PATIENT-SPECIFIC MODELING OF COCHLEAR IMPLANTS

By

Ziteng Liu

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

Computer Science

December 16, 2023

Nashville, Tennessee

Approved:

Jack H. Noble, Ph.D.

Benoit Dawant, Ph.D.

Ipek Oguz, Ph.D.

René H. Gifford, Ph.D.

Shunxing Bao, Ph.D.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Jack H. Noble, for his invaluable guidance, support, and mentorship throughout the entire journey over the years. I am grateful that he provided me the opportunity to work with him in the Biomedical Image Analysis for Image-Guided Interventions Laboratory (BAGL). His expertise, encouragement, and dedication have been instrumental in shaping the direction of this research.

I am also immensely grateful to my committee members, Dr. Benoit Dawant, Dr. Ipek Oguz, Dr. René H. Gifford, and Dr. Shunxing Bao, for serving on my dissertation committee. Their valuable insights, constructive feedback, and expertise are deeply appreciated.

I extend my heartfelt appreciation to my colleagues and fellow researchers, Ahmet Cakir, Rueben A. Banalagay, Jianing Wang, Dongqing Zhang, Yubo Fan, Srijata Chakravorti, Erin L. Bratu, and Ange Lou. They have provided assistance, collaboration, and meaningful discussions. Their inputs and perspectives have played an important role in shaping the ideas and concepts presented in this dissertation.

I am deeply thankful for my family and friends, Ziran Min, Coco (my dog), Yujie Chi, Yi Chi, Zhongwei Teng, Yike Zhang, Yanmin Ji, Tianyu Wang, Ju Song, and many others. They have stood by me, offering their support and encouragement throughout the highs and lows of this research endeavor. Their presence and words of encouragement have been a source of inspiration, and I am grateful for their unwavering belief in me.

Finally, I would like to ackonwledge the National Institute for Deafness and Other Communication Disorders for their grant R01DC014037.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Overview

Hearing is the outcome of a series of complex steps that translate sound wave signals into electrical signals. In normal hearing, sound waves induce pressure oscillations in the cochlear fluids, which in turn initiate a traveling wave of displacement along the basilar membrane (BM) from the base to the apex. This membrane divides the cochlea along its length and produces a maximal response to sounds at different frequencies. The range of frequencies is distributed along the cochlear duct from base to apex. The base of the cochlea is narrow and stiff, so higher frequencies are transduced. The apical end or apex is wide and flexible, so lower frequencies are transduced Santi and Tsuprun (1988). Because the motion of BM is then sensed by hair cells that are attached to the BM, these sensory cells are tonotopically mapped, e.g. they are fine-tuned to respond to different characteristic frequencies of the received sounds. The activation of the corresponding hair cells releases chemical transmitters to electrically stimulate the spiral ganglion nerve cells Raphael and Altschuler (2003); Rask-Andersen et al. (2012). And the electrical signal is propagated along the auditory nerve fibers (ANFs), traveling through the brain stem, and finally reaching the auditory cortex allowing the brain to sense and process the sounds. In summary, the sound signal is decomposed by BM and hair cells, and the ANFs send this information to the brain to hear the sound. For patients suffering sensorineural hearing loss, which is principally caused by damage or destruction of the hair cells, however, the decomposition process of incoming sound waves cannot be performed. In this situation, direct stimulation of the ANFs is possible if they are intact and this can be done with a neural prosthesis called a cochlear implant (CIs) Wilson and Dorman (2008); Greenberg et al. (2004).

With over 500,000 recipients worldwide, CIs are considered the standard-of-care treatment for profound sensory-based hearing loss. A CI replaces the hair cells with an externally worn signal processor that decomposes the incoming sound into signals sent to an electrode array that is surgically implanted into the cochlea (see Figure 1.1). Electrode arrays have up to 22 contacts depending on the manufacturer, dividing the available ANFs to, at most, 22 frequency bands or stimulation areas when using monopolar stimulation. Because there are around 30,000 auditory nerve fibers and each fiber serves as an independent neural channel corresponding to a certain characteristic frequency in natural hearing Spoendlin and Schrott (1989), the electrodes on a CI array stimulate nerves corresponding to a wide range of frequencies and have a limited spectral resolution. Thus, after surgery, CI recipients need to undergo many programming sessions with an

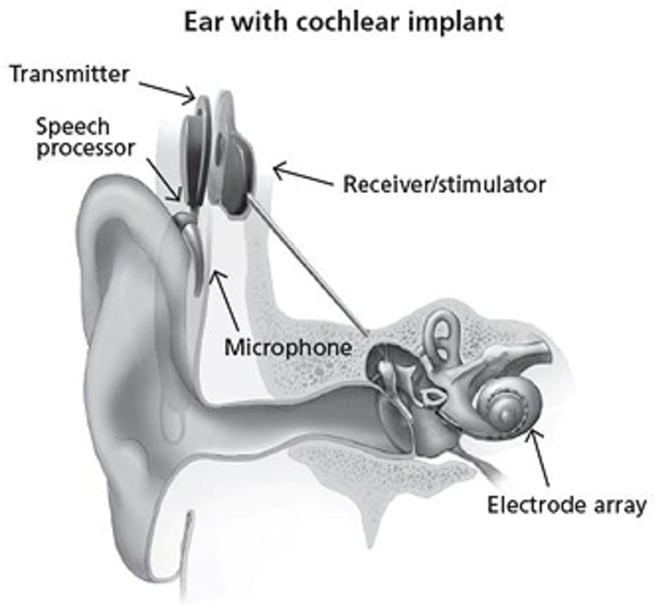


Figure 1.1: Example of a cochlear implant's external and internal parts. Image retrieved from NIDCD (2019)

audiologist who adjusts the settings for every single electrode to improve overall hearing performance, resulting in a so-called MAP. Mapping involves determining patient-specific settings for the CI processor including stimulation levels assigned to each active electrode, sound frequency bands assigned to each active electrode, which electrodes will be activated or deactivated, etc.

The vast majority of CI recipients can benefit from the implant with a satisfactory average postoperative word and sentence recognition rate Buss et al. (2008); Dorman et al. (2011); Gifford et al. (2008, 2014); Litovsky et al. (2006). Yet a significant number of CI recipients experience poor outcomes and restoration to normal fidelity is rare even among the best performers Moberly et al. (2016); Pisoni et al. (2017); Drennan and Rubinstein (2008); Lenarz et al. (2012). This is partially due to a trial-and-error process used in the programming sessions. Since the patients need weeks of experience with given settings before hearing performance stabilizes due to learning effects, this procedure can be frustratingly long. And because it relies heavily on subjective feedback from patients, the programming sessions generally lead to sub-optimal outcomes. Besides that, two other factors also have a high impact on the effectiveness of the CI: the channel overlapping problem among CI electrodes and the variable health status of patients' ANFs.

The focus of this doctoral research is on the development of patient-specific models of CI recipients to provide unique and objective information to audiologists in the programming session. The key areas in which we seek to improve the existing literature are the estimation of the inner ear electrical potential, learning from patients' clinical measurements to predict patient-specific electro-anatomical parameters, prediction of the

auditory nerve health, auditory nerve fiber segmentation, and a super-resolution method on conventional CT images to generate high-resolution tissue electrical classification maps.

1.2 Physiological CI Measurements

Several physiological measurements are available nowadays via CI devices in the clinical setting. These measurements are used to verify the device and electrode function. Some of them also allow us to better understand patient-specific information such as tissue resistivity, voltage distribution, and auditory nerve health after the insertion of the electrode carrier. Because the data are measured using intracochlear electrodes, patients do not need to keep still or be sedated during the measurements.

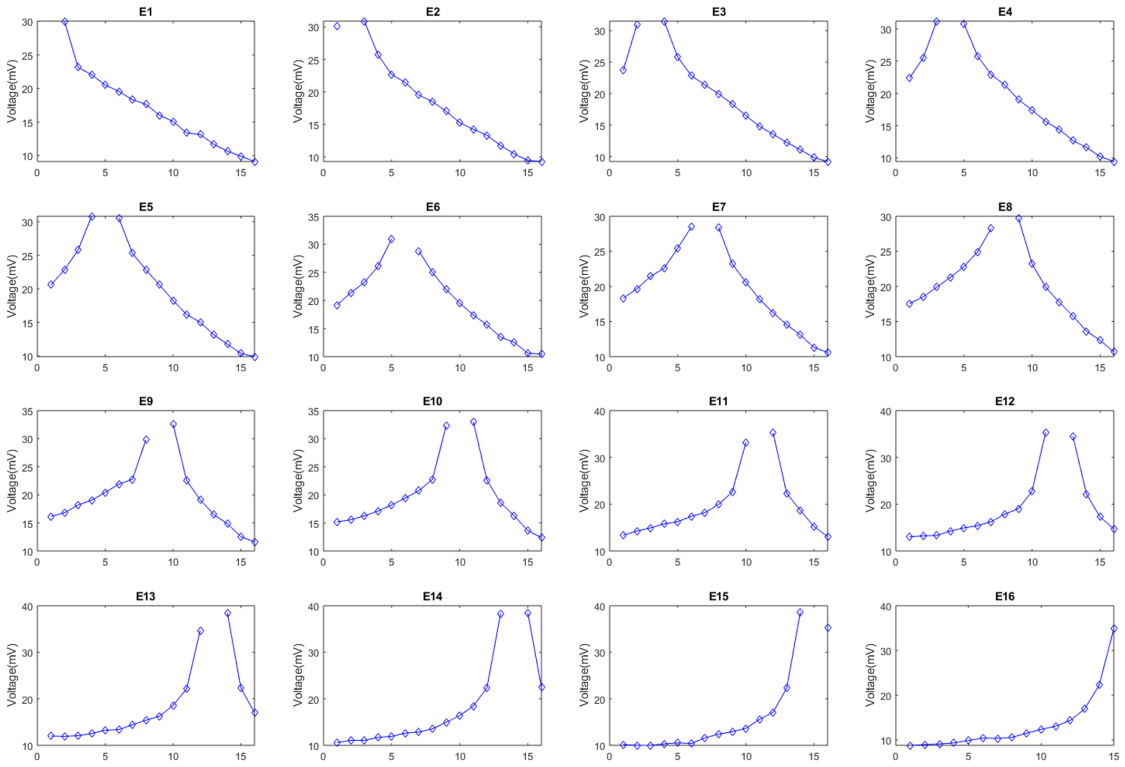


Figure 1.2: Electrical field imaging of a 16 electrode CI array. Each subplot corresponds to a different probe electrode and the x-axis indicates the recording electrode.

One of the widely used measurements is called electrical field imaging (EFI). It is obtained by activating one CI electrode at a time (probe electrode) while measuring the voltage at each of the remaining electrodes in the cochlea. All the probe electrodes will use the same current level and pulse width to initiate a stimulus. Although referred to as "electrical field imaging" by the manufacturer of the Advanced Bionics device, it obviously has the limitation that only the voltage gradient at the location of the CI electrode is known.

Basically, EFI measures have shown a voltage gradient that is about three times larger toward the base (with apical stimulation) than toward the apex (with basal stimulation) Mens (2007). Figure 1.2 shows an example of EFI which captures the electrical potential distribution of a CI array that has 16 electrodes where each subplot corresponds to a different probe electrode.

Electrically evoked compound action potential (ECAP) is another frequently used physiological measurement in the clinical setting. It is a synchronous physiological response from an aggregate population of auditory nerve fibers in response to electrical stimulation characterized by a negative deflection followed by a positive peak Hughes (2012); Bourien et al. (2014). The amplitude of an ECAP is measured by the difference between the negative and positive peaks. Because the ECAP is an early-latency response, artifacts created by the stimulating CI electrode need to be separated from the physiological potential Briaire and Frijns (2005). To isolate the neural response from the stimulus artifact, there are several methods to measure ECAPs using CI. The first of which is the alternating polarity method, where ECAP is measured as the average response from the auditory nerves when both cathodic-leading and anodic-leading bi-phasic current pulses are used (Figure 3a). Another method is the forward-masking subtraction (FMS) method. In FMS, two distinct pulses are used, namely masker and probe pulses, which are injected into the cochlea with a masker-probe interval (MPI). When the interval is short enough, the ANFs that are activated by the masker will be in a refractory state when the probe pulse is injected. Thus, if the masker and probe stimulate the same group of ANFs, the addition of the recorded responses from the individual masker and probe pulses subtracting the recorded responses for when masker and probe are used together with a sufficiently short MPI will result in the ECAP signature of probe (Figure 1.3(b)). And by keeping the probe electrode constant and varying the masker electrode, the overlapping stimulation area can be estimated by measuring the ECAP magnitude extracted using FMS.

Two frequently used ECAP-based functions are the amplitude growth function (AGF) and the spread of excitation (SOE) function. AGF samples how the amplitude of recorded ECAPs (μV) grows as the injecting current increases for the stimulation pulse signal. And SOE measures the overlapping fraction of ECAP responses for two stimulating electrodes that are generated from the same ANFs Hughes (2012). The AGF measurements were mainly collected using the alternating polarity method while the SOE data is commonly obtained using the FMS method. AGF functions should increase as the current is increased, but the gradient implies the electrical characteristics of the inner ear. SOE function is expected to increase as the spatial distance between the masker and the probe electrode decrease and vice versa because the closer those two electrodes are, the bigger simulating area they will share.

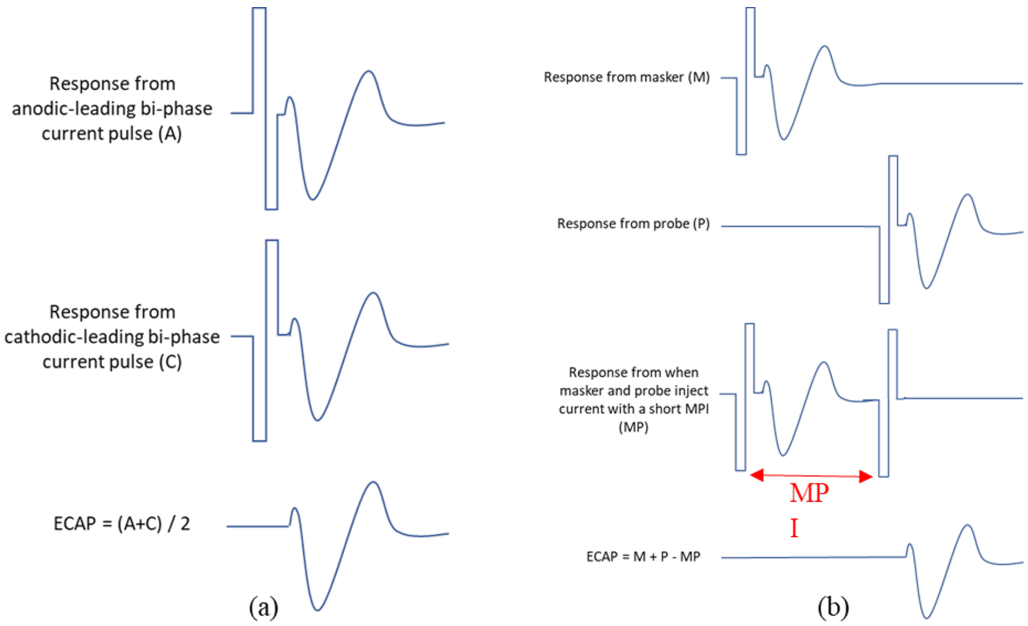


Figure 1.3: Explanation of the artifact reduction methods we used to collect clinical ECAPs via CI and simulate ECAPs via our ANF model. With (a) alternation polarity and (b) forward masking subtraction.

1.3 Image-guided CI Programming Techniques

CIs induce hearing sensation by stimulating auditory nerve pathways within the cochlea using the CI electrode array, which is designed to stimulate nerve pathways corresponding to a predefined frequency bandwidth. Because the CI array is inserted blindly during the implantation, the placement of CI arrays is usually not optimal and the distance between the electrodes and ANFs varies. Electrodes that are distant to the stimulation sites of healthy nerves usually share a broader stimulation area, e.g., recruit a similar group of ANFs with its neighboring electrodes. The nerves in this overlapped stimulation area may fail to recover from their refractory period in the consecutive electrical stimulus and negatively affect CI performance Boëx et al. (2003); Fu and Nogaki (2005). Although audiologists can provide a deactivation plan for electrodes, only suboptimal outcomes are achievable in most cases due to lacking objective information about the neural stimulation patterns of electrodes. To solve this problem, our lab has developed the first image-guided CI programming technique (IGCIP) Noble et al. (2013). IGCIP aims at providing unique and objective information that audiologists can use to define patient-specific CI processor settings. More specifically, we are able to obtain the spatial relationship between the electrodes on the CI array and the spiral ganglion nerves in IGCIP via accurate CT-based cochlear anatomy segmentation and CI electrode localization techniques. IGCIP also proposed to use the electrode distance-versus-frequency (DVF) curves to visualize the programming-relevant spatial relationship, which described the Euclidian distance from each electrode to the spiral ganglion sites

characterized by the corresponding frequencies. By picking a subset of electrodes that consists of as many electrodes as possible and minimizes competition in their peak activation region as a CI activation plan, IGCIP leads to significant improvement in speech recognition, spectral resolution, and subjective hearing quality Noble et al. (2015, 2016). Because the activation/deactivation plan needs to be manually selected, methods that can automatically generate the activation electrode set Zhao et al. (2016); Zhang et al. (2018) were proposed to improve the original approach.

1.4 Electro-anatomical Model of the Inner Ear

The success of IGCIP shows the potential of customized CI electrode programming using image-based technologies. However, IGCIP simulates the nerves' activation pattern using only the spatial information between electrodes and SG nerve sites while the electrical characteristic also varies from patient to patient Malherbe et al. (2015). It is possible that the method could be improved with a more comprehensive estimation of the interaction between the CI electrode and auditory nerves. So, as an extending work to IGCIP, Cakir et al. Cakir et al. (2017b,a) introduced a patient-specific high-resolution electro-anatomical model (EAM) which allows us to customize patient-specific tissue resistivities and estimate intra-cochlear electric potential (EP) created by the CI for individual patients. Compared to other groups who also investigate the voltage distribution and neural activation within the cochlea taking advantage of three-dimensional EAMs Whiten (2007); Kalkman et al. (2015); Goldwyn et al. (2010); Hanekom (2001), the method proposed by Cakir et al. Cakir et al. (2017b) incorporates patient-specific differences and can be applied in vivo. The EAM proposed by Cakir et al. utilizes an active shape model to register nine μCT images with tissue labels with patients' CT images and create high-resolution segmentations for individual patients. And by assigning resistivity values to the bone, neural tissue, soft tissue, and electrolytic fluid in patients' label maps, a system of linear equations can be created by solving Poisson's Equation for electro-statistics using the finite difference method (FDM). The resulting electric potential map can then be used to simulate nerve activation patterns, which provide a physics-based estimation compared to the coarse distance-based method used in IGCIP. However, in order to generate a reasonable estimation of EP in the inner ear for different patients, we first need to find the patient-specific resistivity set for those four tissue types. Cakir et al. Cakir et al. (2017a) proposed to use a grid search and compare the resulting EPs with the clinical measurement of EFI. And the resistivity values that minimize the mean squared error between simulated EFI and clinical EFI are chosen as the patient-specific electrical parameters.

1.5 Auditory Nerve Model

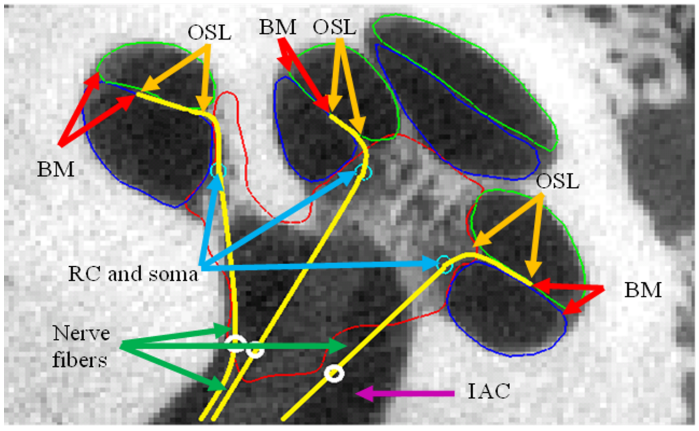
EAMs allow a physics-based estimation of the electrical potential within a given anatomical structure, and auditory nerve models permit the estimation of neural stimulation patterns due to the electric potential produced by CI electrodes. By directly simulating the current spread and neural activation along the auditory nerve fibers, auditory nerve models lead to a more accurate estimation of neural stimulation patterns than considering only the voltage distribution at SG sites. We need to investigate two critical questions before coupling an auditory nerve model to the EAM: (1) Where are the nerves? (2) How does an auditory nerve react to a stimulus? To solve these two questions, we need a nerve segmentation method and biological model.

1.5.1 Auditory Nerve Segmentation

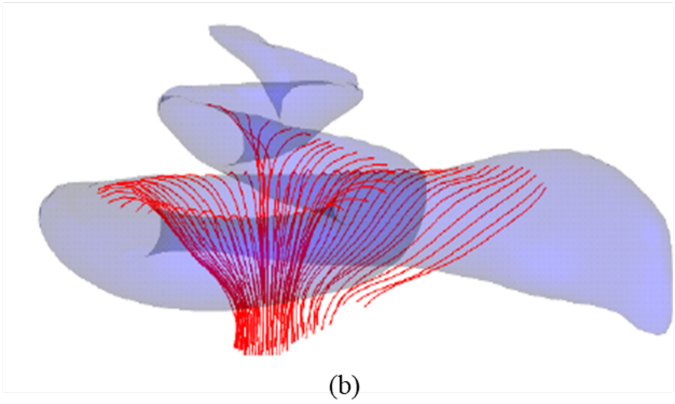
Auditory nerve fibers are invisible in traditional CT or μCT images due to lack of resolution. The spatial resolution of CTs is commonly 0.3 to 0.6 mm and the resolution of μCT we used to build the EAMs is usually 0.03mm, while auditory nerve fibers are approximately 2μm in width. Therefore, our lab proposed a novel auditory nerve fiber segmentation method that relies on prior knowledge of the morphology of the fibers Cakir et al. (2019). This approach treats the fiber localization problem as a path-finding problem, where the aim is to find a path connecting the unmyelinated terminal to the soma to the internal auditory canal (IAC) endpoint with a shape that matches the expected shape of the fiber. Several landmarks are provided, including the starting points between the basilar membrane (BM) and the osseous spiral lamina (OSL) where the unmyelinated terminals are located, the Rosenthal's Canal (RC) where somas are located, and the IAC endpoints where the central axons terminated (see Figure 1.4a). Paths representing 75 fiber bundles that are evenly spaced along RC was found using Dijkstra's algorithm DIJKSTRA (1959) which gives the shortest path connecting all the landmarks and splines were fit to them in order to smooth the shape (see Figure 1.4b). Because the paths are computed independently and in close proximity, sometimes they overlap or cross. As a post-processing step, manual edits to some of the paths are required.

1.5.2 Biological Model

We need a biological model for the auditory nerve fibers to describe how electrical current spreads along the auditory nerve when a stimulus is injected by some CI electrodes. In a 2000 study, Rattay et al. Rattay et al. (2001b) developed a compartmental auditory nerve fiber model using a modified Hodgkin-Huxley (HH) formulation and used this model to study the influence of the effect of different nerve subunits on the excitation of the nerve fibers Litovsky et al. (2006). This model introduced three major features that differ from other nerve models.



(a)



(b)

Figure 1.4: An overview of ANF bundle segmentation: (a) shows the landmarks used during automatic path finding to localize ANF bundles. (b) shows the segmentation results after being fitted to a spline and manual adjustment.

First, they use a compartment model which consists of several subunits with individual geometric and electric parameters. As shown in Figure 1.5(a), a nerve fiber consists of peripheral nodes and internodes, somatic, pre-somatic, and post-somatic regions, and central nodes and internodes. Each of these subunits can be thought of as a compartment that is modeled by an electrical circuit with distinctive electrical properties. The detailed geometry parameters are shown in Figure 1.5 and electrical parameters are listed in Table 1.1.

Second, Ion channel dynamics are described by a modified Hodgkin-Huxley (HH) model, namely, the 'warmed' HH (wHH) model. wHH includes sodium, potassium, and leakage currents and has the following form:

$$\frac{dV}{dt} = \left[-g_{Na}m^3h\left(V-V_{Na}\right) - g_K n^4\left(V-V_K\right) - g_L\left(V-V_L\right) + i_{stimulus}\right]/c \tag{1.1}$$

$$\frac{dm}{dt} = \left[-\left(\alpha_m+\beta_m\right)m+\alpha_m\right]k \tag{1.2}$$

$$\frac{dh}{dt} = \left[-\left(\alpha_h+\beta_h\right)h+\alpha_h\right]k \tag{1.3}$$

$$\frac{dn}{dt} = \left[-\left(\alpha_n+\beta_n\right)n+\alpha_n\right]k \tag{1.4}$$

$$k = 3^{T-6.3} \tag{1.5}$$

$$V = V_i - V_e - V_{rest} \tag{1.6}$$

where V, V_i, V_e, and V_{rest} are the membrane, internal, external, and resting voltages, and V_{Na}, V_K, and V_L are the sodium, potassium, and leakage battery voltages, respectively. g_{Na}, g_K, g_Lare the maximum conductance and m, h, n are probabilities with which the maximum conductance is reduced with respect to measured gating data, for sodium, potassium, and leakage, respectively. $i_{stimulus}$ is the current produced by electrode stimulation and c is the membrane capacity. Finally, α and β are voltage-dependent variables that were fitted from measured data, and k is the temperature coefficient and T is the temperature degree in Celsius. With wHH, the gating processes are accelerated (m, h, n are multiplied by 12) which best fit to observed temporal behavior of human auditory nerves compared to the original HH model, and leakage conductance are multiplied by the factor 10 to simulate 10-fold channel density. Third, the influence of membrane noise is also taken into account in their approach by adding ion channel current fluctuations in the active compartments (unmyelinated terminal, node of Ranvier, pre- and post-somatic compartment, soma). These features allow the model to simulate the electrically excited auditory nerves in the human cochlea more accurately than models based on animals.

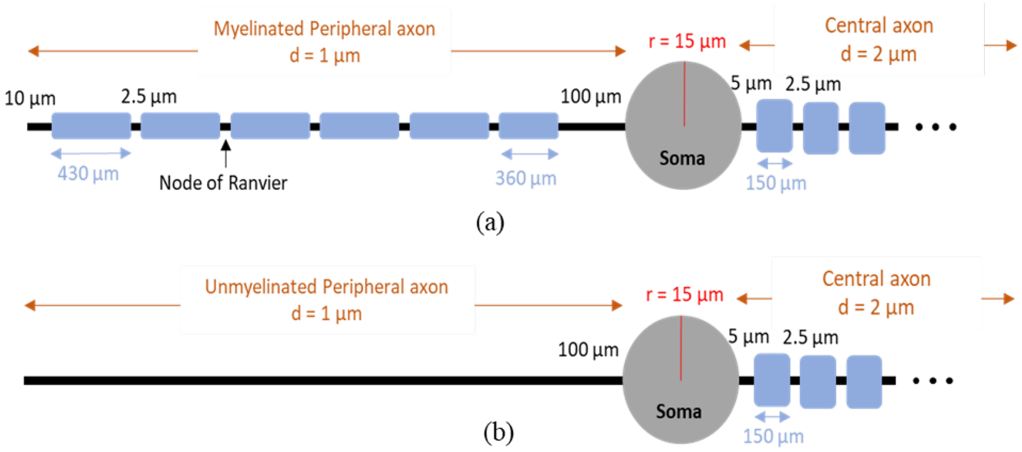


Figure 1.5: ANF biological models. (a) A healthy ANF has a myelinated peripheral axon. Ion channels and leakage gates are located in the nodes of Ranvier. (b) An unhealthy ANF whose peripheral axon has become unmyelinated but still partially functional. Only passive channels remain on the unmyelinated peripheral axon.

Table 1.1: Electrical properties of our biological human ANF model

Electrical properties	Topology	Values
Resistivity (Ω cm)	Intracellular	50
	Peripheral terminal	1
	Presomatic region	1
	Soma	1/3
Capacitance (µF cm^{-2})	Postsomatic region	1
	Peripheral internode	1/40
	Central internode	1/80
	Node of Ranvier	1
	Internode	1
Conductance (mS cm^{-2})	Soma	1
	Other	10

Since the CI electrodes do not directly contact any fraction of auditory nerve fibers, the neural activations are caused by the rapid change of external electrical potential. We implemented our biological model based on Rattay's model Rattay et al. (2001b) using the Neuron library Carnevale and Hines (2006) which supports the simulation of external stimulus of customized nerve models. Thus, we are able to simulate the neural activation patterns using the resting potential and stimulation potential map from EAMs.

1.6 Deep learning techniques

Deep learning has emerged as a transformative technology in the field of medical imaging, revolutionizing the way healthcare professionals analyze and interpret complex medical images. With its ability to automatically learn hierarchical representations from large datasets, deep learning has shown remarkable potential in improving accuracy, efficiency, and diagnostic outcomes in medical imaging tasks.

Medical imaging plays a critical role in diagnosing and monitoring various diseases and conditions, including cancer, cardiovascular diseases, neurological disorders, and musculoskeletal conditions. Traditionally, the interpretation of medical images has relied on the expertise of radiologists and clinicians, who manually analyze the images to detect abnormalities, classify diseases, and make treatment decisions. However, this process is time-consuming, subjective, and can be prone to human error.

Deep learning-based approaches have overcome many of the limitations associated with traditional methods by leveraging the power of neural networks to automatically learn relevant features and patterns directly from medical images. These algorithms can extract intricate and subtle details from images, enabling more accurate and consistent analysis. By training on large datasets comprising annotated medical images, deep learning models can learn to identify complex patterns and abnormalities that may be imperceptible to the human eye.

Convolutional neural networks (CNNs) have emerged as the backbone of deep learning for medical imaging. CNNs excel at learning spatial hierarchies of features, making them well-suited for tasks such as image classification, segmentation, and detection. With multiple convolutional layers, these networks can capture increasingly complex and abstract representations of medical images, enabling them to identify subtle disease markers and abnormalities.

One of the foundational architectures in deep learning for medical imaging is the U-Net Ronneberger et al. (2015). U-Net is a CNN architecture that has demonstrated remarkable success in medical image segmentation tasks. It comprises a contracting path, which captures contextual information through successive convolutional and pooling layers, and an expansive path, which utilizes upsampling and convolutional layers to generate high-resolution segmentation maps. U-Net has been widely used for various segmentation tasks, such as delineating organs, tumors, and lesions in medical images.

Another influential architecture in deep learning is the generative adversarial network (GAN) Goodfellow et al. (2014). GANs consist of a generator and a discriminator network that work in tandem. The generator network learns to synthesize realistic medical images, while the discriminator network aims to distinguish between real and synthetic images. Through an adversarial training process, GANs can generate highly realistic and diverse medical images that closely resemble real patient data. GANs have been applied in medical image synthesis, reconstruction, and augmentation tasks, offering valuable opportunities for data augmentation, rare pathology simulation, and training set expansion.

Deep learning has demonstrated remarkable potential in numerous medical imaging applications. Computer-aided detection (CAD) systems empowered by deep learning algorithms can assist radiologists in the detection and localization of abnormalities in mammograms, chest X-rays, and computed tomography (CT) scans. Deep learning-based image segmentation methods have achieved exceptional accuracy in delineating structures and organs of interest, facilitating treatment planning, quantitative analysis, and surgical guidance.

Furthermore, deep learning has also enabled advancements in medical image synthesis and reconstruction. By training generative models, such as variational autoencoders (VAEs) **?** and GANs Creswell et al. (2018), on large datasets of medical images, it is possible to generate synthetic images that closely resemble real patient data. These synthetic images can be used to augment limited datasets, generate diverse examples for training, and simulate rare pathological conditions for educational purposes.

In conclusion, deep learning has revolutionized medical imaging by harnessing the power of neural networks to automate and enhance the analysis of complex medical images. With architectures like U-Net and GANs, deep learning has made significant strides in tasks such as segmentation, detection, reconstruction, and synthesis. As the technology continues to advance and more data becomes available, deep learning is expected to play an increasingly pivotal role in shaping the future of medical imaging, ultimately benefiting patients and healthcare professionals alike.

1.7 Challenges

The EAMs of cochlear implants show great potential for improving the hearing outcomes of CI users, however, applying the EAMs in clinical settings is difficult. On the one hand, the effectiveness of the methods derived from EAMs is hard to be verified with limited data acquired during or after the CI surgery. Some physiological characteristics, such as the number of ANFs or the health of ANFs, are impossible to be measured directly. Therefore we need to design novel evaluation metrics to evaluate our proof-of-concept studies. On the other hand, our current implementation of EAMs has several drawbacks that may affect its performance. Our research calls for an effective solution to applying EAMs in clinical settings and improving the effectiveness of EAMs. In this context, we have identified a number of challenges that we address in this

doctoral research:

Challenge 1. There are around 30,000 auditory nerve fibers and each fiber serves as an independent neural channel corresponding to a certain characteristic frequency in natural hearing. However, CI arrays have a very limited spectral resolution depending on the number of CI electrode and their interface with the nerves. The methods we introduced in previous sections all aimed at producing a better deactivation plan that minimizes the channel overlapping problem to improve CI performance Bierer and Litvak (2016). But the drawback is that this reduces the already very limited number of spectral channels, further compressing the frequency spectrum.

Challenge 2. Our current model assumes that the auditory nerve fibers along the length of the cochlea are equally healthy, while the health status of auditory nerve fibers varies substantially across individuals. And auditory nerve fibers health is one of the key factors affecting hearing outcomes Nadol Jr et al. (1989); Lousteau (1987); Khan et al. (2005). Although the EAM is customized electrically and anatomically, it has not taken the differences in nerve health into account yet. One remaining challenge is coupling the biological nerve model with the current implementation of EAM and customizing the nerve model to reflect nerve fibers' health status. And since neural health is impossible to be directly tested, we also need to develop a novel method to evaluate the auditory nerve health estimation.

Challenge 3. The current EAM of cochlear implants is customized for individual patients via patient CT-based segmentation and patient-specific electrical parameters. However, the process of searching for the optimal electrical parameters is very computationally expensive. First, the EP map is calculated by solving Poisson's Equation for electro-statistics using FDM. This is achieved by applying the bi-conjugate gradient optimizer to the high-dimensional linear equations, which require substantial computational resources. Second, the current approach utilizes a grid search to find the optimal resistivity values for each type of tissue. An EP map is calculated for each CI electrode at each iteration and compared with clinical EFI. Thus, it takes days of computation time and may still result in sub-optimal parameters if the grid was not fine enough.

Challenge 4. The high-resolution patient segmentation relies on a thin plate spline transformation that registers nine μCTs to the patient CT image using meshes of scala tympani, scala vestibuli, and modiolus. Although these cochlea substructures can be accurately registered because of the one-to-one point correspondence, the soft tissue, bone, and air in the region of interest are labeled using a majority voting based on the transformed μCTs' segmentations. This results in a less accurate segmentation surrounding the cochlea because of the non-rigid transformation and the anatomical differences between ex-vivo μCTs and in-vivo patient CTs. Existing methods have not well explored how to handle this problem.

Challenge 5. The existing auditory nerve segmentation method uses path-finding algorithms to connect automatically estimated landmarks. However, manual correction is needed as a post-processing step in most

cases because: (1) the paths are computed independently and sometimes they overlap or cross, and (2) the whole path is defined by 3 landmarks and the mesh of scala tympani, so the nerve fibers are not aware of the boundary of IAC. Moreover, the peripheral part of nerve fibers is assumed to extend along the direction that is parallel to the central axis of the modiolus, while the natural never fibers are supposed to spiral along the length of IAC.

1.8 Research contributions

In this thesis, we explore the methods that may improve the current patient-specific EAM and CI performance based on the challenges described in Section 1.6. First, we propose a CI electrode sequence optimization algorithm that may improve the CI performance without losing any spectral resolution (Contribution 1, Challenge 1). Second, we introduce a novel method for auditory nerve health estimation where the nerve health is parameterized for each patient and can be trained and evaluated using available ECAP functions (Contribution 2, Challenge 2). Third, we develop machine learning based methods to generate the patient-specific EP map and electrical parameters within seconds, which usually takes days to calculate using the traditional FDM + grid search method (Contribution 3 and 4, Challenge 3). Fourth, we introduce a super-resolution segmentation method that can generate 8X higher-resolution tissue label maps of the inner ear using patient CT images without the help of μCTs (Contribution 5, Challenge 4). At last, We also proposed a novel fully-automatic auditory nerve segmentation method to generate accurate auditory nerve fiber trajectories (Contribution 6, Challenge 5).

Contribution 1. Cochlear implant electrode sequence optimization using patient-specific neural stimulation models

Due to the refractory behavior of the nerve fibers, nerves become insensitive to further stimulation for a period of time following prior stimulation. As a result, when multiple electrodes stimulate the same fibers, masking artifacts can occur, where one electrode may fail to stimulate certain nerve bundles due to preceding stimulation by another electrode. This phenomenon is also known as the channel overlapping problem. To solve this problem, existing methods usually come up with a deactivation plan to minimize this artifact. But the drawback is that this reduces the already very limited number of spectral channels, further compressing the frequency spectrum.

In this research, we propose a new method to reduce channel overlapping problems by determining a customized firing order of the electrodes on the CI array using image-based models of patient-specific neural stimulation patterns. Our models permit estimating the time delay needed after firing an electrode so that the nerve fibers they stimulate can recover from the refractory period. These predictions allow us to design an optimization algorithm that determines a customized electrode firing order that minimizes the negative

effects of overlapping stimulation between electrodes. The customized order reduces how often nerves that are in a refractory state from previous stimulation by one electrode are targeted for activation by a subsequent electrode in the sequence. Our experiments show that this method is able to reduce the theoretical stimulation overlap artifacts and could lead to improved hearing outcomes for CI recipients.

Contribution 2. Auditory nerve fiber health estimation using patient-specific cochlear implant stimulation models

Cochlear implants restore hearing using an array of electrodes implanted in the cochlea to directly stimulate auditory nerve fibers. The existing inner ear models for cochlear implants like IGCIP and EAM assume that all the ANFs along the length of the cochlea are equally healthy, while the health status of ANFs is one of the key factors affecting hearing outcomes.

In this work, we propose patient-customized, computational ANF stimulation models, which are not only coupled with our patient-specific (EAMs) to ensure electrical and anatomy customization but also estimate neural health status along the length of the cochlea. We implement a biological nerve model proposed by Rattay et al. [37] and drive the nerve model using electrical potentials sampled from the electrical field at auditory nerve fiber locations. We tune neural health parameters for each ANF bundle so that simulated amplitude growth functions for each electrode in the array best match the corresponding clinically measured ones. We then conduct a validation study in which we evaluate our health prediction by simulating the spread of excitation (SOE) functions with the estimated neural health parameters and compare the results to clinically measured SOE. Experiments with 8 subjects show promising model prediction accuracy, which suggests our modeling approach may provide an accurate estimation of ANF health for CI users.

Contribution 3. Cochlear implant electric potential estimation using 3D neural networks

In order to provide objective information to the audiologist for programming, our group has developed EAMs to permit estimating which auditory neural sites are stimulated by which CI electrodes. To do this, we have proposed physics-based models to calculate the electric potential in the cochlea generated by electrical stimulation. However, solving these models requires days of computation time and substantial computational resources.

With the rapid development of machine learning for computer vision tasks, 3D U-net is widely used in the segmentation, enhancement, and diagnosis of medical images. Yet, generating physics-based 3D electric potential maps has not been well explored. In this work, we propose a deep-learning based method to estimate the patient-specific electric potential maps using a 3D U-Net-like architecture with a physics-based loss function that reflects Poisson's equation for electrostatics as a self-supervised loss term. Our network is trained with a dataset generated by solving physics-based models and the results show that the proposed method can achieve similar accuracy to the traditional method and largely improves the speed of estimating

the intra-cochlear electric potential.

Contribution 4. Patient-specific electro-anatomical modeling of cochlear implants using deep neural networks

In previous research, our group has developed methods that use patient-specific electrical characteristics to simulate the activation pattern of auditory nerves when they are stimulated by CI electrodes. However, estimating those electrical characteristics require extensive computation time and resources. In this work, we propose a deep-learning-based method to coarsely estimate patient-specific electrical characteristics using a cycle-consistent network architecture. These estimates can then be further optimized using a limited-range conventional searching strategy. Our network is trained with a dataset generated by solving physics-based models. The results show that our proposed method can generate high-quality predictions that can be used in the patient-specific model and largely improves the speed of constructing models.

Contribution 5. Super-resolution segmentation for inner-ear CT images

In our current implementation of the high-resolution EAM, the tissue label maps are obtained from non-rigid registration of specimens' inner-ear segmentations. A total of nine specimens μCT-level segmentations are used and the voxel labels are obtained via a majority voting. Herein, we proposed a deep learning architecture, SRSegN, that is able to perform super-resolution segmentation for pre-operative inner-ear CT images. The deep learning model combines the advantages of convolutional networks and Transformer encoders, and features a dual-encoder architecture where a convolutional encoder is used to extract low-resolution features and a Transformer-based encoder is used to perform upsampling and extract super-resolution features. The proposed model not only outperforms the traditional registration-based method, but also achieved better segmentations performance than multiple state-of-the-art deep learning based segmentation models in terms of dice coefficient on both cochlear dataset and a large public dataset.

Contribution 6. Automatic auditory nerve fiber segmentation

The traditional auditory nerve fiber segmentation method proposed by Cakir et al. Cakir et al. (2019) relies heavily on manual inspection and adjustment. And the central axons of the resulting nerve fibers may still be localized mistakenly even after manual correction because the landmarks used in that method are not representative enough and the assumption that central axons all proceed straight towards the same direction once entering the modiolus. In this work, we introduced a fully automatic ANF segmentation method. The peripheral and central axon of an ANF will be estimated individually based on five sets of automatically generated landmarks. The fast marching method is used to find the geodesic paths for the peripheral axons between the surfaces of the scala tympani (ST) and scala vestibuli (SV) meshes. Cylindrical coordinate systems are constructed based on the landmarks and are used to smoothly interpolate trajectories for the spiral central axons. Experiments show that our proposed method outperforms the original method and achieves

impressive performance with 0 overlapping ANFs and 0 ANFs passing through the bone. The number of ANFs that pass through ST or SV is also reduced by 36.1

1.9 Thesis Organization

The rest of this dissertation is organized as follows. Section 2 describes the CI electrode sequence optimization method; Section 3 describes the ANF health estimation method and contributions; Section 4 describes an alternative UNet-based method for computing electrical potential maps of the inner ear; Extending this UNet architecture, section 5 introduces the multi-task cycle-consistent neural network for generating electrical potential maps as well as predicting the patient-specific electrical parameters; Section 5 introduce the novel architecture of SRSegN for μCT-level segmentations using conventional pre-operative CT images; and Section 6 describes the new ANF segmentation method which is fully-automatic, fast, robust and outperforms the original method.

CHAPTER 2

Cochlear Implant Electrode Sequence Optimization using Patient Specific Neural Stimulation Models

2.1 Introduction

Cochlear implants (CIs) are an effective treatment for patients who suffer sensory-based hearing loss. In CI surgery, an array of electrodes is implanted into the inner ear to permit electrical stimulation of the auditory nerve. The number of electrodes on an electrode array ranges from 12 to 22 depending on the manufacturer. After surgery, CI recipients undergo many programming sessions with an audiologist who adjusts the CI processor settings to improve performance. However, lacking objective information about what settings will lead to better performance, a trial and error procedure is implemented. Traditionally, the programming process includes determining the set of electrodes that need to be activated as well as the dynamic range of electric current for each electrode Noble et al. (2015). As weeks of experience with given settings are needed to indicate long-term outcome with those settings, this process can be frustratingly long and lead to suboptimal outcomes.

In a previous study, our lab developed a patient-specific model of auditory nerve fiber activation, which relies on CT-based intra-cochlear anatomy segmentation Noble and Dawant (2015) and localization of CI electrodes Zhao et al. (2019) in order to create a high-resolution electro-anatomical model (EAM) Cakir et al. (2017a,b). Experiments have shown that our model accurately estimates intra-cochlear voltage distributions. Therefore, when these voltage estimates are input into neural activation models, we are able to estimate the stimulation status of nerve bundles at different locations in the cochlea.

Typically, when an electrode on the array injects a certain electrical current into the cochlea, a number of nerve bundles get activated. With the help of our patient-specific model, we can estimate which nerve fibers are stimulated by which electrodes and find which fiber populations are stimulated by multiple electrodes (see Figure 2.1). Due to the refractory behavior of the nerve fibers, nerves become insensitive to further stimulation for a period of 1 or more milliseconds (ms) following prior stimulation. As a result, when multiple electrodes stimulate the same fibers, masking artifacts can occur, where one electrode may fail to stimulate certain nerve bundles due to preceding stimulation by another electrode. Masking is a phenomenon known to negatively affect hearing outcomes Fu and Nogaki (2005); Boëx et al. (2003). Therefore, a reduction of these cross-electrode masking effects should lead to improved CI performance.

CIs activate one channel at a time in order to mitigate electric field interactions and pre-define a channel firing order that is executed in each stimulation frame. Most manufacturers use a sequential firing order.

CIs manufactured by Advanced Bionics Corp. (Valencia, CA) will by default select a non-sequential firing order so that adjacent electrodes that are most likely to have overlapping stimulation will not be activated one immediately after another to reduce masking effects. However, this is done in a one-size-fits-all manner and does not account for patient-specific masking, which may be non-uniform across the array. In this work, we propose a new method to detect masking effects using our patient-specific model and to customize the firing order to minimize the masking effect.

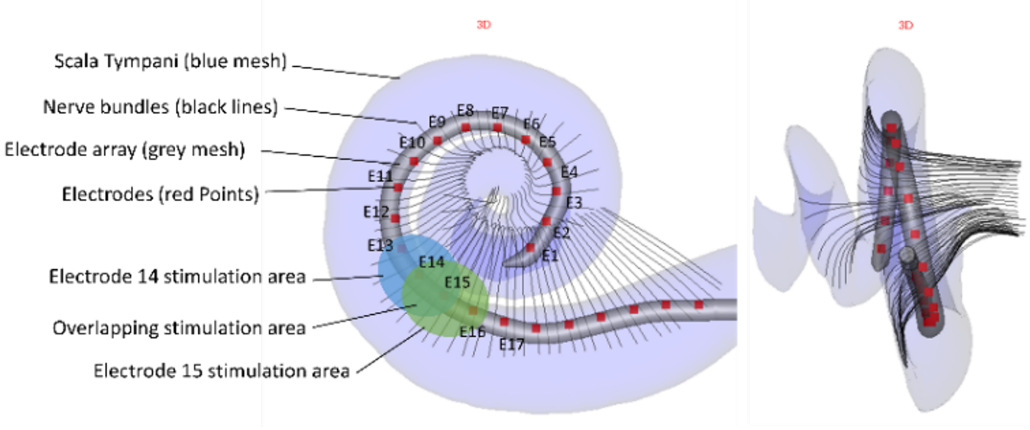


Figure 2.1: A patient-specific model where 75 nerve bundles are segmented around the electrode array.

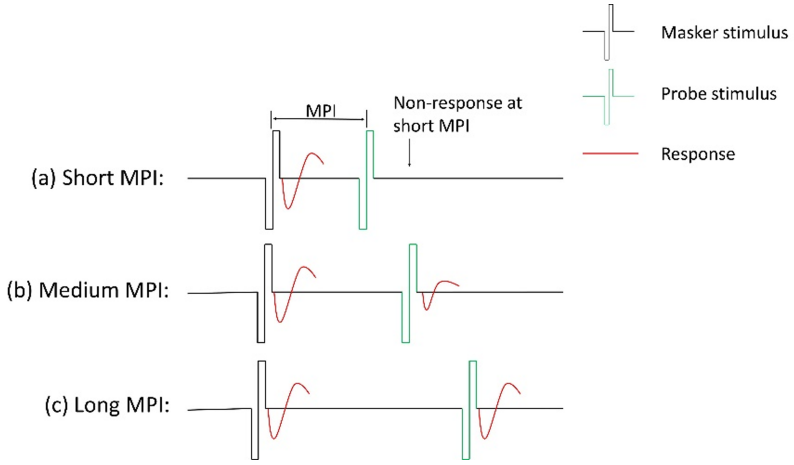


Figure 2.2: Simplified schematic illustration of how the Masker-Probe Interval (MPI) affects the second stimulation of nerve fibers.

2.2 Method

In this section, we will show details of our experimental methods and optimization algorithm. The data we used in experiments are four patient-specific EAMs we created in a previous research study Cakir et al.

(2017b). The resistivity values of different tissue classes have been optimized to match patient-specific impedance measurements. These four models have sixteen electrodes on their CI array. The models provide an estimate of the electric potential generated by each of the electrodes through the cochlear tissue.

2.2.1 Masking effect quantification

In the models we constructed for our four patients, 75 nerve fiber bundles were segmented Cakir et al. (2019) (see Figure 2.1). For each fiber bundle a fiber stimulation model can be created that is driven using the electric potential estimated by the EAM. Using these models, we find that each electrode on the array is able to electrically stimulate several nerve bundles in an area near them. When nerves are activated, they generate an action potential. The compound action potential among all activated fibers can be measured using the implant, similar to the simulated responses shown in Figure 2.2. The groups of bundles recruited by each electrode overlap with each other.

As masking effects prevent nerve fibers to be electrically evoked by two electrodes within the refractory period, a certain amount of time is required for any nerve fiber to return to its resting potential between its last activation and the next one. The time delay after firing a 'masker' electrode and before firing a 'probe' electrode is called the masker-probe interval (MPI) (see Figure 2.2). Therefore, we quantify the masking effect for a masker-probe pair as the smallest MPI required to wait after stimulating a masking electrode before the probe is able to stimulate a required fraction τ of the nerve bundles it recruits with resting state stimulation. We refer to this as the minimum non-masking MPI (mMPI) and define τ to be 0.9 in our experiments.

A binary search is used to find the mMPI for each electrode pair (v, u), where v and u are electrodes on an array, and stimulation with v and u is simulated using our patient-specific models. Notice that mMPI is an asymmetric measure, and the mMPI for pair (v, u) and pair (u, v) could, in general, be different. The mMPI estimated for all electrode pairs can be recorded in a mMPI matrix, where the element at index (v, u) in the mMPI matrix is the mMPI when using masker v to mask probe u.

2.2.2 Electrode sequence optimization

After obtaining the mMPI matrix, we are able to construct a graph that represents the overlapping relationship between electrodes. The nodes in this graph represent each electrode, and the edges between electrodes represent the corresponding mMPI. Since both of the patient models we used in this research have 16 electrodes, their representing graphs will be fully connected weighted directed graphs with 16 vertices.

Determining an optimal electrode firing sequence is equivalent to finding a Hamiltonian path of this graph with the minimum cost. However, it is not possible to assign a single cost locally to each edge as would be

typically done with a graph approach because, given an electrode sequence (or a path in this graph), any electrode in the sequence will not only be affected by its adjacent electrode but also by the second or third electrodes before it in the sequence. This is because the mMPIs are substantially longer than the width of each pulse in the sequence. Thus, considering the stimulation pulse width, the time delay required to ensure non-masking conditions between the ith and jth electrodes in the sequence would be mMPI(e_i, e_j) – $W|i–j|$, where W is the pulse width, to account for the time delay that exists in the sequence between e_i and e_j. We design our cost function for adding a new electrode e_{n+1} to a given electrode sequence $S = \{e_1, e_2, e_3, \ldots, e_n\}$ to be

$$c_{n+1} = \sum_{i=1}^{m} \partial_i M(e_{n-i+1}, e_{n+1}) \tag{2.1}$$

where M is the channel overlap timing matrix, and ∂_i are weighting factors indicating the influence of former electrodes in the sequence. In our experiments, we set m = 2, ∂_1=0.7 and ∂_2=0.3 so that mMPI for the previous two electrodes is included, and the mMPI for the most recent electrode has the greatest importance.

Because the electrode sequence represents one frame, and CIs will stimulate one frame after another, the first electrode stimulated in this frame will be affected by the last two electrodes in the last frame, and similarly, the last two electrodes in this frame will also affect stimulation status in the next frame. To capture the inter-frame interactions, we define the total cost of a full-frame sequence S as

$$C = \sum_{i=3}^{n} c_i + \partial_1 M(e_1, e_2) + \partial_2 M(e_n, e_2) + \partial_1 M(e_n, e_1) + \partial_2 M(e_{n-1}, e_1) \tag{2.2}$$

Because our cost function uses non-local constraints, standard path-finding methods such as Dijkstra's algorithm cannot be used. Thus, the path-finding algorithm we use to attempt to minimize this cost function is based on the optimization method we have used in prior work for electrode localization Noble and Dawant (2015). The algorithm starts with a random electrode as its seed node. Then, the L-1 remaining nodes are determined by growing a list of candidate paths $\{p\}$ that stem from the seed node.

At each iteration, the Grow stage allows the growth of new candidate paths by adding all possible one-node extensions to each candidate path in $\{p\}$ from the last iteration. Then the number of candidate paths is reduced in the Prune stage by sorting and keeping the P paths with the lowest cost while discarding the remaining paths with a higher cost to limit the number of stored paths from growing exponentially in each iteration. This algorithm is shown in Algorithm 1. It is executed with P = 10000 to find the optimal sequence. This algorithm provides no guarantee of finding a global optimum; however, it permits non-local constraints and has controllable computational efficiency.

Algorithm 1 Path finding algorithm

Input: overlapping relationship graph whose nodes are all the electrodes on the array and edges described by overlap timing matrix. **L** = Num. of electrodes in the final sequence, and **P** = maximum num. of candidate path
Initialize list of candidate paths $\{p\} = \{\{e_1, Cost(e_1)\}\}$
for **L**-1 iterations **do**
    Grow
    Initialize new list of candidate paths $\{q\} =$
    for each candidate path in $\{p\}$ **do**
        Find child nodes in $\{p\}$, $\{c\}$ = {n, Cost(n, $\{p\}$)}
        for each node in $\{c\}$ **do**
            Add new path to list $\{q\}$ = $\{q\}|(\{p\}$ + $\{c\})$
        end for
    end for
    Prune
    Sort candidate paths in $\{q\}_1^N$ by increasing path cost
    Update $p = \{q\}_1^{P<<N}$
end for
Output: $\{p_1\}$

Table 2.1: Average mean absolute difference between simulated and measured AGF and SOE.

Patient 1		
	Firing order	Cost
Sequential order	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	17.50 ms
Skip 4 electrode	1 5 9 13 2 6 10 14 3 7 11 15 4 8 12 16	10.01 ms
Optimized order	1 13 5 15 12 7 2 11 6 4 10 9 3 14 8 16	9.01 ms
Patient 2		
	Firing order	Cost
Sequential order	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	17.29 ms
Skip 4 electrode	1 5 9 13 2 6 10 14 3 7 11 15 4 8 12 16	6.95 ms
Optimized order	1 14 11 3 16 9 2 13 8 4 12 7 6 5 15 10	5.38 ms
Patient 3		
	Firing order	Cost
Sequential order	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	16.30 ms
Skip 4 electrode	1 5 9 13 2 6 10 14 3 7 11 15 4 8 12 16	5.28 ms
Optimized order	1 12 8 14 3 6 11 16 2 10 5 15 9 4 13 7	4.70 ms
Patient 4		
	Firing order	Cost
Sequential order	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	18.60 ms
Skip 4 electrode	1 5 9 13 2 6 10 14 3 7 11 15 4 8 12 16	9.46 ms
Optimized order	1 7 12 2 8 5 13 6 16 11 3 15 10 4 14 9	8.81 ms

2.3 Experiments and results

The mMPI matrix for four patient-specific models is shown in Figure 2.3. The y-axis represents the electrodes that serve as "maskers", and the x-axis represents the electrodes that serve as "probes". As we can see in Figure 21, the MPI between the masker and probe is negatively correlated with the distance between electrodes on the array because the further two electrodes are from each other the less overlapping area they share.

Using these mMPI matrices, the optimized electrode sequence results are shown in Table 6. 'Sequential order' indicates firing all the electrodes one by one from the beginning of the array to the end. This method is used by two CI manufacturers and could cause severe masking effects. The 'skip 4 electrodes' order is one that is used by default by Advanced Bionics, where every 5th electrode in the array is activated starting from e_1, followed by e_2, then e_3, then e_4. This method indeed avoids much of the overlap problem, but it is still limited by not taking into account patient-specific mMPI. Compared to the sequential order used by some CI manufacturers and the skipping order used by one manufacturer, the optimized firing order has a cost as defined by equation(2.2) for all patient models in our experiments.

2.4 Conclusion

To the best of our knowledge, this is the first time that an algorithm for patient-specific optimization of cochlear implant electrode firing order has been proposed. Such optimized firing orders could help to improve cochlear implant performance.

In this study, we have used our patient-specific model to predict the stimulation status of nerve bundles. With the help of these predictions, we are able to quantify how much one electrode is affected by another by computing the minimum non-masking masker-probe-interval. we proposed a new method to assist audiologists by determining a customized firing order of the electrodes on the CI array using image-based models of patient-specific neural stimulation patterns. Our models permit estimating the time delay needed after firing an electrode so that the nerve fibers they stimulate can recover from the refractory period. Specifically, we have proposed an algorithm to optimize the firing order using a graph-based path-finding algorithm that permits non-local constraints. Our preliminary experiments show that our optimized firing orders have better theoretical masking performance than traditional methods.

Our future work will include evaluating our method on a larger dataset and implementing the optimized firing order with research subjects to evaluate whether it leads to improved hearing outcomes.

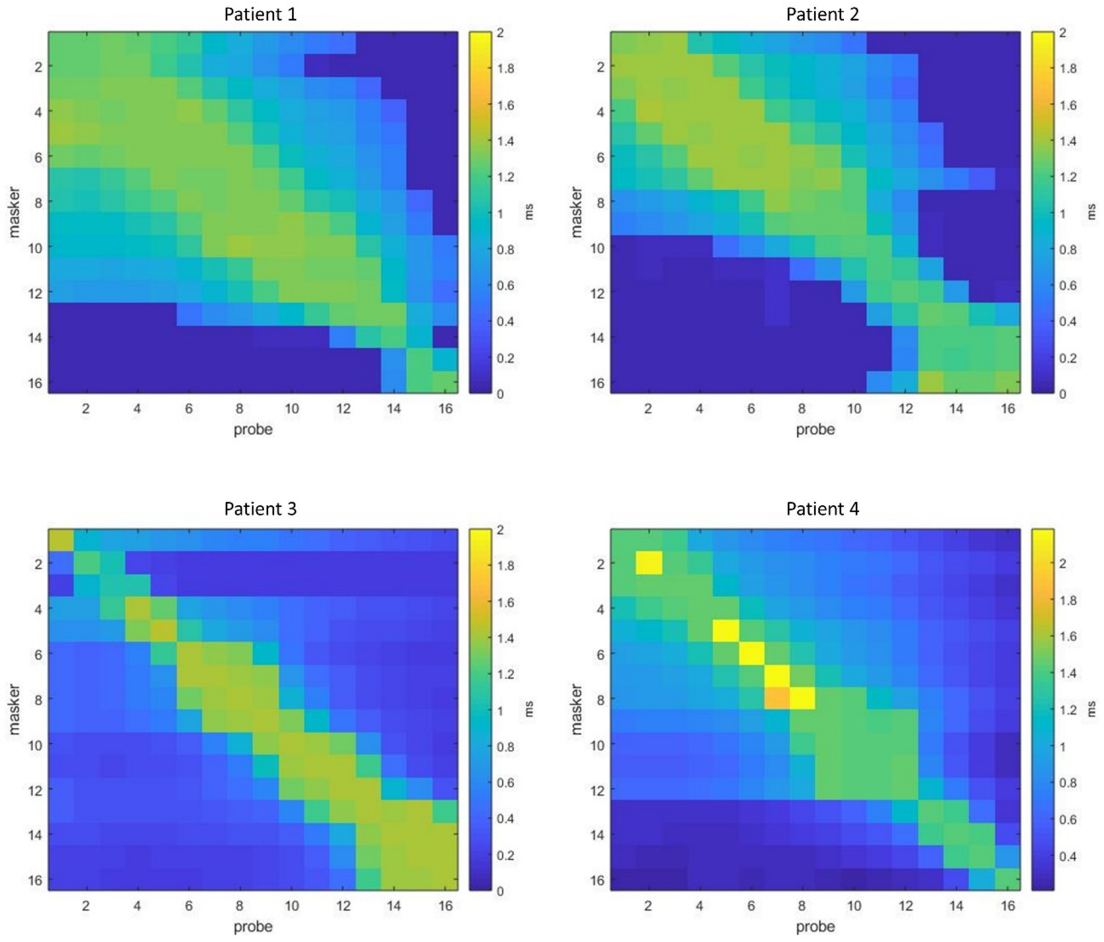


Figure 2.3: mMPI matrix for four patient-specific models. The y-axis represents the electrodes that serve as 'maskers', and the x-axis represents the electrodes that serve as 'probes'. And the brighter the pixel is, the bigger the masking effect will be.

CHAPTER 3

Auditory Nerve Fiber Health Estimation using Patient Specific Cochlear Implant Stimulation Models

3.1 Introduction

Hearing is the outcome of a series of complex steps that translate sound wave signals into electrical signals. In normal hearing, sound waves induce pressure oscillations in the cochlear fluids, which in turn initiate a traveling wave of displacement along the basilar membrane (BM). This membrane divides the cochlea along its length and produces maximal responses to sounds at different frequencies Wilson and Dorman (2008). Because the motion of BM is then sensed by hair cells that are attached to the BM, these sensory cells are fine-tuned to respond to different frequencies of the received sounds. The activation of the corresponding hair cells releases chemical transmitters to electrically stimulate the spiral ganglion nerve cells. And the electrical signal is propagated along the auditory nerve fibers (ANFs), traveling through the brain stem, and finally reaching the auditory cortex allowing the brain to sense and process the sounds. In summary, the sound signal is decomposed by BM and hair cells, and the ANFs send this information to the brain to hear the sound. For patients suffering sensorineural hearing loss, which is principally caused by damage or destruction of the hair cells, however, the decomposition process of incoming sound waves cannot be performed.

In this situation, direct stimulation of the ANFs is possible if they are intact Greenberg et al. (2004). A CI replaces the hair cells with an externally worn signal processor that decomposes the incoming sound into signals sent to an electrode array. Electrode arrays have up to 22 contacts depending on the manufacturer, dividing the available ANFs to, at most, 22 frequency bands or stimulation areas when using monopolar stimulation. Studies have shown that hearing outcomes with CIs are dependent on several factors including how healthy the ANFs are Nadol Jr et al. (1989).

Our group has been developing EAMs in order to provide objective information that can assist audiologists with programming Cakir et al. (2017b,a). Building on those studies, in this study we propose to use these EP models as input to ANF activation models to predict neural activation caused by electrical stimulation with the CI. We also propose the first in vivo approach to estimate the health of individual ANFs for CI patients using these models.

In summary, herein we propose patient-customized, image-based computational models of ANF stimulation. We also present a validation study in which we verify the model's accuracy by comparing its predictions to clinical neural response measurements. Our methods provide a patient-specific estimation of the electro-neural interface in unprecedented detail and could enable novel programming strategies that significantly

improve hearing outcomes with CIs.

3.2 Related works

Several groups have proposed methods for predicting neural activation caused by electrical stimulation Rattay et al. (2001b); Malherbe et al. (2016); Cartee (2000). Most of these methods use physiologically-based active membrane nerve models driven by physics-based estimation of the voltage distribution within a given anatomical structure. However, these studies either lack the capacity to be applied in-vivo or only confine themselves to anatomical customization instead of constructing both anatomically and electrically customized models that take advantage of physiological measurements that are clinically available. It is possible that these models need to be fully customized in order to prove useful for clinical use. Thus, in this work, we are proposing patient-customized, computational ANF stimulation models, which are not only coupled with our patient-specific EAMs to ensure electrical and anatomy customization but also estimate neural health status along the length of the cochlea. Our models permit accurately simulating physiological measurements available via CIs.

Our ANF stimulation models are built on three critical components: the biological auditory nerve model proposed by Rattay et al. (see Chapter I, section 5.2), the CT-based high-resolution EAM of the electrically stimulated cochlea (see Chapter I, section 4), and the auditory nerve fiber segmentation proposed by our group (see Section 1.5.2). These models help to describe auditory nerves from biological, electrical, and spatial features respectively. In the next section, we will illustrate our approach to combining these models and build our novel, health-dependent ANF stimulated models based on them.

3.3 Methods

We start the methods section with an overview of the proposed approach, followed by subsections providing more detail regarding novel components of the work. There are approximately 30,000 ANFs in a healthy human cochlea Spoendlin and Schrott (1989). We represent them using auditory nerve bundles that are segmented along the length of the cochlea as shown in Figure 5a.

To reduce the computational cost of our approach, we represent only 75 distinct bundles, each representing potentially hundreds of fibers. Our proposed nerve bundle action potential model is $P_M * HM + P_U * H * (1 - M)$, where P_M and P_U are the action potential responses of single ANF cell biological nerve models for a myelinated fiber and the degenerated, unmyelinated fiber model, respectively. H is the number of living fibers in the bundle that can be recruited for stimulation. M is the fraction, among those ANFs, of healthy versus degenerated ones. Thus, the bundle action potential is the superposition of the two fiber models' action potential predictions scaled by the number of such fibers we estimate to be present in the bundle. We have

designed an approach, described below, to determine patient-customized values for these two parameters for each of the 75 distinct bundles.

The biological ANF model permits simulating action potentials (APs) created by ANFs as a result of the EP the ANF is subjected to. The EP sampled at discrete locations along the fiber bundle – each node of Ranvier (black nodes between myelinated segments in Figure 1.5b) – is used to drive the ANF activation model. The EP generated by the CI electrodes can drive the ANF models and can be estimated using our CT-based high-resolution EAM of the electrically stimulated cochlea as described previously.

Next, we will use our bundle model to simulate neural response measurements that can be clinically acquired. These measurements include recordings acquired using the CI electrodes of the combined AP signal that is created by the set of ANFs activated following a stimulation pulse created by the CI. Such measurements are called ECAPs. Several ECAP-based functions can be clinically acquired. The most common are the AGF, which samples how the magnitude of recorded ECAPs (μV) grow as the current is increased for the stimulation pulse signal; and the SOE function, which measures the fraction of ECAP responses for two stimulating electrodes that are generated from the same ANFs Hughes (2012); Briaire and Frijns (2005). Both AGFs and SOEs can be simulated using our models and clinically measured using the patient's implant. While both AGF and SOE are rich with information about the electro-neural interface and have been acquirable for CI patients for decades, these metrics are not routinely used for clinical programming because they have been difficult to interpret. Thus, the method we propose provides a unique opportunity to:

(1) estimate neural health by tuning model neural health parameters so that model-predicted ECAP functions match clinically measured ones.

(2) provide a physical explanation for the AGF and SOE measurements. Both of these typically unknown quantities could significantly improve an audiologist's ability to program the CI.

We tune neural health parameters for each ANF bundle so that simulated AGF functions for each electrode in the array best match the corresponding clinically measured ones. Finally, we conduct a validation study in which we evaluate our health prediction by simulating SOE functions using the model with the estimated neural health parameters and compare the results to clinically measured SOE to demonstrate the predictive value of our proposed models. The following subsections detail each step of our approach.

3.3.1 Dataset

N = 8 patients who had undergone CI surgery were used to create neural health estimation models. All the patients underwent pre- and post-implantation CT imaging needed to localize the intra-cochlear position of the electrodes and to create the tissue classification maps for the EAM models. The three clinical electrophysiological measurements critical for tuning and evaluating our models (EFI, AGF, and SOE) were also

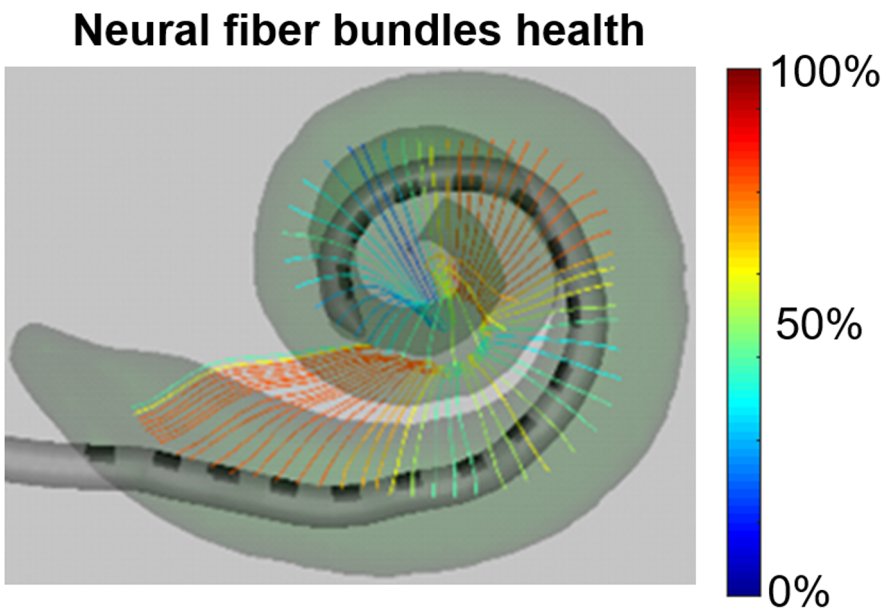


Figure 3.1: Overview of the ANF models. It shows the spatial distribution of ANF bundles colored with a nerve health estimate.

collected for all electrodes, for all patients with institutional review board approval.

3.3.2 Nerve Model

For each nerve fiber model, we follow the approach of Rattay et al. as we described in Chapter I, section 1.5.2. The modeling is done using the NEURON simulation environment. The overview of the auditory nerve fiber used in this study is shown in Figure 5b. As shown in the figure, each nerve model consists of three subunits which are the peripheral axon, the soma, and the central axon. The peripheral axon is located near hair cells in a human cochlea. They are myelinated when the fiber is healthy and fully functional. It is also common in patients with hearing loss that fibers, where the peripheral axon has become unmyelinated, exist and could have a weaker response to stimulation. We define them as functional but 'unhealthy' ANFs. Then we can parameterize the health of each nerve bundle by varying the number of fibers, H, as well as the ratio of myelinated vs unmyelinated fibers, M, for each ANF bundle.

Our bundle model simulates bundle APs to the estimated EP generated by CI electrodes as previously discussed. Subsequently, ECAP measurements can be simulated in the model. To do this, each node of Ranvier for each bundle is treated as a current source, and the same finite difference method used in EAM for estimating EP created by the CI is repurposed for estimating the EP created by the APs generated by all the bundles. This is done by defining bundle nodes as current sources corresponding to cross-membrane

current. Thus, the result of each bundle model drives a new EAM to estimate the EP created by the ANFs in the cochlea. The value of the EP is then recorded at the site where the recording electrode is located. This process directly simulates the clinical ECAP measurement process.

In summary, the ECAP simulation can be divided into three steps: (1) for a given stimulating electrode, we calculate the EP using an EAM and record the resulting EP at the nodes of Ranvier for each nerve bundle; (2) we use those voltages as input to the neural activation models for both myelinated and unmyelinated nerves to compute our combined nerve bundle AP; and (3) we estimate the EP created by the bundle APs using another EAM, permitting simulated ECAP measurement from the position of recording electrode. In practice, in the final step, an EAM can be created independently for each bundle and the compound response at the recording electrode is then given by

$$\text{simulated ECAP} = \sum_{i=1}^{75} P_{M,i}\, H_i\, M_i + P_{U,i}\, H_i\, (1 - M_i) \tag{3.1}$$

where $P_{M,i}$ and $P_{U,i}$ represent the value of the EF sampled at the recording electrode for the simulated ECAP of the myelinated and unmyelinated ANF model in the ith nerve bundle, respectively, and H_i and M_i are the number of fibers and fraction of those fibers that are healthy for the ith nerve bundle.

3.3.3 Optimization process

Spoendlin et al. Spoendlin and Schrott (1989) found that for a healthy human cochlea, the average number of fibers can vary between 500 fibers per millimeter (mm) to 1400 fibers per mm depending on the location within the cochlea. Given that a nerve bundle in our model can represent a region as wide as 0.4 mm along the length of cochlea, we have set the boundary values for the number of functional nerve fibers to be between 0 (all unresponsive) and 550 (all responsive) and the healthy ratio or the myelination ratio from 0 (all responsive nerve fibers are damaged) to 1 (all responsive nerve fibers are healthy).

Instead of determining values for H_i and M_i for each of the 75 nerve bundles independently, a set of control points are used to enforce spatial consistency in parameter values. We define n+1 control points along the length of the cochlea, where n is the total number of active electrodes. The control points are positioned to bracket each electrode. The parameters at those control points were randomly initialized with Hi between 0 to 550 and Mi from 0 to 1. The parameters for each nerve bundle are then linearly interpolated along the length of the cochlea using the control points.

We use the bounded Nelder-Mead simplex optimization algorithm D'Errico (2019) to optimize values at the control points. The cost function is calculated as the mean absolute difference between the simulated and measured AGF values for each electrode. Starting from a random initialization at each control point, our algorithm will iteratively calculate the parameters of every nerve bundle by interpolating control point

values, simulate AGF using those parameters to evaluate the cost function discussed above, and determine new control point parameters using the Nelder-Mead simplex method until a maximum iteration number is reached or the change in error falls below the termination threshold (0.1μV). Algorithm pseudocode is presented in Algorithm 2.

In our implementation, AGF values that were less than 35μV were not included in the optimization process because low AGF values tend to be below the noise floor and are usually excluded from clinical analyses. During our experiments, Algorithm 1 is executed from 250 different random initializations for each patient model. The final fiber count and healthy ratio for every nerve bundle are determined as the median values across the 10 optimization runs that resulted in the lowest average error. This procedure diminishes the likelihood of choosing sub-optimal parameters that are local minima.

Algorithm 2 Estimate the patient-specific neural health parameters

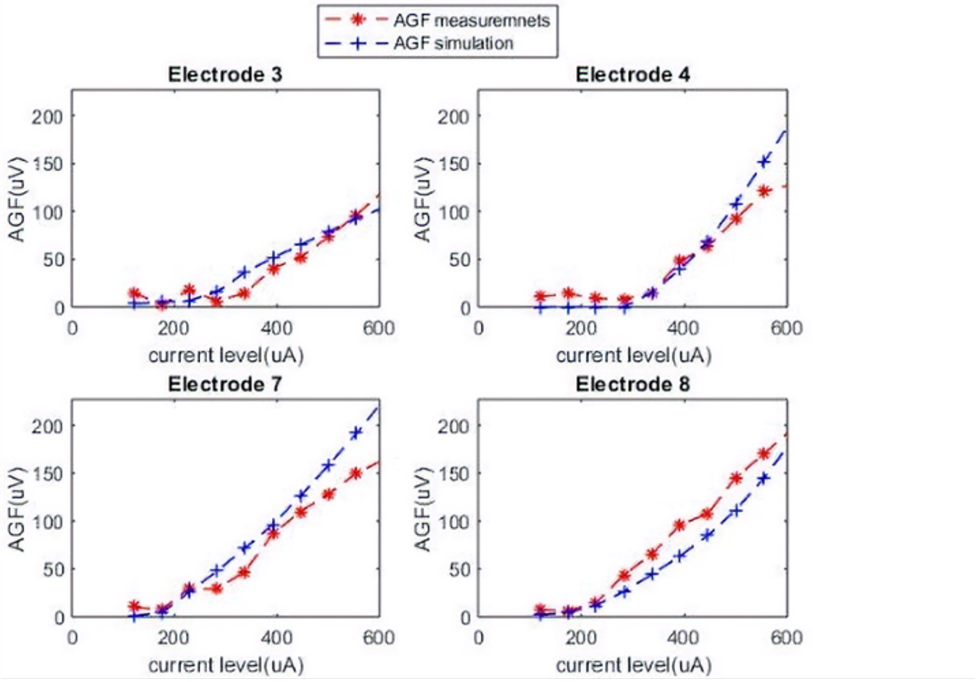
Input: P_{AGF} = Patient AGF measurement
Variable: S_{AGF} = Simulated AGF data, **H** = Number of nerve fibers within bundles, **M** = Myelination ratio of fibers within bundles
Output: **HC** = Fiber count assigned to each control point, **MC** = Myelination ratio assigned to each control point
Start: Assign threshold and maxIteration, randomly assign **HC** and **MC**
while $\Delta|error| > threshold$ and $counter < maxIteration$ **do**
    Interpolate **H** and **M** using **HC** and **MC**
end while
for each electrode i **do**
    $error_{AGF}$[i] = mean(abs(P_{AGF}[i] - S_{AGF}[i]))
end for
$error = mean(error_{AGF})$
Optimize **HC** and **MC** using a constrained nonlinear search based on the Nelder-Mead simplex

3.4 Results

The average absolute differences between the simulated and measured AGF and SOE values for fully customized EAMs are shown on the left side of Table 1. The average absolute difference between the simulated and the measured AGF values could be interpreted as the training error. Mann-Whitney U tests reveal significant improvement in AGF errors after training (p¡0.01). The error between the simulated and the measured SOE can be interpreted as the testing error since SOE was not used to optimize neural health parameters. Further, SOE is likely more sensitive to neural health than AGF because it is much more dependent on the spatial distribution of ANFs that contribute to the neural responses. The average SOE error across all patients after optimizing neural health parameters using our proposed method is 39.5μV.

In Figure 3.3, we plot the simulation and clinical result of both AGF and SOE for subject 1. Both the quantitative and qualitative comparisons show excellent agreement between neural stimulation responses that

(a)



(b)

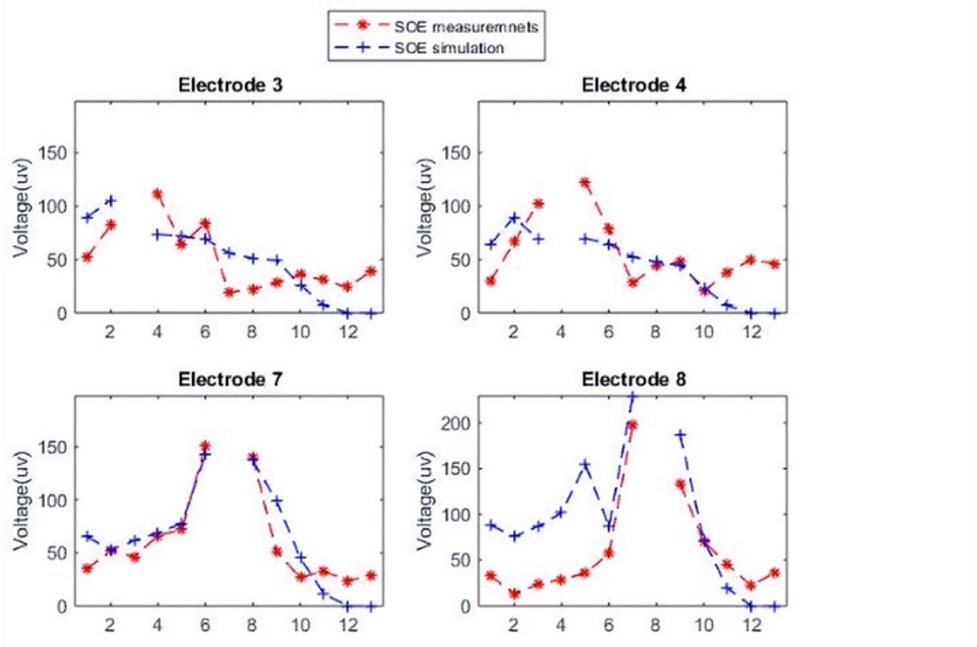


Figure 3.2: (a) Comparison between measured and simulated AGF data. (b) Comparison between measured and simulated SOE data.

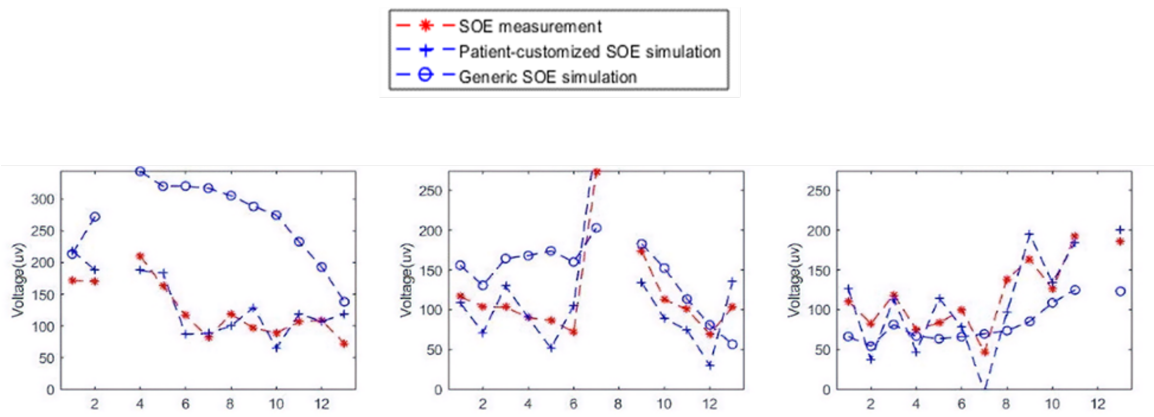


Figure 3.3: SOE testing error for patient-customized versus generic models for Subject 4.

are clinically measured and those that are predicted by our parameter-optimized models. We further compare the difference between neural health estimation using our fully customized models vs. generic models, where default electrical properties are used, for the first six subjects on the right side of Table 3.1.

The AGF error (training error) resulting from the generic and electrically customized models is similar while the testing error with fully customized models is much smaller than generic models. A one-sided Mann-Whitney U test reveals significantly better (p¡0.05) testing error with the fully customized model compared to the generic models. Example plots demonstrating the superiority of SOE simulations using customized for one subject are shown in Figure 3.2. These results imply our patient-specific EAMs are critical, not only for EFI simulation but also for accurate neural health estimation. An example neural health estimation result is shown in Figure 3.1, where the neural health color codes are a combined function of both health parameters equal to $H(0.5+M)$. Varying health of several regions of nerves was identified by the proposed method in order for prediction to match measured AGF.

3.5 Conclusion

In this research, we developed an approach to estimate the health of ANFs using patient-customized, image-based computational models of CI stimulation. The resulting health parameters provide an estimate of the health of ANF bundles. It is impossible to directly measure the number of healthy ANFs in vivo to validate our estimates, however, experiments with 8 subjects show promising model prediction accuracy, with excellent agreement between neural stimulation responses that are clinically measured and those that are predicted by our parameter optimized models. These results suggest our modeling approach may provide an accurate estimation of ANF health for CI users.

With the current IGCIP approach, assumptions are made about electrical current spread to estimate which

Table 3.1: Average mean absolute difference between simulated and measured AGF and SOE.

Subject No.	Fully Customized Models			Generic Models	
	AGF error – before optimiz. health (μV)	AGF error – after optimiz. health (μV)	SOE error-testing error(μV)	AGF error – after optimiz. health(μV)	SOE error-testing error(μV)
1	58	16	31	22	53
2	187	19	32	48	49
3	299	39	37	28	76
4	66	37	44	39	102
5	131	11	29	19	56
6	97	8	21	15	36
7	62	17	48	-	-
8	141	26	59	-	-
Average	134	21.6	39.5	28.5	62.0

fiber groups are activated based on their distance to the electrode. Our estimation of the health of ANFs may improve our estimation of neural stimulation patterns and lead to highly customized IGCIP strategies for patients.

Our future work includes evaluating the effectiveness of novel patient-customized programming strategies that use these models. Further, our methods could provide an unprecedented window into the health of the inner ear, opening the door for studying population variability and intra-subject neural health dynamics.

CHAPTER 4

Cochlear Implant Electric Field Estimation using 3D Neural Networks

4.1 Introduction

Cochlear implants (CIs) are considered the standard-of-care treatment for profound sensory-based hearing loss on Deafness and (NIDCD). In CI surgery, an array of electrodes is implanted into the inner ear to permit electrical stimulation of the auditory nerve. The number of electrodes on an electrode array ranges from 12 to 22 depending on the manufacturer. After surgery, CI recipients undergo many programming sessions with an audiologist who adjusts the CI processor settings to improve performance. However, lacking objective information about what settings will lead to better performance, a trial-and-error procedure is implemented. As weeks of experience with given settings are needed to indicate long-term outcomes with those settings, this process can be frustratingly long and lead to suboptimal outcomes Wolfe and Schafer (2014); Vaerenberg et al. (2014).

In a previous study, our group has proposed an image-guided cochlear implant programming (IGCIP) technology that can estimate auditory nerve activation patterns in order to simplify the traditionally tedious post-operative programming procedure and improve hearing outcomes Noble et al. (2013). It has been proved that our programming strategy does significantly improve hearing outcomes Noble et al. (2015, 2016). However, IGCIP estimates the neural stimulation patterns of the electrodes in a relatively coarse manner using only the spatial relationship between the electrodes and where the nerve should be and assuming all the nerves are healthy. To perform better estimation of neural activation, we further proposed image-based patient-specific electro-anatomical models (EAMs) Cakir et al. (2017a,b), which is an anatomical and electrically customized intra-cochlea. Several groups have used EAMs to study intra-cochlear voltage distribution and its effect on neural activation. However, these models either cannot be applied in vivo, thus, patient-specific differences cannot be incorporated Frijns et al. (2001); Whiten (2007); Kalkman et al. (2015), or lack of accuracy due to using CT images and rigid registration [10]. Our EAMs are created using μCT images of cochlea specimens.

These high resolution models are spatially registered to patient CT data and electrically adapted to patient-specific configuration to ensure fully customization. Based on our patient-specific EAMs, we have developed helpful techniques such as automatic CI electrode sequence optimization Liu et al. (2020b) and auditory nerve fiber health estimation Liu et al. (2020a) to provide unique objective information for programming process with audiologists. However, calculating the electric field (EF) generated by each CI electrode is a bottle-neck in creating EAMs, which requires days of computation time and consumes substantial computational

resources.

In our current system, the EP is calculated by solving the finite difference method (FDM) solution to Poisson's equation for electrostatics, which is given by

$$\nabla \cdot J = -\sigma \nabla^2 \Phi \tag{4.1}$$

where Φ is the EP, J is the electric current density and σ is the conductivity. Such system can be solved as an optimization problem using the biconjugate gradient method with an average time of 135 seconds on a CPU for a volume of 251×251×151. The EPs of different electrical characteristics (i.e. the conductivities for each tissue type), and different current sources (different activated electrodes) are calculated to find the conductivity set that best fits the clinical measurements of EFI. EFI is obtained by activating one CI electrode at a time (probe electrode) while measuring the voltage at each of the remaining electrodes in the cochlea, thus sampling the intra-cochlear potentials at the sites where the electrodes sit, which are sensitive to patient-specific tissue conductivities. Thus, voltage values are sampled from each simulated EP at the electrodes' positions and compared to corresponding measurements to find patient-specific tissue conductivities. Since over 1000 EP maps are needed to be simulated for each CI electrode during the searching process to find optimal conductivities for each tissue type Cakir et al. (2017a) and a good initialization is required for each simulation, 1200 to 2400 hours of computation time is needed in the whole process of creating an EAM for a single patient. In this work, we propose a more efficient method to estimate the intra-cochlea voltage distribution based on 3D U-Net-like architecture taking advantage of GPU resources.

4.2 Method

In this section, we will start with an overview of the task followed by details of our dataset and network architecture. The goal of this neural network is to predict the electric potential of a cochlea that is electrically stimulated by CI electrodes. Because the intra-cochlea tissue was assumed to be electrically linear and the impedances of all the tissue types were assumed to be purely resistive, the activated CI electrode is modeled as a unit electric current source as the solution to such can be linearly scaled to simulate varying current levels. The input to our neural network includes a 3D resistivity map of the cochlea and an inverse distance map of the current source (see Figure 4.1). The network is trained to predict an electric potential map with the same size as the input when the current source injects 1 Amp of electrical current.

4.2.1 Data

Three types of images were generated to construct the dataset: the cochlea resistivity maps, the inverse distance map of the current source, and the corresponding electric potential map. To do this, 8 label maps were generated from 8 patient CT images as described in Cakir et al. (2017a). These label maps contain

different labels corresponding to different tissue types including air, electrolytic fluid, soft tissue, neural tissue, and bone. Electrode positions were also localized and encoded in these label maps following the approach proposed in Zhao et al. (2019). The original size of label maps is 251×251×151, and they were downsampled to 96×96×56 to reduce the computational cost. Data augmentation is performed by selecting electrical resistivity values for each tissue type in the range from 50% to 150% of their default values found by other groups, with air, bone, neural tissue, soft tissue, and electrolytic fluid being assigned resistivity values of ∞, 5000, 300, 300, and 50 ΩcmGeddes and Baker (1967), respectively. The air is considered fully insulating to ensure numerical stability in solving the system. We also simulate the far-field ground by modeling the entire border of the volume as ground. Because the number of electrodes on the electrode array is 16 for our 8 subjects, 16 inverse distance maps of current sources are created for every resistivity map. We generated 2000 pairs of resistivity maps and current source distance maps from 5 patients' label maps by varying resistivity values to serve as a training dataset. The corresponding electric potential maps were then calculated for each resistivity map and for each activated electrode using the EAMs. We then generated 1200 sets of images in the same way from another 3 patients as the testing set (see Figure 4.1).

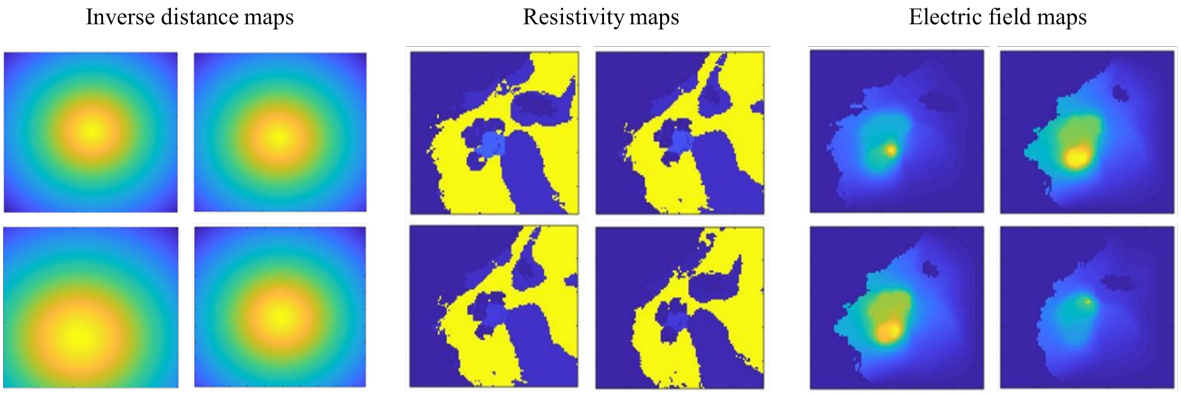


Figure 4.1: Inverse distance maps of the current sources, intra-cochlea resistivity maps, and their corresponding electric potential maps.

4.2.2 Network architecture

Our network is based on the U-net architecture proposed by Ronneberger et al. Ronneberger et al. (2015). As illustrated in Figure 7, each step in the contracting path (left side) consists of two 3D convolutions, each followed by a batch normalization, a rectified linear unit (ReLU), and a 3D max pooling. And in the expansive path (right side), each step consists of a concatenation with the feature map from the contracting path, three 3D convolutional layers, each followed by a ReLU and an up-sampling layer implemented as a transposed convolution operator from Pytorch Paszke et al. (2019). Note that the convolutional layers on the left side are

slightly different from those on the right side since convolutions in the contracting path do not include the addition of a learnable bias which is already included in batch normalization layers.

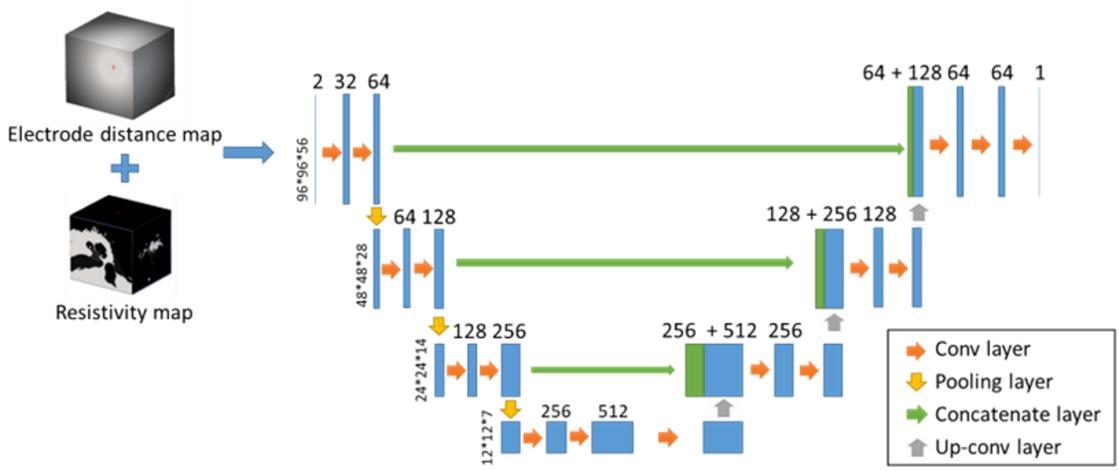


Figure 4.2: Network Architecture. Each blue box corresponds to a multi-channel feature map.

4.2.3 Training

The resistivity map – current source distance map pairs and their corresponding electric potential maps were used to train the network with the Adam optimization method Kingma and Ba (2014) implementation of Pytorch. The input size is 2×96×96×56 and the output volume is 96×96×56. We used a batch size of 1 as recommended in [49] and a learning rate of 0.0001 during the training process. Because the aim of our neural network is to find out the solution to a physical model (as described in equation 1), which is different from other traditional image processing tasks, we propose to use an energy function that encourages the predictions to obey Kirchhoff's current law, which states that for any node in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node. To construct such a loss function, we decided to use a similar method described in Cakir et al. (2017b). In our system, each voxel represents a node that has 6 neighbors in the circuit, and the algebraic sum of currents flowing out of each non-ground voxel should be zero unless that voxel is defined as the current source, which is given by

$$\sum_{i=1}^{6} I_{neib_i} = I_s \tag{4.2}$$

$$I_{neib_i} = \frac{1}{\rho + \rho_{neib_i}} \left(\frac{\Phi_{neib_i} - \Phi}{\Delta x}\right) \Delta y \Delta z \tag{4.3}$$

$$\alpha_0\Phi + \alpha_1\Phi_{neib_1} + \alpha_2\Phi_{neib_2} + \alpha_3\Phi_{neib_3} + \alpha_4\Phi_{neib_4} + \alpha_5\Phi_{neib_5} + \alpha_6\Phi_{neib_6} = I_s \tag{4.4}$$

$$I_s = \begin{cases} 0, & \text{current source} \\ 1, & \text{other non-ground voxels} \end{cases} \tag{4.5}$$

where I_{neib_i} represents the current flow from its ith neighbor and I_s equals 0 if the voxel is not a current source and 1 if it is. And according to Ohm's law, the current flow from one voxel to another can be illustrated by equation (4.3) where ρ,Φ, ρ_{neib_i} and Φ_{neib_i} represent the resistivity and voltage of the central node and its ith neighbor's, respectively. Δx, Δy and Δz are the voxel size in centimeters. By substituting equation (4.3) into (4.2), we can get (4.4) where α_i can be calculated using ρ, ρ_{neib_i} and the voxel size. As shown in equation (4), I_s can be represented as a linear combination of the voltage of the central node and its neighbors, and this is true for every voxel in our system. Thus, we are able to construct a system of linear equations for all the voxels which is given by

$$A\Phi = b \tag{4.6}$$

where A is a coefficient matrix consisting of α_{ij}, Φ is the electric potential map, and b is a column vector whose elements equals 1 when its corresponding voxel is the current source and 0 otherwise. A and b are then modified by adding the following boundary conditions:

(1) Air nodes are forced to have zero values in Φ as they have infinite resistivity;

(2) If a non-air node is adjacent to at least one air nodes, Neumann boundary conditions are implemented;

(3) Rows representing ground nodes are deleted from A and b so that sinks are not constrained by Poisson's equation. The overall energy function we used during our training is a weighted sum of rooted mean squared error (RMSE) and our customized loss function as described below:

$$Loss = w1 * sqrt\left(mean\left[(\Phi - \Phi_{GT})^2\right]\right) + w2 * mean[(A\Phi - b)^2] \tag{4.7}$$

where Φ and Φ_{GT} are the predicted electric potential maps and ground truth respectively. Since the ground truth does not appear in the customized loss, our loss function is a hybrid of a traditional supervised loss term and a self-supervised term which benefit the estimation of a real physical process as described in the next section.

4.3 Experiments and results

Table 4.1 shows the root-mean-squared error (RMSE) and the physics-based loss we obtained from models trained with different weights in our loss function at their respective best epochs. As shown in the table, a higher weight for the loss derived from Kirchhoff's current law helps the model have better performance on both RMSE and the physics-based loss in a certain range. However, our experiments also indicate that it is a trade-off as putting a weight on physics-based loss (w2) that is larger than 1000 or using the physics-based

loss only will lead to reduced precision in terms of RMSE and even divergence of the training process. To achieve better performance of electric potential estimation, tuning of these weights is needed. And we found that our network works best when w1=1 and w2=500. Figure 4.3. shows examples of the electric potential prediction by the model trained with optimal weights together with their corresponding resistivity maps and ground truth.

Table 4.1: Testing errors of models trained with different weights in loss function

w1, w2	RMSE (μV)	$mean[(A\Phi - b)^2]$ (μA^2)
1, 0	25.65	2.13
1, 100	19.12	1.74E-02
1, 500	17.98	2.79E-03
1, 1e3	24.05	1.94E-03
1, 2e3	29.31	9.22E-04
0, 1	Diverge	-

Table 4.2: MSE of EFI simulation (mV^2)

Subject No.	Deep learning method	Physics-based method
1	1.31	0.81
2	25.36	1.12
3	3.92	0.85

The goal of electric potential estimation is to help build EAMs where EFIs are used to select the optimal electrical resistivity values as we introduced previously. Thus, we demonstrate to estimate the EFI by sampling the electric potential predicted by the neural network at the electrode positions. Those estimated EFIs can be evaluated by comparing them with patients' clinically measured EFIs and EFI simulations produced by our traditional physics-based method. Figure 4.4 shows the EFI of the same patient at the same electrodes with three different methods. And in Table 4.2, we calculated the mean squared error (MSE) between EFI simulation from the deep learning method and clinical measurements. We also do the same calculation for our physics-based method for comparison. These results show that the neural network achieves similar accuracy

to the physics-based model both qualitatively and quantitatively.

The executing time of our network is 1.371 seconds for a single input, which is 100 times faster than the physics-based model. This largely speeds up the process of calculating electric potential and thus the process of creating EAMs that helps to improve cochlea implant performance.

4.4 Conclusion

In this work, we have used our 3D neural networks to predict the intra-cochlea electric potential generated by CI electrodes. We proposed to add a physics-based self-supervised loss term in the loss function to penalize the prediction results when Kirchhoff's current law is violated. By using proper weights in our loss function, the accuracy of the model can also be improved in terms of RMSE. The results we have obtained show that our network is able to predict electric potential maps as well as EFI with similar accuracy to our physics-based method. The proposed neural network architecture largely speeds up the process of calculating electric potential and thus the process of creating EAMs that helps to improve cochlea implant performance. To the best of our knowledge, this is the first time that a 3D Unet-like architecture is used for intra-cochlea electrical field estimation.

Our future work includes improving the accuracy by enlarging our dataset and performing augmentation, evaluating the sensitivity of the results, and expanding the architecture to a model that predicts optimal resistivity values.

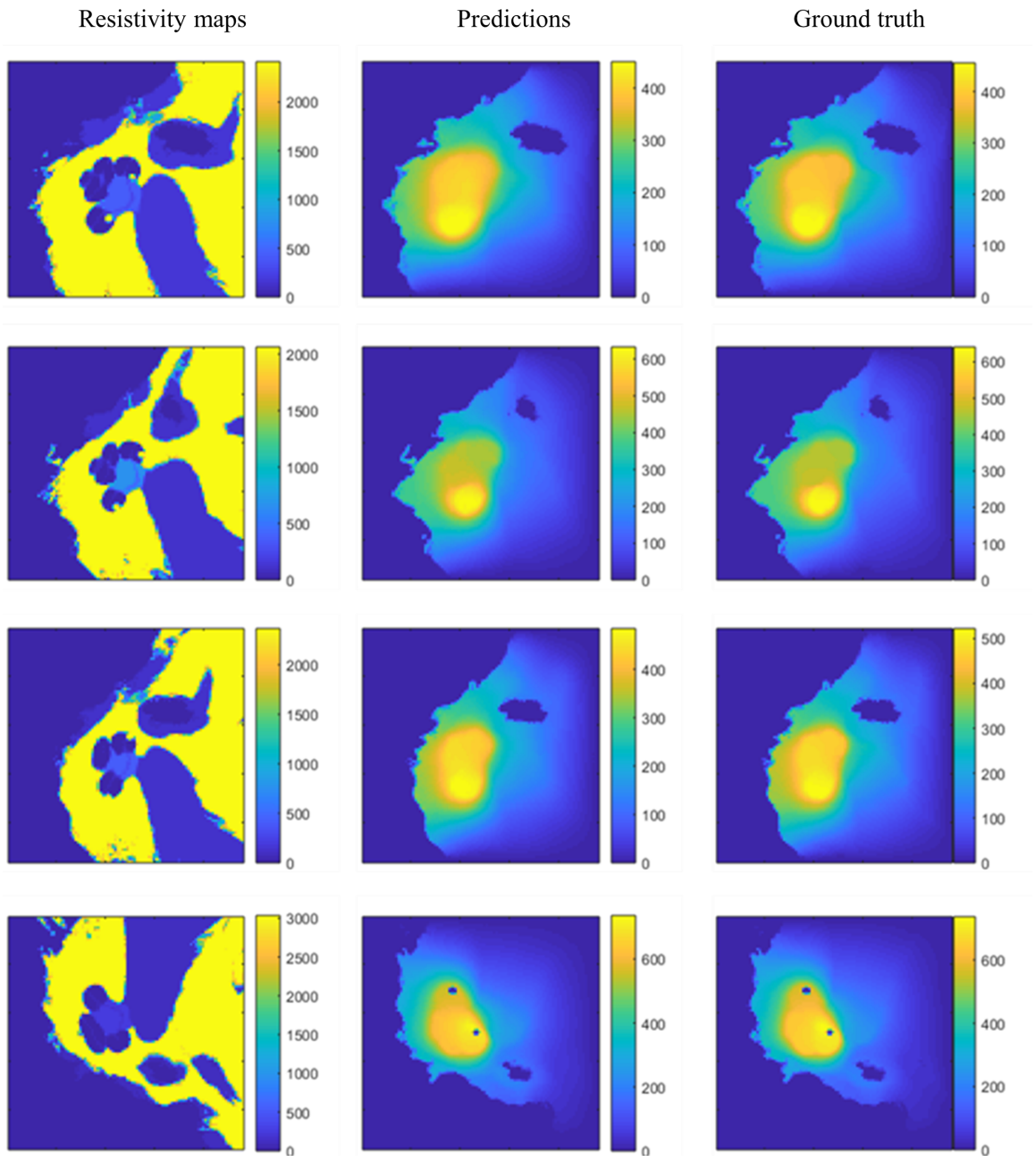


Figure 4.3: Electric field prediction results by models trained with optimal hyperparameters compared with ground truth.

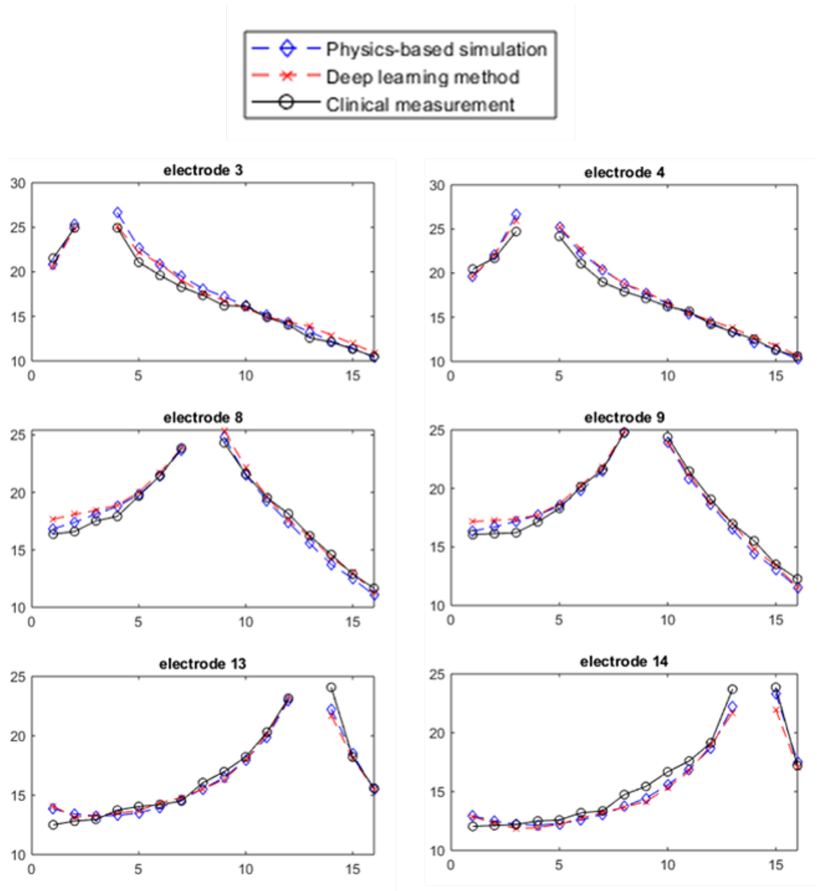


Figure 4.4: EFI simulation results by proposed method and physics-based method compared with clinical measurements.

CHAPTER 5

Patient-specific Electro-anatomical Modeling of Cochlear Implants Using Deep Neural Networks

5.1 Introduction

Cochlear implants (CIs) are considered the standard-of-care treatment for profound sensory-based hearing loss NIDCD (2019). In CI surgery, an electrode array is implanted into the inner ear to permit direct electrical stimulation of the auditory nerve fibers (ANFs). Hearing outcomes with CIs depend heavily on the electro-neural interface, which is a function of several patient-specific factors, including the anatomical structure of the patient's cochlea, the health status of ANFs, the placement of the CI array, and the patient-specific tissue resistivity characteristics. An audiologist will adjust the CI processor settings for each CI recipient after surgery to attempt to improve overall hearing performance. However, because the factors that affect the electro-neural interface are traditionally not known by the audiologist, a trial-and-error procedure is implemented for programming. Further, since the patients need weeks of experience with given settings before hearing performance stabilizes due to learning effects, this programming procedure can be frustratingly long and generally lead to suboptimal outcomes.

Several groups have studied the intra-cochlear voltage distribution and its effect on hearing outcomes using different method. Frijns et al. Frijns et al. (1995) used a rotationally symmetric model of a guinea pig cochlea and human cochlea to solve for the voltage distribution using boundary element method. Whiten et al. Whiten (2007) create human cochlea model from histological images using finite difference method (FDM). Although these models have been shown to be useful, they cannot be customized to show the patient-specific differences in volume, shape, number of turns and length etc Avci et al. (2014). Malherbe et al. Malherbe et al. (2016) rigidly register a high-resolution photomicrograph of a single cochlea to patient CT image in order to estimate these anatomical differences. However, we have shown that a rigid registration is less accurate than a high-resolution non-rigid model to estimate patient-specific anatomy Cakir et al. (2017b).

Our group has developed image-guided CI programming (IGCIP) techniques Noble et al. (2013), where CT-based electrode localization is performed to coarsely estimate the electro-neural interface by simply estimating the distance from each electrode to the sites where the nerves should be. Audiologists can infer the electro-neural interface using this information by assuming that electrodes that are more distant to the nerves create broader, more highly overlapping excitation patterns. This permits providing simple, yet unique, electro-neural interface information to audiologists for use in adjusting patient-specific processor settings and leads to significant improvement in hearing outcomes as shown in several studies (e.g., Noble et al. (2015,

2016)). Extending this work, to attempt to perform more accurate estimation of neural activation, we have proposed a system in which ANF bundles can be directly modeled Cakir et al. (2019), and their stimulation by the implanted CI can be simulated on a patient-specific basis Cakir et al. (2017a). With these models, it is possible to simulate numerous electrical measurements that can be acquired from the patient's CI. Further, by parameterizing these models with patient-specific tissue resistivity parameters as well as patient-specific ANF health, we can estimate these critical electro-neural interface parameters for individual patients by optimizing them to obtain a better agreement between measurements from the CI and the model simulation estimations. Thus, this direct modeling approach permits a far more comprehensive estimation of the electro-neural interface than was possible using simple electrode distances and could lead to further improvement in hearing outcomes when used by the audiologist for programming Cakir et al. (2017b).

In our current system, the ANF activation model, which is constructed following the biological auditory nerve model proposed by Rattay et al. Rattay et al. (2001b,a), is driven by estimates of the electric potentials (EPs) along the length of the ANF that result from activation of a CI electrode [17]. The EPs for each electrode are estimated by solving the FDM solution to Poisson's equation for electrostatics. The inputs to the FDM are the tissue class map of the cochlea and surrounding tissues, the location of the active CI electrode current source within the map, and the resistivity values of the tissue classes. The FDM outputs a resulting EP map, which can be sampled at the ANF locations to drive the ANF activation models (Figure 5.1). The EP map can also be sampled at the locations of the inactive electrodes, and this measurement (termed "Electric Field Imaging" (EFI) by one CI manufacturer) can similarly be acquired from the patient's CI. EFI captures the electric potential at inactive electrodes when one electrode is activated and is sensitive to tissue resistivity. Thus, it provides a means to optimize the estimated resistivity set for each patient. To do this, we have found a grid search to be effective, where tissue resistivity parameters for air, bone, soft tissue, electrolytic fluid, and neural tissue are searched on a grid. For each set of unique tissue resistivity values, the FDM is used to estimate an EP map, and the optimal set is determined as the one that results in minimum mean squared error between actual EFI and simulated EFI sampled from the resulting EP map. Once the EP map that results in simulated EFI that best fits clinical measurements is found, we can use it to drive the ANF models.

While effective in finding plausible tissue resistivity parameters, the grid search process requires substantial computational resources. Each FDM evaluation requires an average time of 135 seconds on a CPU for a volume of 251 x 251 x151 voxels if a reasonable initialization is used. Since over 2000 FDM evaluations are needed to be simulated for each CI electrode, and there are 16-22 electrodes, 1200 to 2400 hours of computation time are needed in the whole process of creating an EP map for a single patient with a search grid fine enough to achieve mean absolute errors in EFI fitting of 0.3mV.

In this study, our hypothesis is that deep learning methods can alleviate or reduce the need to use FDM

and speed up the generation of the EP maps Liu and Noble (2021). We propose a method that predicts patient-specific resistivity values and EP maps at the same time based on cycle-consistent adversarial networks.

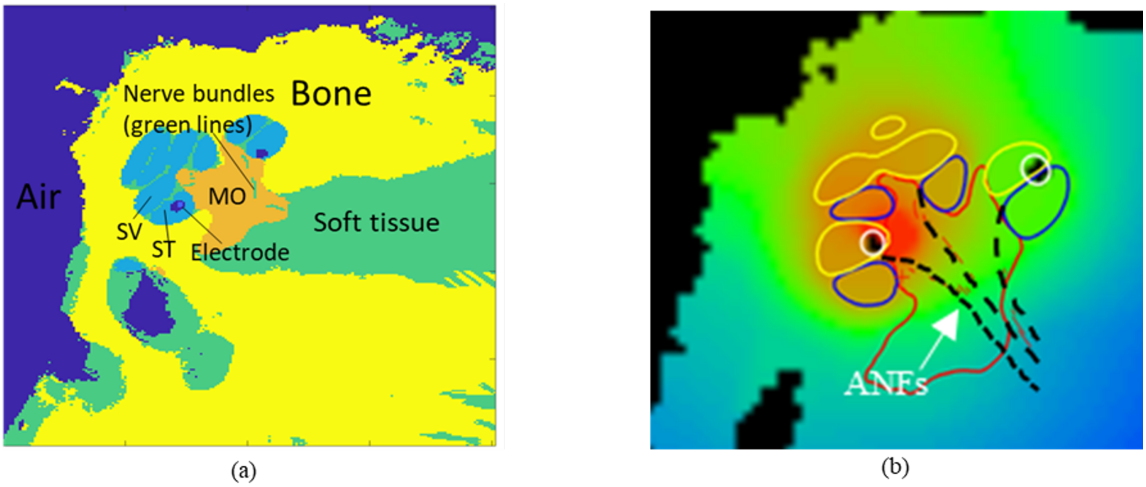


Figure 5.1: A slice of tissue label map (a): scala vestibuli (SV) and scala tympani (ST) are electrolytic fluid (blue), modiolus (orange), soft tissue (green), CI electrode (dark blue circles), bone (yellow) and air (dark blue area); And electrical potential map (b): SV (yellow contour), ST (blue contour), modiolus (red contour), auditory nerve fibers (black dashed- lines), and electrodes (black circles).

5.2 Method

The primary goal of our neural network is to estimate the patient-specific resistivity set. The inputs to the network are the target EFI, the tissue class map of the same patient, and the electrode locations. We employ a multitask learning approach, where in addition to the resistivity set, a secondary goal of the network is to estimate the EP map. Our results demonstrate the multitask approach results in more accurate resistivity estimates. Due to the limitation of GPU memory, the input volumes to our network are all downsampled to 101 x 101 x 61 voxels.

5.2.1 Network architecture

Our network is built based on the cycle-consistent adversarial networks (Cycle GAN) Zhu et al. (2017) as shown in Figure 11. Cycle GAN has been widely used in various image-to-image translation tasks with unpaired datasets. In this work, we propose a novel variant of it. Our network consists of two sub-networks: sub-network A predicts the patient-specific resistivity values using the tissue label map of the inner ear, a distance map of the current source (an activated electrode), and the actual EFI measured from the patient; sub-network B estimates the corresponding EP map given the tissue resistivity map and the same distance map of the current source as used in A. This architecture is cycle consistent because the tissue resistivity

map, which is one of the inputs to part B, can be obtained by assigning predicted resistivity values from sub-network A to corresponding tissue labels available in the tissue label map. And the EFI, which is one of the inputs to A, can also be estimated by sampling potentials from the predicted EP map from sub-network B (as shown in Figure 5.2). The architectures of those two sub-networks mainly consist of a 3D version of resNet-18 He et al. (2016) and a 3D Wasserstein GAN (WGAN) Arjovsky et al. (2017) respectively, which are depicted in Figure 5.3(a) and Figure 5.3(b). The Wasserstein GAN uses a U-net-like generator based on the architecture proposed by Ronneberger et al. Ronneberger et al. (2015) and a critic composed of 4 convolutional layers followed by a linear layer. Our implementation is different from the original Cycle GAN implementation as follows: (1) Sub-network A is not an adversarial network. (2) Sub-network B is a WGAN-GP Gulrajani et al. (2017) instead of conditional GAN because we found it to be more effective in pilot testing. (3) We use paired datasets during training. In section 3, we will compare the performance of our architecture to an individually trained sub-network A and a multi-task U-net-like network, which are more commonly used with paired datasets, to illustrate the advantage of our cycle-consistent architecture.

5.2.2 Dataset

Four types of data were generated to construct the dataset: the electrical resistivity set, the tissue label map, the distance map of the current source, and the corresponding EP map. The resistivity set contains the resistivity values of different tissue types including electrolytic fluid, soft tissue, neural tissue, and bone. We generate the resistivity set by randomly selecting resistivity values for each tissue type in the range from 50% to 150% of their default values found by other groups, with bone, neural tissue, soft tissue, and electrolytic fluid being assigned resistivity values of 5000, 600, 300, and 50 Ωcm Geddes and Baker (1967), respectively. The tissue label maps have labels for the four tissue types above as well as a label for air, which is considered fully insulating during the computation of corresponding EP to improve numerical stability. Label maps were generated from 20 patient CT images using active shape models generated from μCT specimens to follow the method described in Cakir et al. (2017a). The distance maps compute the normalized Euclidean distance to the current source – the activated electrode on the CI array. The activated CI electrode is modeled as a unit electric current source (1 μAmp) as the solution to such can be linearly scaled to simulate varying current levels. Because the number of electrodes on the electrode array is 16 for all our 20 subjects, 16 distance maps of current sources are created for every label map. By assigning a simulated resistivity set to a patient label map and selecting one electrode as the current source, a corresponding EP map is calculated using FDM following the method in Cakir et al. (2017a). In our training set, 10 patient label maps are used, and a total of 100 resistivity sets and corresponding EP maps are simulated for each of the 16 electrodes serving as the current source. Then, all the resulting EP maps are sampled at the site of the CI electrode to obtain the

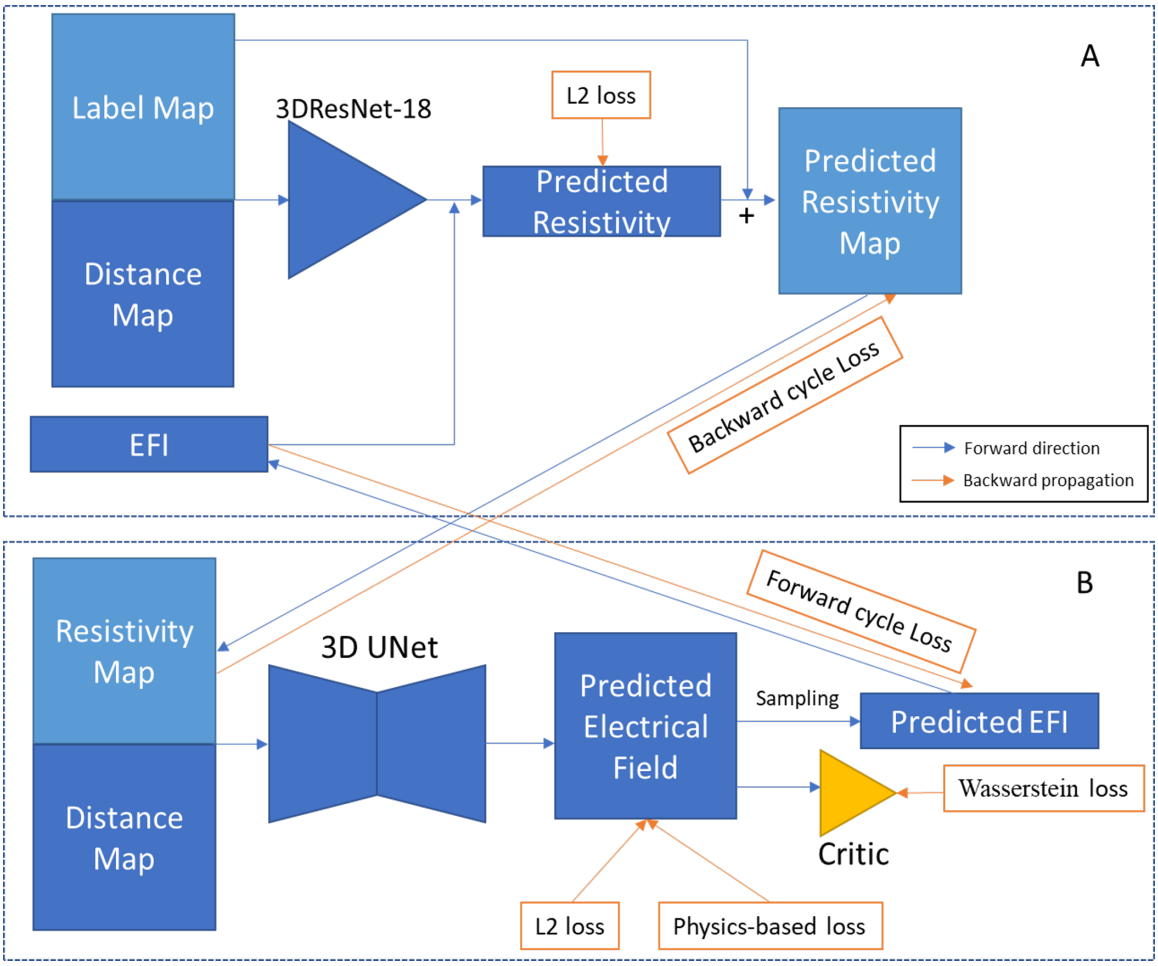


Figure 5.2: Proposed network architecture: blue lines show the forward direction and red lines show the back propagation direction labeled with different loss terms.

simulated EFIs, which consist of a 16-element vector for each current source. Two patient label maps and 320 resistivity sets are used to generate the validation set in a similar manner. For the testing set, 8 patient label maps were used, and 10 simulated sets of data were constructed for each patient. During testing, we run our network to estimate the resistivity set for each of the 16 electrodes and determine the final estimate to be the mean across the 16 runs. While our multitask network also estimates an EP map, we find the potentials in the estimated EP map are unstable around electrode sites, which makes the sampled voltage values not accurate enough to simulate EFI directly. Thus, in our approach to simulate EFI, we sample EP maps produced by FDMs that use the estimated resistivity set from our network, rather than sampling the network-estimated EP maps.

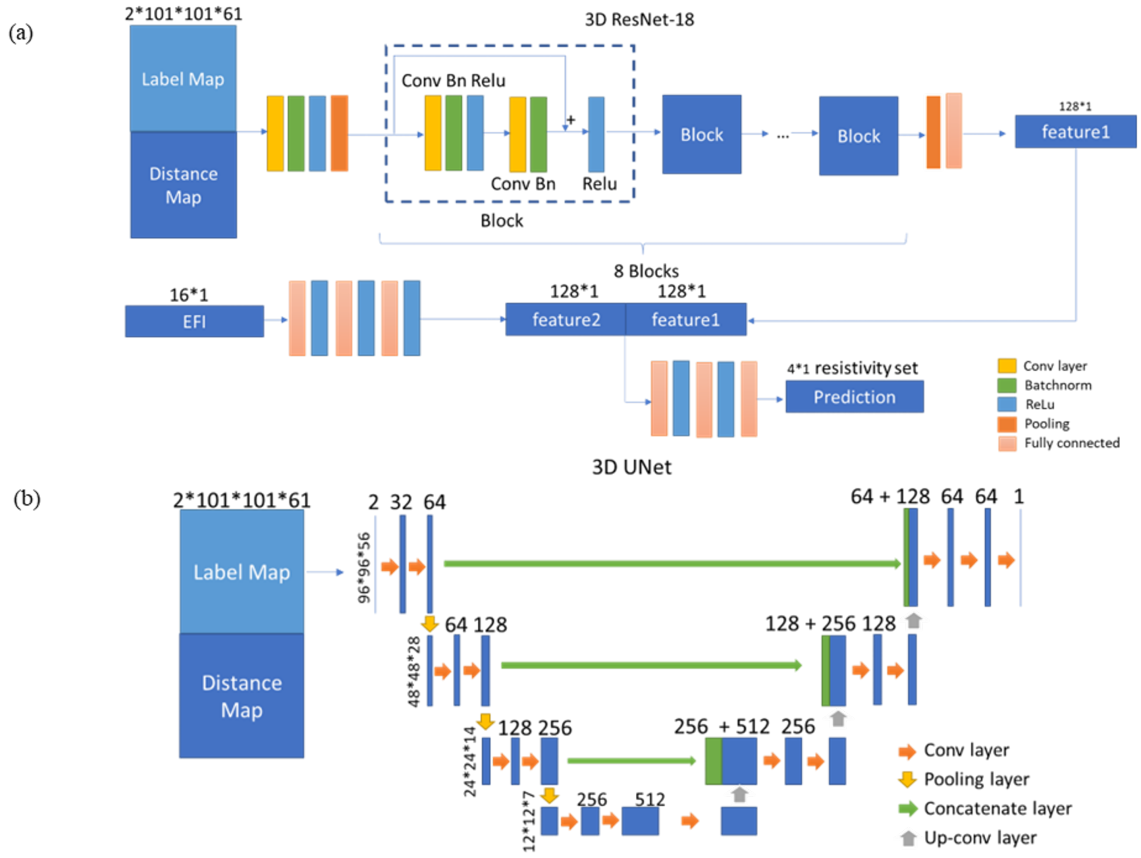


Figure 5.3: The sub-structures in the proposed network: (a) shows the implementation of 3D ResNet-18 used in sub-network A. (b) shows the 3D U-net used as the generator in the sub-network B.

5.2.3 Training

We first consider a forward cycle in our network. Let EFI denote the input EFI data, which is a 16x1 float vector in our case, and $res = f_A(labelMap, distMap, EFI)$ denotes the predicted resistivity set, where f_A is the

sub-network A, *labelMap* and *distMap* are input tissue label map and distance map of the current source respectively. We can use *res* to generate a fake resistivity map $\widehat{resMap_{fake}}$ by assigning predicted resistivity values to *labelMap* and use the fake resistivity map as input to sub-network B: $\widehat{EP} = f_B(\widehat{resMap_{fake}}, distMap)$, where $\widehat{EP}$ is the reconstructed electrical potential map using the fake resistivity map.

The forward cycle loss is:

$$Lossforward_{cyc} = |EFI - \widehat{EFI}|^2 \tag{5.1}$$

where $\widehat{EFI}$ is the EFI produced by sampling from $\widehat{EP}$ at CI electrode positions. Similarly, in a backward cycle, we will have $EP = f_B(resMap, distMap)$, where EP is the prediction of sub-network B with a real resistivity Map denoted *resMap*. We can sample EF in the same way and obtain the fake EFI: $\widehat{EFI_{fake}}$. Then a resistivity set and resistivity map can be estimated by sub-network A:

$$\widehat{res} = f_A(labelMap, distMap, \widehat{EFI_{fake}}) \rightarrow \widehat{resMap} \tag{5.2}$$

We define the backward cycle loss by

$$Lossbackward_cyc = |resMap - \widehat{resMap}|^2 \tag{5.3}$$

Besides the cycle-consistent loss, an L2 loss is also added for both sub-network as we use a paired dataset that has ground truths:

$$L2_{res} = |res - res_{GT}|^2 \tag{5.4}$$

$$L2_{EP} = |EP - EP_{GT}|^2 \tag{5.5}$$

where res_{GT} and EP_{GT} are target values of the resistivity set and electrical potential map. Moreover, we also introduced the same self-supervised physics-based loss function $Loss_{Phy}$, which penalizes estimated potential maps that do not preserve Kirchhoff's current law as described in Liu and Noble (2021), and the Wasserstein loss LossW as described in [62] to our network. In summary, the loss for the generators is defined by

$$Loss = w_1 * Lossforward_{cyc} + w_2 * Lossbackward_{cyc} + w_3 * L2_{res} + w_4 * L2_{EF} + w_5 * Loss_{Phy} + w_6 * Loss_W \tag{5.6}$$

where the weights were 0.5, 0.5, 0.5, 1, 0.002, and 1 respectively. We use the Adam optimizer with a learning rate of 0.0001 and default coefficients used for computing running averages of gradient and its square (0.9, 0.999) in its PyTorch implementation. As for hyperparameters of the critic, which provides the adversarial loss to WGAN-GP, we use the same setting as described by Gulrajani et al. (2017) where the

gradient penalty coefficient $\lambda = 10$ and the number of critic iterations per generator iteration $n_{critic} = 5$. In the next section, we will compare the cycle-consistent model to individually trained sub-network A and a multi-task U-net. These two models are trained with $Loss_A = L2_{res}$ and $Loss_{Unet} = w_4 * L2_{EF} + w_5 * Loss_{Phy} + w_6 * Loss_W$ respectively using Adam with the same parameters. And all the models are trained with a batch size of 1 for 800 epochs. The following experiments are conducted using the epoch with minimum validation error for each model.

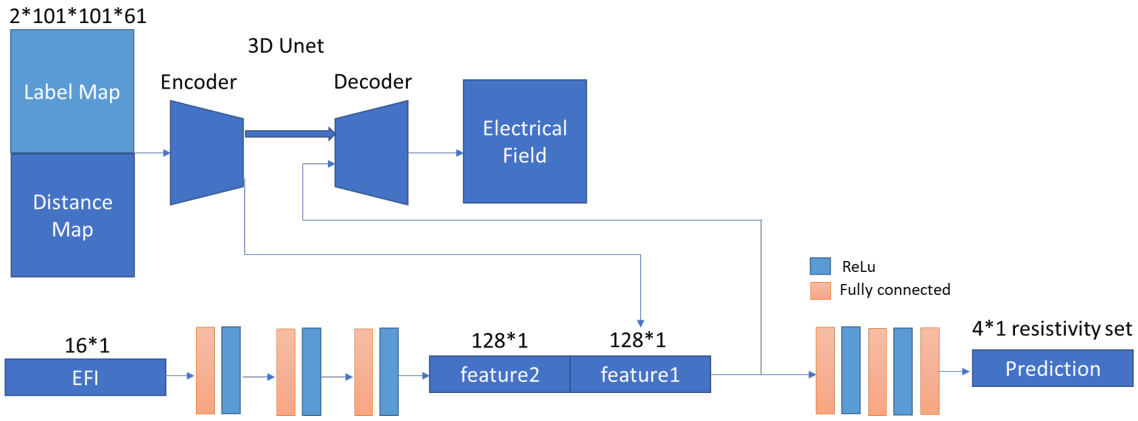


Figure 5.4: A multi-task 3D U-net architecture: the encoder and decoder are the same as shown in Figure 3.3(b).

5.3 Experiments and results

We conducted three sets of experiments. The first set of experiments compares the performance of resistivity set prediction of our cycle-consistent network to an individually trained 3D ResNet-18 (as described in Figure 5.3(a)) and a multi-task 3D U-Net (as described in Figure 5.3(b)). To evaluate the self-supervised physics-based loss term $Loss_{Phy}$ as discussed in the previous section, w_5, the weight of that term is set to 0.002 and 0 for Cycle-Net1 and Cycle-Net2 respectively. We also compared our cycle-consistent network to the multi-task network in terms of EP map estimation. As shown in Table 5.1, the cycle-consistent network outperformed the other two architectures in resistivity set prediction and has a similar accuracy in EP prediction compared to multi-task U-Net. In our second set of experiments, we generate EFIs using FDM with the resistivity values estimated by the cycle-consistent network. We calculate the mean absolute error (MAE) of the reconstructed EFIs. As is shown in Figure 5.5, The average MAE of all testing cases is 0.51 mV. Figure 5.6 shows examples of the reconstructed EFI that has an MAE at 0.48 mV. This allows us to visualize how the error in resistivity value may affect our patient-specific EAMs quantitatively. The result shows that the predicted resistivity set can already produce a reasonably similar EFI simulation compared to the target. The network is able to achieve this level of accuracy using only 0.84 seconds of computation for the inference.

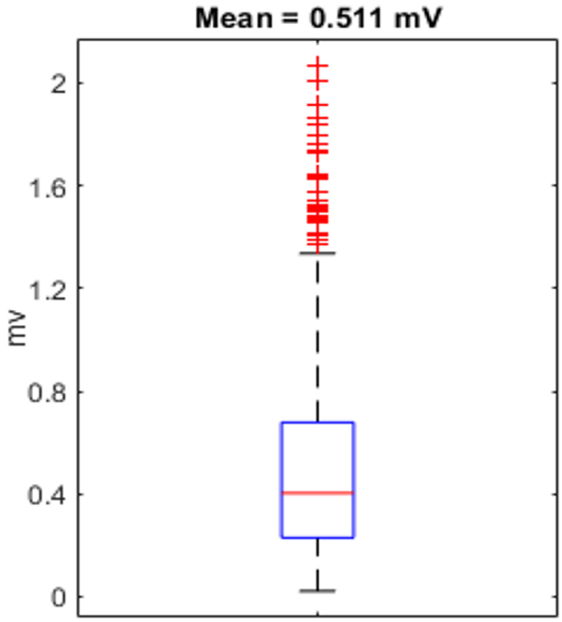


Figure 5.5: Mean absolute difference (MAE) between the reconstructed EFIs and ground truths.

Table 5.1: Testing errors of different architectures. The mean absolute errors (MAE) of four tissue types are displayed in the order of electrolytic fluid, soft tissue, neural tissue, and bone.

Architecture	MAE of resistivity set values (Ωcm)	MAE of resistivity values for each tissue type (Ωcm)	MAE of EP (mV)
Cycle-Net1	157.5	9, 102,99, 420	0.202
Cycle-Net2	168.9	8, 118, 112, 438	0.255
Sub-Net A	184.4	7, 108, 103, 520	NA
Multi-task Net	167.1	10, 102, 103, 453	0.190

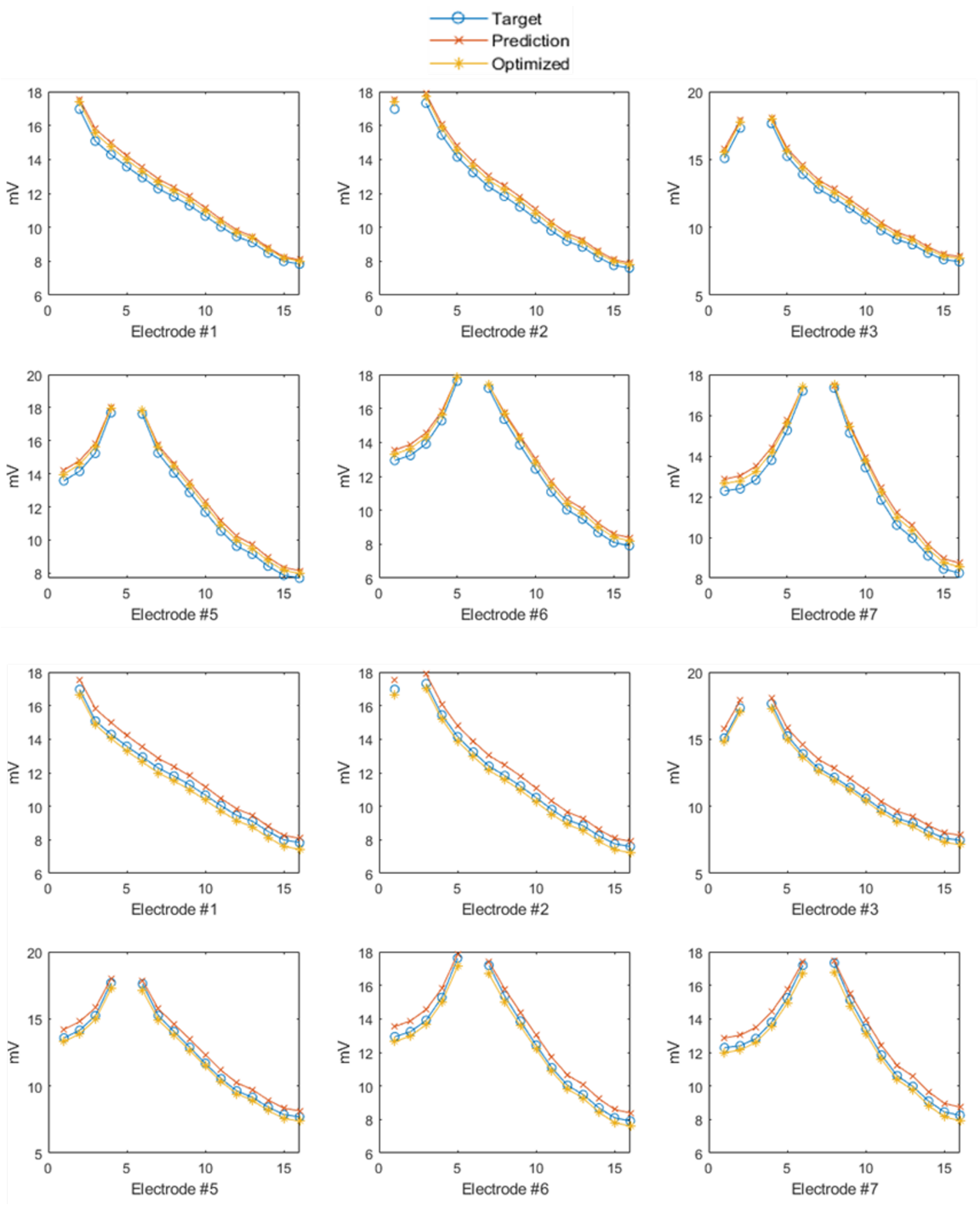


Figure 5.6: An example of the target EFI (in blue), reconstructed EFI using predicted resistivity values (in red), and reconstructed EFI using optimized resistivity set generated by a searching strategy starts from network prediction (in yellow). The MAE between red lines and blue lines is 0.48. Yellow lines in (a) show the result of an optimized resistivity set using the Nelder-Mead method, which lead to an MAE of 0.29 compared to the target. Yellow lines in (b) show the optimized resistivity set given by grid search, which leads to an MAE of 0.11.

In our last set of experiments, we propose to use a searching strategy in which the resistivity values estimated by our deep learning method are used to initialize a refinement optimization step to further improve accuracy. Two strategies are evaluated for refining the resistivity values: grid search and the Nelder-Mead simplex algorithm [63] implemented in MATLAB. Using Nelder-Mead simplex, reduction in errors plateaus at 0.3mV at an average of 67 iterations, which is equal to about 40 hours of computation time. Using a grid search to reach the same level of error requires more computation time (an average of around 600 hours, which is less than half of the original searching time used in our traditional method), but grid search is able to lead to even better results (errors of 0.1mV) when the searching grid is finer, requiring an average of 2400 hours computation time. Example results are shown in Figure 3.6(b).

5.4 Conclusion

In this work, we aim to jump over the most computationally expensive process of building a patient-specific EAM – the grid searching by predicting patient-specific resistivity parameters directly using deep neural networks. We introduced a novel cycle-consistent network that predicts resistivity values for patient-specific cochlear implant neural activation models. Experiments show that our network can generate high-quality predictions that can largely improve the speed of constructing models.

We also showed that the predicted result can be further refined by applying traditional optimization methods. When cluster parallel computing is available, grid search can lead to an optimized resistivity set more rapidly (4 hours for coarse grid, 16 hours for fine grid using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University) because Nelder-Mead simplex algorithm cannot decide the next searching direction until finishing the calculation in the last iteration. However, the Nelder-Mead simplex algorithm uses fewer computation resources and is still able to achieve reasonable accuracy. Because the EP estimations of our network are overall good but still unstable around electrode sites, they cannot be used to simulate EFI.

So, our future work includes investigating multi-resolution grid search as a means to further reduce computation while still achieving accurate results, improving the performance of EP estimation in the regions around the electrodes, and applying this method for patient-customized programming in a clinical setting.

CHAPTER 6

Super-resolution segmentation for inner ear CT images

6.1 Introduction

Performing accurate patient-specific high-resolution segmentation for inner-ear tissue is essential for building electro-anatomical models (EAMs). Thanks to the rapid development of deep learning, fully convolutional neural networks (FCNNs) have been widely applied for processing and analyzing medical image modalities and have achieved state-of-the-art performance with leading network infrastructures, especially in the area of image semantic segmentation Hesamian et al. (2019); Ronneberger et al. (2015); Çiçek et al. (2016); Oktay et al. (2018); Abdollahi et al. (2020); Isensee et al. (2021); Yu and Koltun (2015); Zhao et al. (2017); Peng et al. (2017). Despite the success of these FCNN models, Transformer-based approaches have gradually dominated computer vision tasks in both medical and non-medical scenarios recently due to their capability of learning from a larger receptive field Dosovitskiy et al. (2020); Wang et al. (2021); Zheng et al. (2021); Liu et al. (2021); Wu et al. (2021). Several deep-learning-based cochlear anatomy segmentation methods have been proposed using CT images Zhang et al. (2019); Fan et al. (2020), but they mainly focus on the intra-cochlear anatomy instead of the tissue type surrounding the cochlea. Image super-resolution (SR) is another area that greatly benefits from the rapid development of deep neural networks. Deep-learning-based single image SR models Dong et al. (2014); Tai et al. (2017); Shocher et al. (2018); Iglesias et al. (2021); Zeiler et al. (2010) and learning-based upsampling methods have outperformed traditional machine learning methods and statistical methods in both medical and non-medical scenarios. However, SR methods for medical images usually need a post-segmentation, diagnosis, or classification algorithm to prove useful.

In this work, we propose the Super-Resolution Segmentation Network (SRSegN) for Computed Tomography images (CTs). In particular, we leverage the power of Transformer backbones and convolutional encoders for volumetric medical image segmentation while simultaneously performing multi-scale super-resolution on the segmentation. We introduced a novel architecture, Super-Resolution Transformer (SRTrans), that is tailored to extract and upsample features from volumetric data. The contributions of this work include:

1. We propose an end-to-end architecture that performs super-resolution segmentation for volumetric Computed Tomography images.

2. We propose a novel super-resolution transformer encoder together with a hierarchical convolutional encoder-decoder architecture that perform multi-scale image segmentation.

3. We evaluated the proposed architecture on our cochlear dataset and the AMOS public dataset. Our

model outperforms competing approaches on both datasets.

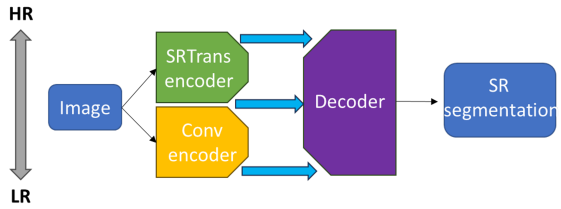


Figure 6.1: Overview of SRSegN.

6.2 Related Works

6.2.1 Semantic segmentaion

FCNN-based segmentation. Semantic segmentation identifies the class label of each pixel in the image. One of the most widely used and well-known architectures for medical image segmentation is the U-Net Ronneberger et al. (2015). This FCNN employs an elegant encoder-decoder design where the encoder extracts features using convolutional layers and downsamples them with pooling layers, and the decoder is symmetric to the encoder and upsamples the features with transposed convolution operators. By introducing the skip connection between the layers of equal resolution in the encoder and decoder, good localization and the use of context are possible at the same time. The success of U-Net has attracted a lot of attention in the area of image segmentation since 2015, and a number of variants such as 3D U-Net Çiçek et al. (2016), Attention U-Net Oktay et al. (2018) and VNet Abdollahi et al. (2020) were proposed to further improve the performance of U-Net. Compared to datasets in other scenarios, datasets in the medical domain are usually small, noisy, or lack diversity. Therefore, nnUNet Isensee et al. (2021) was proposed to provide a standard pipeline for training 2D and 3D UNet, which automatically configures the pre-processing, data augmentation, network architecture, and post-processing approaches based on the fingerprint of each dataset. The nnUNet has been tested on 23 public datasets of biomedical images and has outperformed most competitors since 2020. FCNNs preserve spatial information throughout the forward propagation because of the nature of convolution operators. However, the size of convolutional kernels also limits the receptive field of FCNNs, which further limits their performance in learning long-range dependencies for high-resolution images. Researchers have been focused on improving FCNN by enlarging the receptive field Yu and Koltun (2015); Zhao et al. (2017); Peng et al. (2017). However, these methods introduced empirical modules, making the resulting framework computationally demanding and complicated. Different from FCNNs, Transformer-based architectures may benefit from a larger receptive field and have recently gained traction for semantic segmentation tasks.

Transformer-based backbones. Transformers were first proposed for natural language processing Vaswani

et al. (2017). To make use of Transformers for computer vision tasks, Dosovitskiy et al. proposed Vison Transformer (ViT) in 2020 Dosovitskiy et al. (2020), where images are reshaped into a sequence of flattened 2D patches and then projected to an embedding space with additive positional information. By doing so, the global spatial information of images can be learned using the same multi-head self-attention (MSA) mechanism as text embeddings. ViT proved that a pure Transformer can achieve state-of-the-art performance on 2D image classification datasets. More recently, a number of methods were proposed that explored the potential of using Transformer backbones for 2D image segmentation. Wang et al. first introduced Pyramid Vision Transformer (PVT) Wang et al. (2021), demonstrating the potential of a pure Transformer backbone in dense prediction tasks. Beyond PVT, multiple Transformer models were proposed for the 2D image segmentation task Zheng et al. (2021); Liu et al. (2021); Xie et al. (2021a). SETR Zheng et al. (2021) directly adopts ViT as a backbone to extract features from images and achieves impressive performance with a simple decoder. Swin Liu et al. (2021) proposed a hierarchical Transformer whose representation is computed with Shifted windows in order to improve the efficiency of self-attention computation and enhance the connection among non-overlapping patch embeddings. Segformer Xie et al. (2021a) focuses on another type of hierarchical Transformer encoder, featuring overlapping patch merging and efficient self-attention.

Hybrid networks. Extending pure Transformer backbones, methods such as Cvt Wu et al. (2021) and Coat Xu et al. (2021) are proposed to combine the advantages of FCN and Transformer. In the domain of medical images, many attempts have been made to integrate Transformer encoders with the elegant architecture used by UNet. TransUNet Chen et al. (2021) replaces the last convolutional stage in UNet with a ViT encoder to improve the quality of low-resolution high-level features. CoTr Xie et al. (2021b) adds a Transformer-based encoder between the FCNN encoder and decoder, enhancing the ability of skip connections in traditional U-Net variants. UNETR Hatamizadeh et al. (2022) and Swin-UNETR Hatamizadeh et al. (2021) use tailored ViT Dosovitskiy et al. (2020) and Swin Liu et al. (2021) that capture multi-scale features to replace typical convolutional encoders and demonstrate state-of-the-art performance on medical image semantic segmentation datasets.

6.2.2 Image super-resolution

Super-resolution (SR) refers to the process of increasing the spatial resolution of images. Inherently, image SR is an ill-posed inverse problem since there are always multiple high resolution (HR) images corresponding to a single low-resolution (LR) image. In recent years, deep-learning-based SR models have achieved state-of-the-art performance on various benchmarks of SR Wang et al. (2020). In this section, we will mainly focus on supervised deep learning models for SR. Most of the SR models can be attributed to three classes according to how upsampling is performed: pre-upsampling framework, post-upsampling framework, and

progressive upsampling framework.

The pre-upsampling framework utilizes predefined interpolation methods to upscale LR images before refining them using deep neural networks. Dong et al. first adopted this framework and proposed SRCNN Dong et al. (2014) in 2014, which consists of only three convolutional layers but still outperformed other non-deep-learning SR approaches at that time. This straightforward framework has inspired many SR models Tai et al. (2017); Shocher et al. (2018) since then. Similarly, in the area of medical imaging SR, Iglesias et al. proposed SynthSR Iglesias et al. (2021) in 2021, which uses a combination of trilinear interpolation and U-Net to synthesize HR MRIs. Although pre-upsampling architectures enable neural networks to be designed independently from image sizes and scaling factors, computational inefficiency prevents the implementation of complex network architectures.

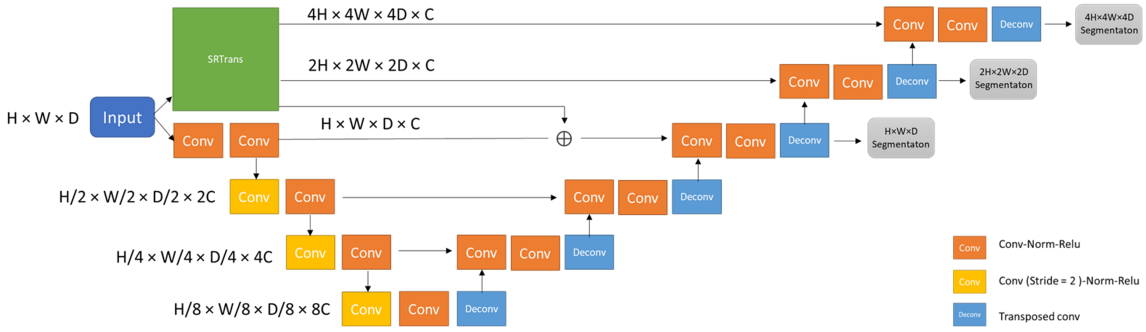


Figure 6.2: Architecture of SRSegN.

The post-upsampling framework makes use of upsampling algorithms to increase image resolution at the end of the network. These upsampling algorithms can be traditional interpolation methods or learning-based upsampling layers such as transposed convolution (deconvolution) Zeiler et al. (2010) or sub-pixel convolution Shi et al. (2016). Since the upscaling of image resolution only happens after the computationally expensive feature extraction process, multiple novel models were proposed without introducing much computational complexity, such as SRGAN Ledig et al. (2017), EDSR Lim et al. (2017) and DSRN Han et al. (2018). However, the major drawback of post-upsampling framework is that the upsampling is performed in only one step, greatly increasing the learning difficulty for scaling factors larger than 2. The progressive upsampling framework upscales the image to a higher resolution and refines the image using a similar network module at each step. Therefore, the learning difficulty can be greatly reduced by decomposing a difficult task into simple tasks, especially with large upsampling factors. Ahmad et al. adopted this framework and proposed a GAN-based SR model Ahmad et al. (2022), which achieved impressive results on four medical image datasets.

6.3 Method

An overview of the proposed architecture is shown in Figure. 6.1. SRSegN consists of three components: (1) A convolutional encoder that extracts features from an LR image; (2) Our novel and efficient Transformer-based encoder, SRTrans, which progressively upsamples images and extracts features to refine HR segmentations; and (3) A convolutional decoder that merges multi-resolution features from both encoders and performs dense predictions.

6.3.1 Convolutional Encoder

Although Vision Transformers enjoy a larger receptive field than FCNNs, they are unable to properly capture localized information. We propose to combine the advantages of both architectures by introducing a dual-path encoder. The convolutional encoder is used to extract features from the input LR images. We define the 3D input volume $x \in \mathbb{R}^{H \times W \times D}$ with a resolution (H, W, D). For each step of the encoder, the feature map will be downsampled by a convolutional layer whose stride is two, while the number of channels will be doubled.

6.3.2 Super-Resolution Transformer Encoder

Convolutional patch embedding. Traditional Transformer-based encoders create a 1D sequence of a 2D image by dividing the image into flattened, uniform, non-overlapping patches. Each sequentialization patch will be projected to a higher-dimension space, and a pre-calculated positional encoding will be added in order to preserve the spatial information of the original image. In SRTrans, we propose to perform the volumetric patch embedding using a convolutional operator as follows:

$$\hat{x} = Norm(Flatten\,(Conv_{embed}\,(x))) \tag{6.1}$$

$$x_v = Linear_{C \to \hat{C}}\,(\hat{x}) \tag{6.2}$$

where $Conv_{embed}\,(\cdot)$ refers to a 3D convolution whose kernel size, stride, and padding are 2+p, p, and 1, respectively. $Flatten(\cdot)$ refers to reshaping features of (Hp, Wp, Dp, C) into (Hp·Wp·Dp, C). And $Linear_{C \to \hat{C}}$ refers to the linear projection taking in a C-dimensional vector and generating a $\hat{C}$-dimensional vector. The resulting embedding is equivalent to splitting the image into non-overlapping patches of (p, p, p). By applying this convolutional patch embedding, we are able to replace manually designed positional encoding with learnable positional features. At the same time, each patch embedding now includes the information from nearby voxels due to the sliding convolution kernel, and adjacent patches overlap with each other because the kernel size is larger than the sliding stride. In general, convolutional patch embedding improves Transformer's ability to learn intra-patch features.

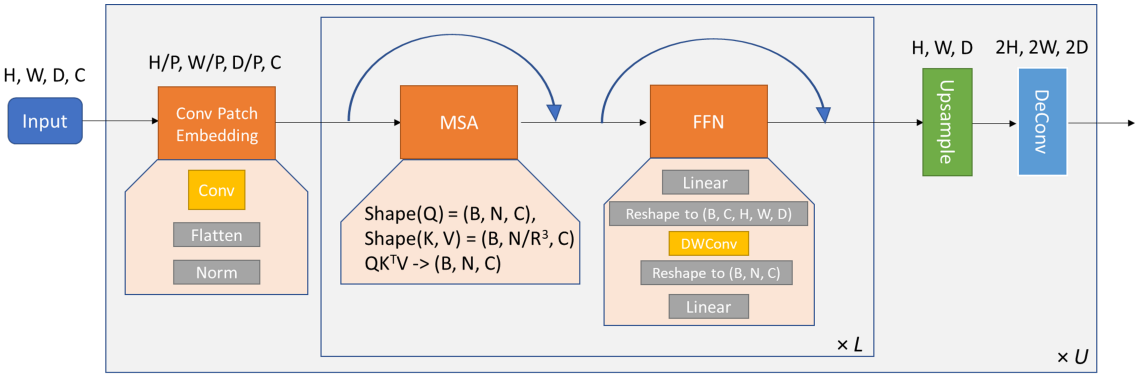


Figure 6.3: Overview of the SRTrans encoder, which consists of convolutional overlapping patch embedding, sequence reduced multi-head self-attention, and channel-wise attention feed-forward network. SRTrans upsamples the input feature maps progressively.

Efficient self-attention and channel attention. Self-attention calculates three matrices in each head: Q, K, and V. Each row vector of these matrices represents the learnable query, key, and value vector for a corresponding patch embedding. All of these matrices have the same dimensions $(N, \hat{C})$, where $N = \frac{H \times W \times D}{p^3}$ is the length of the sequence, and the self-attention is calculated by:

$$Q,\ K,\ V = Linear_{\hat{C} \to C_{hidden}}(x_v) \tag{6.3}$$

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_{head}}}\right)V \tag{6.4}$$

where d_{head} is a constant scaling factor, which is usually equal to the hidden size of an individual head.

Since MSA is a very computing intensive process, especially for HR volumetric data, we propose to reduce the cost of time and space by applying a sequence reduction process. Inspired by PVT Wang et al. (2021) and SegFormer Xie et al. (2021a), the sequence-reduced matrix K is calculated as follows:

$$x_r = Reshape\left(x_v, \frac{H}{p}, \frac{W}{p}, \frac{D}{p}\right) \tag{6.5}$$

$$\hat{x_r} = Flatten(Downsample(x_r,\ R)) \tag{6.6}$$

$$K,\ V = Linear_{\hat{C} \to C_{hidden}}(\hat{x_r}) \tag{6.7}$$

$$Q = Linear_{\hat{C} \to C_{hidden}}(x_v) \tag{6.8}$$

where $Reshape(\cdot, \frac{H}{p}, \frac{W}{p}, \frac{D}{p})$ restores the flattened tensor of $(N, \hat{C})$ to a 3D feature map of $(\frac{H}{p},\ \frac{W}{p},\ \frac{D}{p},\ \hat{C})$. $Downsample(\cdot,\ R)$ refers to a downsampling process with a reduction ratio of R^3. It is implemented using

a convolutional layer whose kernel size and stride are both equal to R, while the input and output channel numbers remain the same. After reducing the sequence length as described in equation (6.6), the dimensions of $\widehat{x_r}$ are $(\frac{N}{R^3}, \hat{C})$ and the dimensions of K, V are $(\frac{N}{R^3}, C_{hidden})$ while the shape of Q is the same as the original MSA. By doing so, the number of parameters used in the MSA mechanism is reduced by R^3, while the calculation of attention described in equation (6.4) remains unchanged.

Another drawback of the original MSA mechanism is that only spatial attention is considered. Similar to SegFormer Xie et al. (2021a) and SegNext Guo et al. (2022), we add a depth-wise convolutional layer to the feed-forward network (FFN) in order to introduce channel-wise attention to the features (as shown in Figure 6.3).

Progressive upsampling block. Although the MSA and FFN modules preserve the resolution of flattened 3D features, the patch embedding process downscales the input image from (H, W, D) to $(\frac{H}{p}, \frac{W}{p}, \frac{D}{p})$, where p is the constant patch size. Reducing the patch size to 1 may improve the quality of features, and the output tensor can be easily reshaped to match input dimensions, but the time and space complexity will increase exponentially. Therefore, we propose to use two consecutive upsampling layers that increase the spatial resolution of feature maps generated by the Transformer encoder. The first upsampling layer interpolates the feature map so that its dimensions become the same as the input image or feature before patch embedding. The second upsampling layer upsamples the feature map with a ratio of 2 using a deconvolutional operator, achieving a 2X SR and refining the high-frequency information. We define the input image as y_0. As a result, the proposed SRTrans encoder is able to perform SR with a hierarchical architecture as follows:

$$x_i = PatchEmbed\left(y_{i-1}\right) \mid i \in [1, U] \tag{6.9}$$

$$z_j^i = FFN\left(MSA\left(z_{j-1}^i\right)\right) \mid z_0^i = \dot{x}_i,\ j \in [1, L] \tag{6.10}$$

$$y_i = DeConv_{ratio=2}(Upsample_{ratio=p}\left(z_L^i\right)) \mid i \in [1, U] \tag{6.11}$$

where L and U are the number of transformer encoder layers and the number of 2-times SR we would like to perform, respectively. $Upsample_{ratio=p}(\cdot)$ refers to a trilinear interpolation that upsamples the feature with a ratio of p, where p is the patch size used in patch embedding. The multi-scale output from SRTrans includes $z_0^1,\ y_1,\ y_2,\ \ldots,\ y_U$, where z_0^1 provides features extracted from the original-resolution image.

6.3.3 Decoder

The proposed decoder adopts a fully convolutional architecture in order to refine features from both of our encoders. We define the convolutional encoder, the SRTrans encoder, and the decoder as $\mathscr{F}$, $\mathscr{H}$, and $\mathscr{D}$,

respectively. At each step of the decoder, it performs feature fusion according to:

$$f_{out}^{1},\ f_{out}^{2},\ldots,\ f_{out}^{D} = \mathscr{F}(x) \tag{6.12}$$

$$h_{out}^{D},\ h_{out}^{D+1},\ldots,\ h_{out}^{D+U} = \mathscr{H}(x) \tag{6.13}$$

$$d_i = \begin{cases} \mathscr{D}(Concat(d_{i-1}, f_{out}^{i})), & i < D \\ \mathscr{D}(Concat(d_{i-1}, f_{out}^{D} + h_{out}^{D})), & i = D \\ \mathscr{D}(Concat(d_{i-1}, h_{out}^{i})), & i > D \end{cases} \tag{6.14}$$

where D is the number of pooling layers in the convolutional decoder and U is the number of deconvolutional layers in the SRTrans encoder. $\{f_{out}^{1},\ f_{out}^{2},\ldots,\ f_{out}^{D}\}$ and $\{h_{out}^{D},\ h_{out}^{D+1},\ldots,\ h_{out}^{D+U}\}$ refer to the feature maps whose resolution get higher as their superscripts get larger. For example, f_{out}^{1} is the most LR feature from the bottom of our convolutional encoder, and h_{out}^{D+U} is the most HR feature from the SRTrans. f_{out}^{D} and h_{out}^{D} refer to the features that have the same dimensions as the input image, we use an addition to merge them before they are fed to the decoder.

6.3.4 Loss function

The loss function we used is a combination of dice coefficient loss and cross-entropy loss. The dice loss is implemented as follows:

$$\mathscr{L}_{dc} = -\frac{1}{K}\sum_{i=1}^{K}\frac{2|U_k \cap V_k|}{|U_k| + |V_k|} \tag{6.15}$$

where $U \in X$, $V \in Y$, are the softmax output of the network and one-hot encoding of the ground truth respectively. K refers to the number of classes or output channels. We also adopted the deep-supervision technique introduced by Lee et al. Lee et al. (2015) in order to provide additional gradient signals and facilitate the learning process according to:

$$\mathscr{L}_{total} = \sum_{i=0}^{U}\frac{1}{2^i}(\mathscr{L}_{dc}^{i} + \mathscr{L}_{CE}^{i}) \tag{6.16}$$

where U is the number of deconvolutional layers in SRTrans and $i = U$ refers to the lowest-resolution segmentation. Therefore, the weights get higher as the resolution gets closer to our target SR segmentation. We use trilinear interpolation to downsample ground truth segmentations to obtain the ground truths used by deep-supervision losses.

Table 6.1: Summary of datasets.

Dataset name	Total Samples	Testing set	Num. of classes	SR ratio
Cochlea	8	4-folds cross-validation	3	8
Amos	500	200	15	4

6.4 Experiments

6.4.1 Dataset

Two CT datasets are used to validate the effectiveness of the proposed architecture, as shown in Table 6.1. Our cochlear dataset contains 8 samples obtained from 8 cadaveric cochlea specimens. Each of them has a CT scan and an HR CT scan (referred to as μCT) that have been rigidly registered with each other. The conventional CTs were acquired using a Xoran XCAT scanner and resampled to an isotropic voxel spacing of 0.3mm. The μCTs are obtained using a ScanCo μCT scanner and resampled into an isotropic voxel spacing of 0.0375 mm. Our manual labels of μCTs contain 3 classes excluding the background: air, soft tissue (including neural tissue, electrolytic fluid, and soft tissue), and bone. Thanks to the hierarchical upsampling architecture of the proposed method, we can simply define U = 3 in SRTrans and train the network to generate 8X SR segmentation. During training, input CTs are cropped to 16×16×16, so that the size of target HR segmentation is 128×128×128. During inference, images are predicted using a sliding window that overlaps by half the size of a patch for adjacent predictions. We use 4-fold cross-validation to evaluate our architecture on this dataset. With limited data, a large SR ratio, and labels acquired from HR μCTs, this dataset provides a challenging task.

To better test the generalizability of our method, we also evaluated SRSegN on the AMOS dataset Ji et al. (2022). AMOS is a public, large-scale, clinical dataset released in 2022 for abdominal organ segmentation. The winner of the segmentation challenge used a modified version of nnUNet Isensee et al. (2023). This dataset includes 500 CT scans and voxel-level labels for 15 organs. We resampled all the CTs into their median spacing of (2mm, 0.68mm, 0.68mm). Since there are multiple small organs, we decided to use an SR ratio of 4. Therefore, LR CTs are obtained by downsampling the original CT four times using a trilinear interpolation and then resampling the synthesized LR CT into a voxel spacing of (8mm, 2.72mm, 2.72mm). As all CTs and their corresponding LR version are resampled, LR CTs are cropped to 16×40×40 and the network is trained to produce 64×160×160 SR segmentations. We follow the official train/test set split [47] where 300 CTs are used for training and 200 CTs are used for testing, and 122 out of the 200 CTs in the test set are out of distribution data. We randomly selected 100 samples from the training set to serve as our

validation set.

During pre-processing, we applied clipping using 0.5 and 99.5 percentiles of the foreground voxels as well as normalization using the global foreground mean and a standard deviation on all images.

Table 6.2: Comparison to state-of-the-art methods on cochlear dataset.

Models	Num of params	DSC-Air	DSC-Soft tissue	DSC-Bone	Mean
SRGAN	9.2M	0.856	0.691	0.905	0.817±0.026
nnUNet	31.17M	0.871	0.736	0.929	0.845±0.014
VNet	22.4M	0.884	0.726	0.901	0.837±0.019
UNETR	128M	0.861	0.735	0.916	0.852±0.018
SRSegN(ours)	29.3M	0.887	0.747	0.932	0.865±0.008

6.4.2 Results

The performance of SRSegN is evaluated by the dice similarity coefficient (DSC) on both datasets. We also present the mean normalized surface dice (NSD) on the AMOS dataset. For both metrics, a higher score indicates a better result. Besides, we also provide a comparison of the number of parameters for each method in Table II. Since our architecture performs SR and segmentation at the same time, while most other segmentation models expect the dimensions of input and output to be the same, we interpolate the LR input image so that it matches the resolution of the target HR label map before feeding it to non-SR models. Because we are also interested in exploring the tense prediction ability of normal SR models, we trained the SRGAN Ledig et al. (2017) model with our loss functions. TABLE 6.2 shows the results with the cochlea dataset, and TABLE 6.3 shows the results with the AMOS dataset. The best scores are shown in bold. As shown in Table II, SRSegN outperforms SRGAN, nnUNet, VNet and UNETR by 5.9%, 2.4%, 3.4%, and 1.5%, respectively, on the cochlea dataset, while the capacity of the model stays reasonable. On the AMOS dataset, our method also outperforms the competitors in all fifteen organ classes by 74.3%, 1.1%, 4.3%, and 3.4%, respectively, in terms of NSD. Figure 6.4 presents qualitative comparisons of inner-ear tissue segmentation and multi-organ segmentation.

Table 6.3: Comparison to state-of-the-art methods on AMOS dataset.

Organs \ Models	SRGAN	nnUNet	VNet	UNETR	SRSegN(ours)
SPL	0.886	0.948	0.946	0.947	**0.950**
RKI	0.885	0.953	0.950	0.950	**0.953**
LKI	0.869	0.955	0.955	0.955	**0.955**
GBL	0.731	0.832	0.798	0.787	**0.836**
ESO	0.713	0.857	0.841	0.844	**0.863**
LIV	0.937	0.971	0.969	0.969	**0.971**
STO	0.861	0.923	0.918	0.921	**0.926**
AOR	0.866	0.950	0.946	0.949	**0.951**
IVC	0.852	0.916	0.908	0.911	**0.918**
PAN	0.773	0.862	0.854	0.854	**0.868**
RAG	0.568	0.727	0.717	0.721	**0.730**
LAG	0.513	0.754	0.737	0.744	**0.758**
DUO	0.742	0.831	0.807	0.815	**0.835**
BLA	0.782	0.870	0.864	0.865	**0.883**
PRO/UTE	0.753	0.822	0.806	0.809	**0.825**
DSC Mean	0.782	0.878	0.868	0.869	**0.882**
NSD Mean	0.432	0.745	0.722	0.728	**0.753**

Table 6.4: Data augmentation details.

SRTrans Modifications	Cochlea dataset			AMOS dataset	
	DSC-Air	DSC-Tissue	DSC-Bone	DSC Mean	NSD Mean
Original	0.887	0.747	0.932	0.882	0.753
None	0.819	0.609	0.835	0.763	0.446
FCNN	0.850	0.719	0.919	0.827	0.692
ViT	0.840	0.700	0.923	0.849	0.714

6.4.3 Ablation study

We analyze the effect of our dual-encoder and SRTrans with an ablation study, as shown in Table 6.4. The first row presents the performance of the original SRSegN. The second and third rows show the model where SRTrans is totally removed and SRTrans is replaced by an FCNN, respectively. In the last row, we replaced the convolutional patch embedding, efficient MSA, and our FFN with traditional patch embedding, MSA and MLP. Experiments showed that: (1) the upsampling encoder improves network performance; and (2) the traditional ViT modules are not as good as the proposed modules used in SRTrans.

6.4.4 Implementation details

We implemented SRSegN in Pytorch. As for other models, they are obtained from the latest-released official implementations. All of the models were trained on a NVIDIA A5000 GPU, and they were trained for 200K iterations with a cosine decay learning rate that is initialized to 0.001 and decays towards zero every 250 iterations. We used a batch size of 2 and the AdamW optimizer for all models except for nnUNet, which used an SGD optimizer. As the SR ratio, patch size, and task difficulty are different with the cochlea dataset and AMOS dataset, we list the detailed configuration of SRSegN for both tasks in TABLE 6.5. It should be noted that SRTrans extracts multi-scale features from the input image, which includes one set of features of the original resolution and N set of features of SR resolution, where N equals the number of upsampling steps in SRTrans. Therefore, when the number of upsampling steps is 3, SRTrans contains four stages, and each of the stages can be individually configured. Also, we set the hidden size of patch embeddings and MLPs equal to the base number of features.

6.5 DISCUSSION AND CONCLUSIONS

In this work, we introduced a hybrid architecture, SRSegN, for super-resolution segmentation of volumetric CT images. We proposed to use a dual-encoder architecture where the convolutional encoder downsamples feature maps and the transformer encoder upsamples feature maps. Unlike FCNNs, which lack the ability to process HR images due to a relatively small receptive field, the convolutional encoder in SRSegN extracts features from very low-resolution input, which leads to a large receptive field with a $3 \times 3 \times 3$ convolutional kernel. Also, because traditional pure transformer-based methods are unable to properly capture localized information, we proposed to use transformer encoders only for the upsampling process of feature maps, where global information is more important. The multi-scale features learned by both encoders are merged via a decoder, where HR details are refined and SR segmentation can be generated. Unlike many models that were designed for a specific SR task or SR ratio, the hierarchical design of SRSegN allows it to be easily applied to any SR task and generate multiple SR segmentations with different resolutions all at once.

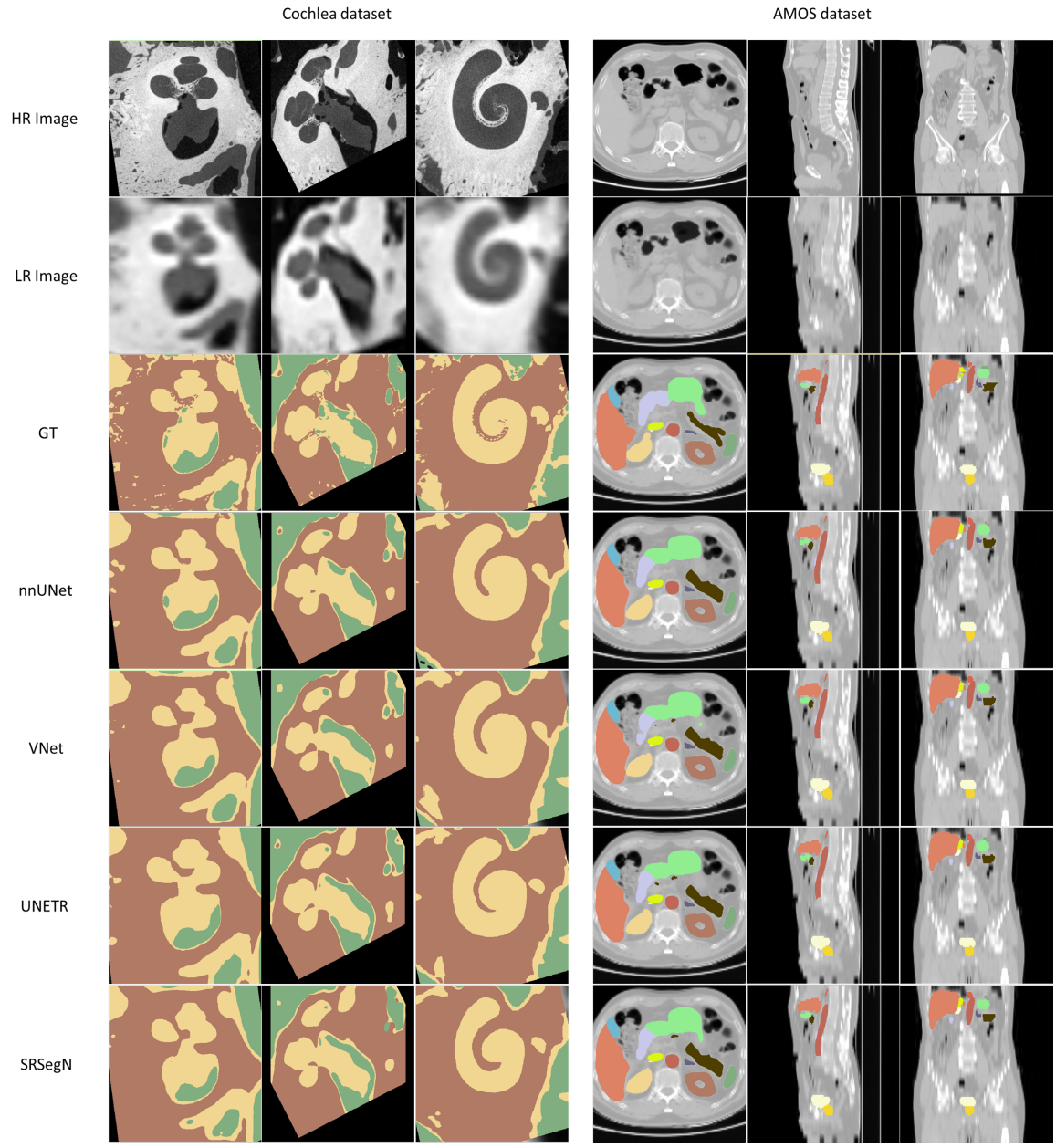


Figure 6.4: Qualitative comparison of different models in the cochlea dataset (left three columns) and AMOS dataset (right three columns).

Table 6.5: Implementation details.

Module	Cochlear dataset	AMOS dataset
Num. upsampling	3	2
Num. downsampling	3	3
Dropout rate	0	0.1
Base num. features	64	72
Patch embed size	[2, 2, 2, 4]	[2, 4, 4]
Num. Trans head	[6, 6, 8, 8]	[8, 8, 8]
Num. Trans layers	[3, 6, 6, 3]	[6, 6, 6]
Sequence reduction ratio	[1, 2, 2, 4]	[1, 2, 4]

The proposed method is also tailored for 3D volumetric data in terms of efficiency and performance. We introduced the novel SRTrans encoder as one of the core components of SRSegN. Convolutional patch embedding and channel-wise attention are used to improve the shortcomings of the original ViT encoder. Considering the huge dimensions of SR volumetric medical data, we also applied the sequence reduction technique.

We validated the SRSegN on two different datasets. The cochlear dataset contains real HR μCTs for each corresponding conventional CT. The μCTs cannot be obtained from subjects in-vivo, but μCT-level label maps are necessary to build computational CI models. Therefore, SRSegN provides a novel and accurate method to help provide important information for CI modeling. Due to the limited number of samples in the cochlea dataset, we further explored the generalizability of SRSegN using the AMOS dataset. Although the LR images are simulated by simply downsampling CTs with a trilinear interpolation, we are able to test SRsegN's effectiveness in solving ill-posed problems on this large dataset that includes 500 samples. Experiments show that the proposed method outperforms SRGAN, nnUNet, VNet, and UNETR by 5.9%, 2.4%, 3.4%, and 1.5% in terms of DSC, respectively, on the cochlea dataset. And it also has the best performance for all 15 organ segmentation tasks in terms of both DSC and NSD on the AMOS dataset.

To the best of our knowledge, this is the first method tailored for super-resolution segmentation of CTs. Such models could be critical for building high-resolution CI computational models. Although SRSegN demonstrates impressive results, there are opportunities for further enhancement. The introduction of semi-

supervised or weakly-supervised learning techniques could address the challenges associated with acquiring high-resolution medical images and voxel-level labels, providing a more cost-effective and accessible solution. Multi-task, multi-modality variants also hold potential for improving the overall performance. In our future work, we will focus on addressing these limitations and exploring the aforementioned variations.

CHAPTER 7

Automatic auditory nerve fiber segmentation

7.1 Introduction

A cochlear implant (CI) is a sophisticated auditory prosthetic device that consists of an externally worn sound signal processor and a surgically implanted electrode array. The purpose of the CI is to directly stimulate the auditory nerve fibers (ANFs), bypassing the damaged cochlea and restoring auditory perception in individuals with hearing loss. Since the CI electrode array is blindly inserted into the cochlea through a small opening during CI surgery, the intra-cochlear position of the array is generally unknown. The placement of the array in the scalae and the distance from each CI electrode to the ANFs are critical factors that determine hearing outcomes Rubinstein (2004); Holden et al. (2013); Wanna et al. (2014). In previous research, a semi-automatic ANF segmentation method Cakir et al. (2019) was introduced in order to sample electric potentials along ANF segmentations and permit the simulation of neural activation patterns due to the CI electrode array. However, there are a few drawbacks to the original ANF segmentation method.

First of all, the peripheral axons of the estimated nerve fiber trajectories sometimes overlap with each other. The peripheral axon of an ANF refers to the segment from the unmyelinated terminal adjacent to hair cells to the nerve body (soma) located within Rosenthal's Canal (RC). In the original segmentation method, this part is localized by finding the shortest path between two landmarks on the patient-specific mesh of the Scala Tympani (ST). Since the resolution of the ST mesh is not fine enough, the shortest paths found by Dijkstra's algorithm are not the geodesic shortest paths on ST's surface, resulting in jagged trajectories. And since every path is calculated independently, adjacent paths may share the same vertex on the mesh and overlap with each other. Another problem raised by limited mesh resolution and Dijkstra's algorithm is that the nerve fibers may pass through the scala structures, including ST and Scala Vestibuli (SV). Besides the peripheral part, the central axon of nerve fibers may also be localized incorrectly. The original segmentation method assumes that all the nerve fibers will proceed straight and parallel outward from the modiolus into the Internal Auditory Canal (IAC). However, the last set of landmarks that determine the proceeding direction of nerve fibers is located in the modiolus, providing inaccurate directions for nerves. As a result, almost half of the number of ANF segmentations would pass through the bone instead of gathering in the IAC. In order to solve the problems described above, the original method required manual inspection and correction of each individual nerve fiber.

In this paper, we will introduce a novel, fully automated ANF segmentation method. We propose to use

five sets of automatically generated landmarks located in ST, RC, Modiolus and IAC to perform stable and reliable path finding. The peripheral axon of a nerve fiber is calculated using an image-based fast marching method, and the central axon trajectory is localized in a cylindrical coordinate, forming into a bundle spirally in the modiolus and extending naturally through the IAC. We evaluated the proposed method on 10 subjects, and experiments show that our new segmentation method has better performance both qualitatively and quantitatively.

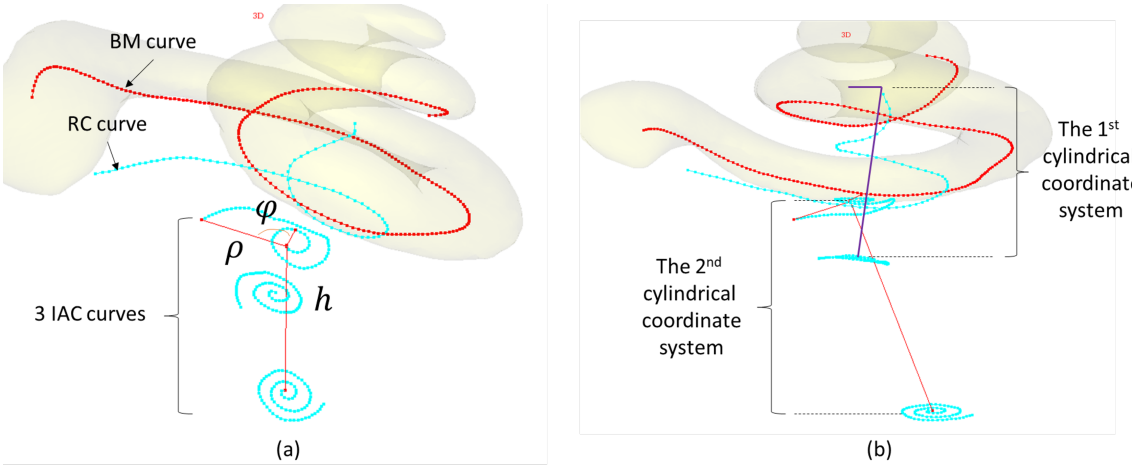


Figure 7.1: Automatically generated landmarks. (a) Five sets of landmarks. The BM curve and RC curve define the peripheral axons, and the RC curve and IAC curves define the central axons of ANFs. (b) Two cylindrical coordinate systems are defined based on those landmarks. We use cylindrical coordinates to interpolate ANFs' paths between landmarks.

7.2 Method

7.2.1 Landmarks

We create high-resolution EAMs for CI patients using techniques described in Cakir et al. (2017a) where patient-specific ST, SV meshes are generated by an active shape model (ASM) using pre-operative CT images. Since conventional CT images do not have adequate resolution to directly visualize nerve fibers, which are approximately 2µm in width, their positions need to be inferred from prior knowledge of the morphology of the fibers. We introduce five sets of landmarks that help to localize an ANF from the unmyelinated terminal to the IAC, as shown in Figure 7.1.

The first set of landmarks is located where the osseous spiral lamina meets the Basilar Membrane (BM) between ST and SV. We refer to these landmarks as the BM curve. They serve as the starting points of each nerve fiber and are manually selected by an expert on the ST mesh of our ASM. Because the ST meshes produced by the same ASM are all point-to-point correlated, we only need to perform this manual selection once, and it can be applied to any EAM in the future. The second set of landmarks aims to indicate the

location of ANF's soma and is referred to as the RC curve. To obtain this curve, we first defined landmarks representing the RC within nine specimen cochleae. Each specimen cochlea has been scanned using a ScanCo scanner, and the resulting μCT images have a voxel size of approximately 0.036mm isotropic, which is about 10 times higher resolution than conventional CT images and can help us localize the RC accurately. We then project each RC curve of specimens to the patient space through a thin-plate-spline (TPS) transformation that registers the specimen's ST mesh to the patient's ST mesh because the RC is closely adjacent to the ST. The last three sets of landmarks are located in Modiolus and IAC. They provide a cylindrical coordinate system for the central axon of ANFs, allowing them to proceed spirally outwards from the modiolus into the IAC where they form a nerve bundle and ultimately proceed to the auditory cortex. These three landmarks are generated in a similar way as the RC curve, but only one specimen cochlea is used as a reference for the TPS transformation. We refer to them as the IAC curves.

In summary, the landmarks can be automatically generated for CI users based on pre-defined vertices in the ASM or landmarks of specimens. With their help, we are able to accurately localize the starting point, soma, and the proceeding direction of the central axon of each ANF. In the following sections, we will describe in detail how to determine the final segmentation of the peripheral and central axons using these landmarks respectively.

7.2.2 Peripheral axon

The peripheral part of an ANF is defined by the BM curve and the RC curve. The RC curve is divided into 80 evenly spaced points that determine the location of the somas for each ANF path. Morphological studies have shown that most of the peripheral axons of ANF extend radially towards the center Li et al. (2021). Therefore, in order to find the corresponding points on the BM curve, we project both the BM curve and the RC curve to the same plane along the direction of the "mid-modiolar axis". The middle modiolus axis is defined by connecting the most apex point in the RC and the first IAC curve (as shown in Figure 7.2(a)). We define the circle center according to the points in the last 180 degrees of the projected RC curve and calculate the radian of each point on it. We assume that the first point on the BM curve always corresponds to the first point on the RC curve and define an upper boundary for the distance between two adjacent nerves so that ANFs located at the most basilar part are not too sparse. For each RC point, we find its corresponding point on the BM according to the radians.

Each pair of points on the BM curve and RC curve determines the starting point and endpoint of an ANF's peripheral part. We use the fast marching method to find the geodesic path that connects these two landmarks. The fast marching algorithm Sethian (1999) is a numerical method that is able to solve boundary value problems of the Eikonal equation $\nabla D(x) = P(x)$, which describes a level set function whose front

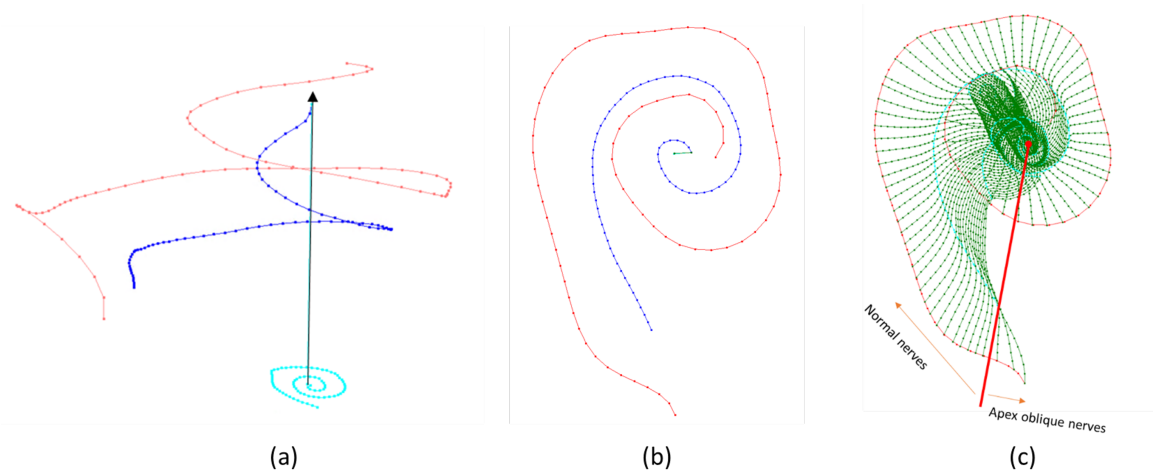


Figure 7.2: The peripheral axons of ANFs. (a) We project the RC curve and BM curve along the direction defined by connecting the most apex point on the first IAC curve (in cyan) and RC curve (in blue). (b) We define the circle center based on the projected RC curve (in blue) and calculate radians accordingly. (c) Nerves at the basilar part are oblique and will be calculated differently.

advances with speed P(x). The resulting function D is a time function, and if the speed P is constant, it can be regarded as the distance function to the starting point. Once we obtain the time function D, we are able to extract a good approximation of the geodesic path by backtracking from the endpoint using the gradient descent of D. In our case, the nerve fibers are expected to proceed along the surface of ST or SV and should never pass through them. Therefore, the traveling front is supposed to be slow near the ST and SV, and totally stopped in the normal direction of their surfaces. Herein, we generate the speed map P using binary masks of ST and SV. The masks are combined and smoothed by a 3D Gaussian filter whose size is 9 and whose standard deviation equals 2. In our implementation, the fast marching algorithm computes the time function to the point on the RC curve according to the speed map P, and the computation terminates as soon as the marching front reaches the corresponding point on the BM curve. Then we calculate the gradient descent and extract the path that represents the peripheral axon by backtracking from the BM curve (as shown in Figure 7.3).

Although most of the peripheral axons can be segmented successfully following the method described above, ANFs located at the most basilar part of the cochlea are oblique Sethian (1999) and need to be treated differently. We identify these oblique nerves by finding the points on the BM curve whose radians are smaller than their corresponding points on the RC curve. For these oblique nerves, we manually add a vector field that points towards the tangent direction to the BM curve to the gradient vector field of D, forcing those nerves to bend slightly. The vector field is calculated for each ANF as follows:

$$V_x^i = (1 - Normalize\,(Dist_{BM}\,(x,i))) * \frac{(i_{max} - i)^2}{{i_{max}}^2},\ i \in 1,\ 2,\ \ldots,\ i_{max} \tag{7.1}$$

where $x \in \mathbb{R}^3$ is a point in space, V_x^i is the additive vector field for the $i-th$ oblique ANF, assuming there are a total of i_{max} oblique ANFs, and $Dist_{BM}(x,i)$ represents the Euclidean distance from x to the tangent line of BM curve at the i-th point. As a result, the vector field has more impact on the peripheral part of peripheral axons, and the weights will decrease as they come closer to the non-oblique normal ANFs.

7.2.3 Central axon

The RC curve and three IAC curves determine the paths of central axons. Unlike the original ANF segmentation method which assumes that all the central axons proceed straight through the modiolus and IAC, our predefined landmarks in the IAC will guide the central axons to proceed spirally along the direction of the central axis of the IAC. In order to smoothly interpolate the nerve fibers between those landmarks, we define two cylindrical coordinate systems. The first cylindrical axis is defined by connecting the last node on the RC curve and the last node in the middle IAC curve. The zero-degree vector is defined by making a straight line that go through the first node on the RC curve and is perpendicular to the cylindrical axis. Similarly, the second cylindrical axis is found using the first and last IAC curves, with its zero-degree pointing to the direction of the first node on the first IAC curve. We interpolate the ANFs that are located between the RC curve and the first IAC curve using the first cylindrical coordinate system, and the remaining central axons are interpolated under the second cylindrical coordinate system. We define the number of control points as m. The first cylindrical coordinate system uses the corresponding points on the RC curve and IAC curve to interpolate a nerve fiber, and the second one uses one control point from each of the three IAC curves, so m = 2 and 3, respectively. In both cylindrical coordinate systems, the interpolation is performed according to:

$$Landmarks_i = \left\{ \left(x_1^i,\ y_1^i,\ z_1^i\right),\ \ldots, \left(x_m^i,\ y_m^i,\ z_m^i\right) \right\} \mid i\ \varepsilon\ [1,N] \tag{7.2}$$

$$\left(\rho_j^i, \varphi_j^i, h_j^i\right) = Cylin_Coord\left(x_j^i,\ y_j^i,\ z_j^i\right) \mid i\ \varepsilon\ [1,N],\ j\ \varepsilon\ [1,m] \tag{7.3}$$

$$\hat{H} = \left\{ \widehat{h_1^i}, \widehat{h_2^i}, \ldots,\ \widehat{h_p^i} \right\} = Linear\left(p,\ z_1^i, \ldots, z_m^i\right) \tag{7.4}$$

$$\left\{ \widehat{\rho_1^i}, \widehat{\rho_2^i}, \ldots,\ \widehat{\rho_p^i} \right\} = Linear\left(\hat{H},\ \rho_1^i, \ldots, \rho_m^i\right) \tag{7.5}$$

$$\left\{ \widehat{\varphi_1^i}, \widehat{\varphi_2^i}, \ldots,\ \widehat{\varphi_p^i} \right\} = Linear\left(\hat{H},\ \varphi_1^i, \ldots, \varphi_m^i\right) \tag{7.6}$$

$$\left(\widehat{x_j^i}, \widehat{y_j^i}, \widehat{z_j^i}\right) = Cart_Coord\left(\rho_j^i, \varphi_j^i, h_j^i\right) \mid i\ \varepsilon\ [1,N],\ j\ \varepsilon\ [1,m] \tag{7.7}$$

where i indicates the $i-th$ nerve fiber out of a total of N ANFs, $Cylin_Ccoord(\cdot)$ transfers a point from the Cartesian coordinate to the cylindrical coordinate, and ρ, φ, and h are the radius, radian, and height in the cylindrical coordinate system, respectively. $Linear(p, \cdot)$ performs a linear interpolation that creates p evenly sampled points out of the input, while $Linear\left(\hat{H}, \cdot\right)$ sampling the input values according to the relative distance indicated by the vector $\hat{H}$. In the end, $Cart_Ccoord(\cdot)$ calculates the Cartesian coordinates for each input in the cylindrical coordinate. In summary, we constructed two cylindrical coordinate systems to interpolate the points between the RC and IAC, and within the IAC, separately. The landmarks under the Cartesian coordinate will be transferred to a cylindrical coordinate, and the radius, radian, and height will be linearly interpolated, respectively. Finally, we transform the coordinate back to the Cartesian coordinate and obtain the spiral central axons that proceed through the IAC.

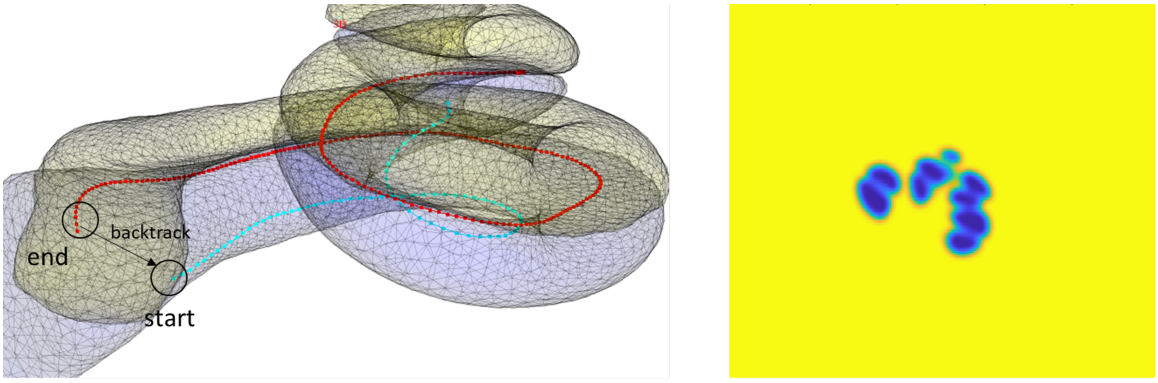


Figure 7.3: The setup for the fast marching algorithm. (a) The marching front starts from the RC curve (in cyan) and ends at the BM curve (in red). Then the path is found by backtracking from the BM curve to the RC curve. (b) The speed map P that is used by the fast marching algorithm. It is generated using ST (blue mesh) and SV (yellow mesh) masks, where they are combined and smoothed by a Gaussian filter.

7.3 Experiments

Since the original method is a semi-automatic approach, we compare the result of the proposed method with both initial ANF trajectories and manually optimized trajectories from the original method. As shown in Table 1, the result of our proposed method outperforms the semi-automatic method for all ten CI subjects used in the experiment. For each approach, the first column indicates the number of ANFs that overlap with one or more adjacent ones, and the second column and the third column show the number of ANFs that pass through the ST or SV, or the bone surrounding the IAC. We edit the peripheral axons during the semi-automatic method's manual correction by viewing the ANF segmentation in a 3D view using customized software, and we then modify those automatically generated paths in accordance with the ST mesh so that they never cross over one

another and always travel along the surface of ST. As for central axons, we attempted to prevent them from passing through the bone by manually moving the last set of landmarks that decide the proceeding direction in the modiolus. Experiments show that the manual adjustment indeed improves the ANF segmentation results for the original method, however, our proposed method is able to achieve better performance fully automatically, with zero overlapping ANFs, zero ANFs that go into the bone, and 36.1% fewer ANFs that go through the ST or SV.

We also present the voltage distribution along ANFs due to the stimulation of the CI electrode array (as shown in Figure 7.4), where Nerve 1 is located at the basilar part of the cochlea and Nerve 65 is located at the apex region, and each figure corresponds to the electrical field due to a certain CI electrode. The results of the proposed method are plotted in solid lines while the results of the original method are plotted in dashed lines. For nerves that can be properly adjusted in the original method (such as Nerve 45 and Nerve 65), the voltage distributions along the new nerve trajectory and old nerve trajectory are similar. In contrast, for nerves like Nerve 1 and Nerve 25, the electrical potentials are significantly larger at the central axons of the new nerve segmentation. This is because those ANFs pass through the bone in the original nerve segmentation while they can be constrained correctly in the IAC with our proposed method. In Figure 7.5, we present the trajectories of Nerve 1 generated by the original method (green line) and the proposed method (purple line). The red circle shows the location where the ANF estimated by the old method goes into the bone from the modiolus (red mesh/contour), while our proposed method is able to address this problem. Such improvement may be critical for increasing the accuracy and stability of EAMs.

Table 7.1: Comparison of ANF segmentation result using the original semi-automatic method and the proposed method.

Subject	Semi-automatic method (*Before* manual correction)			Semi-automatic method (*After* manual correction)			The proposed method		
	Num. overlap	Num. cross STSV	Num. cross bone	Num. overlap	Num. cross STSV	Num. cross bone	Num. overlap	Num. cross STSV	Num. cross bone
1	29	10	45	0	5	42	0	1	0
2	33	5	34	0	1	32	0	2	0
3	33	20	39	0	10	35	0	11	0
4	22	2	32	0	0	32	0	0	0
5	26	10	38	0	8	35	0	5	0
6	30	8	46	0	8	41	0	4	0
7	27	2	34	0	2	32	0	3	0
8	26	11	36	0	10	33	0	5	0
9	19	7	41	0	7	37	0	2	0
10	32	12	36	0	10	34	0	6	0
Avg.	27.7	8.7	38.1	**0**	6.1	35.3	**0**	**3.9**	**0**

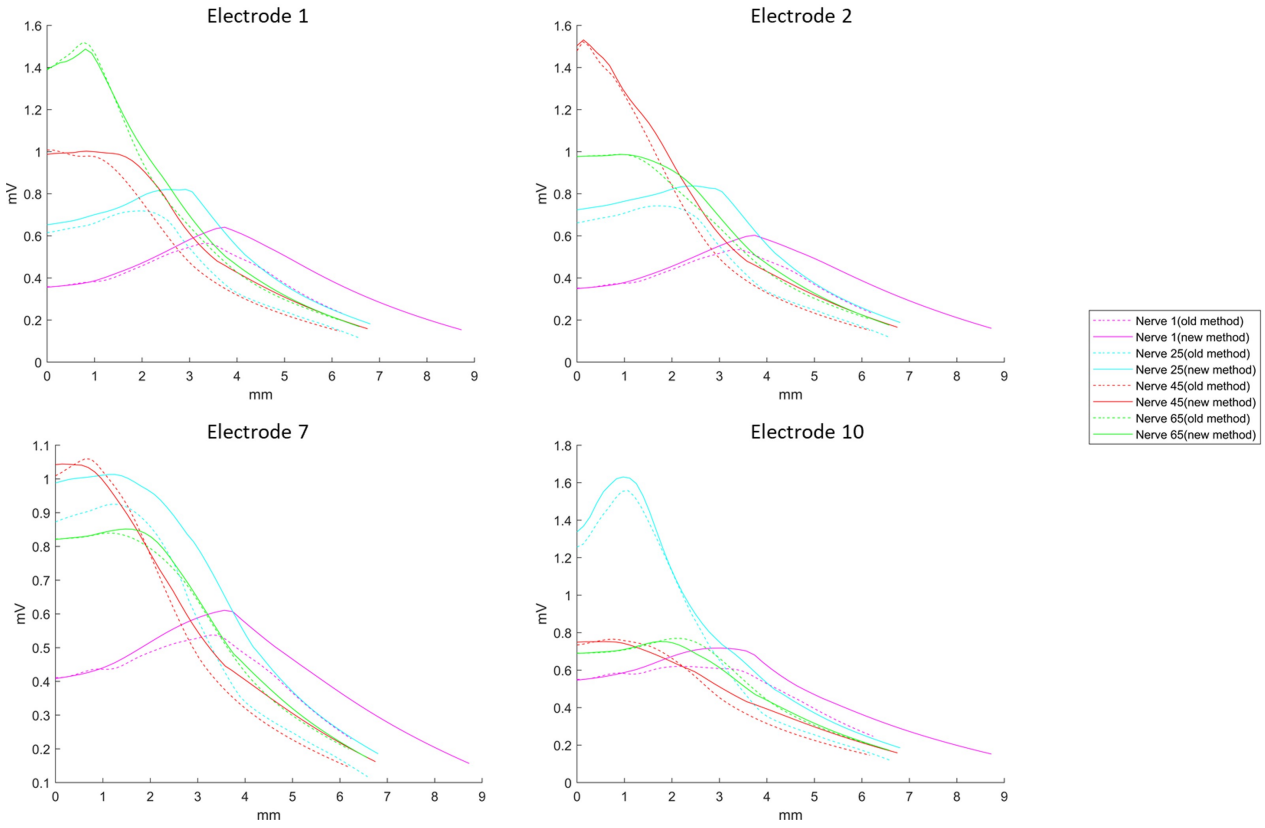


Figure 7.4: Voltage distribution along ANFs. Dashed lines represent ANF segmentation obtained using the original semi-automatic method proposed in Cakir et al. (2019) and solid lines represent our proposed method.

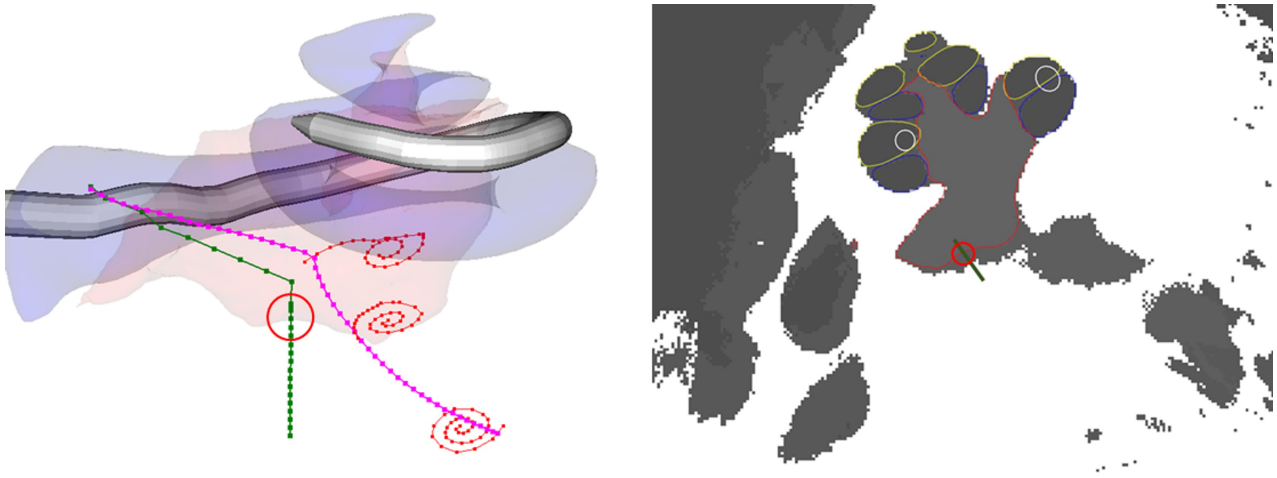


Figure 7.5: Visualication of ANFs generated by the original method (green line) and the proposed method (purple line).

7.4 Conclusion

Localizing ANFs is essential for building EAMs and improving hearing outcomes for CI users. In this paper, we introduce a patient-specific, fully automatic ANF segmentation method. Unlike the traditional method which requires a lot of effort to manually inspect and adjust each individual ANF trajectory, the proposed method achieved better performance by utilizing automatically generated landmarks, the fast-marching algorithm, and cylindrical coordinate systems. We also include the oblique ANFs in the high-frequency region that do not exist in the original method so that the EAMs can simulate neural activations across a broader spectrum. We evaluated the new segmentation method using 10 cochlea subjects. Experiments show that the proposed method achieved impressive performance with zero overlapping ANFs and zero ANFs proceeding towards the bone. The number of ANFs that pass through ST or SV is also reduced by 36.1%. We also visualize the ANF segmentation together with ST, SV, and Modiolus in Figure 7.5, which shows good agreement with real human ANF imaging obtained using recent techniques Li et al. (2021). In conclusion, we introduce a fully automatic, fast, and robust ANF segmentation method, that outperforms the original segmentation method both qualitatively and quantitatively. Our future work includes comparing the activation thresholds of the original and newly obtained ANF segmentations, creating EAMs using the proposed method, and collecting subjective scores for the original and new ANF shapes from experts.

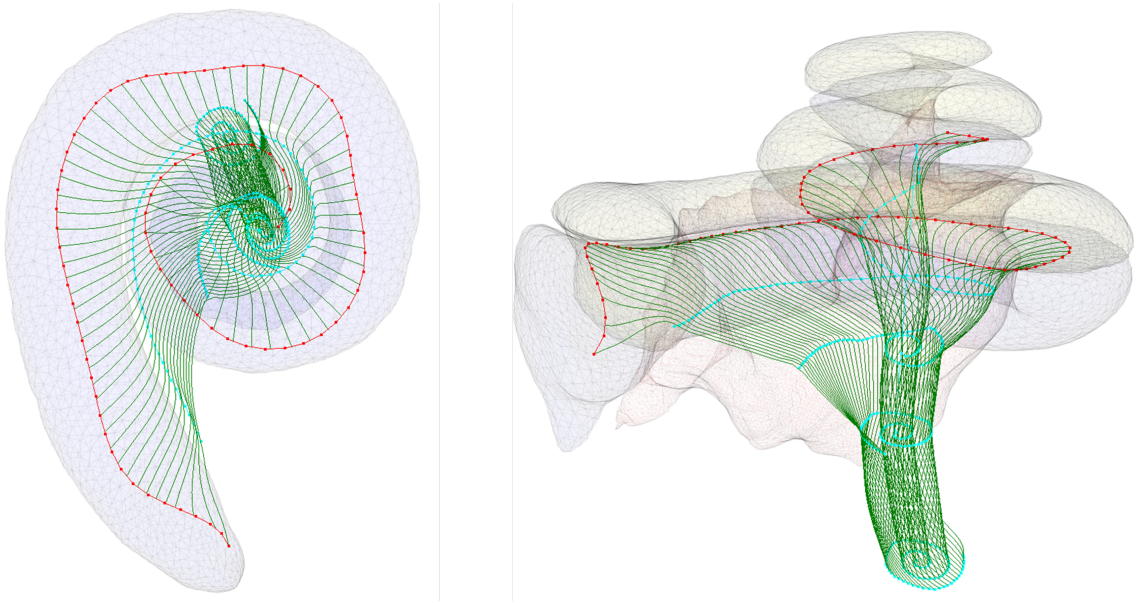


Figure 7.6: ANF trajectories generated by the proposed method.

CHAPTER 8

Concluding remarks

This dissertation has focused on improving the performance and customization of CIs through various research projects. The goal was to address the challenges faced in providing patient-specific EAM and optimizing CI outcomes. By addressing these challenges, this research contributes to the advancement of CI technology and has the potential to improve the hearing outcomes for CI recipients.

8.1 Summary of Research Contributions

This doctoral research has made several key contributions to the field of cochlear implants:

Customized Electrode Firing Order: One major contribution was the development of an optimization algorithm for the firing order of electrodes in CIs. By utilizing patient-specific neural stimulation models, this method minimized the negative effects of overlapping stimulation, enhancing CI performance without compromising spectral resolution.

Auditory Nerve Fiber Health Estimation: Another significant contribution was the development of computational models to estimate the health of auditory nerve fibers. By incorporating patient-specific EAMs and the biological ANF model, this approach provided an estimation of nerve health along the cochlear length based on post-operative physiological measurements that are available in clinical settings, offering valuable insights for personalized CI programming and improved hearing outcomes.

Efficient Electric Potential Estimation: A key challenge in CI modeling is the computational complexity of estimating patient-specific electric potentials. This research proposed a deep learning-based method that significantly reduced computation time while accurately estimating intra-cochlear electric potentials. This approach enables faster and more efficient EAM construction.

Patient-Specific Electrical Characteristic Estimation: This research introduced a deep learning-based method for estimating patient-specific electrical characteristics. By combining a cycle-consistent network architecture with conventional searching strategies, this approach provided high-quality predictions, enhancing the speed and accuracy of CI modeling and programming.

Super-Resolution Segmentation: A critical aspect of CI modeling is accurate segmentation of inner-ear CT images. This research proposed a deep learning architecture, SRSegN, which outperformed traditional methods and provided high-quality segmentation, enabling more precise modeling of CIs. Additionally, we conducted experiments on a large multi-class segmentation dataset, showcasing the architecture's impressive performance. These findings indicate the potential of SRSegN to be utilized in various medical image datasets

beyond cochlear imaging.

Automatic Auditory Nerve Fiber Segmentation: This research developed a fully automatic method for auditory nerve fiber segmentation. By estimating peripheral and central axons individually based on automatically generated landmarks, this approach significantly improved segmentation accuracy, enhancing the fidelity of auditory nerve fiber trajectories in CI models.

8.2 Discussion and future work

While this dissertation has made improvements and contributions to our current computational models of CIs, it is important to acknowledge the current limitations of the research:

(1) Limited clinical validation: Although the proposed methods have shown promising results in simulation and experimental setups, further validation and clinical studies are necessary to assess their real-world effectiveness and impact on CI recipients. Extending the research to larger and more diverse patient cohorts will provide more comprehensive insights into the generalizability and robustness of the developed techniques.

(2) Data availability: Acquiring high-resolution medical images and voxel-level labels for training deep learning models remains a challenge. The limited availability of such data hinders the scalability and generalizability of the proposed methods. Exploring approaches like semi-supervised or weakly-supervised learning could help alleviate this limitation by leveraging existing data more effectively.

(3) Model interpretability: Deep learning models, such as SRSegN, are often considered black boxes due to their complex architectures. While these models achieve impressive results, understanding the underlying decision-making process and ensuring their reliability and safety in clinical applications are essential. Developing techniques for model interpretability and explainability would enhance the trustworthiness and adoption of the proposed methods.

To address the current limitations and pave the way for further research, the following future works are suggested:

(1) Clinical validation: Conduct extensive clinical studies to evaluate the performance and impact of the proposed methods in real-world CI recipients.

(2) Data augmentation and Transfer Learning: Augmenting the available data through techniques like data synthesis, domain adaptation, and transfer learning can help alleviate the data scarcity issue. Leveraging pre-existing annotated datasets or leveraging knowledge from related medical imaging domains can enhance the training and generalization capabilities of the developed models.

(3) Model optimization and efficiency: Further optimize the proposed models to improve their efficiency and reduce computational resource requirements. Exploring model compression, quantization, and hard-

ware acceleration techniques can enable the deployment of these models on resource-constrained platforms, making them more accessible and practical for real-time applications.

(4) Exploration of different biological models of ANF: it would be valuable to investigate and incorporate alternative biological models of auditory nerve fibers (ANFs) into the CI research. While the current research utilizes the Rattay et al. model Rattay et al. (2001b), which has shown promising results, exploring different ANF models can provide a more comprehensive understanding of the complexities and variations in neural responses.

8.3 List of Publications

1. Liu, Ziteng, Ahmet Cakir, and Jack H. Noble. "Cochlear implant electrode sequence optimization using patient specific neural stimulation models." In Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 11315, pp. 726-731. SPIE, 2020.

2. Liu, Ziteng, Ahmet Cakir, and Jack H. Noble. "Auditory nerve fiber health estimation using patient specific cochlear implant stimulation models." In Simulation and Synthesis in Medical Imaging: 5th International Workshop, SASHIMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 5, pp. 184-194. Springer International Publishing, 2020.

3. Liu, Ziteng, and Jack H. Noble. "Cochlear implant electric field estimation using 3D neural networks." In Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 11598, pp. 373-379. SPIE, 2021.

4. Liu, Ziteng, and Jack H. Noble. "Patient-specific electro-anatomical modeling of cochlear implants using deep neural networks." In Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 12034, pp. 96-103. SPIE, 2022.

5. Bratu, Erin L., Ziteng Liu, and Jack H. Noble. "Influence of auditory nerve fiber model parameters on electrical stimulus thresholds." In Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 12466, pp. 549-557. SPIE, 2023.

6. Lou, Ange, Kareem Tawfik, Xing Yao, Ziteng Liu, and Jack Noble. "Min-Max Similarity: A Contrastive Semi-Supervised Deep Learning Network for Surgical Tools Segmentation." IEEE Transactions on Medical Imaging (2023).

References

Abdollahi, A., Pradhan, B., and Alamri, A. (2020). Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *IEEE Access*, 8:179424–179436.

Ahmad, W., Ali, H., Shah, Z., and Azmat, S. (2022). A new generative adversarial network for medical images super resolution. *Scientific Reports*, 12(1):9533.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Avci, E., Nauwelaers, T., Lenarz, T., Hamacher, V., and Kral, A. (2014). Variations in microanatomy of the human cochlea. *Journal of Comparative Neurology*, 522(14):3245–3261.

Bierer, J. A. and Litvak, L. (2016). Reducing channel interaction through cochlear implant programming may improve speech perception: Current focusing and channel deactivation. *Trends in hearing*, 20:2331216516653389.

Boëx, C., de Balthasar, C., Kós, M.-I., and Pelizzone, M. (2003). Electrical field interactions in different cochlear implant systems. *The Journal of the Acoustical Society of America*, 114(4):2049–2057.

Bourien, J., Tang, Y., Batrel, C., Huet, A., Lenoir, M., Ladrech, S., Desmadryl, G., Nouvian, R., Puel, J.-L., and Wang, J. (2014). Contribution of auditory nerve fibers to compound action potential of the auditory nerve. *Journal of neurophysiology*, 112(5):1025–1039.

Briaire, J. J. and Frijns, J. H. (2005). Unraveling the electrically evoked compound action potential. *Hearing research*, 205(1-2):143–156.

Buss, E., Pillsbury, H. C., Buchman, C. A., Pillsbury, C. H., Clark, M. S., Haynes, D. S., Labadie, R. F., Amberg, S., Roland, P. S., Kruger, P., et al. (2008). Multicenter us bilateral med-el cochlear implantation study: Speech perception over the first year of use. *Ear and hearing*, 29(1):20–32.

Cakir, A., Dawant, B. M., and Noble, J. H. (2017a). Development of a ct-based patient-specific model of the electrically stimulated cochlea. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 773–780. Springer.

Cakir, A., Dwyer, R. T., and Noble, J. H. (2017b). Evaluation of a high-resolution patient-specific model of the electrically stimulated cochlea. *Journal of Medical Imaging*, 4(2):025003–025003.

Cakir, A., Labadie, R. F., and Noble, J. H. (2019). Auditory nerve fiber segmentation methods for neural activation modeling. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10951, pages 396–402. SPIE.

Carnevale, N. T. and Hines, M. L. (2006). *The NEURON book*. Cambridge University Press.

Cartee, L. A. (2000). Evaluation of a model of the cochlear neural membrane. ii: Comparison of model and physiological measures of membrane properties measured in response to intrameatal electrical stimulation. *Hearing research*, 146(1-2):153–166.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65.

DIJKSTRA, E. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271.

Dong, C., Loy, C. C., He, K., and Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer.

Dorman, M. F., Yost, W. A., Wilson, B. S., and Gifford, R. H. (2011). Speech perception and sound localization by adults with bilateral cochlear implants. In *Seminars in hearing*, volume 32, pages 073–089. © Thieme Medical Publishers.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Drennan, W. R. and Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of rehabilitation research and development*, 45(5):779.

D'Errico, J. (2019). Bound constrained optimization using fminsearch.

Fan, Y., Zhang, D., Wang, J., Noble, J. H., and Dawant, B. M. (2020). Combining model-and deep-learning-based methods for the accurate and robust segmentation of the intra-cochlear anatomy in clinical head ct images. In *Medical Imaging 2020: Image Processing*, volume 11313, pages 315–322. SPIE.

Frijns, J., De Snoo, S., and Schoonhoven, R. (1995). Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea. *Hearing research*, 87(1-2):170–186.

Frijns, J. H., Briaire, J. J., and Grote, J. J. (2001). The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants. *Otology & neurotology*, 22(3):340–349.

Fu, Q.-J. and Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, 6:19–27.

Geddes, L. A. and Baker, L. E. (1967). The specific resistance of biological material—a compendium of data for the biomedical engineer and physiologist. *Medical and biological engineering*, 5:271–293.

Gifford, R. H., Dorman, M. F., Sheffield, S. W., Teece, K., and Olund, A. P. (2014). Availability of binaural cues for bilateral implant recipients and bimodal listeners with and without preserved hearing in the implanted ear. *Audiology and Neurotology*, 19(1):57–71.

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neurotology*, 13(3):193–205.

Goldwyn, J. H., Bierer, S. M., and Bierer, J. A. (2010). Modeling the electrode–neuron interface of cochlear implants: Effects of neural survival, electrode placement, and the partial tripolar configuration. *Hearing research*, 268(1-2):93–104.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Greenberg, S., Ainsworth, W. A., Popper, A. N., Fay, R. R., and Clark, G. (2004). Cochlear implants. *Speech Processing in the Auditory System*, pages 422–462.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. *Advances in neural information processing systems*, 30.

Guo, M.-H., Lu, C.-Z., Hou, Q., Liu, Z., Cheng, M.-M., and Hu, S.-M. (2022). Segnext: Rethinking convolutional attention design for semantic segmentation. *Advances in Neural Information Processing Systems*, 35:1140–1156.

Han, W., Chang, S., Liu, D., Yu, M., Witbrock, M., and Huang, T. S. (2018). Image super-resolution via dual-state recurrent networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1654–1663.

Hanekom, T. (2001). Three-dimensional spiraling finite element model of the electrically stimulated cochlea. *Ear and hearing*, 22(4):300–315.

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., and Xu, D. (2021). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI Brainlesion Workshop*, pages 272–284. Springer.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., and Xu, D. (2022). Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hesamian, M. H., Jia, W., He, X., and Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32:582–596.

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and hearing*, 34(3):342.

Hughes, M. L. (2012). *Objective measures in cochlear implants*. Plural Publishing.

Iglesias, J. E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., González, R. G., Alexander, D. C., Golland, P., Edlow, B. L., Fischl, B., et al. (2021). Joint super-resolution and synthesis of 1 mm isotropic mp-rage volumes from clinical mri exams with scans of different orientation, resolution and contrast. *Neuroimage*, 237:118206.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211.

Isensee, F., Ulrich, C., Wald, T., and Maier-Hein, K. H. (2023). Extending nnu-net is all you need. In *BVM Workshop*, pages 12–17. Springer.

Ji, Y., Bai, H., Ge, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., et al. (2022). Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *Advances in Neural Information Processing Systems*, 35:36722–36732.

Kalkman, R. K., Briaire, J. J., and Frijns, J. H. (2015). Current focussing in cochlear implants: an analysis of neural recruitment in a computational model. *Hearing research*, 322:89–98.

Khan, A. M., Handzel, O., Burgess, B. J., Damian, D., Eddington, D. K., and Nadol Jr, J. B. (2005). Is word recognition correlated with the number of surviving spiral ganglion cells and electrode insertion depth in human subjects with cochlear implants? *The Laryngoscope*, 115(4):672–677.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z. (2015). Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570. Pmlr.

Lenarz, M., Sönmez, H., Joseph, G., Büchner, A., and Lenarz, T. (2012). Long-term performance of cochlear implants in postlingually deafened adults. *Otolaryngology–Head and Neck Surgery*, 147(1):112–118.

Li, H., Helpard, L., Ekeroot, J., Rohani, S. A., Zhu, N., Rask-Andersen, H., Ladak, H. M., and Agrawal, S. (2021). Three-dimensional tonotopic mapping of the human cochlea based on synchrotron radiation phase-contrast imaging. *Scientific reports*, 11(1):4437.

Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144.

Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study. *Ear and hearing*, 27(6):714.

Liu, Z., Cakir, A., and Noble, J. H. (2020a). Auditory nerve fiber health estimation using patient specific cochlear implant stimulation models. In *Simulation and Synthesis in Medical Imaging: 5th International Workshop, SASHIMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 5*, pages 184–194. Springer.

Liu, Z., Cakir, A., and Noble, J. H. (2020b). Cochlear implant electrode sequence optimization using patient specific neural stimulation models. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315, pages 726–731. SPIE.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022.

Liu, Z. and Noble, J. H. (2021). Cochlear implant electric field estimation using 3d neural networks. In *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11598, pages 373–379. SPIE.

Lousteau, R. J. (1987). Increased spiral ganglion cell survival in electrically stimulated, deafened guinea pig cochleae. *The Laryngoscope*, 97(7):836–842.

Malherbe, T., Hanekom, T., and Hanekom, J. (2016). Constructing a three-dimensional electrical model of a living cochlear implant user's cochlea. *International journal for numerical methods in biomedical engineering*, 32(7):e02751.

Malherbe, T. K., Hanekom, T., and Hanekom, J. J. (2015). The effect of the resistive properties of bone on neural excitation and electric fields in cochlear implant models. *Hearing research*, 327:126–135.

Mens, L. H. (2007). Advances in cochlear implant telemetry: evoked neural responses, electrical field imaging, and technical integrity. *Trends in amplification*, 11(3):143–159.

Moberly, A. C., Bates, C., Harris, M. S., and Pisoni, D. B. (2016). The enigma of poor performance by adults with cochlear implants. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 37(10):1522.

Nadol Jr, J. B., Young, Y.-S., and Glynn, R. J. (1989). Survival of spiral ganglion cells in profound sensorineural hearing loss: implications for cochlear implantation. *Annals of Otology, Rhinology & Laryngology*, 98(6):411–416.

NIDCD (2019). Nidcd (2019) nidcd fact sheet, hearing and balance: cochlear implants. Accessed 10 Jan 2023.

Noble, J. H. and Dawant, B. M. (2015). Automatic graph-based localization of cochlear implant electrodes in ct. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18*, pages 152–159. Springer.

Noble, J. H., Gifford, R. H., Hedley-Williams, A. J., Dawant, B. M., and Labadie, R. F. (2015). Clinical evaluation of an image-guided cochlear implant programming strategy. *Audiology and Neurotology*, 19(6):400–411.

Noble, J. H., Hedley-Williams, A. J., Sunderhaus, L., Dawant, B. M., Labadie, R. F., Camarata, S. M., and Gifford, R. H. (2016). Initial results with image-guided cochlear implant programming in children. *Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 37(2):e63.

Noble, J. H., Labadie, R. F., Gifford, R. H., and Dawant, B. M. (2013). Image-guidance enables new methods for customizing cochlear implant stimulation strategies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(5):820–829.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

on Deafness, N. I. and (NIDCD), O. C. D. (2011).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Peng, C., Zhang, X., Yu, G., Luo, G., and Sun, J. (2017). Large kernel matters–improve semantic segmentation by global convolutional network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361.

Pisoni, D. B., Kronenberger, W. G., Harris, M. S., and Moberly, A. C. (2017). Three challenges for future research on cochlear implants. *World Journal of Otorhinolaryngology-Head and Neck Surgery*, 3(04):240–254.

Raphael, Y. and Altschuler, R. A. (2003). Structure and innervation of the cochlea. *Brain research bulletin*, 60(5-6):397–422.

Rask-Andersen, H., Liu, W., Erixon, E., Kinnefors, A., Pfaller, K., Schrott-Fischer, A., and Glueckert, R. (2012). Human cochlea: anatomical characteristics and their relevance for cochlear implantation. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, 295(11):1791–1811.

Rattay, F., Leao, R. N., and Felix, H. (2001a). A model of the electrically excited human cochlear neuron. ii. influence of the three-dimensional cochlear structure on neural excitability. *Hearing research*, 153(1-2):64–79.

Rattay, F., Lutter, P., and Felix, H. (2001b). A model of the electrically excited human cochlear neuron: I. contribution of neural substructures to the generation and propagation of spikes. *Hearing research*, 153(1-2):43–63.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Current opinion in otolaryngology & head and neck surgery*, 12(5):444–448.

Santi, P. A. and Tsuprun, V. L. (1988). Cochlear microanatomy and ultrastructure. *Physiology of the Ear*, 2:257–283.

Sethian, J. A. (1999). Fast marching methods. *SIAM review*, 41(2):199–235.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883.

Shocher, A., Cohen, N., and Irani, M. (2018). "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3126.

Spoendlin, H. and Schrott, A. (1989). Analysis of the human auditory nerve. *Hearing research*, 43(1):25–38.

Tai, Y., Yang, J., and Liu, X. (2017). Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155.

Vaerenberg, B., Smits, C., De Ceulaer, G., Zir, E., Harman, S., Jaspers, N., Tam, Y., Dillon, M., Wesarg, T., Martin-Bonniot, D., et al. (2014). Cochlear implant programming: a global survey on the state of the art. *The scientific world journal*, 2014.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., and Shao, L. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 568–578.

Wang, Z., Chen, J., and Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387.

Wanna, G. B., Noble, J. H., Carlson, M. L., Gifford, R. H., Dietrich, M. S., Haynes, D. S., Dawant, B. M., and Labadie, R. F. (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *The Laryngoscope*, 124(S6):S1–S7.

Whiten, D. M. (2007). *Electro-anatomical models of the cochlear implant*. PhD thesis, Massachusetts Institute of Technology.

Wilson, B. S. and Dorman, M. F. (2008). Cochlear implants: current designs and future possibilities. *J Rehabil Res Dev*, 45(5):695–730.

Wolfe, J. and Schafer, E. C. (2014). *Programming cochlear implants*. Plural Publishing.

Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., and Zhang, L. (2021). Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 22–31.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021a). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090.

Xie, Y., Zhang, J., Shen, C., and Xia, Y. (2021b). Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 171–180. Springer.

Xu, W., Xu, Y., Chang, T., and Tu, Z. (2021). Co-scale conv-attentional image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9981–9990.

Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zeiler, M. D., Krishnan, D., Taylor, G. W., and Fergus, R. (2010). Deconvolutional networks. In *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, pages 2528–2535. IEEE.

Zhang, D., Banalagay, R., Wang, J., Zhao, Y., Noble, J. H., and Dawant, B. M. (2019). 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*, volume 10949, pages 45–52. SPIE.

Zhang, D., Zhao, Y., Noble, J. H., and Dawant, B. M. (2018). Selecting electrode configurations for image-guided cochlear implant programming using template matching. *Journal of medical imaging*, 5(2):021202–021202.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890.

Zhao, Y., Chakravorti, S., Labadie, R. F., Dawant, B. M., and Noble, J. H. (2019). Automatic graph-based method for localization of cochlear implant electrode arrays in clinical ct with sub-voxel accuracy. *Medical image analysis*, 52:1–12.

Zhao, Y., Dawant, B. M., and Noble, J. H. (2016). Automatic selection of the active electrode set for image-guided cochlear implant programming. *Journal of Medical Imaging*, 3(3):035001–035001.

Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P. H., et al. (2021). Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.