Daniilidis, P. Maragos, and N. Paragios, eds.), Lecture Notes in Computer Science, pp. 112–125, Springer Berlin Heidelberg, 2010.

[122] P. Yan, W. Zhang, B. Turkbey, P. L. Choyke, and X. Li, "Global structure constrained local shape prior estimation for medical image segmentation," *Computer Vision and Image Understanding*, vol. 117, pp. 1017–1026, Sept. 2013.

[123] J.-G. Kim, J. A. Noble, and J. M. Brady, "Probabilistic Models for Shapes as Continuous Curves," *Journal of Mathematical Imaging and Vision*, vol. 33, pp. 39–65, Jan. 2009.

[124] A. M. Quispe and C. Petitjean, "Shape prior based image segmentation using manifold learning," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 137–142, Nov. 2015.

[125] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," *Image and Vision Computing*, vol. 17, pp. 567–573, June 1999.

[126] A. Gooya, K. Lekadir, I. Castro-Mateos, J. M. Pozo, and A. F. Frangi, "Mixture of Probabilistic Principal Component Analyzers for Shapes from Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 891–904, Apr. 2018.

[127] H. Zhang, S. Zhang, K. Li, and D. N. Metaxas, "Robust shape prior modeling based on Gaussian-Bernoulli restricted Boltzmann Machine," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 270–273, Apr. 2014.

[128] S. M. Eslami, "Generative probabilistic models for object segmentation," 2014.

[129] F. Chen, H. Yu, R. Hu, and X. Zeng, "Deep Learning Shape Priors for Object Segmentation," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1870–1877, June 2013.

[130] T. Albrecht, M. Lüthi, T. Gerig, and T. Vetter, "Posterior shape models," *Medical Image Analysis*, vol. 17, pp. 959–973, Dec. 2013.

[131] K. Petersen, M. Nielsen, and S. S. Brandt, "Conditional Point Distribution Models," in *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging* (B. Menze, G. Langs, Z. Tu, and A. Criminisi, eds.), Lecture Notes in Computer Science, pp. 1–10, Springer Berlin Heidelberg, 2011.

[132] A. Kirillov, M. Gavrikov, E. Lobacheva, A. Osokin, and D. Vetrov, "Deep Part-Based Generative Shape Model with Latent Variables," in *27th British Machine Vision Conference (BMVC 2016)*, (York, United Kingdom), Sept. 2016.

[133] S. M. A. Eslami, N. Heess, C. K. I. Williams, and J. Winn, "The Shape Boltzmann Machine: A Strong Model of Object Shape," *International Journal of Computer Vision*, vol. 107, pp. 155–176, Apr. 2014.

[134] C. Nash and C. K. I. Williams, "The shape variational autoencoder: A deep generative model of part-segmented 3D objects," *Computer Graphics Forum*, vol. 36, no. 5, pp. 1–12, 2017.

[135] Y. Xu, *Statistical shape analysis for deep brain structures*. Ph.D., University of California, Los Angeles, United States – California, 2008.

[136] R. Bhalodia, S. Y. Elhabian, L. Kavan, and R. T. Whitaker, "DeepSSM: A Deep Learning Framework for Statistical Shape Modeling from Raw Images," in *Shape in Medical Imaging* (M. Reuter, C. Wachinger, H. Lombaert, B. Paniagua, M. Lüthi, and B. Egger, eds.), Lecture Notes in Computer Science, pp. 244–257, Springer International Publishing, 2018.

[137] M. Shakeri, H. Lombaert, S. Tripathi, and S. Kadoury, "Deep Spectral-Based Shape Features for Alzheimer's Disease Classification," in *Spectral and Shape Analysis in Medical Imaging* (M. Reuter, C. Wachinger, and H. Lombaert, eds.), Lecture Notes in Computer Science, pp. 15–24, Springer International Publishing, 2016.

[138] J. Duan, J. Schlemper, W. Bai, T. J. W. Dawes, G. Bello, C. Biffi, G. Doumou, A. De Marvao, D. P. O'Regan, and D. Rueckert, "Combining Deep Learning and Shape Priors for Bi-Ventricular Segmentation of Volumetric Cardiac Magnetic Resonance Images," in *Shape in Medical Imaging* (M. Reuter, C. Wachinger, H. Lombaert, B. Paniagua, M. Lüthi, and B. Egger, eds.), Lecture Notes in Computer Science, pp. 258–267, Springer International Publishing, 2018.

[139] K. D. Toennies and P. Benedix, "A weight-adaptive dynamic model for shape segmentation," in *IEEE International Conference on Image Processing 2005*, vol. 2, pp. II–814, Sept. 2005.

[140] A. Eguizabal and P. J. Schreier, "A weighting strategy for Active Shape Models," *arXiv:1707.09233 [cs]*, July 2017. arXiv: 1707.09233.

[141] C. Santiago, J. C. Nascimento, and J. S. Marques, "2D Segmentation Using a Robust Active Shape Model With the EM Algorithm," *IEEE Transactions on Image Processing*, vol. 24, pp. 2592–2601, Aug. 2015.

[142] G. Ruiz, E. Ramon, J. García, M. A. G. Ballester, and F. M. Sukno, "Weighted regularized ASM for face alignment," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2906–2910, Sept. 2016.

[143] M. Amberg, M. Lüthi, and T. Vetter, "Local Regression Based Statistical Model Fitting," in *Pattern Recognition* (M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, eds.), Lecture Notes in Computer Science, pp. 452–461, Springer Berlin Heidelberg, 2010.

[144] S. S. Chandra, J. A. Dowling, P. B. Greer, J. Martin, C. Wratten, P. Pichler, J. Fripp, and S. Crozier, "Fast automated segmentation of multiple objects via spatially weighted shape learning," *Physics in Medicine and Biology*, vol. 61, pp. 8070–8084, Oct. 2016.

[145] A. Gooya, C. Davatzikos, and A. Frangi, "A Bayesian Approach to Sparse Model Selection in Statistical Shape Models," *SIAM Journal on Imaging Sciences*, vol. 8, pp. 858–887, Jan. 2015.

[146] C. Shi, Y. Cheng, F. Liu, Y. Wang, J. Bai, and S. Tamura, "A hierarchical local region-based sparse shape composition for liver segmentation in CT scans," *Pattern Recognition*, vol. 50, pp. 88–106, Feb. 2016.

[147] B. Gutierrez, D. Mateus, E. Shiban, B. Meyer, J. Lehmberg, and N. Navab, "A sparse approach to build shape models with routine clinical data," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 258–261, Apr. 2014.

[148] S. G. H. Erbou, M. Vester-Christensen, R. Larsen, L. B. Christensen, and B. K. Ersbøll, "Comparison of sparse point distribution models," *Machine Vision and Applications*, vol. 21, pp. 999–1008, Oct. 2010.

[149] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Medical Image Analysis*, vol. 16, pp. 1385–1396, Oct. 2012.

[150] B. Wang, C. Fan, H. Xie, and L. Gu, "Improved Sparse Shape Composition Model for Multi-shape Prior," 2014.

[151] F. Bernard, P. Gemmar, F. Hertel, J. Goncalves, and J. Thunberg, "Linear Shape Deformation Models with Local Support Using Graph-Based Structured Matrix Factorisation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5629–5638, June 2016.

[152] Y. Wang, Q. Zheng, and P. A. Heng, "Online Robust Projective Dictionary Learning: Shape Modeling for MR-TRUS Registration," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1067–1078, Apr. 2018.

[153] K. Sjostrand, E. Rostrup, C. Ryberg, R. Larsen, C. Studholme, H. Baezner, J. Ferro, F. Fazekas, L. Pantoni, D. Inzitari, and G. Waldemar, "Sparse Decomposition and Modeling of Anatomical Shape Variation," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1625–1635, Dec. 2007.

[154] M. B. Stegmann, K. Sjöstrand, and R. Larsen, "Sparse modeling of landmark and texture variability using the orthomax criterion," in *Medical Imaging 2006: Image Processing*, vol. 6144, p. 61441G, International Society for Optics and Photonics, Mar. 2006.

[155] J. C. Nascimento and G. Carneiro, "Top-Down Segmentation of Non-rigid Visual Objects Using Derivative-Based Search on Sparse Manifolds," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1963–1970, June 2013.

[156] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou, "Towards robust and effective shape modeling: Sparse shape composition," *Medical Image Analysis*, vol. 16, pp. 265–277, Jan. 2012.

[157] J. Chen and L. G. Shapiro, "PCA vs. tensor-based dimension reduction methods: An empirical comparison on active shape models of organs," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5838–5841, Sept. 2009.

[158] J. Ma, *Self-learning shape recognition in medical images*. Thesis, Mar. 2019.

[159] B. Romaniuk and M. Desvignes, "Missing data estimation with statistical models," in *Advanced Concepts for Intelligent Vision Systems 2004*, Citeseer, 2004.

[160] C. Zhang, B. Heeren, M. Rumpf, and W. A. P. Smith, "Shell PCA: Statistical Shape Modelling in Shell Space," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1671–1679, Dec. 2015.

[161] D. Alcantara, O. Carmichael, E. Delson, W. Harcourt-Smith, K. Sterner, S. Frost, R. Dutton, P. Thompson, H. Aizenstein, O. Lopez, J. Becker, and N. Amenta, "Localized Components Analysis," in *Information Processing in Medical Imaging* (N. Karssemeijer and B. Lelieveldt, eds.), Lecture Notes in Computer Science, pp. 519–531, Springer Berlin Heidelberg, 2007.

[162] J. Ma, F. Lin, J. Honsdorf, K. Lentzen, S. Wesarg, and M. Erdt, "Weighted Robust PCA for Statistical Shape Modeling," in *Medical Imaging and Augmented Reality* (G. Zheng, H. Liao, P. Jannin, P. Cattin, and S.-L. Lee, eds.), Lecture Notes in Computer Science, pp. 343–353, Springer International Publishing, 2016.

[163] M. Abboud, A. Benzinou, and K. Nasreddine, "A robust tangent PCA via shape restoration for shape variability analysis," *Pattern Analysis and Applications*, May 2019.

[164] K. B. Hilger, R. Larsen, and M. C. Wrobel, "Growth modeling of human mandibles using non-Euclidean metrics," *Medical Image Analysis*, vol. 7, pp. 425–433, Dec. 2003.

[165] T. Liu, D. Shen, and C. Davatzikos, "Predictive modeling of anatomic structures using canonical correlation analysis," in *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pp. 1279–1282 Vol. 2, Apr. 2004.

[166] J. J. Cerrolaza, R. M. Summers, and M. G. Linguraru, "Soft Multi-organ Shape Models via Generalized PCA: A General Framework," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds.), Lecture Notes in Computer Science, pp. 219–228, Springer International Publishing, 2016.

[167] J. Koikkalainen, T. Tolli, K. Lauerma, K. Antila, E. Mattila, M. Lilja, and J. Lotjonen, "Methods of Artificial Enlargement of the Training Set for Statistical Shape Models," *IEEE Transactions on Medical Imaging*, vol. 27, pp. 1643–1654, Nov. 2008.

[168] S. Alimohamadi Gilakjan, J. Hasani Bidgoli, R. Aghaizadeh Zorofi, and A. Ahmadian, "Artificially enriching the training dataset of statistical shape models via constrained cage-based deformation," *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, pp. 573–584, June 2019.

[169] T. Tölli, J. Koikkalainen, K. Lauerma, and J. Lötjönen, "Artificially Enlarged Training Set in Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, pp. 75–82, Springer, Berlin, Heidelberg, Oct. 2006.

[170] T. Cootes and C. Taylor, "Combining point distribution models with shape models based on finite element analysis," *Image and Vision Computing*, vol. 13, pp. 403–409, June 1995.

[171] J. Ehrhardt, M. Wilms, and H. Handels, "Patch-Based Low-Rank Matrix Completion for Learning of Shape and Motion Models from Few Training Samples," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), Lecture Notes in Computer Science, pp. 712–727, Springer International Publishing, 2016.

[172] M. Pereañez, *Enlargement, subdivision and individualization of statistical shape models: Application to 3D medical image segmentation*. Ph.D. Thesis, Universitat Pompeu Fabra, Sept. 2017.

[173] Y. Zheng, "Cross-modality medical image detection and segmentation by transfer learning of shapel priors," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 424–427, Apr. 2015.

[174] C. Lindner, J. Thomson, T. a. Consortium, and T. F. Cootes, "Learning-Based Shape Model Matching: Training Accurate Models with Minimal Manual Input," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), Lecture Notes in Computer Science, pp. 580–587, Springer International Publishing, 2015.

[175] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 187–198, Apr. 1997.

[176] G. K. Rohde, A. Aldroubi, and B. M. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 1470–1479, Nov. 2003.

[177] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The Importance of Skip Connections in Biomedical Image Segmentation," in *Deep Learning and Data Labeling for Medical Applications* (G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and J. Cornebise, eds.), Lecture Notes in Computer Science, (Cham), pp. 179–187, Springer International Publishing, 2016.

[178] T. C. Lee, R. L. Kashyap, and C. N. Chu, "Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 56, pp. 462–478, Nov. 1994.

[179] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017. arXiv: 1412.6980.

[180] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Oct. 2017. ISSN: 2380-7504.

[181] J. H. Noble, F. M. Warren, R. F. Labadie, B. Dawant, and J. M. Fitzpatrick, "Determination of drill paths for percutaneous cochlear access accounting for target positioning error," vol. 6509, p. 650925, Mar. 2007. Conference Name: Medical Imaging 2007: Visualization and Image-Guided Procedures ADS Bibcode: 2007SPIE.6509E..25N.

[182] J. M. Fitzpatrick, P. E. Konrad, C. Nickele, E. Cetinkaya, and C. Kao, "Accuracy of customized miniature stereotactic platforms," *Stereotactic and Functional Neurosurgery*, vol. 83, no. 1, pp. 25–31, 2005.

[183] M. Caversaccio, W. Wimmer, J. Anso, G. Mantokoudis, N. Gerber, C. Rathgeb, D. Schneider, J. Hermann, F. Wagner, O. Scheidegger, M. Huth, L. Anschuetz, M. Kompis, T. Williamson, B. Bell, K. Gavaghan, and S. Weber, "Robotic middle ear access for cochlear implantation: First in man," *PLoS ONE*, vol. 14, p. e0220543, Aug. 2019.

[184] J. T. Holder, D. M. Kessler, J. H. Noble, R. H. Gifford, and R. F. Labadie, "Prevalence of Extracochlear Electrodes: Computerized Tomography Scans, Cochlear Implant Maps, and Operative Reports," *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 39, pp. e325–e331, June 2018.

[185] J. H. Noble, R. F. Labadie, R. H. Gifford, and B. M. Dawant, "Image-guidance enables new methods for customizing cochlear implant stimulation strategies," *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 21, pp. 820–829, Sept. 2013.

[186] B. P. O'Connell, J. B. Hunter, R. Gifford, A. Rivas, D. S. Haynes, J. H. Noble, and G. B. Wanna, "Electrode location and audiologic performance after cochlear implantation: a comparative study between Nucleus CI422 and CI512 electrode arrays," *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 37, pp. 1032–1035, Sept. 2016.

[187] B. P. O'Connell, A. Cakir, J. B. Hunter, D. O. Francis, J. H. Noble, R. F. Labadie, G. Zuniga, B. M. Dawant, A. Rivas, and G. B. Wanna, "Electrode Location and Angular Insertion Depth are Predictors of Audiologic Outcomes in Cochlear Implantation," *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 37, pp. 1016–1023, Sept. 2016.

[188] B. P. O'Connell, J. B. Hunter, D. S. Haynes, J. T. Holder, M. M. Dedmon, J. H. Noble, B. M. Dawant, and G. B. Wanna, "Insertion depth impacts speech perception and hearing preservation for lateral wall electrodes," *The Laryngoscope*, vol. 127, pp. 2352–2357, Oct. 2017.

[189] K. A. Berg, J. H. Noble, B. M. Dawant, R. T. Dwyer, R. F. Labadie, and R. H. Gifford, "Speech recognition with cochlear implants as a function of the number of channels: Effects of electrode placement," *The Journal of the Acoustical Society of America*, vol. 147, p. 3646, May 2020.

[190] J. H. Noble, B. M. Dawant, F. M. Warren, and R. F. Labadie, "Automatic Identification and 3-D Rendering of Temporal Bone Anatomy," *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 30, pp. 436–442, June 2009.

[191] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, p. 29, Aug. 2015.

[192] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945. Publisher: Ecological Society of America.

[193] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. Publisher: [International Biometric Society, Wiley].

[194] C. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilita'," 1936.

[195] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Adaptive computation and machine learning, Cambridge, Massachusetts: The MIT Press, 2016.

[196] X. Yang, J. Cerrolaza, C. Duan, Q. Zhao, J. Murnick, N. Safdar, R. Avery, and M. G. Linguraru, "Weighted Partitioned Active Shape Model for Optic Pathway Segmentation in MRI," in *Clinical Image-Based Procedures. Translational Research in Medical Imaging* (M. G. Linguraru, C. Oyarzun Laura, R. Shekhar, S. Wesarg, M. González Ballester, K. Drechsler, Y. Sato, and M. Erdt, eds.), Lecture Notes in Computer Science, (Cham), pp. 109–117, Springer International Publishing, 2014.

[197] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[198] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for medical image analysis and computer vision," in *Medical Imaging 2001: Image Processing*, vol. 4322, pp. 236–248, International Society for Optics and Photonics, July 2001.