Fall 2018

Multi-scale convolutional neural networks for segmentation of pulmonary structures in computed tomography

Sarah E. Gerard
University of Iowa

MULTI-SCALE CONVOLUTIONAL NEURAL NETWORKS FOR SEGMENTATION OF PULMONARY STRUCTURES IN COMPUTED TOMOGRAPHY

by

Sarah E. Gerard

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Biomedical Engineering
in the Graduate College of
The University of Iowa

December 2018

Thesis Supervisor: Professor Joseph M. Reinhardt

To my family, Michael, Diane, Alex, Betty, and Wilma -
without whom this work would not have been possible.

All the knowlege is in the connections

David Rumelhart

ACKNOWLEDGEMENTS

This work would not have been possible without the mentorship and collaboration of many professors and graduate students. First and foremost, I would like to express my sincere gratitude to my PhD advisor Joseph M. Reinhardt. I have had the privilege of working with Professor Reinhardt for many years, starting in my undergraduate studies where he provided guidance as my academic mentor and professor, through five and a half years of graduate studies. Professor Reinhardt's expertise in medical image analysis, ability of asking interesting research questions, and patience with graduate students have been instrumental in my development as a researcher. I will forever be grateful for the mentorship I have received from him; in research, career, and life. I could not imagine a better mentor to guide me through my PhD.

Next, I would like to sincerely thank Gary E. Christensen, Hans J. Johnson, John E. Bayouth, and Eric A. Hoffman for serving on my PhD committee. I would like to acknowledge Professor Christensen and Professor Johnson, who have provided invaluable mentorship from the beginning of my graduate studies. Their willingness to help and commitment to teaching has been truly admirable; this has greatly influenced my decision to pursue a career in academia. I am thankful for the many graduate students have contributed to this work: Taylor Patton, Yue Pan, Wei Shao, Jacob Herrmann, and Sandeep Bodduluri. A special thanks to Taylor Patton, for providing his expertise on physics of imaging and radiation therapy, as well as constant support and encouragement of my research abilities. I am very grateful for our daily insightful

discussions via Google Hangouts.

Several research grants and scholarships have provided financial support throughout my graduate studies. The Wisconsin functional avoidance radiation therapy grant (NIH CA166703) made much of this work possible. The University of Iowa Presidential Graduate Research Fellowship and Seashore and Ballard Dissertation Year Fellowship have provided person financial support through the entire duration of my graduate studies. The NASA Iowa Space Grant Consortium Fellowship has further provided personal financial support.

Last but not least, I am thankful for the unconditional support of my close friends and family. Many thanks to my friends, Jacob, Srivats, Pratik, and Sid, for the insightful conversations, as well as the many laughs and memories. I am forever grateful for my parents, Michael and Diane, who will always be my biggest role models; I would not be where I am today without their support and guidance. I am sincerely thankful for my favorite brother, Alex, whom I have had the great pleasure of attending undergraduate and graduate school with; I will forever cherish these experiences. Lastly, I am grateful for my puppies, Betty and Wilma, for all the much-needed cuddles.

ABSTRACT

Computed tomography (CT) is routinely used for diagnosing lung disease and developing treatment plans using images of intricate lung structure with submillimeter resolution. Automated segmentation of anatomical structures in such images is important to enable efficient processing in clinical and research settings. Convolution neural networks (ConvNets) are largely successful at performing image segmentation with the ability to learn discriminative abstract features that yield generalizable predictions. However, constraints in hardware memory do not allow deep networks to be trained with high-resolution volumetric CT images. Restricted by memory constraints, current applications of ConvNets on volumetric medical images use a subset of the full image; limiting the capacity of the network to learn informative global patterns. Local patterns, such as edges, are necessary for precise boundary localization, however, they suffer from low specificity. Global information can disambiguate structures that are locally similar.

The central thesis of this doctoral work is that both local and global information is important for segmentation of anatomical structures in medical images. A novel multi-scale ConvNet is proposed that divides the learning task across multiple networks; each network learns features over different ranges of scales. It is hypothesized that multi-scale ConvNets will lead to improved segmentation performance, as no compromise needs to be made between image resolution, image extent, and network depth. Three multi-scale models were designed to specifically target segmentation of three pulmonary

structures: lungs, fissures, and lobes.

The proposed models were evaluated on a diverse datasets and compared to architectures that do not use both local and global features. The lung model was evaluated on humans and three animal species; the results demonstrated the multi-scale model outperformed single scale models at different resolutions. The fissure model showed superior performance compared to both a traditional Hessian filter and a standard U-Net architecture that is limited in global extent.

The results demonstrated that multi-scale ConvNets improved pulmonary CT segmentation by incorporating both local and global features using multiple ConvNets within a constrained-memory system. Overall, the proposed pipeline achieved high accuracy and was robust to variations resulting from different imaging protocols, reconstruction kernels, scanners, lung volumes, and pathological alterations; demonstrating its potential for enabling high-throughput image analysis in clinical and research settings.

PUBLIC ABSTRACT

Convolution neural networks (ConvNets) are largely successful at performing image segmentation with the ability to learn complex and discriminative patterns in image data. Constraints in hardware memory do not allow deep networks to be trained with large images, such as volumetric medical images. Restricted by memory constrains, current applications of ConvNets on volumetric medical images use a subset of the full image; limiting the capacity of the network to learn informative global patterns. Local patterns, such as edges, are necessary for precise boundary localization, however, they suffer from low specificity. Global information can help disambiguate structures that have similar local appearance. The central thesis of this doctoral work is that both local and global image patterns are important for segmentation of anatomical structures in medical images. A novel multi-scale ConvNet is proposed that divides the learning task across multiple networks; each network learns features over different ranges of scales. Three multi-scale models were designed to target segmentation of three pulmonary structures: lungs, fissures, and lobes. The results demonstrated that multi-scale ConvNets improved pulmonary CT segmentation by incorporating both local and global features using multiple ConvNets within a constrained-memory system. Overall, the proposed pipeline achieved high accuracy and was robust to variations resulting from different imaging protocols, reconstruction kernels, scanners, lung volumes, and pathological alterations; demonstrating its potential for enabling high-throughput image analysis in clinical and research settings.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

1.1 Motivation and Specific Aims

Medical imaging applies principles of electromagnetic radiation and inverse problem solving to noninvasively reconstruct spatially varying physical properties of the anatomy. Medical imaging is an invaluable tool for diagnosing diseases, developing and guiding treatment interventions, and monitoring progression of disease. There are various imaging modalities which make use of physical phenomena to measure structural or functional information inside the body. Computed tomography (CT) imaging is commonly used to image the lungs as it is capable of producing high-resolution images with soft tissue contrast. Technological advancements in CT imaging have made it possible to generate images with submillimeter spatial resolution in less than a second. Furthermore, the emergence of 4DCT imaging enables acquisition of temporally resolved images for tracking thoracic motion during breathing. With the increasing prevalence of lung disease there is an increase in the number of thoracic CT scans being acquired in clinical and research environments. This provides a rich source of information for characterizing lung disease and its progression. However, the amount of data that is being produced imposes a high demand on radiologists to manually read and extract relevant information in these large scans. Computer-aided algorithms have great potential in assisting radiologists in efficiently processing and fully utilizing the information present in CT images.

Segmentation of the region of interest is an initial step for analyzing CT

images. While manual segmentation is possible, it is prohibitively time-consuming, tedious, and subject to intra- and inter-observer variations. Automated segmentation is critical to foster efficient workflows in clinical and research settings. There has been extensive efforts in developing automated segmentation algorithms for lung, fissures, and lobes in thoracic CT scans. The majority of these methods are rule-based systems which consist of a pipeline of steps targeting a specific anatomy. These methods are generally evaluated on small datasets of 10-30 scans from one study and may not be robust to the variations seen in large datasets, e.g. multi-institutional clinical trials. Different scanners, image reconstruction kernels, and imaging dose can greatly effect the appearance of pulmonary CT images. Designing rule-based systems that are robust to all these factors has been a major challenge.

Recently, in the computer vision field there has been a paradigm shift from designing rule-based algorithms to allowing computers to learn from data without being explicitly programmed. Deep learning using convolutional neural networks (ConvNets) have been successful at solving tasks in computer vision including image classification, object detection, and segmentation. The majority of the development and application of these tools has been on natural 2D images. A major barrier to applying these techniques to medical images is the size of 3D datasets and ability to train on full images given limitations in GPU memory. The majority of ConvNet methods for medical imaging applications train on 2D slices or local 3D patches. Such methods sacrifice global context and spatial smoothness in 3D. Furthermore, patch-based approaches are not as efficient since the model needs to be evaluated for

all patches in the image. The hierarchical design of ConvNets give them the ability to learn multi-scale features with different levels of abstraction. This is a distinguishing feature which makes ConvNets so powerful compared to feature engineering and rule-based methods. However, training ConvNets using 2D slices or 3D patches greatly limits the global context that can be learned.

For segmentation of the pulmonary anatomy local appearance is necessary for precision, however, it is not sufficient for discriminating various structures and textural patterns. Local appearance of injured lung can be indistinguishable from surrounding soft tissue without the anatomical cues from the ribcage. Many image features can locally resemble the characteristic plate-like appearance of fissures. However, when viewing the entire lung the fissure can be readily identified using prior knowledge of the fissure orientation and proximity to blood vessels. Global context from the entire image is critical for learning anatomical variations in shape and relative location of the anatomy.

The overarching theme of this work is to design ConvNet models which enable learning of global and local features for segmentation in large medical images. **The motivation of this work is to allow for high-throughput regional image analysis of lungs in thoracic CT scans. This is accomplished through development and extensive evaluation of a pipeline consisting of a series of multi-scale ConvNets for pulmonary segmentation in CT images.** The following **specific aims** were achieved in this doctoral thesis:

1. Segmentation of lungs across multiple species and pulmonary diseases.

2. Segmentation of pulmonary fissures in COPD and lung cancer subjects.

3. Segmentation of pulmonary lobes in COPD and lung cancer subjects.

1.2 Thesis Organization

This dissertation is presented in manuscript style, therefore each chapter can be read independently. As a result, there is some repetition between chapters for completeness. The contributions of this dissertation are presented in Chapters 4, 5, 6,and 7 and Appendices A and B of the Appendix. Chapters 4- 7 are works which fall under the multi-scale ConvNet theme of this dissertation. Appendices A and B are preliminary works that fall under the rule-based segmentation category. Below is a brief summary of remaining chapters is provided.

Chapters 2 and 3 present background information relevant to this dissertation work. Chapter 2 presents background information on physiology of the respiratory system, CT imaging of the lungs, and clinical trials which use imaging to study pulmonary diseases. Chapter 3 presents a brief overview of deep learning models including artificial neural network (ANNs) and convolutional neural networks (ConvNets).

Chapter 4 introduces a novel multi-resolution ConvNet for 3D image segmentation. The model was designed to enable ConvNet models to first learn global features and then learn local features in large medical images. The model was applied to segmentation of pathological lungs in human datasets. Extensive evaluation of the method was performed on a diverse dataset of 899 CT images of subjects with COPD, IPF, and lung cancer. The method achieved high performance and repeatable quantitative computed tomography (qCT) measurements when compared to manual

segmentations.

Chapter 5 extends the multi-resolution ConvNet model presented in Chapter 4 to segment lungs in three animal species with severe lung injury. Transfer learning was utilized to exploit features learned from the human trained model. The multi-resolution ConvNet showed superior performance compared to single resolution models. The proposed method was evaluated on four datasets consisting of three animal species using a five fold cross-validation.

Prepared for submission to Medical Image Analysis. Modified from: ***SE Gerard****, J Herrmann, DW Kaczka, JM Reinhardt: Transfer Learning for Segmentation of Injured Lungs Using Coarse-to-Fine Convolutional Neural Networks. Image Analysis for Moving Organ, Breast, and Thoracic Images, 2018.* [47]

Chapter 6 presents a novel ConvNet-based model for fissure segmentation. The model initially learns a fissure region of interest (ROI) and subsequently refines the precise fissure location within the ROI. This model is designed to alleviate the class-imbalance between fissure voxels and non-fissure voxels. The proposed method was extensively evaluated on 7412 images from COPDGene and 20 images from 4DCT scans of lung cancer subjects.

Published in: ***SE Gerard****, TJ Patton, GE Christensen, JE Bayouth, JM Reinhardt: FissureNet: A Deep Learning Approach for Pulmonary Fissure Detection in CT images. IEEE Trans Med Imaging, 2018.* [49]

Chapter 7 presents a multi-resolution ConvNet model for lobe segmentation. The model has the same design as the multi-resolution model proposed in Chapter

5. To allow for aggressive downsampling without loss of the fissure signal, the fissure prediction from the model proposed in Chapter 7 serves as an input to the lobe segmentation model. The proposed method was evaluated on 1152 images from the COPDGene dataset.

Submitted as: ***SE Gerard*** *and JM Reinhardt: Pulmonary Lobe Segmentation Using a Sequence of Convolutional Neural Networks for Marginal Learning. International Symposium on Biomedical Imaging (ISBI), 2019.*

Chapter 8 presents an overall discussion and conclusion to this dissertation work.

Appendix A presents a 4D lung segmentation algorithm. The method initially performs a rough registration to align all phases of a 4DCT scan. A 4D optimal surface finding graph search is used which incorporates temporal constraints.

Published in: ***SG Yeary****, GE Christensen, JE Bayouth, S Bodduluri, Y Pan, J Guo, K Du, JH Song, B Zhao, I Oguz, JM Reinhardt: 4D Lung CT Segmentation for Radiation Therapy Applications. ICART: Imaging and Computer Assistance in Radiation Therapy, 2015.* [159]

Appendix B presents a segmentation algorithm for inclusion of large tumors. The proposed method starts with an intensity-based segmentation algorithm which identifies normal lung tissue. Alpha shapes are applied to the intensity-based segmentation to include large tumors.

Published in: ***SE Gerard****, HJ Johnson, JE Bayouth, GE Christensen, K Du, J Guo, JM Reinhardt: Alpha shapes for lung segmentation in the presence of Large Tumors.*

6^{th} *International Workshop on Pulmonary Image Analysis, 2016.* [48]

CHAPTER 2

BACKGROUND

2.1 Respiratory System

The function of the respiratory system is to facilitate the transfer of oxygen and carbon dioxide between an organism and the environment. The respiratory system consists of lungs, airways, pulmonary vasculature, and respiratory muscles (Figure 2.1). The lungs are located in the chest cavity and are surrounded by the rib cage for protection. Humans have a left and a right lung which are enclosed in a double layer membrane called the pleural sac. Invagination of the inner membrane forms the pulmonary fissures which anatomically separate the lungs into five lobes. The left lung has two lobes, the lower and upper lobe which are separated by the oblique fissure. The right lung has three lobes, the lower and middle lobe are separated by the oblique fissure and the middle and upper lobe are separated by the horizontal fissure. The accessory fissures further separate the lobes into bronchopulmonary segments. Each lobe is supplied by separate vasculature and airway trees.

The pulmonary airways form a tree like structure consisting of approximately twenty-three generations of branching in humans. The trachea is the primary conduit that transfers air between the environment and the lungs. The trachea splits into a left and right main bronchi which enter the left and right lung, respectively. Within the lungs the main bronchi subdivide into lobar bronchi, segmental bronchi, bronchioles, terminal bronchioles, respiratory bronchioles, and finally terminate with alveolar sacs containing tiny hollow structures called alveoli. The alveolus is the functional unit

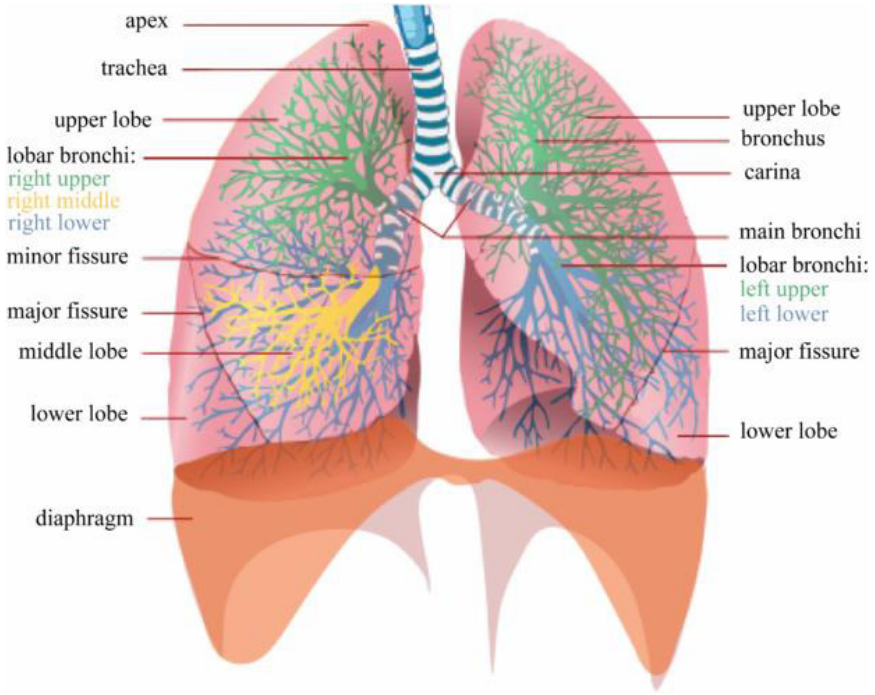


Figure 2.1. Anatomy of the human respiratory system. Image taken from [147].

of the lung where gas exchange occurs. The capillaries are a network of small blood vessels that surround the alveoli, through diffusion oxygen is transferred from the alveoli to the capillary blood and carbon dioxide is transferred from the blood to the alveoli.

The lungs have their own vascular system which allows blood to flow from the heart, to the lungs, and back to the heart, a process called pulmonary circulation. Deoxygenated blood leaves the heart through the pulmonary artery and travels to the lungs. Within the lungs the blood travels through arteries, arterioles, and capillaries where it becomes oxygenated. The oxygenated blood travels through venules, veins, and finally back to the heart via the pulmonary vein. The heart supplies the oxygen rich blood to the rest of the body, a process called systemic circulation.

2.2 Computed Tomography

Computed tomography (CT) imaging is used to create high-resolution and high-contrast volumetric images of the pulmonary anatomy. This enables the visualization of the intricate lung structures, such as vasculature and airways, as well as the heterogeneous textures of the lung parenchyma. A CT image is produced by measuring the attenuation of X-rays at different angles around a body. Each angle produces a projection of the anatomy, which can be used to reconstruct a 3D image. A CT image has intensity values of Hounsfield Units (HU). HU is calculated by applying a linear transformation to the measured linear attenuation coefficient, such that distilled water is 0 HU and air is -1000 HU. High-density tissues attenuate the X-rays more than low-density tissues; tissues with high attenuation appear bright, such as bone and blood, and tissues with low attenuation appear dark, such as air. Figure 2.2 displays the three standard anatomical cross-sections of a CT image: axial, coronal, sagittal.

Thoracic CT scans are commonly acquired while the subject is performing a breath-hold maneuver. Standard lung volumes are commonly used such as total lung capacity (TLC), functional residual capacity (FRC), or residual volume (RV). If multiple scans are acquired at different lung volumes for a given subject, image registration can be used to obtain quantitative measurements of lung dynamics [155], lung mechanics [3], and regional lung ventilation [31]. Parametric response mapping (PRM) also utilizes image registration and is used to characterize different phenotypes of COPD by measuring the extent of functional small airways disease (fSAD) and emphysema [43].

The use of four-dimensional CT (4DCT) has increased greatly since the early 2000s [34] as it is capable of imaging the lungs over the respiratory cycle providing spatial-temporal information. Keall et al. [66] define 4DCT imaging as "The acquisition of a sequence of CT image sets over consecutive segments of a breathing cycle". The use of 4DCT during radiotherapy planning allows consideration of temporal changes of the anatomy caused by intrafraction motion, or anatomical motion during a treatment session due to breathing, and thus greater precision of treatment. A 4DCT scan is acquired while the subject is breathing and the respiratory trace is simultaneously being recorded. A schematic of an idealized respiratory trace is illustrated in Figure 2.3. A strain gauge belt or a reflective cube on the chest wall are used as an external surrogate for respiratory volume. Retrospectively-gated 4DCT sorts the image data into different "phase" or "amplitude" bins post-acquisition using the external respiratory signal. Amplitude-based sorting has been shown to produce less artifacts for nonperiodic motion [154].

2.3 Respiratory Pathologies

Respiratory pathologies are the number three leading cause of death in the United States [40]. Subjects afflicted with pulmonary disease have difficulty breathing normally. There are many types of diseases which affect the lungs, including obstructive diseases like chronic obstructive pulmonary disorder (COPD) which lead to difficulties in expelling air in the lungs; and restrictive diseases like interstitial lung disease (ILD) which leads to stiffening of the lung tissue. Figure 2.4 depicts the appearance of different lung diseases in cross-sectional views of CT images. The motivation of this

(a) Axial (b) Coronal (c) Sagittal

Figure 2.2. Standard cross-sectional views of a volumetric thoracic CT scan. CT images are oriented with the left lung on the right side of the images in axial and coronal cross-sections.

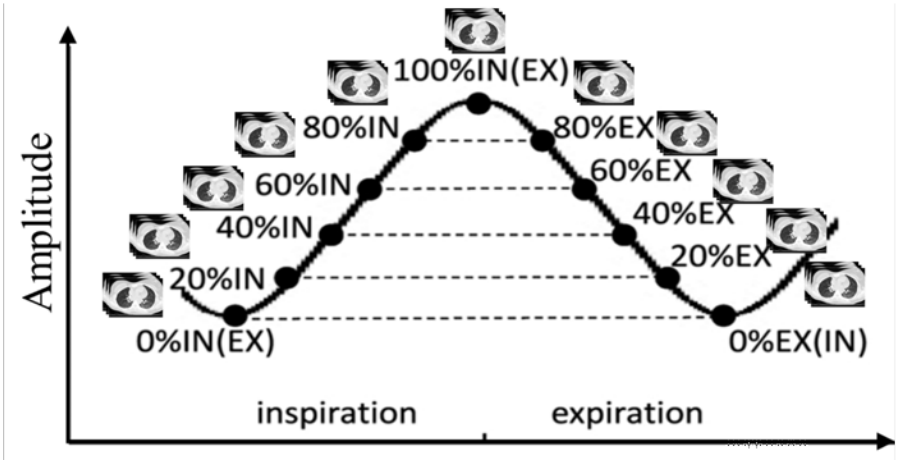
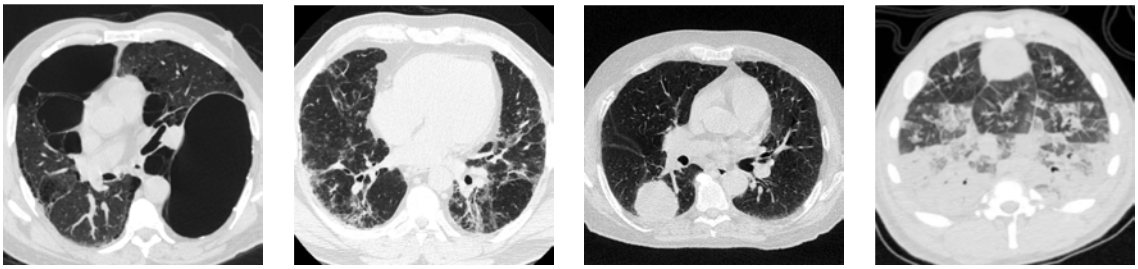


Figure 2.3. Schematic of idealized external respiratory signal with amplitude-sorted bins of 4DCT image.

work is to enable high-throughput regional analysis on large-scale pulmonary CT datasets with pathologies. This work includes five large-scale datasets: lung cancer clinical trial (NIH CA166703) [101], COPDGene clinical trial [115], SPIROMICS [23, 129, 22], PANTHER-IPF [93, 94], and acute respiratory distress syndrome study with multiple animal species study.



(a) COPD (b) IPF (c) Lung Cancer (d) ARDS

Figure 2.4. Axial slices of CT scans of various pulmonary diseases and injuries.

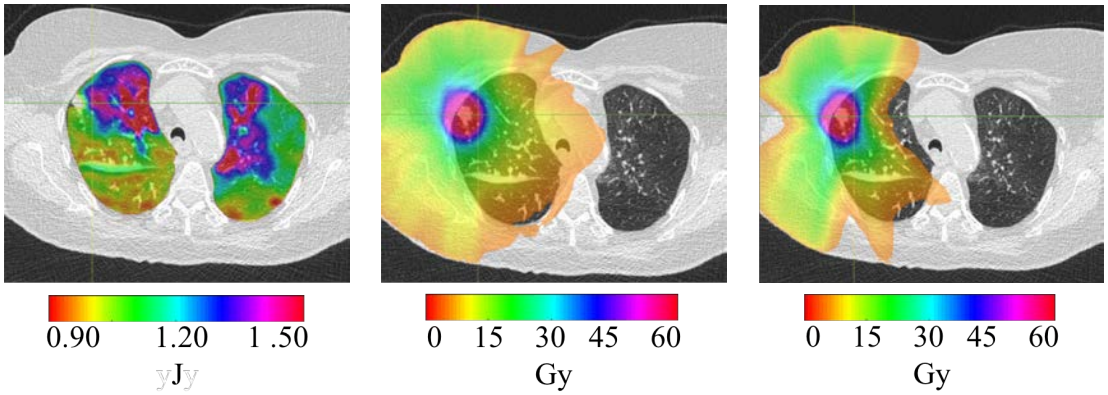
2.3.1 Lung Cancer

Lung cancer is the leading cause of cancer related deaths world wide, accounting for 1.59 million deaths annually. In the United States it is estimated that in 2016 alone there will be nearly 225,000 new cases and 158,000 deaths due to lung and bronchus cancer [1]. The five-year survival rate is 54 percent if detected at an early stage but only 4 percent for advanced stages. Unfortunately, only 15 percent of lung cancer cases are detected at an early stage. Early detection and treatment can greatly increase the chances of survival. Much effort has been put into lung cancer screenings using computed tomography (CT) imaging for high risk individuals such smokers.

Screening with low-dose CT scans has shown to increase the chance of survival for heavy smokers [2]. CT imaging is used in all stages of lung cancer screening, detection, diagnosis, and treatment. With the increasing resolution and quantity of CT images there is a high demand for computer-based analysis systems to automatically extract quantitative measurements from the large datasets.

The lung cancer dataset used in this study is from a clinical trial (NIH CA166703) which is using functional avoidance radiation therapy for lung cancer treatment. Radiation therapy is used on approximately 85 percent of lung cancer patients to help manage lung cancer. Radiotherapy uses high energy X-rays to kill cancer cells, however, the X-rays also damage healthy cells. The effectiveness of radiation therapy has been shown to be superior when high doses are administered. However, current protocols limit the amount of dose to sub-therapeutic levels to avoid lung toxicity. Currently, lung toxicity is evaluated based on the dose-volume relationship of the lung tissue being treated. This simple metric ignores the complex interplay of the spatial and temporal heterogeneity of lung function and anatomy and its response to dose. The novelty of this clinical trial is that high functioning regions are avoided in the dose plan. Functional information is obtained from performing image registration on the different volumes of a 4DCT image. Image registration produces a dense transformation that matches corresponding points between two images. The determinant of the Jacobian matrix of the transformation (J) represents the local volumetric expansion or contraction. The determinant of the Jacobian is used as a surrogate for lung ventilation. While this method ignores the perfusion component

of lung function, these metrics are readily obtainable as a 4DCT scan is necessary for conventional treatment. Therefore, there is no additional imaging or cost to the patient. Figure 2.5 shows a Jacobian image, which can be used as a surrogate for lung function [116, 29], with the conventional RT dose plan and the functional avoidance dose plan. The high functioning region in the anterior right lung receives high dose in the conventional plan, however, this region is avoided in the avoidance plan.



(a) Jacobian (b) Conventional (c) Functional Avoidance

Figure 2.5. Jacobian and radiation therapy treatment plans.

In total, 120 subjects will be enrolled in the clinical trial with 60 randomized to the control arm (conventional RT) and 60 subjects randomized to the experimental arm (function avoidance RT). For each subject, two 4DCT scans are acquired at baseline prior to RT, which are used to develop the treatment plan. The two scans are acquired in order to evaluate repeatability of measurements; with the assumption that the anatomy and function should not change between the two scans. Duplicate

scans are acquired post-treatment at 3 month, 6 months and 9 months. Therefore, each subject has eight 4DCT scans. Each 4DCT scan consists of ten 3D images. This produces a huge amount of data which needs to be analyzed: 9600 3D high-resolution images.

2.3.2 Chronic Obstructive Pulmonary Disease

In 2008, COPD was the third leading cause of death in the United States [40]. The main cause of COPD is smoking, however, smoking is neither necessary nor sufficient to develop COPD [8, 106, 63]. COPD is a combination of emphysema and chronic bronchitis [89]. Emphysema results from the destruction of the alveoli and chronic bronchitis occurs when the lining of the bronchial tubes become inflamed. Radiographically, emphysema is characterized by large airspaces resulting from parenchymal tissue destruction, and chronic bronchitis is characterized by airway wall thickening and enlarged blood vessels. The extent and spatial patterns of emphysema and air trapping can be measured from CT images.

COPDGene is a large multi-center clinical trial with over 10,000 subjects enrolled [115]. CT images were acquired across 21 imaging centers using a variety of scanner makes and models. Each subject had two breath-hold 3D CT scans acquired, one at TLC and one at FRC. A subset of of subjects had an additional RV scan.

SPIROMICS is a large multi-center clinical trial studying subpopulations and intermediate outcomes of COPD subjects [23, 129, 22]. CT images of subjects in this study were acquired at TLC and RV.

2.3.3 Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease that leads to scarring of the lung tissue. Subjects with IPF have difficulty taking deep breaths and cannot get enough oxygen to the lungs. IPF is a form of ILD, i.e., it is a disease that affects the interstitium.

The IPF dataset was obtained from an ancillary study of PANTHER-IPF [93, 94]. The ancillary study study image derived IPF textural patterns in CT images and their relations to disease progression [122].

2.3.4 Acute Respiratory Distress Syndrome

Acute respiratory distress syndrome (ARDS) is a type of respiratory failure that leads to inflammation and fluid accumulation in the lungs [41]. Radiographically this condition presents with diffuse bilateral consolidation in the dependent lung region as shown in Figure 2.4(d). ARDS subjects require mechanical ventilation for survival until recovery. CT imaging can be used for diagnosis of ARDS and to perform quantitative analysis of spatial aeration during mechanical ventilation.

A multi-species dataset consisting of animal models of ARDS was used in this work. The dataset consists of canine, porcine, and ovine with lung injury mimicking ARDS. The porcine subjects have both 3DCT scans at constant pressures and 4DCT phase-gated scans acquired during different mechanical ventilation protocols.

CHAPTER 3

DEEP LEARNING AND NEURAL NETWORKS

Machine learning is a form of data analytics which enables computers to learn from data rather than being explicitly programmed. Machine learning can be divided into supervised learning and unsupervised learning; supervised learning uses labeled training data and aims to find the mapping between inputs and label, whereas unsupervised learning aims to find patterns and clusters in unlabeled data.

Traditional machine learning uses explicitly defined features extracted from raw data. These features are manually engineered based on domain knowledge of the problem at hand and are typically limited to low-level features. Many research efforts have worked on the design of informative features in images. Examples of such image features include scale-invariant feature transform (SIFT) [84], histogram of oriented gradients (HOG) [26], Gabor filters [87], and local binary patterns (LBP) [98].

Deep learning is a subgroup of machine learning algorithms which combines the feature extraction and the output prediction in one integrated system. This eliminates the need to design features and instead enables automated learning of features which are most useful mapping inputs to outputs. This key difference between traditional machine learning and deep learning is illustrated in Figure 3.1. Neural networks are the primary model used for deep learning frameworks, which have a hierarchical design allowing for features with different levels of abstraction to be learned. This chapter will introduce artificial neural networks (ANNs), convolutional

neural networks (ConvNets), popular ConvNet architectures, and the process of "training" neural networks.

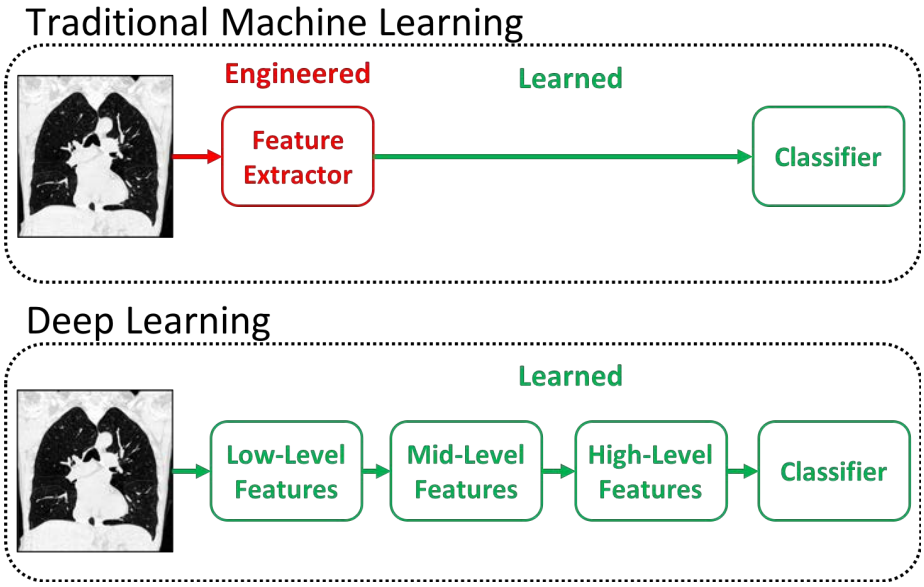


Figure 3.1. Comparison of traditional machine learning and deep learning.

3.1 Artificial Neural Network

A multilayer perceptron (MLP) or artificial neural network (ANN) is a computation model that was inspired by the biological neural networks in animal brains. The fundamental unit of an ANN is a neuron. Each neuron has a set of inputs and the neuron computes a weighted sum of its inputs and applies a nonlinear function to produce an output. Multiple neurons are organized into layers and layers are stacked hierarchically, i.e., the output of layer i is the input to layer $i+1$. Each layer can be represented as a vector and the weights connecting layers i and $i+1$ can

be represented as a matrix where each element represents the connection strength between each neuron in a layer i to each neuron in layer $i+1$. The output of the layer can efficiently be computed as a matrix-vector multiplication between the weight matrix and the vector input producing a vector output. A schematic of an ANN with four input units, two hidden layers with five neurons each, and an output layer with two units is illustrated in 3.2. The functional form of this neural network is

$$f(x, W) = W_2(\sigma(W_1(\sigma(W_0 x)))), \tag{3.1}$$

where x is the input vector, $\sigma(\cdot)$ is a nonlinear function, and W is the set of weights consisting of W_0, W_1,and W_2, which are weight matrices for the first, second, and third layers, respectively. The number of hidden layers and number of neurons in each hidden layer are hyperparameters which are experimentally determined; more hidden layers and hidden neurons results in more free parameters used to fit the model.

3.2 Convolutional Neural Networks

Convolutional neural networks (ConvNets or CNNs) are a specialized neural network designed for learning patterns in spatially correlated data, such as images and videos. The input to a ConvNet has is multidimensional tensor with elements that are spatially correlated. These spatial relationships are preserved throughout the ConvNet. A ConvNet consists of a hierarchy of layers. Each layer i takes an input image representation I_i and transforms it to an output image representation I_{i+1}. The output of a layer serves as the input to the next layer in the hierarchy. I_i and I_{i+1} are both 4-dimensional (4D) tensors, with three spatial dimensions and

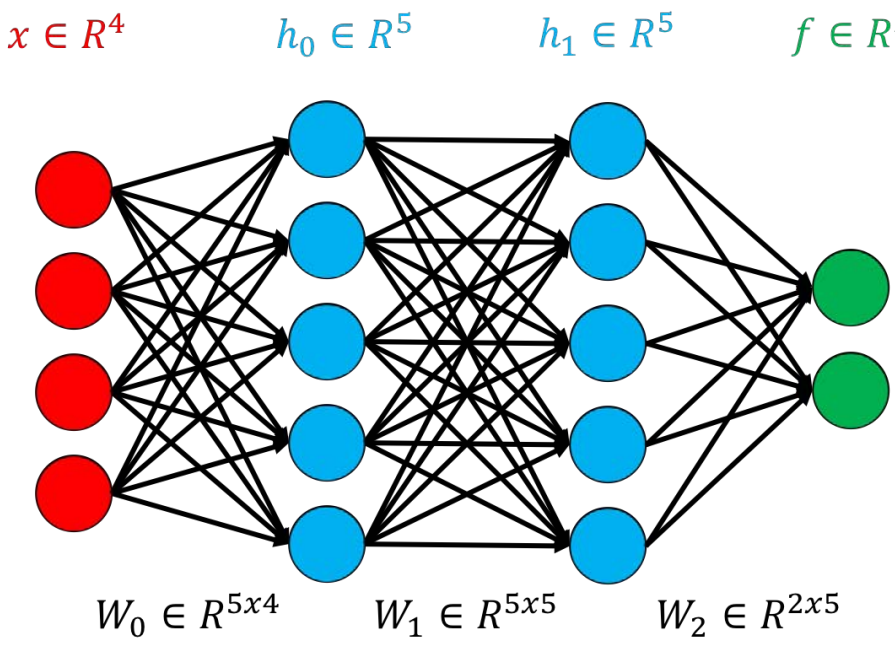


Figure 3.2. Multi-layer perceptron with two hidden layers each with five neurons.

one channel dimension. The channel dimension represents different feature types and is analogous to the red, green, and blue channels in a RGB image. The sizes of I_i and I_{i+1} are $X_i \times Y_i \times Z_i \times C_i$ and $X_{i+1} \times Y_{i+1} \times Z_{i+1} \times C_{i+1}$, respectively, where X, Y, Z are the spatial dimensions and C is the channel dimension. There are different layer types which perform various transformations, the most common layer types are: convolutional, nonlinearity, pooling, and transposed convolution. Each layer type has a set of learned parameters and a set of hyperparameters. The learned parameters will be optimized during training. The hyperparameters are fixed parameters that do not change.

A convolutional layer performs image convolution with a kernel and the input image. The kernels are also 4D tensors, with local spatial extent and a channel extent

which equals the input channel size. The kernel can be thought of as a feature detector and the convolution of the kernel with the image produces a spatial activation map of that feature. This is repeated with N kernels, each producing a 3D tensor. The 3D tensors are concatenated along the channel dimension to produce the output 4D tensor with $C = N$ channels. If zero-padding and convolution with a stride of one are used, a convolutional layer preserves the spatial size of the input, i.e., $X_{i+1} = X_i$, $Y_{i+1} = Y_i$, and $Z_{i+1} = Z_i$. The learned parameters are the kernels, therefore the features detectors are being learned rather than designed. The convolution operation gives the local connectivity and weight sharing properties to ConvNets, which distinguishes them from MLPs. The hyperparameters in a convolutional layer are the number of kernels N_i, the spatial extent of each kernel, and the stride of the kernel.

Pooling layers are used to reduce the spatial size of an image representation. The channel dimension size is preserved through a pooling layer, i.e., $C_{i+1} = C_i$. Pooling layers have no learned parameters. The hyperparameters in a pooling layer are the kernel size, the stride, and the pooling function. It is common to have a kernel size of $2 \times 2 \times 2$ and stride of $2 \times 2 \times 2$, which results in downsampling by a factor of two along each spatial dimension, i.e., $X_{i+1} = \frac{X_i}{2}$, $Y_{i+1} = \frac{Y_i}{2}$, and $Z_{i+1} = \frac{Z_i}{2}$. Max-pooling and mean-pooling are common pooling functions.

Nonlinearity layers are used after each convolutional layer and perform an elementwise nonlinear function to an image representation. Nonlinearity layers do not change the size of the image representation. There are no learned parameters in a nonlinearity layer, and the only hyperparameter is the function type. The most

common nonlinearity functions, sigmoid, hyperbolic tangent (TanH), and rectified linear units (ReLu), are displayed in Figure 3.3.

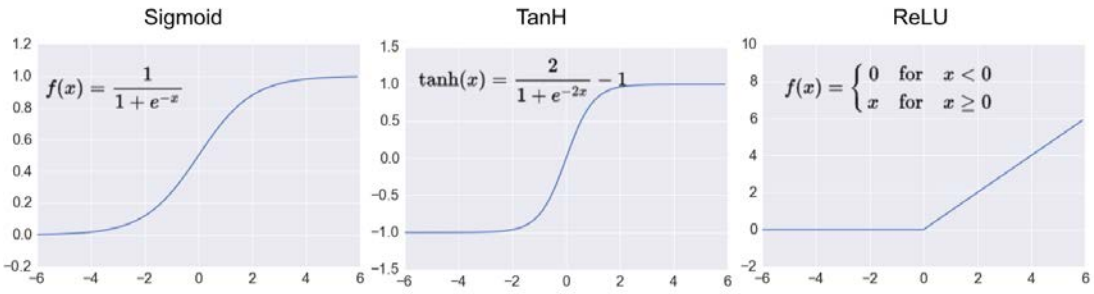


Figure 3.3. Common nonlinearity functions utilized in neural networks.

Transposed convolutional layers are used to achieve learnable upsampling. Segmentation networks produce an output image with the same size as the input image, so if pooling is used the image representation needs to be upsampled to the original size. Conventional upsampling uses interpolation with nearest neighbor, linear, bi-linear, or bi-cubic functions. Instead of defining the interpolation function, transposed convolution learns the interpolation through a reverse convolution operation. The learned parameters and hyperparameters in a transposed convolutional layer are the same as a convolutional layer.

In addition to the core layers discussed above, in recent years batch normalization [62] and dropout [134] have been highly successful and widely used. Batch normalization is used to make the training process more stable and prevent "zero gradients". Zero gradients refer to the partial derivatives of the loss function with respect to

the weights. When these gradients are close to zero the parameters do not get updated and learning cannot proceed. Batch normalization is commonly used after convolutional layers to normalized the output of the layer. Dropout layers are used to prevent overfitting on the training data. A predefined percentage of output neurons are randomly "dropped out", i.e., assigned a value of zero, during the training process. This is a similar idea to training an ensemble of models. Each pass through the network a new model is being trained.

3.3 Architectures

There are different architectures designed for classification, regression, and segmentation tasks. Classification architectures predict the probability that a input image belongs to each of $|Y|$ classes, therefore the number of neurons in the last layer is $|Y|$. To transform a multi-dimensional tensor to a vector with $|Y|$ neurons, these networks use fully-connected layers which are equivalent to a layer in an ANN. The first ConvNet architecture, LeNet [77], is illustrated in Figure 3.4, and is designed for classification tasks. Regression architectures predict a continuous value rather than a discrete class label. A regression architecture is the same as a classification architecture, however, there is just one output unit. Segmentation architectures predict the probability that each voxel belongs to each of $|Y|$ classes, therefore the number of neurons in the last layer is the input image size times $|Y|$. A popular segmentation architecture, U-net [119], is illustrated in Figure 3.5. The main difference between classification and regression architectures compared to segmentation architectures is segmentation architectures use learnable upsampling layers to recover the spatial

resolution lost from pooling layers. Furthermore, segmentation architectures do not use fully-connected layers.

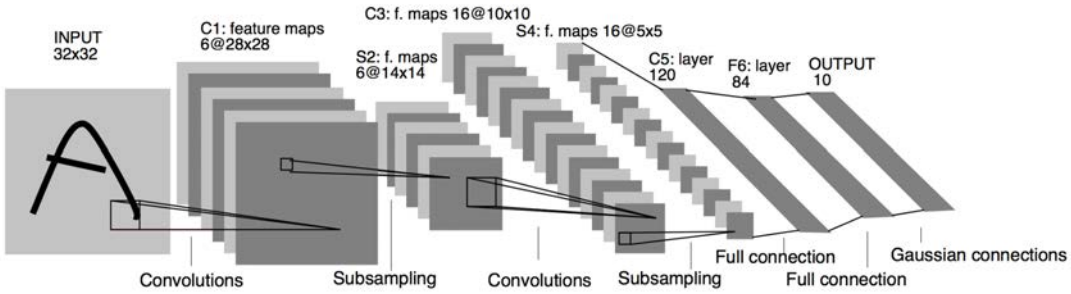


Figure 3.4. LeNet architecture from [77]

3.4 Supervised Training

Supervised training of a neural network requires training data which have inputs and corresponding target output labels. Training is the process of optimizing the weights of a neural network to find the mapping from inputs to outputs. A loss function is defined that represents the dissimilarity between the predicted output of the network and the desired output. The model is optimized to minimize the dissimilarity. This is essentially fitting a model to the training data points; a model which has millions of parameters. Training the network is a four step process: forward-pass, loss function evaluation, backward-pass, and weight update. The forward-pass propagates the input data through the layers of the network. The loss function compares the output of the network with the input data label. The backward-pass, or backpropagation, calculates the partial derivative of the loss function with respect

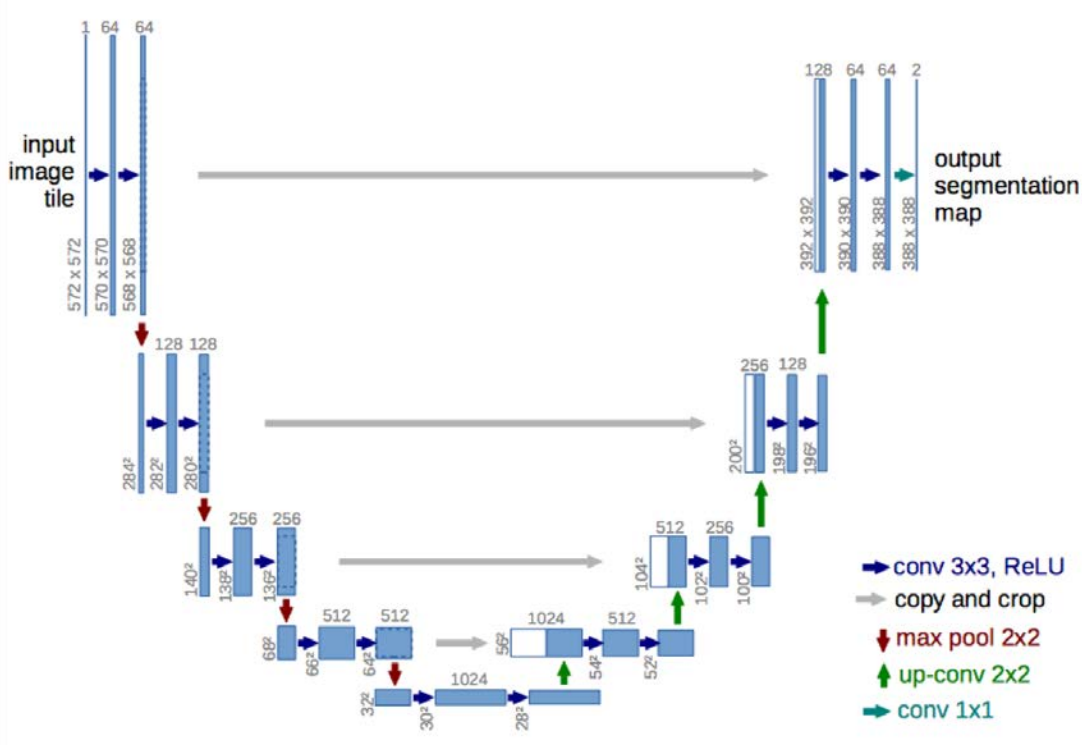


Figure 3.5. U-Net architecture from [119].

to each weight in the network. Backpropagation is a recursive application of the chain rule that allows for straightforward calculation of the analytical gradient of a complex function with millions of parameters. Finally, the weights are updated using gradient descent optimization, or some variant thereof.

3.5 Challenges

Deep learning using ConvNet models have been largely successful in computer vision. However, these techniques have been developed for applications in natural images. Many works are developed and evaluated on the publicly available ImageNet dataset. ImageNet is a large-scale dataset for evaluating object detection and image classification algorithms. The dataset consists of over 14 million RGB images with

annotated labels belonging to more than 20 thousand classes.

Theoretically, ConvNets seamlessly extend from 2D networks to 3D networks for handling medical images. However, there are challenges that need to be addressed to allow practical usage of ConvNets in medical imaging. First, training ConvNets is performed on a graphics processing unit (GPU) card which have thousands of cores. This greatly increases the throughput of performing the matrix operations required for training neural networks. However, current GPU cards have a limited amount of memory; the memory demand for high-resolution volumetric images and deep ConvNets exceed the available GPU memory. Second, annotated medical images are not as readily available as natural images. Annotation of medical image requires expert analysts, therefore high-throughput techniques such as crowdsourcing the annotations are not available. Third, in medical datasets there is often a class-imbalance when dealing with rare diseases making it a challenge to learn patterns in the underrepresented class. Class-imbalance also often occurs for segmenting small structures in large images. In the remaining chapters we address these challenges to enable usage of deep learning with ConvNet models towards segmentation of pulmonary structures in CT images.

CHAPTER 4

LUNG SEGMENTATION IN HUMANS

Lung segmentation is a necessary pre-processing step in quantitative computed tomography (qCT) analysis. Developing segmentation algorithms to robustly handle different lung pathologies has been a challenging endeavor that has been studied for several decades due to its necessity in clinical and research settings. Deep learning frameworks using convolutional neural networks (ConvNets) have shown exceptional performance at performing tasks in computer vision including image classification and segmentation. In this work, a deep learning approach is used with a novel multi-scale ConvNet model for learning segmentation in large volumetric images. The proposed method was extensively evaluated on a diverse dataset consisting of 899 CT images of subjects with chronic obstructive pulmonary disorder (COPD), idiopathic pulmonary fibrosis (IPF), and lung cancer. The dataset consisted of images acquired at multiple lung volumes including total lung capacity (TLC), functional residual volume (FRC), residual volume (RV), and intermediate phases in 4DCT images. Overall, the proposed method achieved an average symmetric surface distance of 0.234 mm and a median Jacaard index of 0.984 when compared to manual segmentations.

4.1 Introduction

Computed tomography (CT) produces 3-dimensional (3D) reconstructions of the anatomy by measuring the attenuation of x-rays through a body at multiple angles. CT is the modality of choice for imaging the intricate structures of the lungs.

CT imaging produces high-contrast and high-resolution images which are necessary for characterization of anatomical alterations in lung parenchyma, vasculature, and airways. Technological advancements in CT scanners have enabled the acquisition of images with sub-millimeter slice thickness in less than a second. This produces large datasets with more than 500 slices that need to be reviewed. Automated image analysis pipelines are critical for extracting and understanding the information in these large datasets.

An initial step for automated pulmonary analysis is to generate a lung segmentation that distinguishes pulmonary tissue from non-pulmonary tissue. It is important that the lung segmentation includes both healthy and pathological regions so these regions are not excluded from quantitative analysis. Quantitative CT (qCT) can be used to characterize spatial patterns of lung aeration and extent of diseases such as emphysema [86] and asthma [95]. Parametric response mapping (PRM) uses image registration between inspiration and expiration images to quantify amount of emphysema and airways disease [43]. These methods rely on an accurate lung segmentation to produce reliable measurements. Image registration requires a lung segmentation to limit define where the cost function should be evaluated. Furthermore, varifold-based registration algorithms align surface representations which require a segmentation to produce [100].

The low-density lung tissue is easily separated from the surrounding dense soft tissue using simple threshold-based techniques [60, 54, 113, 6, 76]. The threshold can automatically be determined for each image with optimal thresholding [60]. The

trachea and main bronchi are removed from the threshold result using region growing with automatic seed detection. Morphological operations are subsequently utilized to fill any holes in the segmentation result.

Intensity-based lung segmentation algorithms are highly successful and widely used in both clinical and research settings as these methods are computationally efficient and easy to implement. However, these simplistic methods can be fragile and not well suited for highly diseased lungs. Intensity-based algorithms fail to include high-density pathologies, such as tumors and fibrosis. Peripheral pathologies are especially challenging since there can be little or no contrast between the pathology and the non-pulmonary tissue. Air in the stomach or intestines can erroneously be included in intensity-based methods; this is especially problematic when there is blurring near the diaphragm causing the air to appear connected to the lungs. Metal artifacts resulting from pacemakers can result in both over- and under-segmentation in the affected region.

To account for the short-comings of intensity-based methods, several model-based methods have been developed to target pathological lung segmentation. Atlas-based segmentation methods use image registration to map a known segmentation to the image to be segmented [144, 130, 78, 162]. A collection of atlases were used in [130] to form a probabilistic atlas to model the many variations in lung shape and appearance. Uncertain voxels in the probabilistic atlas are classified by a kNN classifier. A disadvantage of atlas-based methods is the image registration is time consuming and a difficult problem in itself. In [144] an automated lung segmentation error detection

algorithm was developed to detect failed segmentations by an intensity-based method. If an error is detected an advanced atlas-based segmentation was used.

Statistical shape models (SSMs) and active appearance models (AAMs) are another model-based segmentation method [19]. SSMs estimate shape variations from a training dataset and subsequently fit the model to test cases using local appearance information. A robust SSM was developed to segment pathological lungs in CT images [136]. In [153], a 4D SSM was developed to segment lungs in 4DCT images. While SSMs have been successful across a wide range of applications, they are sensitive to model initialization and can fail in cases with weak edges. A review of SSMs used in medical imaging analysis is given in [57].

Anatomical information is valuable for segmenting the lungs, especially in presence of dense pathologies. Curvature of the ribcage closely follows the curvature of the lung boundary, this information has been exploited for diseased lung segmentation [109, 88]. An anatomical model consisting of the chest wall, mediastinum, and large airways was developed to guide lung segmentation in [12]. In [131], anatomical landmarks are automatically detected at the carina, ribs and spine for initialization of a SSM.

Recently, the success of deep learning in medical imaging applications there has been a paradigm shift from using rule-based image analysis pipelines to learning directly from raw image data. Several works have proposed convolutional neural networks for lung segmentation in CT images [56, 5]. A limitation of these methods is the ConvNets are trained using 2D slices and not the full anatomy. Segmentation in 3D incorporates spatial smoothness and contextual information that can help

differentiate ambiguous regions on 2D slices.

In this work we propose a multi-scale ConvNet model for 3D segmentation of lungs in CT images. The multi-scale model consists of a series of ConvNets which are trained with images of increasing resolution. The motivation of this design was to allow ConvNets to learn both global contextual features and local high-resolution features in large volumetric medical images. The model was extensively evaluated on 899 CT images of subjects with chronic obstructive pulmonary disorder (COPD), idiopathic pulmonary fibrosis (IPF), and lung cancer.

4.2 Datasets and Reference Standards

4.2.1 COPDGene

The first dataset was obtained from the COPDGene clinical trial - a large multi-center trial studying genetics and imaging phenotypes in COPD subjects [115]. All subjects used in this study had scans acquired at total lung capacity (TLC) and functional residual volume (FRC) and a subset of subjects had scans acquired at residual volume (RV). The dataset consisted of 2888 CT images from 1945 subjects.

4.2.2 SPIROMICS

The second dataset was obtained from the SPIROMICS clinical trial - a multi-center trial studying subpopulations and intermediate outcomes of COPD subjects [23, 129, 22]. The dataset used in this work consisted of 225 CT images from 60 subjects acquired at TLC and RV, see [129] for full imaging protocol.

4.2.3 PANTHER-IPF

The IPF dataset was obtained from an ancillary study of PANTHER-IPF [93, 94]. The ancillary study study image derived IPF textural patterns in CT images and their relations to disease progression [122]. The dataset used in this work consisted of 305 images from 190 subjects. All datasets were acquired at TLC, see [122] for complete imaging protocol.

4.2.4 Lung Cancer

The lung cancer dataset was acquired in coordination with a large clinical trial studying functional avoidance radiation therapy (NIH CA166703) [101]. Each subject has two 4DCT scans prior to treatment and two scans at 3 months, 6 months and 9 months post-treatment. 4DCT were reconstructed using retrospective amplitude based sorting. Voice guidance was used to improve breathing repeatability. A subset of data with manual segmentations was used in this work consisting of 1620 3D images (162 4DCT images with 10 phases each) from 40 unique subjects.

4.2.5 Manual Segmentations

All manual lung segmentations were generated semi-automatically. Automated segmentations were initially generated using Pulmonary Analysis Software Suite (PASS, University of Iowa Advanced Pulmonary Physiomic Imaging Laboratory [55]). The automatically generated segmentations were edited by a human analyst when necessary.

4.3 Methods

4.3.1 Overview

A multi-scale ConvNet model is proposed for learning features with multiple scale ranges. The model consists of a series of two Seg3DNet architectures that are trained sequentially. Training and evaluation datasets consisted of human subjects with COPD, IPF, and lung cancer.

4.3.2 Seg3DNet

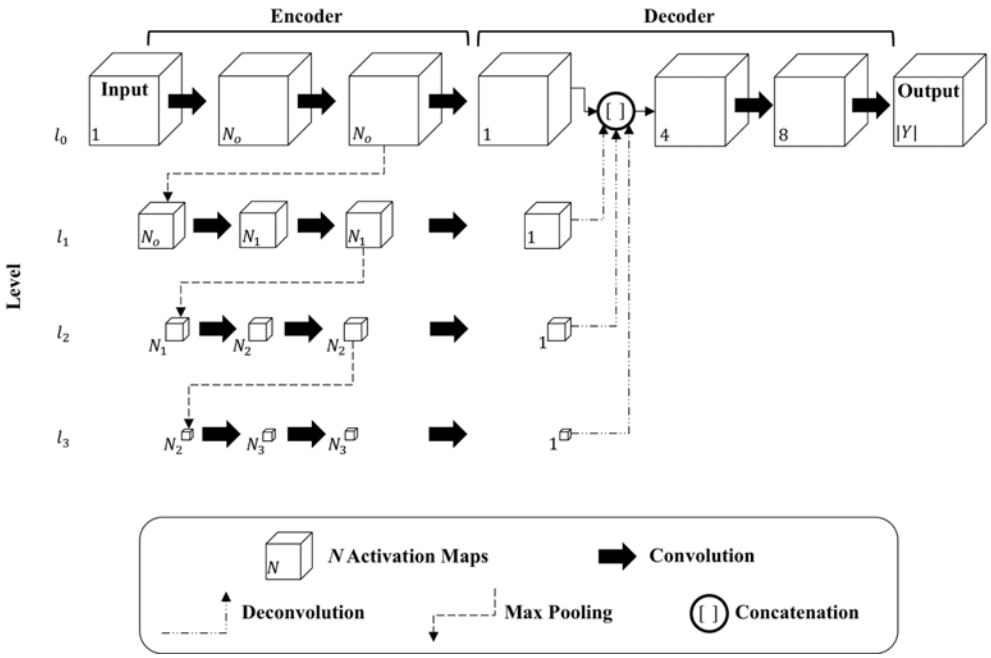


Figure 4.1. Seg3DNet architecture. The number of channels for each image representation is denoted in the lower left corner of each cube. For the encoder module, we define $N_i = 2^{i+5}$ so that the number of activation maps increases by a factor of two at each level. The number of kernels used in each convolutional layer can be inferred by the number of activation maps in the layer's output representation, i.e., the first convolutional layer has $N_0 = 2^{0+5} = 32$ kernels. The relative spatial size of the activation maps are drawn to scale. At each level the spatial dimensions of the feature representation gets downsampled by a factor of two.

Seg3DNet is a fully-convolutional network for volumetric image segmentation with encoder and decoder modules, see Figure 4.1. The encoder network consists of L resolution levels; each resolution level has two convolutional layers followed by a max-pooling layers. Max-pooling is used to achieve a spatial downsampling with a factor of two between resolution levels. The decoder network produces a single-channel representation for each level using convolutional layers with a single kernel. The final representations from each resolution level are integrated by concatenating the representations followed by two convolutional layers. The output of Seg3DNet is a probability map corresponding to the probability that each voxel is in the lung.

4.3.3 Multi-Resolution Model

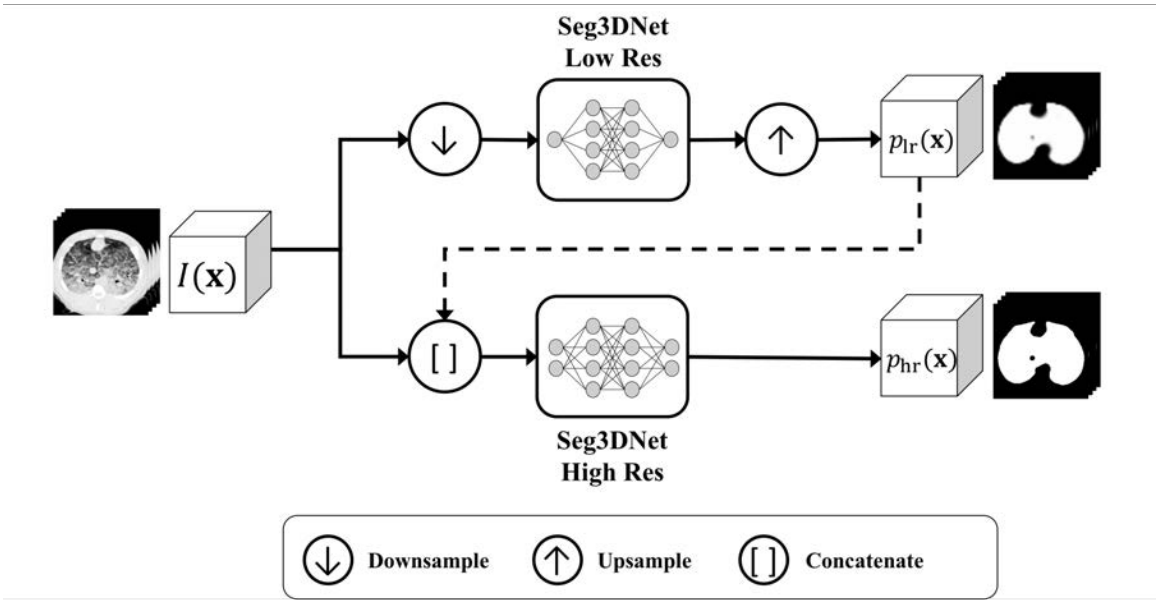


Figure 4.2. Multi-resolution model. The upper pipeline corresponds to the low-resolution model and the lower pipeline corresponds to the high-resolution model. The dashed line connection links the two models, allowing the global information learned in the low-resolution model to be used in the high-resolution model.

The multi-resolution model is composed of a series of two Seg3DNet's, see Figure 4.2. The two networks have the same architecture and hyperparameters, however, the first network is trained with low-resolution images and the second network is trained with high-resolution images. This allows the two networks to learn different ranges of features scales without having to tradeoff global context and spatial smoothness for high-resolution.

The first Seg3DNet is trained using the entire 3DCT image giving the network the capacity to utilize global contextual features. To accommodate the entire image, the CT images and target segmentations are aggressively downsampled to $64 \times 64 \times 64$ voxels. This corresponds to a downsampling factor of approximately eight in each spatial direction.

The second network is trained on high-resolution images; all images are resampled to have 1mm^3 isotropic voxels for consistency. The input to the second network is a two channel image consisting of the CT image and the prediction from the first network. This allows the network to learn precise boundary information while still integrating global information learned from the first network. At this resolution, the network cannot train on the full images due to limitations in GPU memory. Axial slabs of size $256 \times 256 \times 32$ are extracted at various positions to cover the entire lung field.

4.3.4 Training

The multi-resolution model is learned by first training the low-resolution network and subsequently training the high-resolution network. Initially, only the COPD and

IPF datasets are used for training the model. Subsequently the model is fine-tuned using the lung cancer dataset to allow inclusion of large lung tumors. Each dataset was split into 80% training and 20% validation.

A binary cross entropy loss function is used, the loss for each voxel $\mathbf{x}$ is given by

$$L(\mathbf{x}) = -(y(\mathbf{x}) \log(\hat{y}(\mathbf{x})) + (1 - y(\mathbf{x})) \log(1 - \hat{y}(\mathbf{x}))) \tag{4.1}$$

where $y(\mathbf{x})$ is the true class label for voxel $\mathbf{x}$, $y(\mathbf{x}) = 1$ for lung and $y(\mathbf{x}) = 0$ for background, and $\hat{y}(\mathbf{x})$ is the predicted probability that voxel $\mathbf{x}$ belongs to the lung class. The total loss is given by the average loss for all voxels. The loss function is optimized with respect to the free parameters (the convolution kernels) using standard backpropagation.

Adam optimization [67] is used for training with a learning rate of 5×10^{-4}. Parameters are initialized using Xavier normal initialization [50]. Training was performed using a P40 NVIDIA GPU with 24 GB RAM. Total training time is approximately 48 hours.

4.3.5 Post-Processing

The ConvNet predicts probability image indicating the probability that each voxel is lung, with no distinction between the left and right lungs. Post-processing is performed to obtain a final binary segmentation and then separate the binary segmentation into left and right lungs. First a threshold of $p = 0.5$ is applied to the probability image to obtain a binary image. Next connected component analysis is performed and all components except the two largest are removed. The volumes

of the two largest components are compared; if the volume of the second largest component is less than half the volume of the first largest component, the second largest component is removed. If two components remain, these are identified as left and right lungs and no further processing is required. If only the first component is retained the left and right lungs are assumed to be connected and lung separation is performed.

4.4 Experiments and Results

4.4.1 Metrics

Performance of the proposed method was evaluated using the Jacaard index for volume overlap and average symmetric surface distance (ASSD). The Jacaard index is defined as

$$J(\mathrm{P}, \mathrm{M}) = \frac{|\mathrm{P} \cap \mathrm{M}|}{|\mathrm{P} \cup \mathrm{M}|}, \tag{4.2}$$

where $|\cdot|$ is the set cardinality and $P \cap M$ and $P \cup M$ are the intersection and union, respectively, of the set of voxels predicted to be lung in the automated segmentation P and the set of voxels defined as lung in the manual segmentation M. The Jacaard index has values ranging from zero to one, with one indicating perfect agreement. ASSD was used to measure the distance between the predicted lung boundary $\mathrm{B_P}$ and manually generated lung boundary $\mathrm{B_M}$. The distance between a voxel $\mathbf{x}$ and a set of voxels on boundary B is defined as

$$D(\mathbf{x}, \mathrm{B}) = \min_{\mathbf{y} \in \mathrm{B}} d(\mathbf{x}, \mathbf{y}), \tag{4.3}$$

where $d(\mathbf{x}, \mathbf{y})$ is the Euclidean distance between voxels $\mathbf{x}$ and $\mathbf{y}$. The ASSD between B_P and B_M is defined as

$$\text{ASSD} = \frac{1}{|\mathrm{B_P}| + |\mathrm{B_M}|} \times \left(\sum_{\mathbf{x} \in \mathrm{B_P}} D(\mathbf{x}, \mathrm{B_M}) + \sum_{\mathbf{y} \in \mathrm{B_M}} D(\mathbf{y}, \mathrm{B_P}) \right). \tag{4.4}$$

ASSD is greater or equal to zero, with zero being perfect agreement.

4.5 Results

The model was evaluated on 899 images: 576 COPDGene images, 51 SPIROMICS images, 62 PANTHER-IPF images, and 210 lung cancer images. Segmentation results of lung cancer subjects with large tumors are illustrated in 4.5. Each image corresponds to a single phase of a 4DCT scan.

Quartiles of the Jacaard index and ASSD distributions for each dataset are illustrated in Figure 4.6. The COPDGene and SPIROMICS datasets achieved higher performance with median ASSD (Jacaard index) of 0.221 mm (0.990) and 0.222 mm (0.987), respectively. The IPF and lung cancer datasets achieved median ASSD (Jacaard index) of 0.383 mm (0.979) and 0.342 mm (0.978), respectively. Performance was consistent for the three lung volumes in the COPDGene study as illustrated in Figure 4.7. Figure 4.8 illustrates the distribution of ASSD and Jacaard index for each subject. The tight distribution of errors within each subject demonstrate the repeatability of segmentation performance across volumes of 4DCT and different 4DCT scans of same subjects.

Lung segmentations were used to limit the region of interest for qCT analysis. For quality control, the mean HU calculated within the manual lung segmentation was compared to the mean HU calculated within the predicted lung segmentation

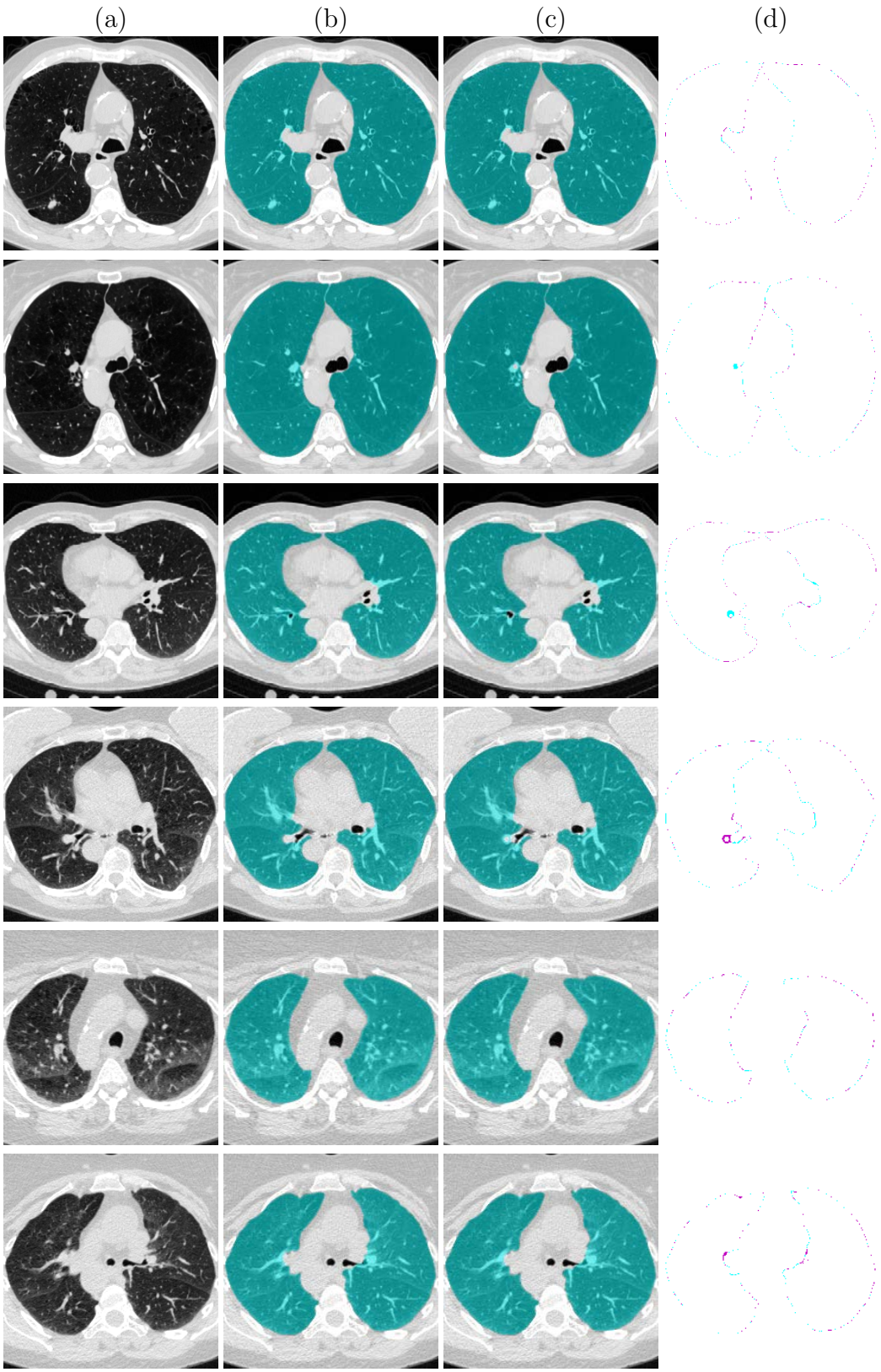


Figure 4.3. Representative results for six cases from the COPDGene dataset. a) CT image, b) predicted mask, c) manual mask, d) difference image with under- and over-segmentation in magenta and cyan, respectively. Rows 1-2 are TLC, rows 3-4 are FRC, and rows 5-6 are RV images.

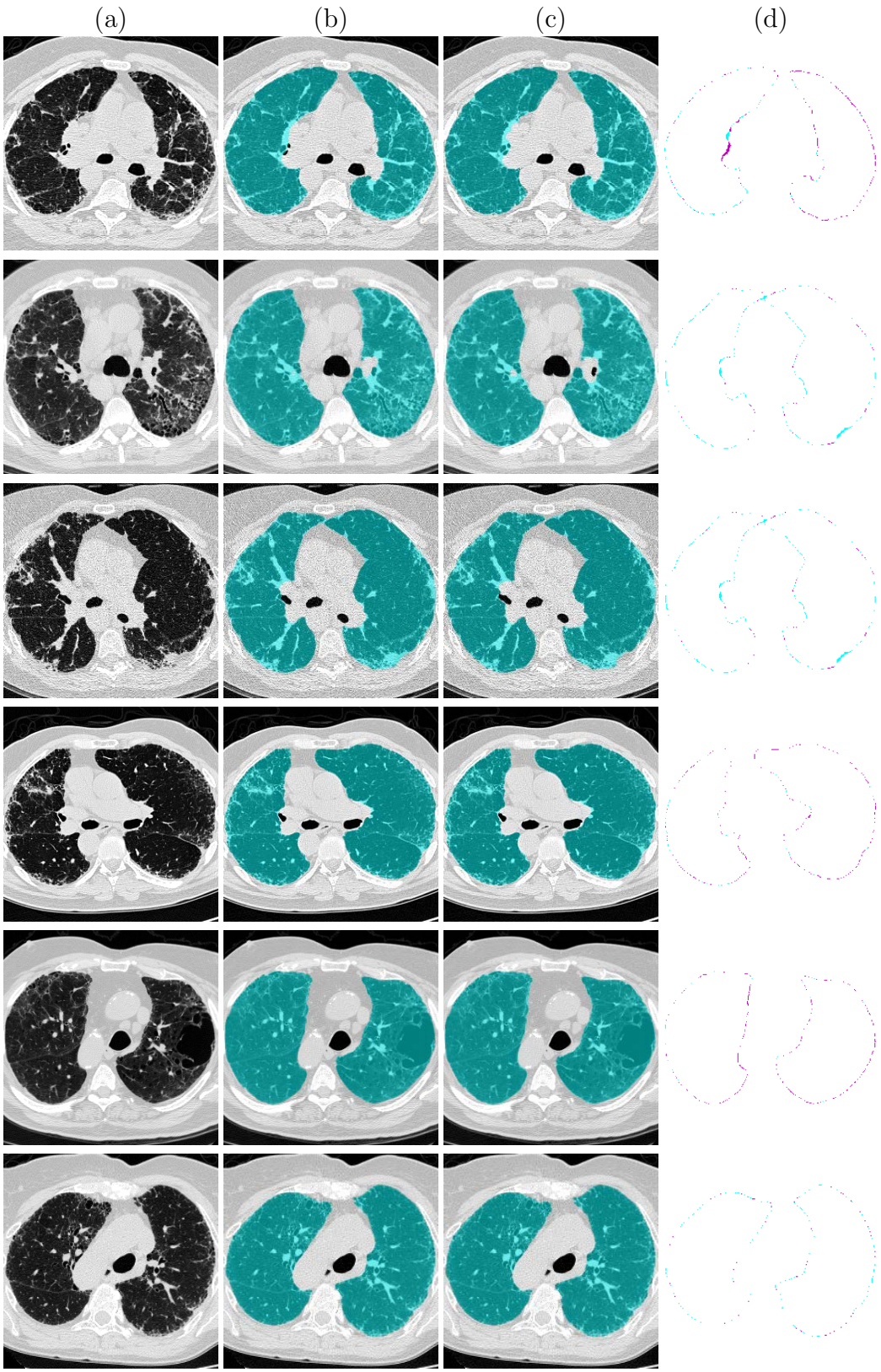


Figure 4.4. Representative results for six cases from the PANTHER-IPF dataset. a) CT image, b) predicted mask, c) manual mask, d) difference image with under- and over-segmentation in magenta and cyan, respectively.

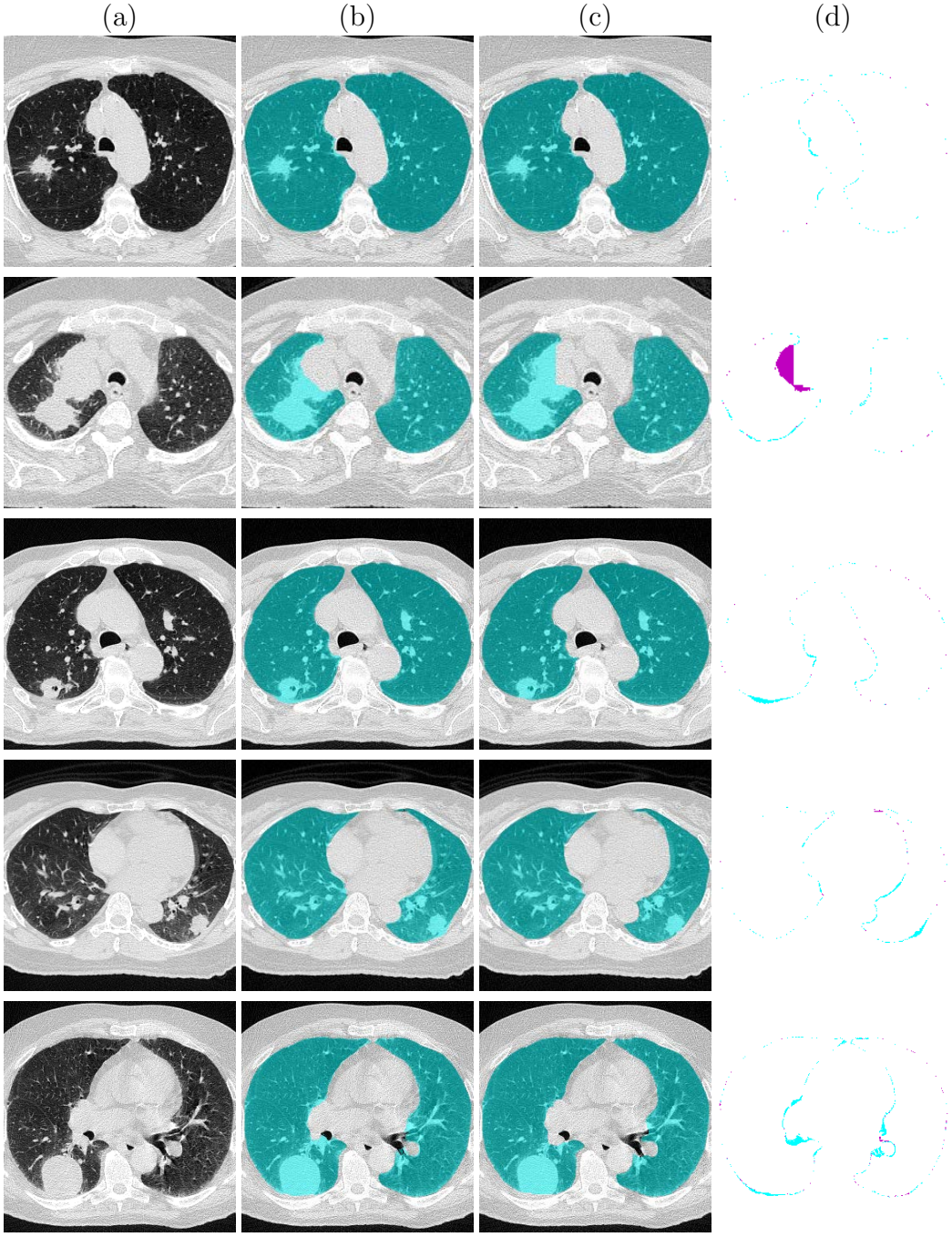


Figure 4.5. Representative results for five cases from the lung cancer dataset. a) CT image, b) predicted mask, c) manual mask, d) difference image with under- and over-segmentation in magenta and cyan, respectively.

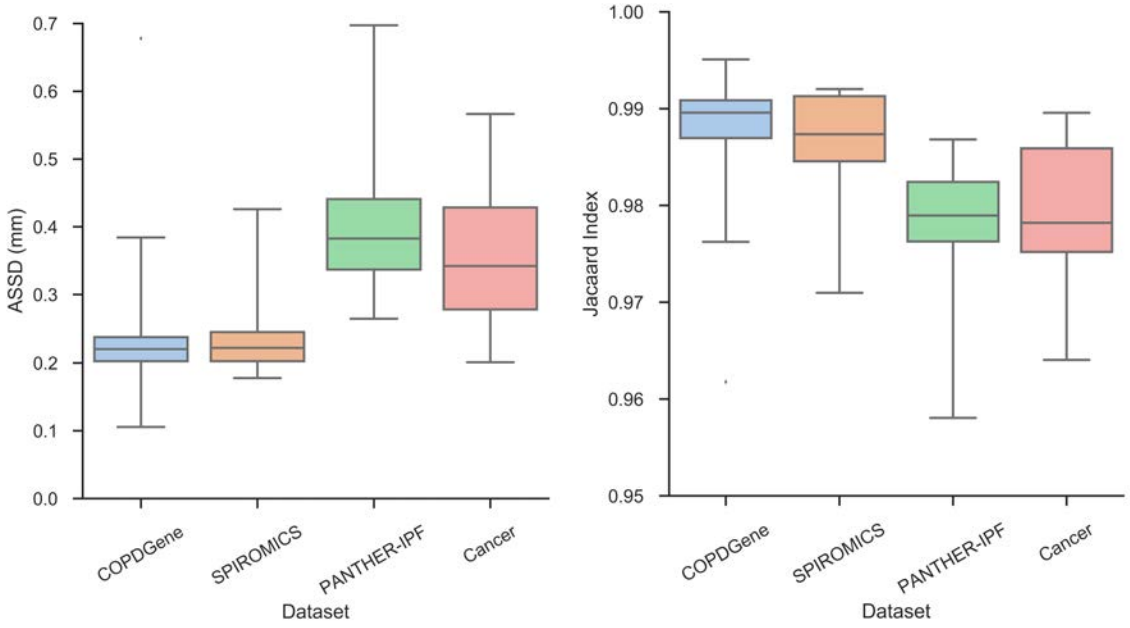


Figure 4.6. Quartiles of ASSD and Jacaard index distributions for the four datasets used in this study.

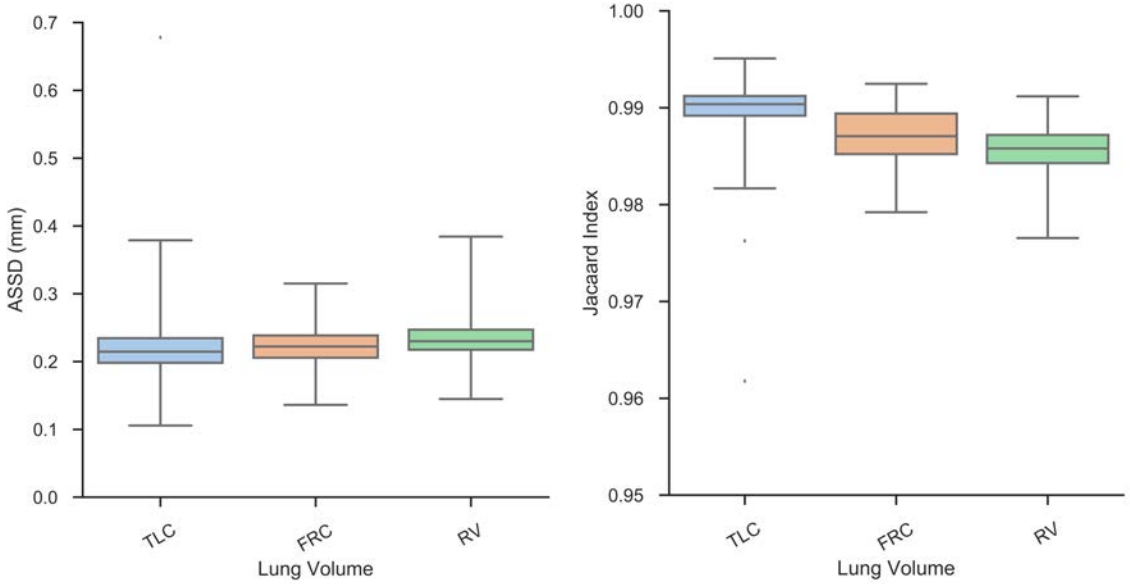


Figure 4.7. Quartiles of ASSD and Jacaard index distributions stratified by lung volumes. Only the COPDGene dataset was included in analysis.

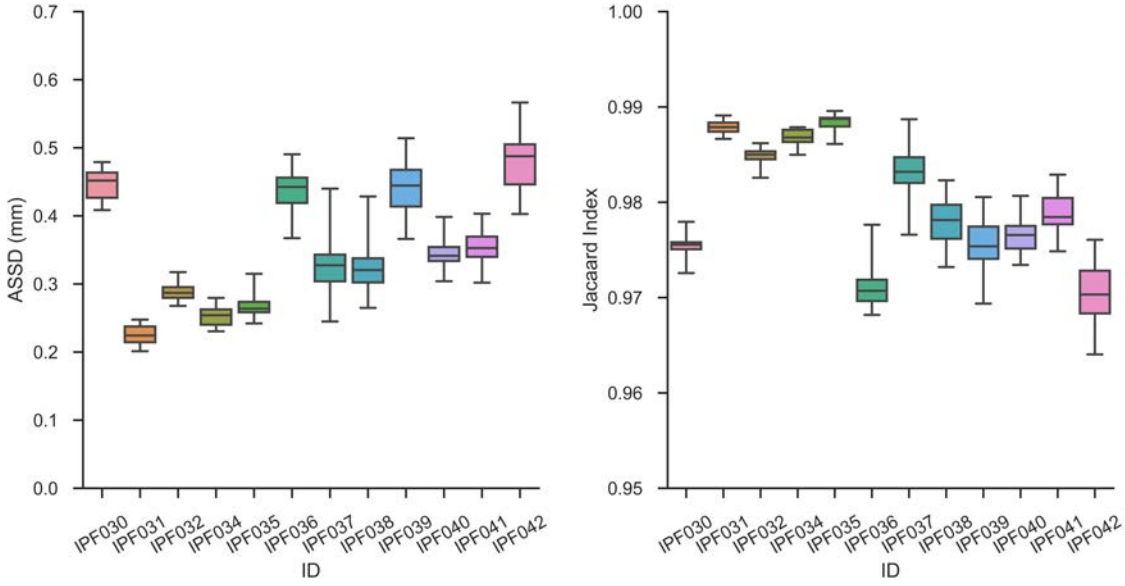


Figure 4.8. Distribution of ASSD and Jacaard index for each subject in 4DCT dataset.

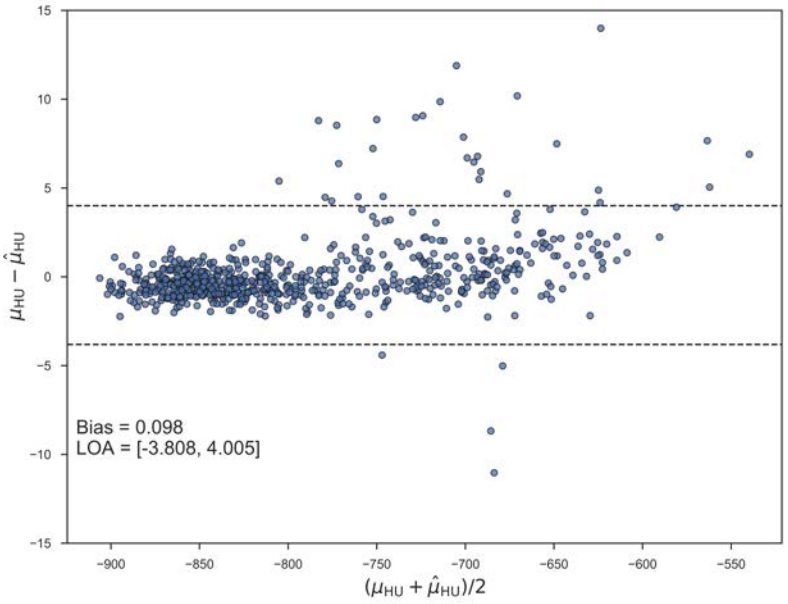


Figure 4.9. Bland-Altman plot comparing mean HU evaluated inside the manual lung segmentation (μ_{HU}) and the predicted lung segmentation ($\hat{\mu}_{\mathrm{HU}}$). Dashed lines represent limits of agreement (LOA).

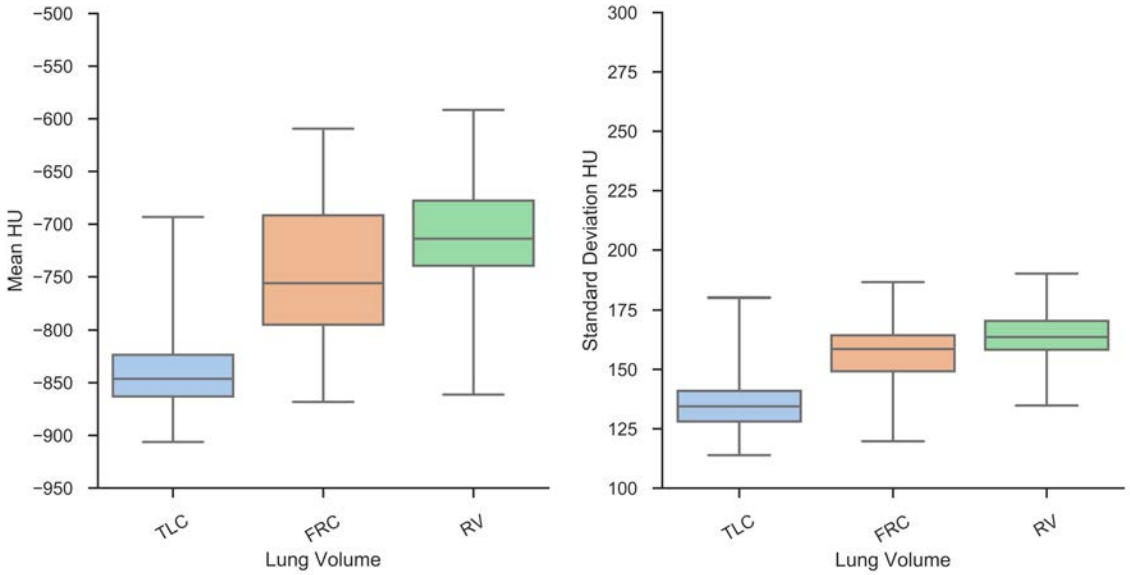


Figure 4.10. Quartiles of the mean HU and standard deviation of HU stratified by lung volume. Predicted lung segmentation is used for calculation. Only COPDGene dataset was included in analysis.

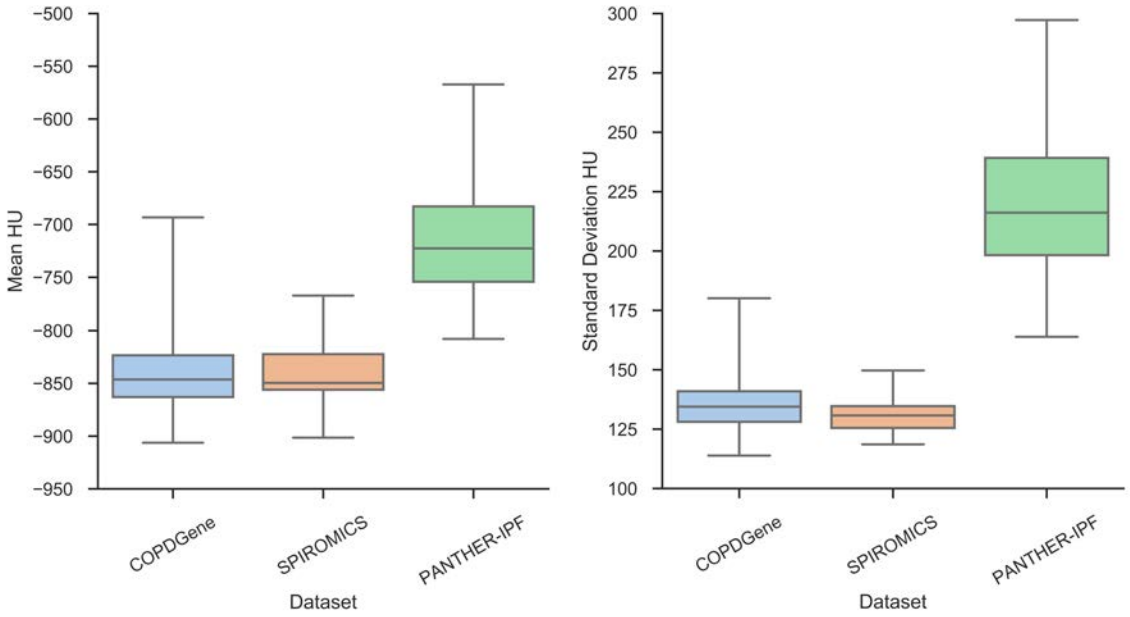


Figure 4.11. Quartiles of mean HU and standard deviation of HU stratified by dataset. Predicted lung segmentation is used for calculation. Only TLC scans were included in analysis.

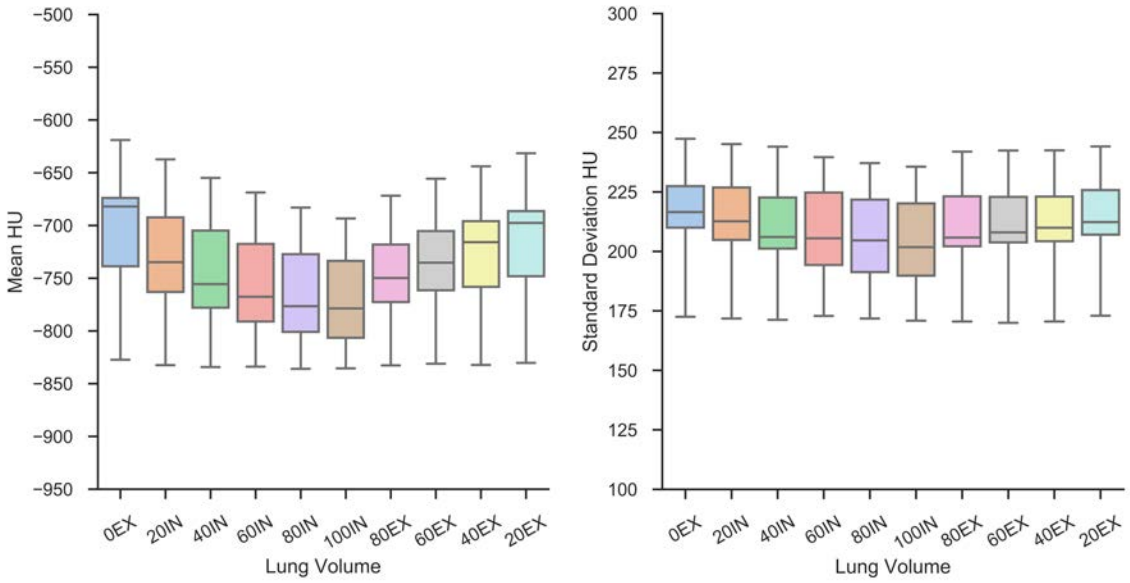


Figure 4.12. Quartiles of mean HU and standard deviation of HU by stratified lung volumes in 4DCT scans.

using a Bland-Altman analysis, results are displayed in Figure 4.9. For the majority of cases, the two lung segmentations produced mean HU calculations that differed by less than 4 HU. The remaining analysis only used the predicted lung segmentation for qCT measurements. Mean HU an standard deviation HU stratified by lung volume are illustrated in Figure 4.10, only COPD datasets were used for this analysis. Similarly, mean HU and standard deviation of HU across the volumes of 4DCT images are illustrated in Figure 4.12 Results stratified by dataset are illustrated in Figure 4.11, only scans acquired at TLC were used for this analysis to allow fair comparison to PANTHER-IPF dataset which only had TLC scans.

4.6 Discussion

A multi-resolution ConvNet model was proposed which consists of two ConvNets for learning different scales of features in large volumetric images. Overall, the method showed high performance on a diverse dataset of 899 CT images. This included images from four large-scale clinical trials: COPDGene, SPIROMICS, PANTHER-IPF, and a clinical trial studying functional avoidance in radiation therapy for lung cancer. Datasets with COPD subjects (COPDGene and SPIROMICS), showed slightly higher performance compared to the IPF and lung cancer datasets. COPD generally does exhibit diffuse dense pathologies as seen in IPF and lung cancer subjects. This makes COPD segmentation fairly straight forward and intensity-based methods perform well. IPF and lung cancer or more difficult to segment, especially cases where the pathology is in the peripheral lung. The proposed method nonetheless demonstrated high performance for these challenging cases.

Images acquired at different lung volumes can show large variations in both shape and appearance. The COPDGene dataset consisted of scans which were acquired at three lung volumes: TLC, FRC, and RV. Furthermore, the intermediate phases of 4DCT scans were included in evaluation. The proposed method was able to segment each of these phases with high accuracy. Consistent segmentation between different lung volumes of the same subjects is important for reliable qCT analysis such as calculating aeration at different lung volumes.

Identifying the lung boundary near the mediastinal region is notoriously difficult and subjective. The blood vessels and airways enter the lung at the mediastinum and

there is no clear definition on how much of these structures should be included in the segmentation. Lung tumors at the mediastinum can further obscure the true boundary. The proposed method was able to consistently segment this region to match the training data. However, it was observed that the largest surface errors were generally in this region.

The initial model trained using only COPD and IPF data failed to include large tumors when applied to the lung cancer dataset. After fine-tuning the model with the lung cancer dataset the model was able to include large tumors. In the future, a single model could be trained using all data. A potential problem with this could be the class imbalance between datasets with and without large tumors. This may result in exclusion of tumors since there is a very low cost for miss-classification of tumor voxels. While the lung cancer training dataset included 1410 3D images, there are only variations from 28 unique subjects. In the future this dataset should be expanded to account for more variations in tumor shape, size, and position.

For segmentation of 4DCT images, the multi-scale model was applied independently to each phase volume. Analogous to how 3D segmentation is advantageous over 2D segmentations, 4D segmentation would allow for incorporation temporal information and may produce more temporally consistent results. This was currently not feasible due to the limited amount of training data with all phase images included. In future work, a recurrent neural network (RNN) will be explored for modeling the temporal patterns in 4DCT datasets.

Many of the 4DCT scans from the lung cancer dataset have large artifacts near

the diaphragm in one or more phases. These artifacts are produced by inconsistent lung volumes between breaths resulting in duplication or missing anatomy. It is not clear what the true lung segmentation should include for these cases, however, the proposed method produced a reasonable segmentation which included the low-density voxels with a smooth transition to high-density voxels. A 4D segmentation algorithm would be ideal to handle these artifacts by utilizing the phase image with the least amount of artifacts and producing temporally consistent lung shapes to match this phase.

Aggressive downsampling is feasible for lung segmentation since it is a large structure with high-contrast. Although the precise boundary location is compromised, the lung field can still be easily distinguished in downsampled images. However, this may not be appropriate for structures that have a smaller scale or lower contrast. For example, the pulmonary fissures are approximately a voxel thin with low-contrast. Fissures can be difficult to identify even in high-resolution images. The fissures cannot be distinguished at all in downsampled images. Small blood vessels and airways are also not visible. The proposed method would need to be modified to segment these structures.

4.7 Conclusion

A multi-scale ConvNet was developed for learning segmentation in large medical images. The proposed method was trained to segment lungs in CT images of diseased subjects. The method achieved high performance on a diverse dataset of 899 CT images consisting of various pathologies, lung volumes, and imaging protocols. In

future work the method will be extended to incorporate temporal information in 4DCT datasets. The proposed method could be applied to segmentation of other anatomical structures in medical images where global contextual information is important.

CHAPTER 5

CROSS-SPECIES LUNG SEGMENTATION

Revision of: SE Gerard, J Herrmann, DW Kaczka, JM Reinhardt: Transfer Learning for Segmentation of Injured Lungs Using Coarse-to-Fine Convolutional Neural Networks. Image Analysis for Moving Organ, Breast, and Thoracic Images, 2018.

Segmentation of lungs with acute respiratory distress syndrome (ARDS) is a challenging task due to the diffuse opacification in the dependent lung which results in little or no contrast at the lung boundary. For segmentation of severely injured lungs, a method which incorporates both local intensity and texture information and global contextual information are important for consistent inclusion of injury. A deep learning framework is proposed which uses uses a novel multi-resolution convolutional neural network (ConvNet) for automated segmentation of lungs in multiple animal species with models of ARDS. The multi-resolution model eliminates the need to tradeoff between high-resolution and global context by using a cascade of low-resolution to high-resolution networks. Transfer learning is used to accommodate the limited number of ARDS training data: the model is initially pre-trained on a human dataset and subsequently fine-tuned on an animal dataset consisting of canine, porcine, and ovine images of subjects with injuries similar to ARDS. For comparison the multi-resolution model is compared to two single resolution ConvNets: a high-resolution ConvNet and a low-resolution ConvNet. The multi-resolution model outperforms both a low-resolution and a high-resolution model. On the animal dataset (N = 287)the multi-resolution model achieves an overall Jacaard index of

0.963 compared to 0.919 and 0.950 for the low-resolution and high-resolution models, respectively. The multi-resolution model achieves an overall average symmetric surface distance of 0.438 mm compared to 0.971 mm and 0.657 mm for the low-resolution and high-resolution modes, respectively.

5.1 Introduction

Computed tomography (CT) imaging produces high-resolution volumetric reconstructions of the anatomy. The intensity values in a CT image reflect the density of the tissue, producing high contrast between low-density lungs and the surrounding soft tissue. High-resolution CT images allow for the intricate lung texture, vasculature, and airway tree to be visualized. CT imaging is routinely utilized for diagnosing lung pathologies, guiding treatment, monitoring progression, and characterizing lung diseases.

Acute respiratory distress syndrome (ARDS) is a severe respiratory failure that leads to diffuse inflammation, increased pulmonary vasculature permeability, and loss of lung tissue aeration [21]. Radiographically, this condition presents with diffuse bilateral opacification in the dependent lung [45]. While chest x-ray can confirm diagnostics of ARDS, it does not capture injury localization and spatial heterogeneity. CT has great potential for imaging ARDS as it can differentiate injury phenotypes and correlate with patient response to mechanical ventilation [44]. CT imaging is increasingly being used to characterize the spatial heterogeneity of injury and regional mechanics of ARDS [65, 103, 38, 44, 102, 14]. Quantitative CT (qCT) enables objective quantification of injury and has been used for evaluating response to

mechanical ventilation protocols [9, 51, 104] and monitoring injury progression [16, 15]. Spatial and temporal heterogeneity of ventilation in ARDS can be measured through registration of dynamically imaged lungs [58] or lungs imaged at multiple inflation levels [65, 103, 15].

A precursor to qCT analysis ARDS is a lung segmentation which distinguishes pulmonary tissue from the surrounding tissue. Intensity-based methods are widely used for lung segmentation in CT images as there is high contrast between the air-filled lungs and surround soft tissue. However, these methods fail to include dense pathologies such as the non-aerated lung regions in ARDS subjects. Lungs with ARDS are particularly challenging to segment as the injury is diffuse and predominately in the posterior lung region. Peripheral injury is more challenging to segment compared to interior injury because there is very little or no contrast between the injury and the non-pulmonary tissue. Furthermore, consolidated regions have no textural features that make it distinguishable from the surround soft tissue.

There have been several works that specifically target segmentation of subjects with ARDS. A semiautomatic approach using segmentation-by-registration to segment longitudinal images of rats with surfactant depletion was proposed in [158]. A limitation of this approach is it requires a manual segmentation for the baseline scan. Additionally, this method relies on a time-consuming image registration (4-6 hours). Similar segmentation-by-registration approaches were proposed in [108, 107]. Anatomical information from the airways and rib cage has been used to identify the boundary between injured lung and soft tissue in injured lungs [25]. A wavelet-based

approach was proposed in [139], however, this method may fail for severely injured case where no edge information is present between the injury and soft tissue.

Manual segmentation is still widely used for segmentation of lungs with ARDS, as current automated methods are not reliable. Manual segmentation is tedious, time consuming, and subject to high intra- and inter-observer variations. Furthermore, manual segmentations are performed on 2D slices which limits global context and produces segmentations that are not smooth in 3D. For large datasets, such as dynamic 4DCT images with multiple phase images, manual segmentation is not feasible. Obtaining accurate lung segmentation for ARDS images is a major barrier that prevents routine use of qCT for ARDS in clinical and research settings.

Recently, deep learning with ConvNet models have dominated across a wide range of applications in computer vision, with the ability to perform image classification, localization, segmentation, and registration at human-level accuracy. Deep learning enables computers to learn directly from raw data rather than explicitly defining a rule-based system or learning from human engineering features. Deep learning based systems have shown to be more robust and computationally efficient. The majority of development of deep learning using ConvNets has been on 2D natural images, however, this technique has also been applied to medical imaging. ConvNets are successfully detecting skin cancer [35], classifying lung nodules [127], and segmenting various anatomical structures and diseases [110, 81, 119, 4, 128]. A survey on deep learning applied to medical imaging is given in [82].

A major barrier to using ConvNets for medical image segmentation is that the

memory requirement for large 3D images exceeds the limited amount of memory on GPU cards. To overcome this barrier, most methods extract 2D slices or 3D patches with local extent. These approaches sacrifice global information and 3D smoothness in favor of high-resolution. Another barrier for using deep learning in medical imaging is the availability of labeled training data. Deep learning methods require large training datasets to fit the millions of free model parameters. It is especially challenging to obtain expert annotated training data for volumetric medical images, which typically have upwards of 500 2D slices in a single image. Manual annotation of these images requires the time and cost of a medical expert. Furthermore, many interesting research involves rare diseases yielding very small cohorts of subjects.

These challenges need to be addressed for successful application of ConvNets for segmentation of ARDS lungs. Global contextual information, such as the surrounding anatomical features, are necessary for segmentation of injured lung as local intensity is non-distinguishable from surrounding tissue. Current methods that use 2D slices or 3D crops are not ideal as this removes global features. Furthermore, limited annotated training data of ARDS lungs is available due to the time necessary to produce manual segmentations. The main contributions of this work are as follows:

- Multi-resolution ConvNet model which has the capacity to learn both local and global features in large volumetric medical images.
- Fully automated and computationally efficient segmentation of lungs with ARDS in CT images.

- Cross-species segmentation of ARDS subjects in a unified model using limited annotated training data.

The novel multi-resolution ConvNet cascade makes use of both low-resolution and high-resolution models to enable multi-scale learning. We hypothesize a method that makes use of both global and local information is is superior to a model that uses only local or global for this type of injury. The importance of global and local information is explored by comparing the proposed model to a conventional high-resolution, and which uses image crops, and a low-resolution model, which uses aggressive downsampling. The high-resolution model sacrifices global contextual information in favor of high-resolution information whereas the low-resolution model sacrifices high-resolution detail in favor of global context. A transfer learning approach is used which allows training the system with a limited amount of ARDS training data. First a model is learned using a large dataset of human subjects, without ARDS. Subsequently this model is tuned for cross-species ARDS segmentation is learned by fine-tuning the human model using a dataset with multiple animal species.

5.2 Datasets and Reference Standards

For this study, we utilized a dataset consisting of CT scans from four species: human, canine, porcine, and ovine. Hereinafter, the collection of human images is referred to as the human dataset and the collection of canine, porcine, and ovine images is referred to as the animal dataset. The human dataset is used for pre-training and the animal dataset is used for fine-tuning. The human dataset consists of 3418 images, including 3113 images of subjects with chronic obstructive pulmonary disorder

(COPD) and 305 images of subjects with idiopathic pulmonary fibrosis (IPF). The animal dataset consists of 301 images of subjects with various experimental models of ARDS: 76 images of canine subjects with an oleic acid model of ARDS, 152 images of porcine subjects with an oleic acid model of ARDS, 27 images of ovine subjects with a saline lavage model of ARDS, and 46 images of ovine subjects with a lipopolysaccharide (LPS) model of ARDS.

5.2.1 Human Dataset

The human dataset consisted of scans acquired from three large-scale clinical trials: COPDGene, SPIROMICS, and PANTHER-IPF. COPDGene is a large multi-institutional clinical trial studying genetics and imaging biomarkers of COPD subjects [115]. The subset of subjects used in this study had images acquired at total lung capacity (TLC), functional residual volume (FRC), and residual volume (RV). TLC scans were acquired at 120 kVp and 200 mAs. FRC and RV scans were acquired at 120 kVp and 50 mAs. See [115] for full imaging protocol. SPIROMICS is also a multi-institutional clinical trial studying subpopulations and intermediate outcomes in COPD subjects [23, 129, 22]. The subjects used in this study had images acquired at TLC and RV, see [129] for full imaging protocol. The IPF dataset was obtained from an ancillary study of PANTHER-IPF [93, 94]. This study used high-resolution CT images to identify IPF textural features and their relations to disease progression [122]. This dataset consisted of subjects with TLC scans, see [122] for full imaging protocol. Lung segmentations for all human images were generated using Pulmonary Analysis Software Suite (PASS, University of Iowa Advanced Pulmonary Physiomic Imaging

Laboratory [55]) followed by manual editing.

5.2.2 Animal Dataset

5.2.2.1 Porcine Dataset

The porcine dataset was obtained from a study of alternative mechanical ventilation modalities to treat ARDS, approved by the University of Iowa Institutional Animal Care and Use Committee. Subjects approximately 10 to 15 kg in size were scanned under baseline conditions and following maturation of acute lung injury induced by infusion of oleic acid into the superior vena cava. 3DCT images were acquired during breath-hold maneuvers at constant airway pressures of 0, 5, 10, 15, 20, 25, and 30 cmH_2O. 4DCT images were acquired during mechanical ventilation using three ventilator modalities: conventional pressure-controlled ventilation, high-frequency oscillatory ventilation, and multi-frequency oscillatory ventilation [58]. All images were acquired using a Siemens Somatom Force 128-slice scanner, with 120 kVp, 90 mA s, and 0.5 mm slice thickness for 3DCT, or 80 kVp, 150 mAs, and 0.6 mm slice thickness for 4DCT. The 4DCT images have a limited axial coverage of 5.76 cm, which excludes the apex and base of the lungs. Manual 3DCT lung segmentations were generated semi-automatically using PASS software and manually corrected. 4DCT lung segmentations were generated semi-automatically using thresholding and connected components analysis followed by manual correction using 3D Slicer software [36].

5.2.2.2 Canine Dataset

The canine dataset was obtained from a study of respiratory mechanics in subjects with ARDS, approved by the Johns Hopkins University Institutional Animal

Care and Use Committee. Subjects approximately 22 to 33 kg in size were scanned under baseline conditions and following maturation of acute lung injury induced by infusion of oleic acid into the pulmonary artery. 3DCT images were acquired during breath-hold maneuvers at constant airway pressures of 0, 5, 10, 15, and 20 cmH_2O. Images were acquired using a Siemens Somatom Sensation 16-slice scanner, with 137 kVp, 165 mA s, and 2.5 mm slice thickness. Manual lung segmentations were generated semi-automatically using PASS software and manually corrected.

5.2.2.3 Ovine Dataset 1

The first ovine dataset was obtained a study of prone vs. supine positioning to treat subjects with ARDS, approved by the Massachusetts General Hospital Institutional Animal Care and Use Committee. Subjects approximately 20 to 30 kg in size were scanned following acute lung injury induced by saline lavage. 3DCT images were acquired during breath hold maneuvers at inflation levels corresponding to end-expiration (PEEP 5 cmH_20), end-inspiration (tidal volume 8 mL/kg), and mean airway pressure during mechanical ventilation. Images of prone subjects were flipped to align anatomical features to a corresponding supine orientation. Images were acquired using a Siemens Biograph combined PET-CT scanner, with 120 kVp, 80 mA s, and 0.5 mm slice thickness. Manual lung masks were generated semi-automatically using PASS and manually corrected.

5.2.2.4 Ovine Dataset 2

The second ovine dataset was obtained a study of subjects with ARDS, approved by the Johns Hopkins University and University of Iowa Institutional Animal Care

and Use Committees [37]. Subjects approximately 25 to 45 kg in size were scanned under baseline conditions and following acute lung injury induced by intravenous infusion of lipopolysaccharide . 3DCT images were acquired using respiratory-gated CT imaging at inflation levels corresponding to end-expiration and end-inspiration during mechanical ventilation. Images were acquired using a Siemens Somatom Sensations 16- or 64-slice scanner, with 120 kVp, 250 or 180 mA s, and 1.5 or 1.2 mm slice thickness. Manual lung segmentations were generated semi-automatically using PASS and manually corrected.

5.3 Methods

5.3.1 Overview

A multi-resolution ConvNet model is proposed for the task of lung segmentation in CT images (Section 5.3.3.3), designed to handle severely injured lungs across multiple animal specials. The multi-resolution model is compared to each individual component of this model: a low-resolution model (Section 5.3.3.1) and a high-resolution model (Section 5.3.3.2). All models make use of the same underlying ConvNet architecture, called Seg3DNet (Section 5.3.2), however, the spatial resolution of the training data is varied. Training using images of different resolutions results in a different range of feature scales learned by each model. Due to the limited number of scans for each species in the animal dataset, transfer learning from the human dataset is used for training all models (Section 5.3.4).

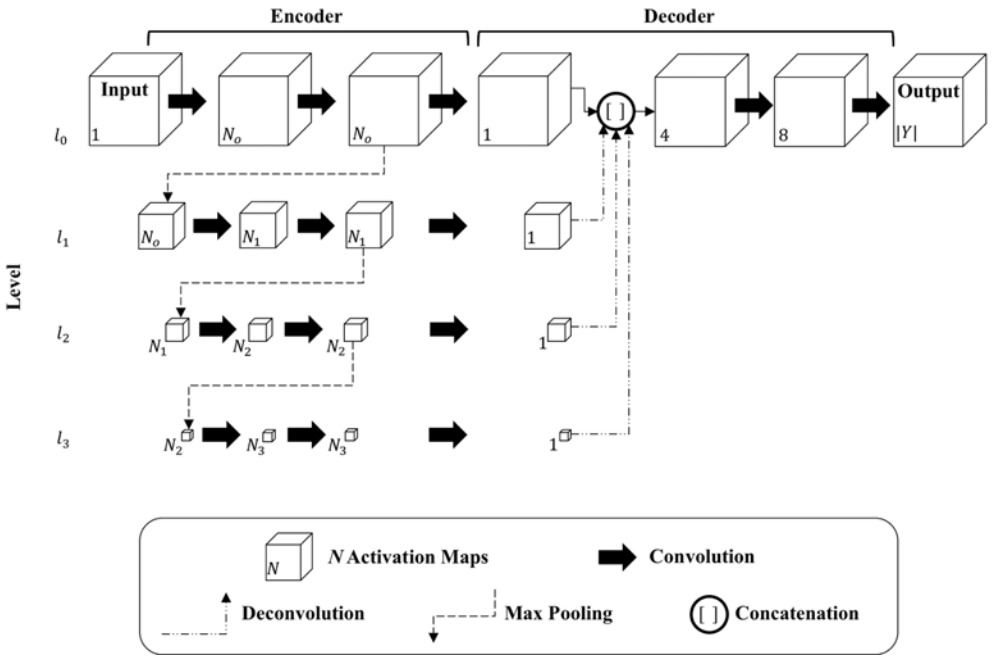


Figure 5.1. Seg3DNet architecture. The number of channels for each image representation is denoted in the lower left corner of each cube. For the encoder module, we define $N_i = 2^{i+5}$ so that the number of activation maps increases by a factor of two at each level. The number of kernels used in each convolutional layer can be inferred by the number of activation maps in the layer's output representation, i.e., the first convolutional layer has $N_0 = 2^{0+5} = 32$ kernels. The relative spatial size of the activation maps are drawn to scale. At each level the spatial dimensions of the feature representation gets downsampled by a factor of two.

5.3.2 Convolutional Neural Network

The underlying ConvNet architecture used in each of the three models is a fully convolutional network (FCN) called Seg3DNet [49], see Figure 5.1. The network has an encoder and decoder module, similar to the popular U-Net architecture [119], however, Seg3DNet is extended to three spatial dimensions and the decoder was designed to use less memory. The input to the network is an image with three spatial dimensions and the output is an image with three spatial dimensions and the same size as the input image. The input image is transformed to increasing

abstract image representations using a hierarchy of network layers. Each intermediate image representation has three spatial dimensions and a fourth dimension representing different feature types. Henceforth, we will refer to the fourth dimension as the channel dimension, analogous to that of an RGB images. The output of Seg3DNet is an image with $|Y|$ channels, where Y is the class set. The task of lung segmentation is treated as a binary segmentation problem, i.e., $|Y| = 2$ where the classes correspond to lung and background.

The encoder network consists L resolution levels. Each resolution level l_i consists of two convolutional layers followed by a max pooling layer. The decoder network upsamples the image representation at the end of each level back to the input image resolution using deconvolution layers and combines the multi-scale features using two subsequent convolutional layers. Each voxel in the output image is a floating point number corresponding to the probability that the voxel is part of the lung field.

Convolutional layers use a kernel with spatial extent of $3 \times 3 \times 3$ voxels and zero-padding is used such that the spatial size of the image representation remains unchanged. Max pooling with a kernel size of $2 \times 2 \times 2$ voxels and stride of $2 \times 2 \times 2$ voxels is used which effectively spatially downsamples the image representation by a factor of two along each spatial dimension, with the number of feature maps remaining unchanged. Batch normalization and a rectified linear unit (ReLu) activation function are used after each convolutional layer, with the exception of the last layer. The last layer uses a softmax vector nonlinearity (Equation 5.1). The output of the softmax function gives the conditional probability distribution that a voxel $\mathbf{x}$ belongs to each

$y \in Y$.

$$P(Y = y|\mathbf{x}) = \frac{e^{f_y(\mathbf{x})}}{\sum_{j \in Y} e^{f_j(\mathbf{x})}}. \tag{5.1}$$

For a binary classification, the predicted probability distribution can unambiguously be represented as a single floating-point number. The predicted probability that $\mathbf{x}$ belongs to the lung is denoted as $\hat{y}(\mathbf{x})$. The predicted probability that $\mathbf{x}$ belongs to the background can be inferred as $1 - \hat{y}(\mathbf{x})$.

5.3.3 Models

5.3.3.1 Low-Resolution Model

The low-resolution model consists of a single Seg3DNet which is trained using aggressively downsampled CT images and lung segmentations. All training images are downsampled to constant image size of $64 \times 64 \times 64$ voxels, regardless of original size. This corresponds to a downsampling factor of roughly eight along each spatial dimension. At this image size, the entire image can be input to the network which allows for global features to be learned, however, exact boundary information is lost with the downsampling. Gaussian smoothing is performed prior to downsampling to avoid aliasing. The output of the low-resolution model is upsampled to the original resolution using b-spline interpolation.

5.3.3.2 High-Resolution Model

The high-resolution model consists of a single Seg3DNet which is trained using high-resolution CT images. The CT images are resampled to isotropic voxels: 1 mm isotropic voxels for humans, dogs, and sheep and 0.6 mm isotropic voxels for pigs. The downsampling is performed to achieve consistent voxel sizes and relative scale

between scans and species. This corresponds to a downsampling factor of less than two along each dimension. At this high-resolution, the entire CT image cannot fit into GPU memory. Therefore, axial slabs of size $256 \times 256 \times 32$ are sampled at multiple axial positions are used for training the model. This limits the amount of global context that can be learned by the high-resolution model, as features from the entire lung field cannot be learned.

5.3.3.3 Multi-Resolution Model

The multi-resolution model consists of two Seg3DNet's, utilizing both a low-resolution model and a high-resolution model, and linking the models to allow information learned by the low-resolution model to by exploited by the high-resolution model. The two Seg3DNet's are trained sequentially. In the first stage, the low-resolution model is trained on aggressively downsampled images as described in Section 5.3.3.1. In the second stage, a high-resolution model is trained, similar to the model described in Section 5.3.3.2, however, the low-resolution model prediction is included in the input in addition to the high-resolution CT image. Combining the low-resolution and high-resolution models eliminates the necessity of choosing between global contextual information and precise boundary detail. The multi-resolution model is illustrated in Figure 5.2.

5.3.4 Training

Transfer learning [99] is used for training all models. A model is pre-trained using the human training dataset. The model learned from the human dataset is used to initialize the animal model and fine-tuning is then performed using the animal

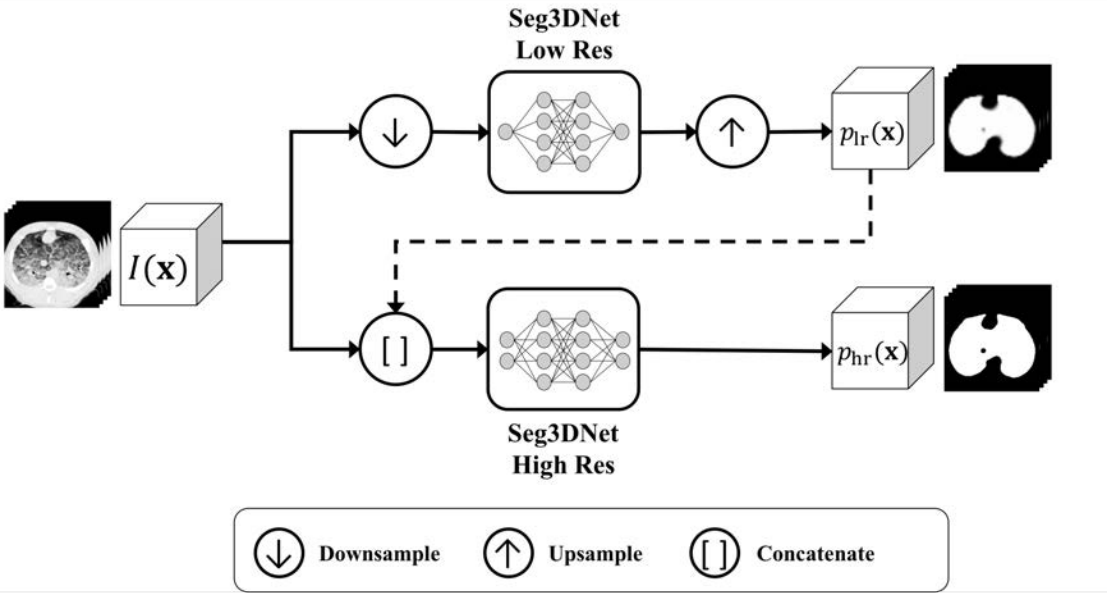


Figure 5.2. Multi-resolution model. The upper pipeline corresponds to the low-resolution model and the lower pipeline corresponds to the high-resolution model. The dashed line connection links the two models, allowing the global information learned in the low-resolution model to be used in the high-resolution model.

dataset.

Due to the limited number of animal scans with ARDS, a five-fold cross validation is performed for training the animal model. The animal dataset is split into five groups (approximately 60 images per group), four of the groups are used for training and the performance is evaluated on the left out group. This is performed five times, leaving out a different group each time and evaluating on that left out group. This allows all images in the animal dataset to be used for both training and evaluation. Each of the five groups have the same number of images and approximately equal representation of each species. The five-fold cross validation is done for training all models, using the same splits for fair comparison.

A binary cross entropy loss function is used for training, the loss for each voxel $\mathbf{x}$ is given by

$$L(\mathbf{x}) = -(y(\mathbf{x}) \log(\hat{y}(\mathbf{x})) + (1 - y(\mathbf{x})) \log(1 - \hat{y}(\mathbf{x}))) \tag{5.2}$$

where $y(\mathbf{x})$ is the true class label for voxel $\mathbf{x}$, $y(\mathbf{x}) = 1$ for lung and $y(\mathbf{x}) = 0$ for background, and $\hat{y}(\mathbf{x})$ is the predicted probability that voxel $\mathbf{x}$ belongs to the lung class. The total loss for each image is given by the average loss over all voxels in the image. The loss function is optimized with respect to the free parameters (the convolution kernels) using standard backpropagation. Adam optimization [67] is used for training, a learning rate of 5×10^{-4} is used for pre-training, and a learning rate of 5×10^{-5} is used for fine-tuning. Prior to pre-training, all free parameters are initialized using Xavier normal initialization [50]. The networks are trained using a P40 NVIDIA GPU with 24 GB RAM. Total training time is approximately 48 hours for each model.

5.3.5 Post-processing

The output predicted by each model is a lung probability image, i.e. an image with floating point values between 0 and 1 representing the probability that each voxel belongs to the lung. We adopt a simple post-processing step to obtain a final binary lung segmentation. A binary threshold is applied to the probability image using a cutoff of $p = 0.5$, which is empirically determined. Subsequently, 3D connected component analysis is performed on the thresholded image. For images from the human dataset, the two largest connected components, are retained and any remaining components are discarded. The two components correspond to the

left and right lungs. In some cases the left and right lung form one connected component, in these cases only the largest connected component is retained. This is determined automatically by taking the ratio of volumes for the two largest connected components. For the animal dataset, only the largest connected component is retained and any other components are discarded. For all animals used in this study, the left and right lungs are connected by a middle lobe.

5.4 Experiments and Results

5.4.1 Metrics

Lung segmentation performance was evaluated by comparing the computer generated segmentation to a manual lung segmentation. Two metrics were used to assess agreement: the Jacaard index and average symmetric surface distance (ASSD). The Jacaard index is a measure of volume overlap given by

$$J(\mathrm{P},\mathrm{M}) = \frac{|\mathrm{P} \cap \mathrm{M}|}{|\mathrm{P} \cup \mathrm{M}|}, \tag{5.3}$$

where $|\cdot|$ is the set cardinality and $P \cap M$ and $P \cup M$ are the intersection and union, respectively, of the set of voxels predicted to be lung in the automated segmentation P and the set of voxels defined as lung in the manual segmentation M. The Jacaard index has values ranging from zero to one, with one indicating perfect agreement. ASSD was used to measure the distance between the predicted lung boundary $\mathrm{B_P}$ and manually generated lung boundary $\mathrm{B_M}$. The distance between a voxel $\mathbf{x}$ and a set of voxels on boundary B is defined as

$$D(\mathbf{x},\mathrm{B}) = \min_{\mathbf{y}\in\mathrm{B}} d(\mathbf{x},\mathbf{y}), \tag{5.4}$$

where $d(\mathbf{x}, \mathbf{y})$ is the Euclidean distance between voxels $\mathbf{x}$ and $\mathbf{y}$. The ASSD between B_P and B_M is defined as

$$\text{ASSD} = \frac{1}{|\mathrm{B_P}| + |\mathrm{B_M}|} \times \left(\sum_{\mathbf{x} \in \mathrm{B_P}} D(\mathbf{x}, \mathrm{B_M}) + \sum_{\mathbf{y} \in \mathrm{B_M}} D(\mathbf{y}, \mathrm{B_P}) \right). \tag{5.5}$$

ASSD is greater or equal to zero, with zero being perfect agreement.

5.4.2 Quantitative Comparison of Models

The proposed multi-resolution model, the low-resolution model, and the high-resolution model were quantitatively evaluated by comparison to manual segmentations. A paired t-test was performed to test for significant differences between the models.

5.4.3 Results

Surface renderings of the different models are illustrated in Figure 5.3. Figure 5.4 illustrates multiple axial slices of a porcine subject with the multi-resolution model result. Surface renderings, minimum intensity projections, and maximum intensity projections are illustrated in Figure 5.5 to emphasize extent of injury and the inclusion of this injury in the predicted segmentation.

The distributions of ASSD and Jacaard for each model are displayed in Figure 5.6. Paired t-tests revealed that the multi-resolution model had a significantly lower (higher) ASSD (Jacaard) compared to both the low-resolution model and the high-resolution model ($p < 0.001$), and the high-resolution model had a significantly lower (higher) ASSD (Jacaard) compared to the low-resolution model. Maximum surface distance (Max SD) distributions for each model are displayed in Figure 5.7. The results show that the multi-resolution model achieved lower Max SD and the high-resolution had

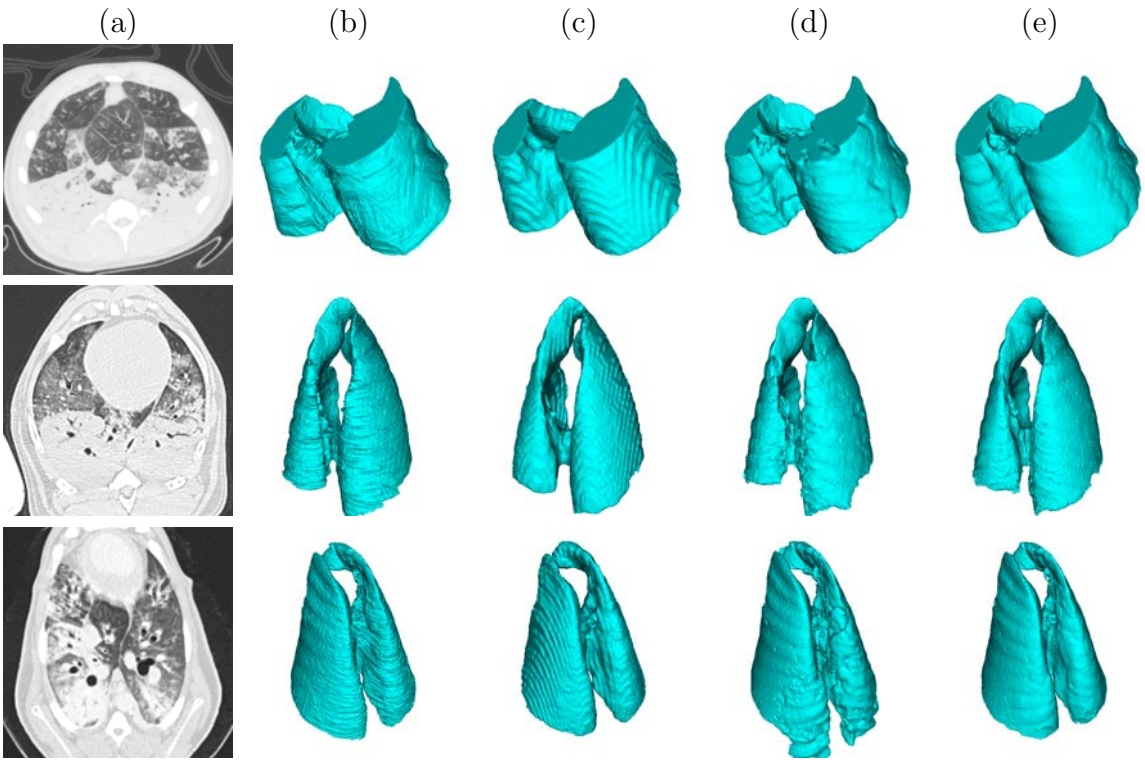


Figure 5.3. Surface rendering or lung segmentations porcine, canine, and ovine subjects in rows one, two, and three, respectively. (a) axial slice of CT, (b) manual (c) low-resolution model, (d) high-resolution model, (e) multi-resolution model.

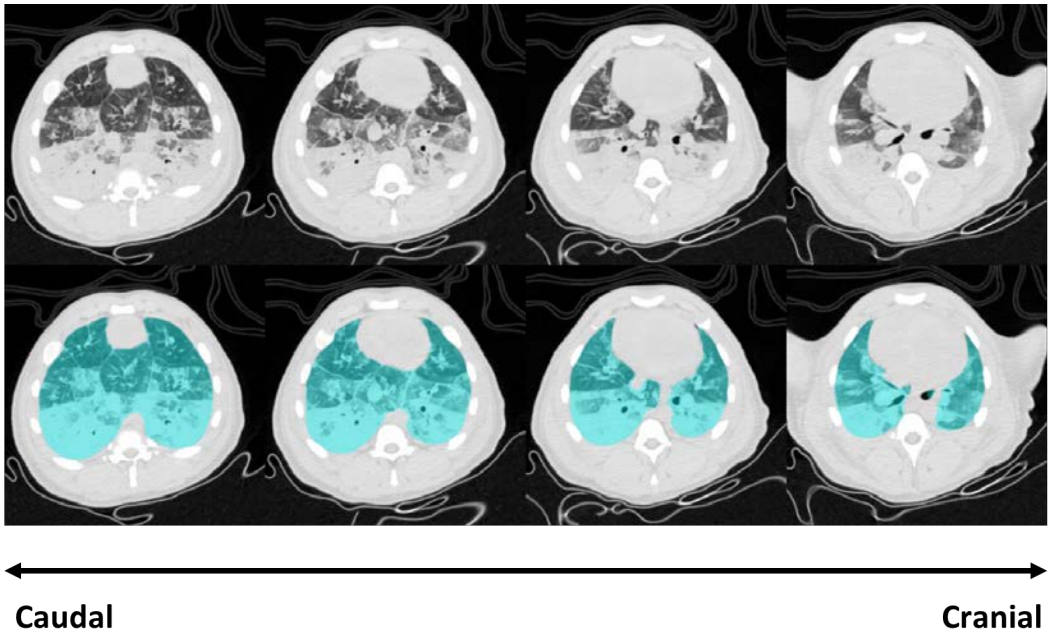


Figure 5.4. Multi-resolution network results for a porcine subjects. Multiple axial slices show the diffuse lung injury.

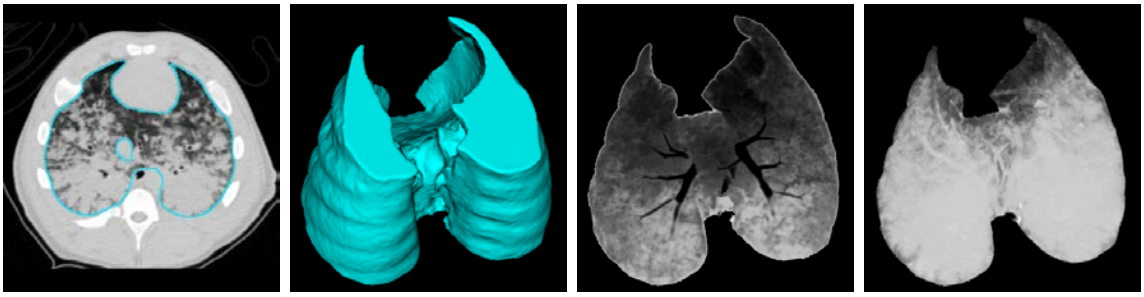


Figure 5.5. Multi-resolution model results. Left to right: contour of predicted segmentation overlaid on CT image, surface rendering of predicted segmentation, minimum intensity projection of voxels included in predicted segmentation, and maximum intensity projection of voxels included in predicted segmentation.

the largest Max SD. Results stratified by species and cross validation fold are displayed in Figures 5.8 and 5.9, respectively. The results show all models performed best on the porcine datasets and worst on the ovine datasets. All folds of the cross validation performed equally well.

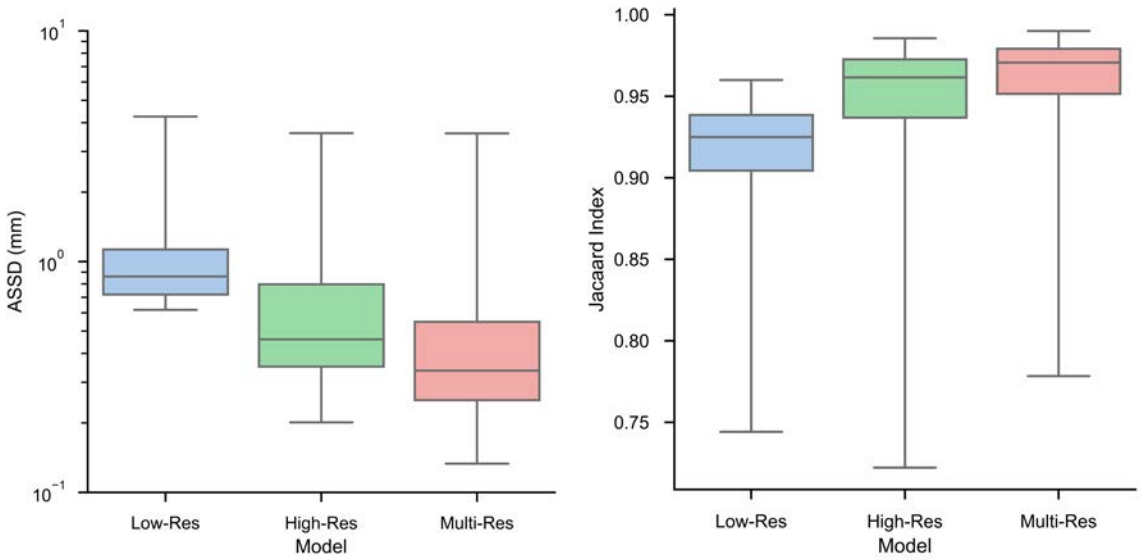


Figure 5.6. ASSD and Jacaard index distributions over all animal species.

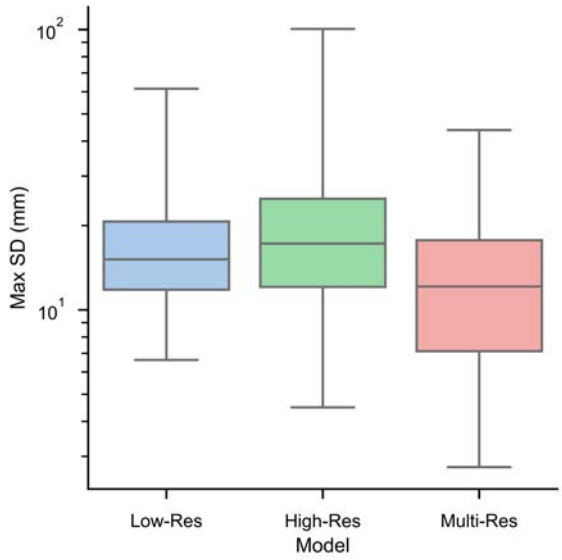


Figure 5.7. Distribution of maximum surface distance for each model.

The percent of lung volume that is non-aerated is used as a surrogate for lung injury severity. Voxels with HU greater than 100 are considered non-aerated. Figure 5.10 show the proposed method performance vs. percent non-aerated lung volume. Linear regression was used to model the relationship between injury severity and segmentation performance. Linear correlation coefficient was calculated to assess the model fit. The model slope for injury severity vs. ASSD was 0.011 ($R^2 = 0.096$, $p < 0.001$) for the low-resolution model, 0.026 ($R^2 = 0.298$, $p < 0.001$) for the high-resolution model, and 0.013 ($R^2 = 0.206$, $p < 0.001$) for the multi–resolution model. The model slope for injury severity vs Jacaard index was -0.001 ($R^2 = 0.201$, $p < 0.001$) for the low-resolution model, -0.002 ($R^2 = 0.437$, $p < 0.001$) for the high-resolution model, and -0.001 ($R^2 = 0.316$, $p < 0.001$) for the multi–resolution model.

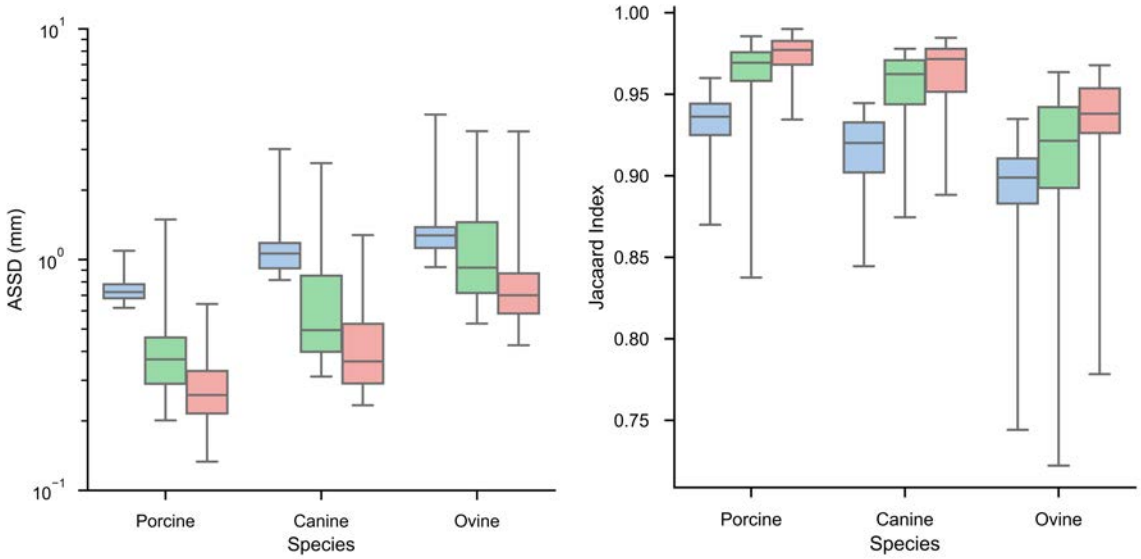


Figure 5.8. ASSD and Jacaard index distributions stratified by species.Blue is low-resolution model, green is high-resolution model, and red is multi-resolution model.

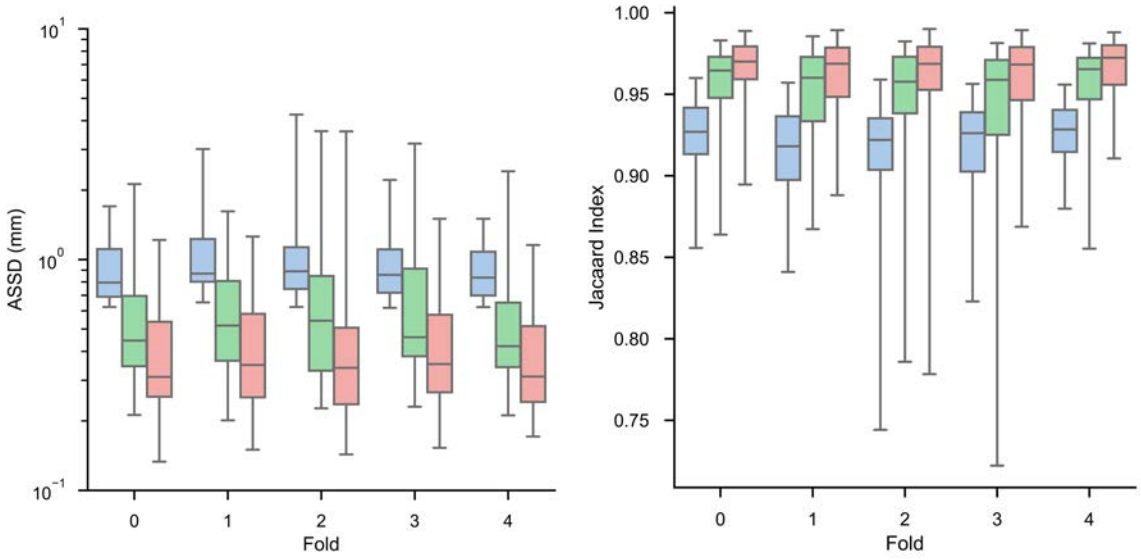


Figure 5.9. ASSD and Jacaard index distribution stratified by cross validation fold. Blue is low-resolution model, green is high-resolution model, and red is multi-resolution model.

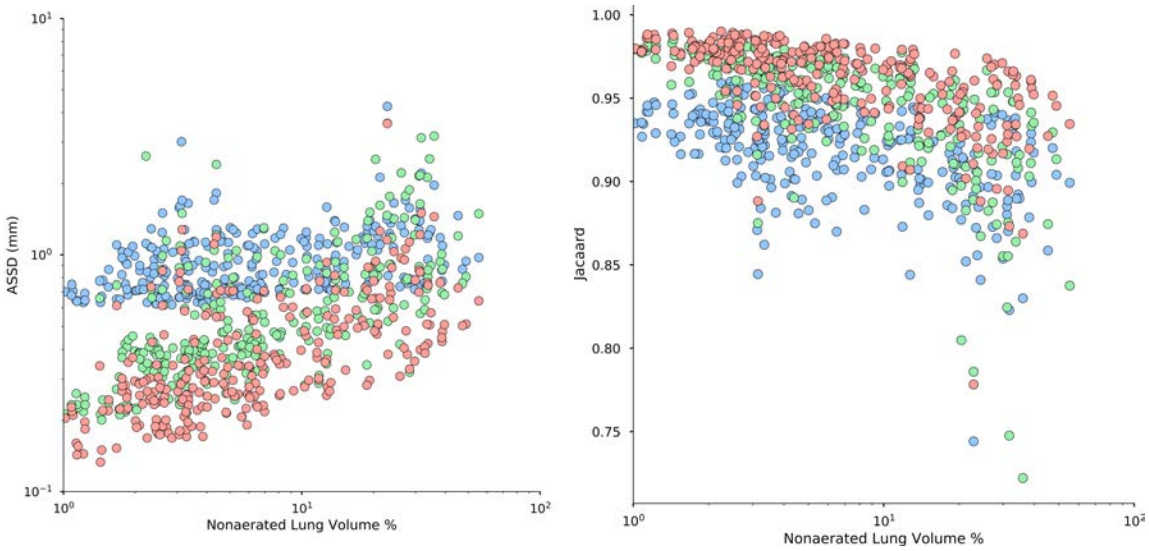


Figure 5.10. ASSD and Jacaard index stratified by percent of lung volume that is non-aerated. A lower ASSD and higher Jacaard index corresponds to better segmentation performance. Blue is low-resolution model, green is high-resolution model, and red is multi-resolution model.

5.5 Discussion

The motivation of this work was to develop a method for cross species lung segmentation targeted towards subjects with a lung injury similar to ARDS. This injury is particularly challenging to segment due to the diffuse injury in the dependent lung, resulting in little or no contrast between the posterior lung boundary and the surrounding soft tissue. A novel multi-resolution ConvNet model was proposed to incorporate both high-resolution local features and global contextual information.

Current paradigms in medical imaging do not favor sacrificing spatial resolution, and instead preserve the spatial resolution at the cost of removing global context. Training ConvNets is done using high-resolution 2D slices or 3D crops. In this work a multi-resolution model is proposed which has the capacity to incorporate both

global and local resolution. The multi-resolution model was compared to each of its components - a high-resolution model and a low resolution model. The high-resolution model is trained on crops, thus removing some global information. The low-resolution model is trained on aggressively downsampled images. All models showed high performance in terms of ASSD and Jacaard Index, with the multi-resolution model out-performing the other models.

The high-resolution model has less global context compared to the low-resolution and multi-resolution models, however, this model is still trained on large 3D crops. To further evaluate the influence of global information, the proposed method could be compared to training on 2D slices and small 3D patches which have less global context compared to the high-resolution model used in this study. The low-resolution model performed the worst of the three models, however, the model still performed well considering the downsampling factor that was used (approximately a factor of eights). Furthermore, compared to the high-resolution model, the low-resolution model requires less training and inference time and can be trained using a GPU with less memory. Therefore, this model may be preferred over a 2D slice model which has very limited global context and sacrifices 3D smoothness.

All models were evaluated on three species with models of ARDS. All models performed best on the porcine dataset and worst on the ovine dataset. The worse performance on the ovine dataset can likely be attributed to the fewer number of training data for this species. The proposed method nonetheless showed high performance on all species. Rule-based systems often fail to generalize across species,

requiring the method to be tuned for each dataset. More advanced methods such as statistical shape models also cannot seamlessly handle multiple species, since an underlying assumption is that the data comes from a Gaussian distribution.

Subjects were stratified by injury severity using percent of non-aerated lung volume as a surrogate for injury severity. All models showed a decrease in segmentation performance with increase in injury severity was observed for both Jacaard index and ASSD. The high-resolution model showed the largest decline in performance with increasing injury severity, indicating that global information becomes more important for cases that have more injury. Although there is a small decrease in performance with increasing injury severity, the multi-resolution model still showed high performance for highly diseased subjects, voxel-level ASSD error was achieved for severe cases with 30% non-aeration. Another factor to consider is the validity of the manual segmentations in these severely diseased cases. More injury results in more subjectivity in manual tracing, which could explain the decrease in performance.

An additional benefit of the proposed method is that it generates lung segmentations that have a smooth surface in 3D, compared to manual segmentation which can vary by several millimeters between slices. Manual segmentation was performed on 2D slices, and when the lung boundary is poorly defined manually choosing a consistent boundary between slices is very difficult. Defining the lung boundary in lungs with diffuse consolidation is very subjective. Intra- and inter-observer variation is likely very high on these cases, however, only a single manual segmentation was available per image due to the amount of time required to do the manual segmentation. In

future work it would be interesting to evaluate the variations in manual tracings between different analysts and within the same analyst.

The run-time for executing the proposed coarse-to-fine model on a single 3D image is 40 seconds when run on a GPU. If a GPU is unavailable, the model can be executed on a CPU with a run-time of approximately 2.5 minutes. Manual segmentation of these images takes anywhere from 4 to 6 hours, depending on the severity of the injury. This time savings make quantitative CT analysis on ARDS patients feasible in both clinical and experimental settings.

5.6 Conclusion

In this work, a multi-resolution ConvNet cascade was proposed that enables learning of both local and global features in large 3D images despite limitations in GPU memory. This method was applied to segmentation lung with ARDS in three animal species. We explored the hypothesis that both global and local features are important for segmentation of injured lungs by comparing the multi-resolution model to a low-resolution model and a high-resolution model. The proposed multi-resolution model performed best in terms of Jacaard Index and ASSD, demonstrating the importance of both global and local features for the task of lung segmentation in injured lungs. Furthermore, we have demonstrated that the proposed method works across three animal species including canine, porcine, and ovine subjects using a limited number of training data.

CHAPTER 6

PULMONARY FISSURE SEGMENTATION

Published as: SE Gerard, TJ Patton, GE Christensen, JE Bayouth, and JM Reinhardt. FissureNet: A deep learning approach for pulmonary fissure detection in CT images. IEEE Trans. Medical Imaging, 2018.

Pulmonary fissure detection in computed tomography (CT) is a critical component for automatic lobar segmentation. The majority of fissure detection methods use feature descriptors that are hand-crafted, low-level, and have local spatial extent. The design of such feature detectors is typically targeted towards normal fissure anatomy, yielding low sensitivity to weak and abnormal fissures that are common in clinical datasets. Furthermore, local features commonly suffer from low specificity, as the complex textures in the lung can be indistinguishable from the fissure when global context is not considered. We propose a supervised discriminative learning framework for simultaneous feature extraction and classification. The proposed framework, called FissureNet, is a coarse-to-fine cascade of two convolutional neural networks. The coarse-to-fine strategy alleviates the challenges associated with training a network to segment a thin structure that represents a small fraction of the image voxels. FissureNet was evaluated on a cohort of 3706 subjects with inspiration and expiration 3DCT scans from the COPDGene clinical trial and a cohort of 20 subjects with 4DCT scans from a lung cancer clinical trial. On both datasets, FissureNet showed superior performance compared to a deep learning approach using the U-Net architecture and a Hessian-based fissure detection method in terms of area under the precision-recall

curve (PR-AUC). The overall PR-AUC for FissureNet, U-Net, and Hessian on the COPDGene (lung cancer) dataset was 0.980 (0.966), 0.963 (0.937), and 0.158 (0.182), respectively. On a subset of 30 COPDGene scans, FissureNet was compared to a recently proposed advanced fissure detection method called derivative of sticks (DoS) and showed superior performance with a PR-AUC of 0.991 compared to 0.668 for DoS.

6.1 Introduction

Computed Tomography (CT) measures X-ray projections of the body at different angles to reconstruct a volumetric image of the anatomy. The contrast produced in a CT image reflects differences in X-ray photon attenuation, which in the lungs broadly reflects tissue density. Technological advancements in CT hardware have made it possible to scan the entire thoracic cavity in less than one second and reconstruct images with submillimeter spatial resolution. These properties make CT imaging the standard modality for imaging the intricate structures of the lung. Pulmonary CT is routinely used for diagnostics, treatment planning and delivery, and post-intervention evaluation.

CT images provide a rich source of information regarding the extent and spatial distribution of pulmonary disease. Computer-aided systems are essential for objective quantification and characterization of the complex information present in the image. Algorithms have been developed for detection and classification of nodules [142], texture classification of obstructive disease [133], pulmonary embolism detection [118], and quantitative airway analysis [24]. Although CT is an anatomical

imaging modality, functional information about the lung may be derived from CT scans collected at different inspiration levels using image registration [117].

The human lungs are composed of five lobar compartments, which are separated anatomically by three lobar fissures. The left oblique (major) fissure (LOF) separates the lower and upper lobes of the left lung. The right oblique (major) fissure (ROF) separates the lower lobe from the middle and upper lobes, and the right horizontal (minor) fissure (RHF) separates the middle and upper lobes of the right lung. It is often of clinical interest to perform quantitative analysis within each lobe individually. Boueiz et al. recently identified subgroups of upper-lobe-predominant emphysema and lower-lobe-predominant emphysema and found associations with clinical and imaging outcomes [10]. Accurate knowledge of lobar anatomy is critical for successfully treating severe emphysema with bronchoscopic lung volume reduction [92]. Lobar information also serves as a precursor to other image analysis algorithms including image registration. Currents- and varifolds-based registration algorithms rely on accurate surface representations of the lungs, lobes, and vessel trees [100].

The lobes are generally anatomically independent, but incomplete fissures are possible and the detection of incompleteness may be clinically relevant. An individual's unique lobar structure is likely to influence lung tissue mechanics and patterns of regional ventilation. Fissure incompleteness and the resulting collateral ventilation reduces the efficacy of endobronchial valves [71, 72, 32]. Gopelmann et al. recently showed that apical vs. basal emphysema distribution varies with fissure integrity [52]. However, Pu et al. found no relationship between fissure integrity and

COPD severity [112].

Natural variability in lobar anatomy has impeded the development of robust CT analysis methods for fissure and lobar segmentation. In cross-sectional CT images the fissures appear as thin surface-like structures (less than 1 mm thick) with higher image intensity than the surrounding lung parenchyma. This makes it difficult to identify fissures in low-dose or thick-slice CT scans. Fissure segmentation in pathological lungs is further complicated by diseases that locally resemble fissures, for example, bullous lung disease and fibrosis may locally resemble fissures.

Despite these challenges, many attempts have been made to design automatic methods for lobar segmentation [161, 146, 74, 163, 141, 75, 114]. The majority of these methods consist of four common modules: lung segmentation, fissure detection using local appearance information, removal of falsely identified fissures, and surface fitting to interpolate and/or extrapolate incomplete fissures. Doel et al. presented an extensive review on pulmonary lobe segmentation and proposed that these individual components should be independently developed and evaluated, opposed to comparing entire pipelines [30]. We follow this proposal and focus on the fissure detection in this work.

Several methods have been proposed for the detection of fissures in CT images. Eigenanalysis of the Hessian matrix is commonly used to exploit the characteristic property that plane-like structures have one direction with large curvature in the intensity profile and two orthogonal directions with vanishing intensity curvature [42, 123, 80, 152, 75]. Zhang et al. used a ridgeness operator based on 2D multi-local level

set extrinsic curvature measure with structure tensor analysis (MLSEC-ST) [161, 83]. Other works use knowledge of fissure appearance on 2D cross sections to design a filter bank of 2D line filters to detect fissure structures [70, 156, 157]. Traditional machine learning approaches use domain-specific hand-crafted features and labeled training data to train a classifier. van Rikxoort et al. used a feature set including intensity, Gaussian derivatives, gradient, and Hessian eigenvalues with labeled training data to build a kNN classifier [148]. The authors showed superior performance compared to conventional unsupervised fissure detection. Wei et al. trained an artificial neural network using texture-derived image features. However, a limitation of this method is that it requires extensive post-processing and only works on major fissures [151].

These existing fissure detection methods are limited to local descriptors of fissure shape and appearance. Although local information is necessary for the precise localization of the fissure, we argue that it is not sufficient. Weak and incomplete fissures diminish local response, and pulmonary disease can locally resemble fissures. We hypothesize that knowledge of global and contextual information can improve specificity by providing guidance when the fissure signal is low or noisy. However, it is far more challenging to design abstract features, such as those that capture global context, compared to low-level features, such as edges. Additionally, hand-crafting features requires domain expertise, and generalizing such a framework to other tasks is not trivial. Alternatively, convolutional neural networks (ConvNets or CNNs) are capable of learning abstract features directly from training data.

Several ConvNet architectures have been proposed for semantic segmentation;

the majority are symmetrical networks consisting of an encoder and corresponding decoder [119, 7, 125]. Compared to a classification ConvNet which yields a single prediction for each class, a segmentation ConvNet produces a prediction map that has the same spatial resolution as the input. U-Net and SegNet are notable encoder-decoder networks, each of which incorporate skip connections between corresponding encoder and decoder elements to preserve precise localization information that would otherwise be lost with pooling operations [119, 7]. These symmetrical networks are memory intensive and cannot be trained on entire volumetric medical images due to current GPU memory limitations. The majority of ConvNet methods use either 2D slices or small image crops to accommodate memory limitations thereby compromising the capacity of the network to learn large-scale 3D features or global patterns.

For the task of fissure segmentation, both 3D structure and global context are critical for accurate segmentation. Therefore 2D slices or patchwise approaches are not ideal. Furthermore, directly training a network to segment fissures is challenging due to the large class imbalance between fissure and non-fissure voxels. High accuracy could be achieved by learning the trivial classifier that always predicts the majority class (i.e. non-fissure).

To address these challenges, we propose a new coarse-to-fine deep learning segmentation approach called FissureNet. FissureNet achieves superior segmentation performance compared to other methods by concatenating two Seg3DNet ConvNets. The new Seg3DNet[1] architecture is less memory-intensive compared to U-Net and

[1]While Seg3DNet and the existing SegNet [7] are similar in name, the proposed Seg3DNet

SegNet, enabling it to learn global contextual information from entire lung images. Seg3DNet is a generic 3D segmentation network suitable for many applications. Within FissureNet, the first Seg3DNet is trained to detect an approximate fissure region of interest (ROI) and the second Seg3DNet is trained to detect precise fissure location within the ROI. The coarse-to-fine approach used by FissureNet overcomes the challenges associated with training a network to segment a thin structure that represents a very small fraction of the total voxel count.

6.2 Methods

6.2.1 Overview

We model fissure detection as a probabilistic classification problem. Given a dataset X and a finite class set Y, a probabilistic classifier models the conditional probability distribution $P(Y|X)$. That is, given a feature vector $x \in X$, the classifier predicts a probability distribution over the class set Y. The features and the conditional probability distribution are learned jointly through end-to-end training of a Seg3DNet.

For pulmonary fissure classification the class set Y consists of the three fissures and a non-fissure class, such that all voxels that are not fissure are assigned to the non-fissure class. The number of fissure voxels is very small compared to the number of non-fissure voxels; there is approximately one fissure voxel for every 100 non-fissure voxels within the lung mask (at the image resolution used in this study). FissureNet uses a coarse-to-fine approach by cascading two Seg3DNets (Fig. 6.1). The first Seg3DNet is trained to detect an approximate fissure region of interest (ROI) and

uses a different architecture and is not related to SegNet.

the second Seg3DNet is trained to detect the precise fissure location within the ROI. Separate pipelines are trained for the left and right lungs, yielding four total Seg3DNet classifiers: left fissure ROI, right fissure ROI, left fissure, and right fissure. The proposed Seg3DNet architecture is illustrated in Fig 6.2.

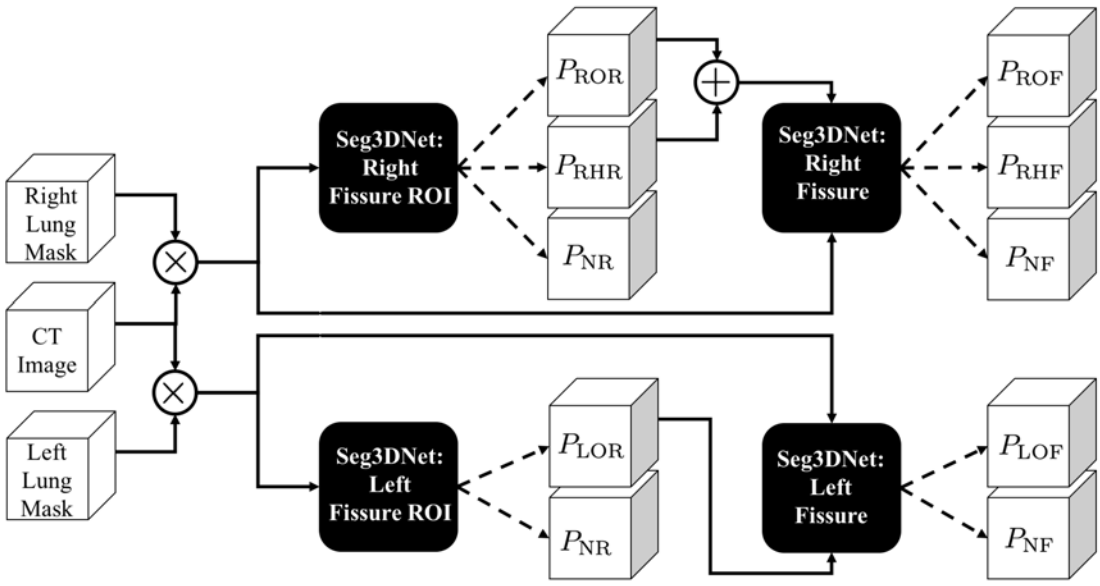


Figure 6.1. FissureNet: coarse-to-fine network cascade. Black boxes represent four Seg3DNet classifiers, each trained for a different classification task as indicated in the box. For the right lung pipeline (top), the CT image is masked with the right lung mask and input to the right fissure ROI Seg3DNet. The output of the ROI Seg3DNet represents the probability that each voxel is right oblique ROI (P_{ROR}), right horizontal ROI (P_{RHR}), and non-fissure ROI (P_{NR}). The input to the right fissure Seg3DNet is the masked CT image and the probability maps P_{ROR} or P_{RHR}. The output of the right fissure Seg3DNet gives the probability that each voxel is right oblique fissure (P_{ROF}), right horizontal fissure (P_{RHF}), and non-fissure (P_{NF}). The left lung pipeline (bottom) is similar, except each classifier only predicts two classes corresponding to left oblique fissure and non-fissure.

6.2.2 Convolutional Neural Network

A ConvNet is a specialized neural network model designed to exploit patterns in spatially correlated data, such as images and videos. At a high level, a ConvNet has multiple layers of learned feature detectors arranged hierarchically. The feature detectors in each layer are local, however, the composition of layers allows the spatial extent defined on the input image, called receptive field, to grow with layer depth. This design gives the network the capacity to learn global features, without the computational overhead and increased number of parameters required for large feature detectors. The feature detectors are shared spatially, making a given feature relevant at any location in the image. Layers with feature detectors are called convolutional layers. The feature detectors, or kernels, are not explicitly encoded but are the parameters being learned through optimization. Convolutional layers are typically followed by an elementwise nonlinearity and interleaved with pooling layers which serve to reduce the spatial resolution.

Each layer in a ConvNet takes a feature representation as input, performs an operation to transform the input, and produces a new feature representation as the output. The layers are arranged hierarchically: the output feature representation of one layer serves as the input to the following layer. Each feature representation consists of a set of spatial activation maps, each representing a different feature type. The activation maps are concatenated along a non-spatial dimension, i.e. the channel dimension, to form the feature representation. Therefore, in a 3D ConvNet the intermediate feature representations are all 4D images. The channel dimension is

analogous to that of an RGB image, i.e., the voxels are vector-valued and each vector element represents a different feature type.

The ConvNet parameters (feature detectors) are learned from labeled training data using backpropagation [120]. Starting with randomly initialized parameters, a training example is propagated through the network and a prediction is made. The dissimilarity between the prediction and the true label is quantified with a loss function. The gradient on the loss with respect to each parameter is calculated using backpropagation, which is a recursive application of the chain rule. All parameters are updated to decrease the error using stochastic gradient descent, or some variant thereof.

6.2.3 Seg3DNet

In this work, we propose a 3D ConvNet architecture for image segmentation called Seg3DNet (Fig. 6.2). Seg3DNet consists of an encoder which generates a high dimensional feature representation of the image, and a decoder which decodes the features to produce a segmentation. Unlike many segmentation architectures, the encoder and decoder modules in Seg3DNet are asymmetrical. The encoder module consists of L resolution levels l_i for $i = 0, 1, ..., L-1$, where the activation maps in level l_i are downsampled by a factor of 2^i relative to the full resolution level l_0. Each level of the encoder has two convolutional layers followed by a max-pooling layer. All convolutional layers use $3 \times 3 \times 3$ voxel kernels, and the number of kernels in level l_i is given by $N_i = 2^{i+5}$. After the second convolution layer of each level, max pooling with kernel size $2 \times 2 \times 2$ and stride of 2 produces the downsampling factor of 2

between levels. While recent ConvNet architectures have eliminated pooling layers, downsampling is necessary to achieve a global receptive field on large input volumes. To mitigate the loss of precise localization information from the pooling layers, the decoder network combines representations from all scale levels.

The decoder module condenses the representation at each scale level to a single activation map using a convolutional layer with a single voxel kernel of size $1 \times 1 \times 1 \times N_i$. The lower resolution activation maps are upsampled to full resolution using nearest neighbor interpolation followed by a convolution with filter size $2^i + 1$, effectively performing a variant of deconvolution [96]. The resulting activation maps, one from each scale level, are concatenated along the feature dimension to form a multi-scale representation. Two more convolutional layers are used to combine information from different scales.

The representation at the last layer of the Seg3DNet has $|Y|$ activation maps each with the same spatial dimensions as the input volume. The output at spatial location $\mathbf{x}$ of activation map y, $f_y(\mathbf{x})$, is interpreted as an unnormalized log probability of $\mathbf{x}$ belonging to class y. The softmax vector nonlinearity is used to obtain the conditional probability distribution, given by:

$$P(Y = y|\mathbf{x}) = \frac{e^{f_y(\mathbf{x})}}{\sum_{j \in Y} e^{f_j(\mathbf{x})}}. \tag{6.1}$$

We denote the probability for each class $y \in Y$ as $P_y(\mathbf{x})$. By construction, Y is a valid probability distribution function with $P_y(\mathbf{x}) \in [0, 1]\ \forall y \in Y$, and $\sum_{y \in Y} P_y(\mathbf{x}) = 1$. For the right lung ConvNets, we define separate classes to distinguish between oblique and horizontal fissures. Therefore, the class set cardinality for the left and right lung

ConvNets is $|Y| = 2$ and $|Y| = 3$, respectively.

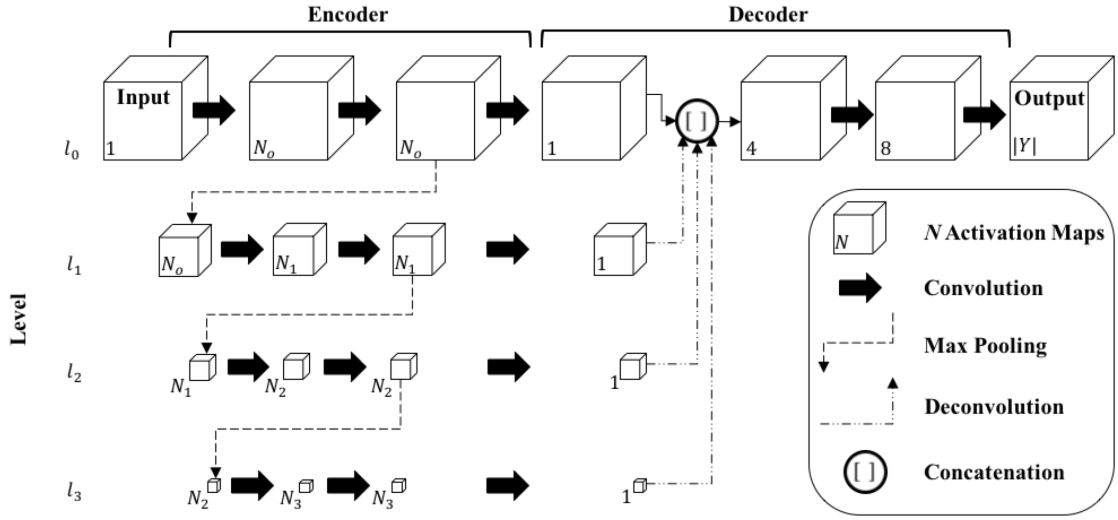


Figure 6.2. Proposed Seg3DNet architecture. Each arrow represents an operation performed by a layer and each cube represents the intermediate feature representations produced by a layer. For visualization purposes, only the spatial dimensions of the feature representations are illustrated. The number of activation maps (size of channel dimension) is denoted in the lower left corner. For the encoder module, we define $N_i = 2^{i+5}$ so that the number of activation maps increases by a factor of two at each level. The number of kernels used in each convolutional layer can be inferred by the number of activation maps in the layer's output representation, e.g., the first convolutional layer has $N_0 = 2^{0+5} = 32$ kernels. The relative spatial size of the activation maps are drawn to scale. At each level the feature representation is spatially downsampled by a factor of two. Batch normalization and ReLU nonlinearity are performed after each convolution except the last.

Batch normalization [62] and ReLU nonlinearities [91] are used after each convolution layer with the exception of the last layer. All convolutional layers use zero-padding to prevent reduction in spatial dimensions.

6.2.4 FissureNet

As shown in Fig. 6.1, FissureNet has two parallel pipelines, each of which is a coarse-to-fine cascade of two Seg3DNets. The first Seg3DNet is trained to detect a fissure ROI. The original ground truth fissure segmentations are modified to produce the fissure ROI training labels. A voxel belongs to the fissure ROI if it is located within 5 mm of the corresponding fissure, otherwise it is non-fissure. This dilation of the single-voxel ground truth reduces the class skewness. Additionally, by dilating the ground truth fissure the network is able to focus on global patterns rather than precise fissure appearance. As a result, the network is more robust to weak and radiographically incomplete fissures. The fissure ROI allows for small imperfections in the training data which are expected due to the nature of manually tracing a single voxel curve.

For training the first Seg3DNet, we define the loss associated with each voxel using categorical cross entropy of the form

$$L(\mathbf{x}, Y) = -\sum_{y \in Y} t_y(\mathbf{x}) \log P_y(\mathbf{x}), \tag{6.2}$$

where $t_y(\mathbf{x})$ represents a one-hot encoding of the target label for voxel $\mathbf{x}$ and class y, i.e., $t_y(\mathbf{x})$ is one when y corresponds to the true class and zero for all other classes The total loss for an input image is given by

$$L_{\text{ROI}} = \frac{\sum_{\mathbf{x} \in \Omega} L(\mathbf{x}, Y_{\text{ROI}})}{|\Omega|}, \tag{6.3}$$

where Ω is the input image domain and Y_{ROI} is ROI classifier class set.

The second Seg3DNet is trained to detect the precise fissure location. The

original ground truth fissure segmentations are used as training labels. The loss associated with each voxel is the same as the first Seg3DNet (6.2). However, the total loss is a weighted average using the probability that the voxel is in a fissure ROI

$$L_{\mathrm{F}} = \frac{\sum_{\mathbf{x}\in\Omega}(1 - P_{\mathrm{NR}}(\mathbf{x}))L(\mathbf{x}, Y_{\mathrm{F}})}{\sum_{\mathbf{x}\in\Omega}(1 - P_{\mathrm{NR}}(\mathbf{x}))}, \tag{6.4}$$

where $P_{\mathrm{NR}}(\mathbf{x})$ is the probability that voxel $\mathbf{x}$ is non-fissure ROI as predicted by the first Seg3DNet and Y_{F} is the fissure classifier class set. This weighting limits the contribution of the large number of non-fissure voxels to the loss function, mitigating the class imbalance problem while allowing for precise fissure localization.

6.2.5 Implementation

FissureNet was implemented using the open source frameworks Theano [140] and Lasagne [28]. Training was performed using a P40 NVIDIA GPU with 24 GB of RAM. Adam optimization was used with an initial learning rate of 5×10^{-4} [67]. All parameters were initialized using Xavier normal initialization [50]. Each network was trained for six epochs, which took approximately 48 hours. Inference time with the trained network is 10 seconds per lung on a consumer grade GPU card.

6.3 Experimental Methods

6.3.1 Datasets and Preprocessing

Training and testing data were acquired from the COPDGene study, a large multi-center clinical trial with over 10,000 subjects with chronic obstructive pulmonary disease (COPD) [115]. COPD which includes emphysema and chronic bronchitis, is characterized by airway inflammation and large regions of trapped gas on CT.

The COPDGene image datasets were acquired across 21 imaging centers using

a variety of scanner makes and models. Each patient had two breath-hold 3D CT scans acquired, one at total lung capacity (TLC) with a dose of 200 mAs and one at functional residual capacity (FRC) with a dose of 50 mAs. Original slice thicknesses ranged from 0.625 mm to 0.9 mm. See [115] for the complete image acquisition protocol.

A subset of 1601 subjects was selected for training and a separate disjoint subset of 3706 subjects was selected for testing. The TLC scan and FRC scan for each subject were utilized, producing a training dataset with 3202 scans and a testing dataset with 7412 scans. All COPD GOLD levels [13] were used for training and evaluation; see Table 6.1 for distribution of disease severity in training and testing datasets. Each subject in COPDGene has a unique identifier consisting of five numbers and one letter, e.g., 10005Q. Subjects with identifiers that begin with values in the range 10-12 were included in the training dataset and subjects with identifiers that begin with values in the range 13-19 were included in the testing dataset.

An additional dataset of 20 4-dimensional computed tomography (4DCT) scans from a lung cancer clinical trial were used were used for evaluation. The 4DCT scans were acquired on a Siemens EDGE CT scanner with parameters of 120 kV tube voltage, 100 mAs tube current, 0.5 second tube rotation period, 0.09 pitch, 76.8 mm beam collimation, 128 detector rows, and a reconstructed slice thickness of 0.6 mm. The 4DCT scans were acquired with audio guidance. The image data was retrospectively sorted and reconstructed into ten phase images. A single phase from

Table 6.1. Disease stratification for the training and testing datasets. GOLD0-GOLD4 defined in [13]. No PFT: spirometry data not available; PRISm: Preserved Ratio Impaired Spirometry [149]

	Training	Testing
GOLD0	583	1625
GOLD1	133	309
GOLD2	350	617
GOLD3	235	360
GOLD4	135	186
No PFT	28	53
Non-Smoker	0	90
PRISm	137	466
Total	1601	3706

each subject was selected for fissure evaluation, the selected phases were chosen to represent an assorted range of tidal volumes and phases. There were no scans of this type included in the training dataset.

Due to memory constraints on the GPU, all images were resampled to isotropic 1 mm^3 voxels. For each lung, subvolumes of size $64 \times 200 \times 200$ voxels were extracted for training. For the majority of subjects, this crop size covers the entire sagittal view of the lung, however, it may not cover all sagittal slices. Therefore, for training we extracted three subvolumes centered at different sagittal slices to ensure the entire lung was covered.

For preprocessing, CT intensity values were clamped to the range of interest for fissure detection (i.e. -1024 HU and -200 HU) which also removes outliers caused by calcification and metal artifacts. Voxels outside of the lung mask were set to -1024 HU. After clamping and masking, the CT image intensities were linearly rescaled according

to $I'(\mathbf{x}) = \frac{I(\mathbf{x})-\mu_{\text{HU}}}{\sigma_{\text{HU}}}$, where μ_{HU} and σ_{HU} are the mean and standard deviation of HU values calculated over the entire training dataset and $I(\mathbf{x})$ and $I'(\mathbf{x})$ are the original HU values and rescaled image intensities, respectively, for a given voxel $\mathbf{x}$.

6.3.2 Ground Truth

Lung and lobar segmentations in the COPDGene study were obtained using a commercial software package (Apollo, VIDA Diagnostics, Coralville, IA), followed by manual inspection and editing as needed. The Apollo software calculates a fissure probability measure using local fissure appearance and anatomical information from the airways and vasculature and then uses surface fitting to interpolate and extrapolate a complete fissure. The manual correction was performed by trained analysts (professional research assistants) with experience levels ranging from 0-4 years. Corrections were supervised by an experienced radiologist. Manual analysts were instructed to interpolate and extrapolate when necessary to completely divide the lung into five lobes. Fissure segmentations were extracted automatically from the lobar segmentations by identifying adjacent voxels with different lobe labels, producing a two-voxel thin fissure segmentation.

Fissure segmentations for the lung cancer 4DCT dataset were defined manually by an experienced medical physics PhD student using MimVista 6.4.7 software (MIM Software, Cleveland, OH). For this dataset, both complete fissures (forming full lobar boundaries) and visible fissures were identified. The visible fissures were first identified and subsequently interpolated and extrapolated to separate lobes when necessary.

6.3.3 Comparison of Fissure Detection Methods

We compared the proposed FissureNet to three other fissure detection methods: a conventional Hessian-based method [152], the derivative of sticks (DoS) method [157], and a deep learning approach using the popular U-Net architecture [119]. The Hessian-based and U-Net methods were evaluated on the entire testing dataset (7412 scans). For comparison to DoS, a subset of 30 scans from the testing dataset was used. For this subset, three subjects were randomly selected from each GOLD level. Both the TLC scan and the FRC scan were used for each subject. The DoS method consists of a fissure filtering step (DoS1) and a post-processing step (DoS2). The aim of the post-processing step is to remove falsely detected fissure voxels, particularly those that are connected to the true fissure. This is done using a pipeline consisting of global multi-thresholding, junction detection and removal, and connected component analysis. The aim of the Hessian-based and DoS methods is the detection of visible fissures and thus no interpolation or extrapolation is performed. For the U-Net method, the depth of the U-Net was set to three levels and the branching factor was set to four to accommodate the increased memory demand of the decoder network. To mitigate the class imbalance, we used the approach from [119], which weights the underrepresented class to have higher misclassification cost.

6.3.4 Evaluation Metrics

Receiver operating characteristic (ROC) curves are commonly used to evaluate the performance of a binary classifier by measuring the tradeoff between true positive rate (TPR) and false positive rate (FPR) at different thresholds. Similarly, precision-recall

(PR) curves measure the tradeoff between precision and recall (recall is the same as TPR). Recent studies have demonstrated that PR curves are better at evaluating and comparing binary classifiers in problems with a skewed prior class probability [121]. In such cases, ROC curves tend to be overly optimistic and do not distinguish between methods with different false positive behavior. The prior probability of fissure voxels is very small, so we report PR results to better discriminate between methods.

For PR evaluation, we use a 3 mm margin as described in [157]. Briefly, for calculating precision, predicted fissure voxels are categorized as true positive (TP) or false positive (FP): TP if the voxel is within 3 mm of the ground truth fissure or otherwise FP. Likewise, for calculating recall, ground truth fissure voxels are categorized as TP or false negative (FN): TP if the voxel is within 3 mm of the predicted fissure or otherwise FN. The 3 mm margin accounts for the subjectivity in manual delineation of a thin structure with low contrast. All voxels within the lung mask are considered for PR calculation and the ground truth is a two-voxel thin fissure. We report the PR curve and the area under the PR curve (PR-AUC). The optimal operating point on a PR curve is the upper right corner, and PR-AUC has values ranging from 0 to 1, with 1 describing a perfect classifier.

We evaluated the surface distance between predicted fissure and nearest ground truth fissure. For this analysis, we obtained hard fissure predictions by thresholding the probability output. The optimal threshold was determined separately for each method and was defined as the threshold on the PR curve where precision is equal to recall using a dataset of 20 COPDGene subjects (using both TLC and FRC scans) that

were not included in the training or testing datasets. For each scan the average surface distance (ASD) and standard deviation of surface distances (SDSD) was calculated.

6.4 Results

We compared FissureNet against three other fissure detection methods: Hessian-based [152], DoS [157], and U-Net [119]. The Hessian and the DoS methods do not distinguish between the right oblique and horizontal fissures. Therefore, only an aggregated right fissure (RF) measure is made for the right lung. For comparison, the RF measure is evaluated on FissureNet and U-Net by adding the ROF and RHF probabilities.

Fig. 6.3 compares PR curves for FissureNet and U-Net methods on 3706 subjects (TLC and FRC scans for each subject) from COPDGene and 20 lung cancer subjects with 4DCT scans. PR-AUCs for FissureNet, U-Net, and Hessian are displayed in Tables 6.2 and 6.3 for the COPDGene and lung cancer datasets, respectively. Overall, PR-AUC for FissureNet, U-Net, and Hessian methods were 0.980, 0.963, and 0.158, respectively, on the COPDGene dataset and 0.966, 0.937, and 0.182, respectively, on the lung cancer dataset. All methods had similar performance on the COPDGene and the lung cancer datasets and FissureNet performed best with regards to PR-AUC. Table 6.4 shows PR-AUCs on the lung cancer dataset using a ground truth which only indicates radiographically visible fissures. FissureNet and U-Net performed slightly better using the visible-only ground truth, while Hessian performed slightly worse. Table 6.5 shows PR-AUCs for FissureNet and DoS evaluated on a subset of 15 subjects (30 scans). The post-processing in the DoS method greatly improves the PR-AUC from 0.177 (DoS1) to 0.668 (DoS2), however, FissureNet

consistently performed better than DoS2 without any post-processing with an overall PR-AUC of 0.991. In Tables 6.2-6.5, only LOF and RF fissures are included in mean calculation to avoid over-weighting right lung results.

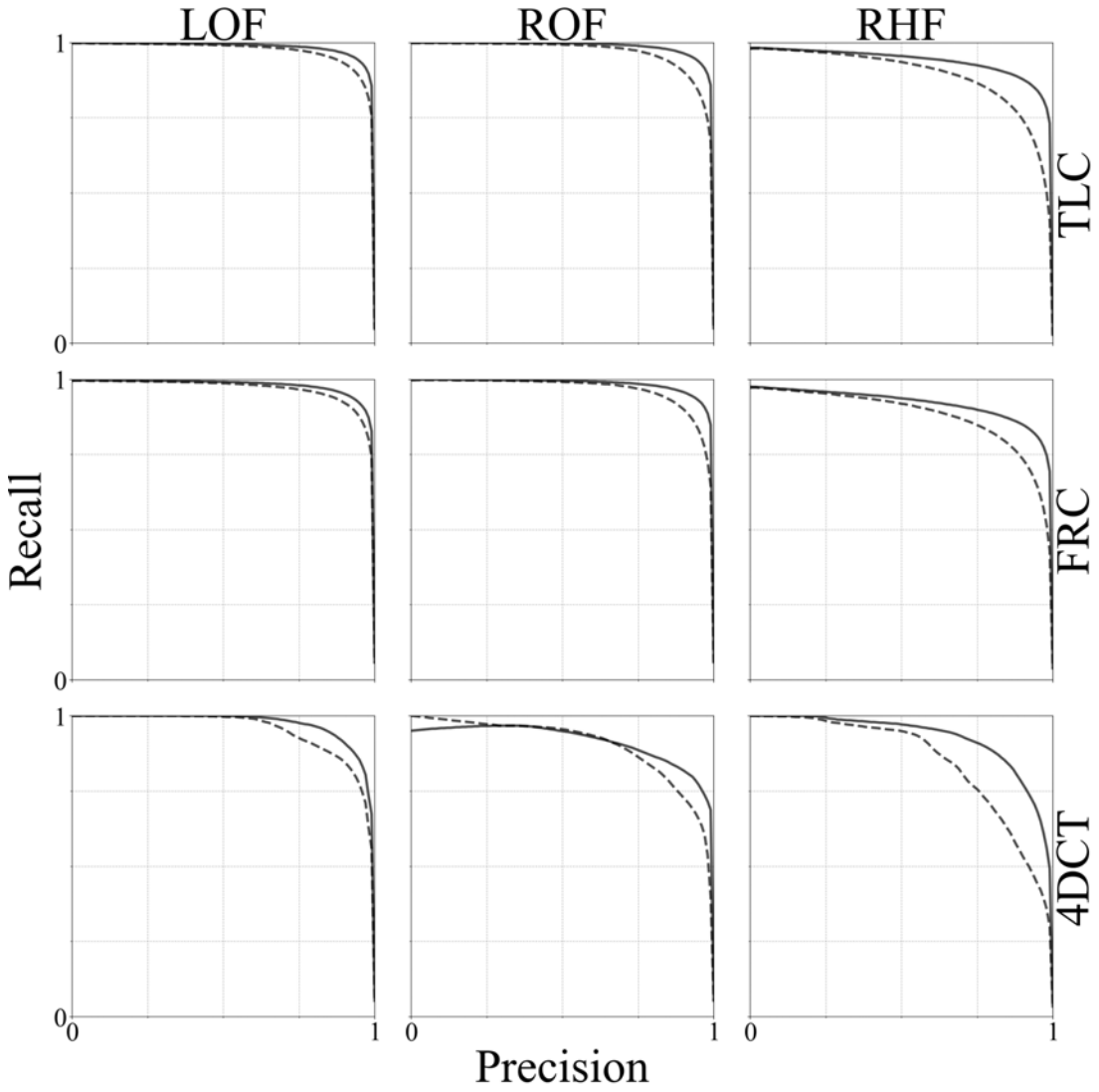


Figure 6.3. Precision-Recall curves for FissureNet (—) and U-Net (- - -) evaluated on the testing dataset of 3706 COPDGene subjects (TLC and FRC scans for each subject) and 20 lung cancer subjects (4DCT).

Table 6.2. PR-AUC for 3706 subject (7412 scan) testing dataset from COPDGene

Phase	Fissure	PR-AUC		
		Hessian	U-Net	FissureNet
TLC	LOF	0.145	0.973	0.985
	RF	0.216	0.959	0.982
	ROF	—	0.967	0.987
	RHF	—	0.891	0.939
FRC	LOF	0.108	0.968	0.979
	RF	0.165	0.952	0.975
	ROF	—	0.964	0.983
	RHF	—	0.878	0.919
Mean		**0.158**	**0.963**	**0.980**

Table 6.3. PR-AUC for 20 lung cancer subjects with 4DCT scans. Non-visible fissures were interpolated and extrapolated to form complete boundaries between lobes

Fissure	PR-AUC		
	Hessian	Unet	FissureNet
LOF	0.171	0.950	0.972
RF	0.193	0.924	0.961
ROF	—	0.899	0.916
RHF	—	0.848	0.926
Mean	**0.182**	**0.937**	**0.966**

Table 6.4. PR-AUC for 20 lung cancer subjects with 4DCT scans. Only visible fissures were marked in the ground truth

Fissure	PR-AUC		
	Hessian	Unet	FissureNet
LOF	0.113	0.978	0.992
RF	0.137	0.985	0.988
ROF	—	0.965	0.953
RHF	—	0.917	0.946
Mean	**0.125**	**0.982**	**0.990**

Table 6.5. PR-AUC for 15 subject (30 scan) testing subset from COPDGene

Phase	Fissure	PR-AUC		
		DoS1	DoS2	FissureNet
TLC	LOF	0.167	0.706	0.993
	RF	0.155	0.652	0.992
	ROF	—	—	0.995
	RHF	—	—	0.968
FRC	LOF	0.191	0.704	0.987
	RF	0.196	0.610	0.990
	ROF	—	—	0.994
	RHF	—	—	0.967
Mean		**0.177**	**0.668**	**0.991**

Table 6.6. Average surface distance (ASD) and standard deviation of surface distances (SDSD) averaged over 3706 subject COPDGene testing dataset (TLC and FRC) and lung cancer testing dataset (4DCT). Distances reported in mm

Dataset	Fissure	ASD		SDSD	
		U-Net	FissureNet	U-Net	FissureNet
TLC	LOF	3.75	0.65	10.56	2.14
	ROF	4.83	0.57	12.78	2.06
	RHF	7.23	2.39	13.64	5.49
FRC	LOF	2.32	0.66	6.93	1.66
	ROF	4.03	0.53	10.38	1.43
	RHF	5.57	1.96	9.79	3.85
4DCT	LOF	2.97	0.62	9.06	1.36
	ROF	6.50	1.97	13.86	2.88
	RHF	9.31	1.90	15.27	4.93

Table 6.6 shows ASD and SDSD averaged over all subjects. On average, the ASD for FissureNet was less than U-Net for all scan types and fissures. Fig. 6.4 shows a histogram of the differences in ASD between U-Net and FissureNet on a subject-by-subject basis; 97% of the histogram area is to the right of the vertical line corresponding to cases where FissureNet has a lower ASD compared to U-Net.

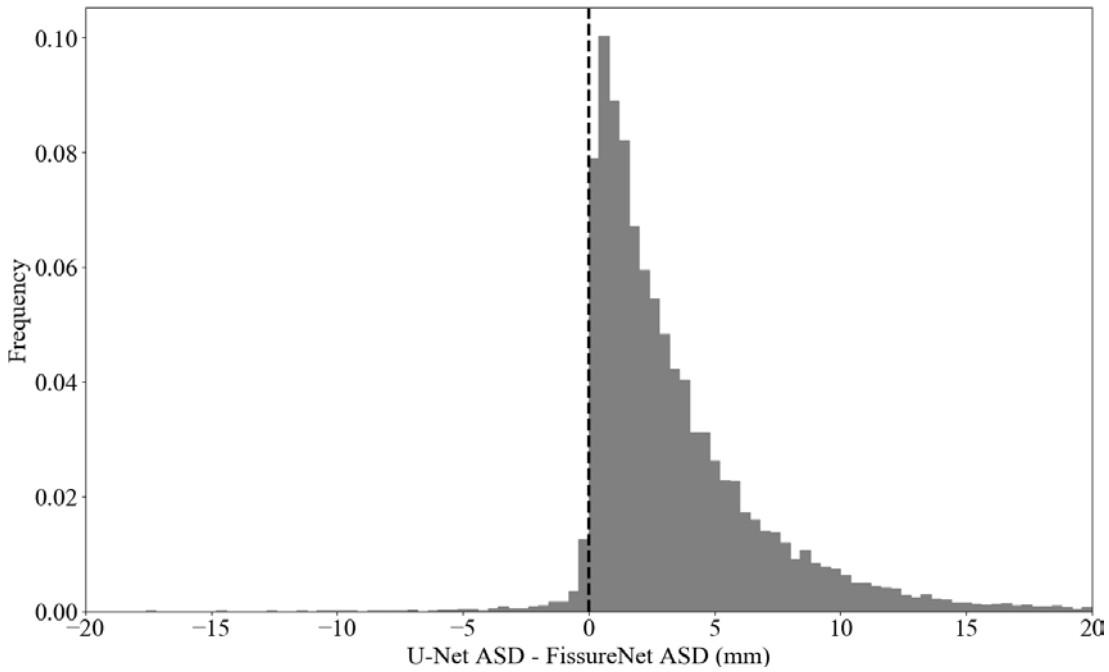


Figure 6.4. Histogram of differences in ASD between U-Net and FissureNet on a subject-by-subject basis.

Statistical testing was performed to test for significant differences in performance between methods with regards to evaluation metrics. Paired t-tests showed that FissureNet had a significantly greater PR-AUC and a significantly lower ASD compared to U-Net on both the COPDGene and lung cancer datasets ($p < 0.001$). Additionally,

FissureNet had a significantly greater PR-AUC compared to Hessian on both the COPDGene and lung cancer datasets ($p < 0.001$).

Representative fissure detection results are displayed in Fig. 6.5 for the COPDGene dataset and Fig. 6.6 for the lung cancer dataset. These results show DoS2 and U-Net have far fewer false positives compared to Hessian, however, FissureNet produces the fewest false positives while maintaining high sensitivity. The difference in false positive behavior between FissureNet and U-Net is further emphasized in Fig. 6.7, where surface renderings are annotated in red to depict false positives. The only post-processing performed to generate the renderings was thresholding at the optimal PR-AUC thresholds.

6.5 Discussion

Existing fissure detection methods are limited to hand-crafted and local features. These features typically suffer from low specificity as it is difficult to differentiate fissures from the other structures in the lung without global context. Additionally, it is difficult to design features that are robust against all fissure variations, especially for global compared to local features. To overcome the challenge of designing robust and discriminative features we use a deep learning approach to learn the feature detectors from labeled training cases. The main challenges associated with training a ConvNet to detect fissures in CT images are the size of the input images and the highly skewed class distributions.

The majority of ConvNets used in medical imaging applications use 2D image slices or use a sliding window approach with small image crops to overcome limitations

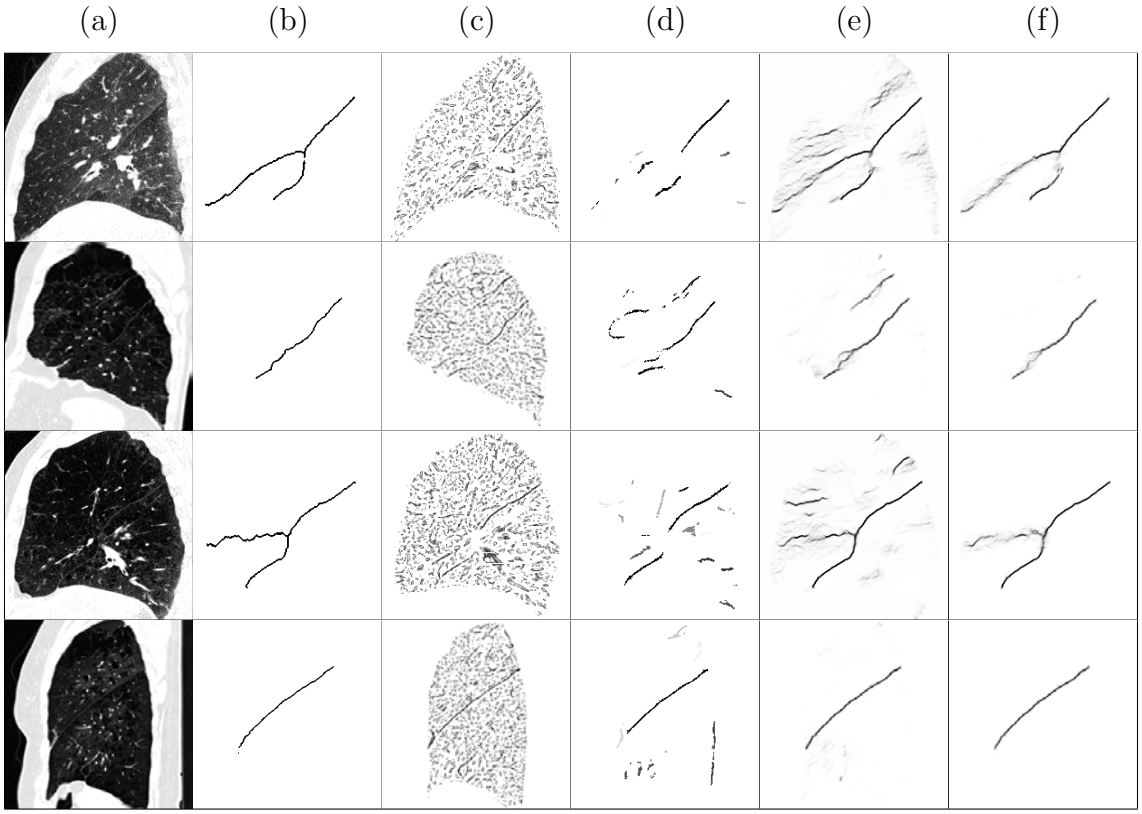


Figure 6.5. Representative results for four COPDGene subjects in rows one to four. (a) CT sagittal slice, (b) ground truth, (c) Hessian, (d) DoS2, (e) U-Net, (f) FissureNet. Top to bottom: GOLD1 FRC, GOLD3 TLC, GOLD3 FRC, GOLD4 TLC. DoS2 greatly reduces the number of false positives compared to Hessian, however, for these cases it fails to detect the horizontal fissure (rows 1 and 3). FissureNet and U-Net results are similar with high sensitivity and few false positives, however, FissureNet has fewer false positives.

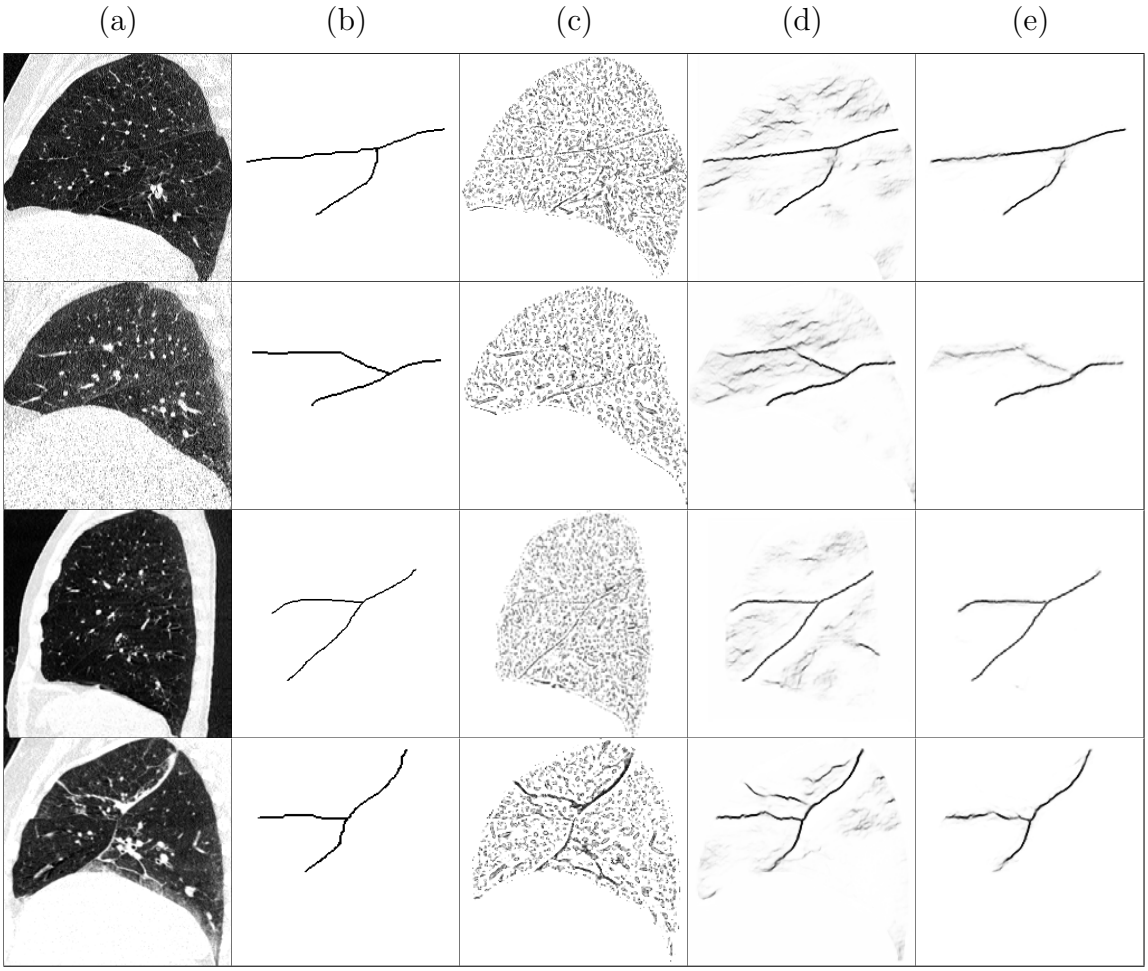


Figure 6.6. Representative results for four lung cancer subjects in rows one to four. (a) CT sagittal slice, (b) ground truth, (c) Hessian, (d) U-Net, (e) FissureNet. Row 4 has a dense pathology superior to the horizontal fissure that is falsely detected by Hessian and U-Net, however, FissureNet correctly classifies this as non-fissure.

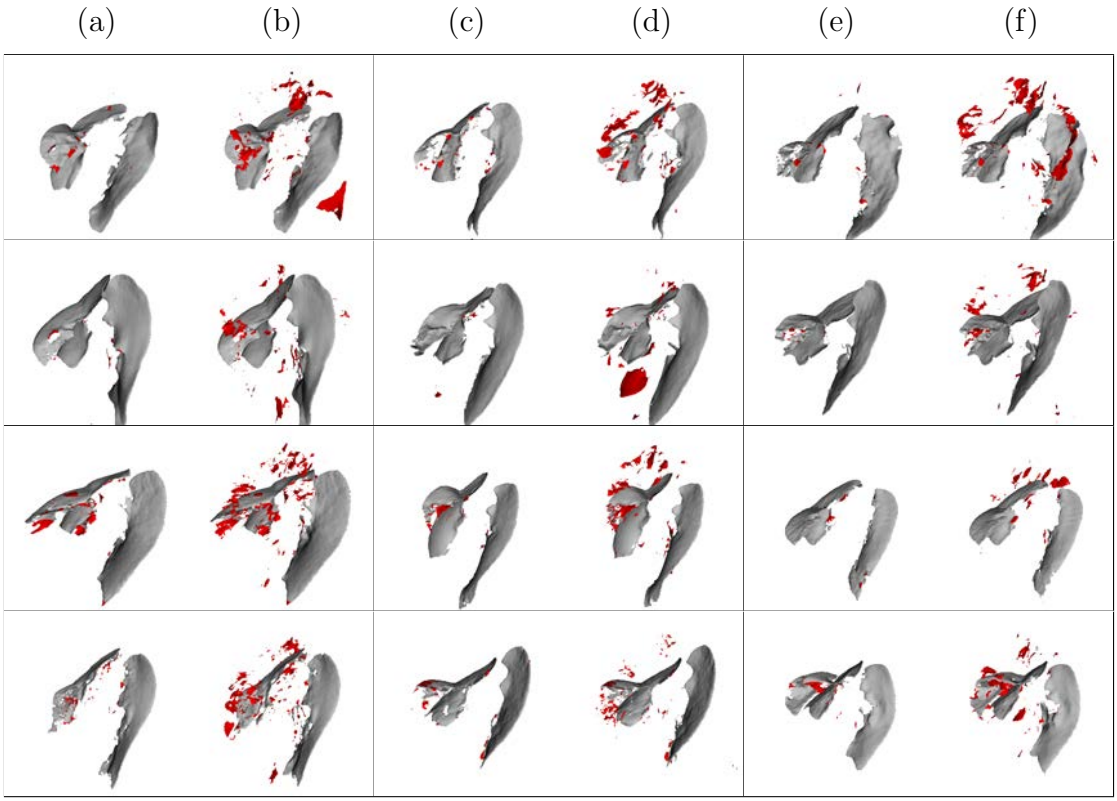


Figure 6.7. Surface renderings of FissureNet (a, c, e) and U-Net (b, d, f) results for six COPDGene subjects (rows 1 and 2) and six lung cancer subjects (rows 3 and 4). True positives and false positives are depicted in gray and red, respectively. Probability output was thresholded at optimal threshold for each method as determined by PR curve.

in GPU memory. While this is a reasonable approach for some tasks, for fissure segmentation it is not desirable. The 3D appearance of a fissure is important to distinguish it from other structures that would otherwise appear similar on 2D slices. Global information provides additional context which is especially important when the fissure signal is weak, however, this information is not considered by patch-based approaches.

Compared to other segmentation architectures, Seg3DNet is an asymmetrical encoder-decoder network which uses less memory in order to accommodate a 3D network, larger input images, and more network levels. This allows for global information to be learned and results in higher specificity. By training separate Seg3DNets for the left and right lungs, we were able to reduce the size of the input image by a factor of two. This optimization does not degrade performance, as information from one lung does not provide global information for fissure detection in the other lung.

To handle the class imbalance, we use a coarse-to-fine ConvNet cascade: the first ConvNet learns the fissure ROI and the second ConvNet learns the precise fissure location. In addition to mitigating the effect of class imbalance, the fissure ROI classifier is more sensitive to weak and incomplete fissures. Since the second training phase weights the voxel misclassification costs by the probability of being in the fissure ROI, the contribution of costs from the large number of non-fissure voxels is limited. Therefore, the class imbalance problem is mitigated while allowing for precise fissure prediction. A similar, and more elegant, approach would be to train a single network with two outputs: one for the fissure ROI and one for precise fissure prediction.

However, current limitations on GPU memory do not allow for this.

Fully-connected layers are not used in Seg3DNet, making it a fully-convolutional network (FCN) [125]. This greatly reduces the number of parameters and makes the network less prone to overfitting; the proposed network has 3 million parameters compared to the popular VGG-16 network which has 138 million parameters. Furthermore, in a FCN the number of parameters is not dependent on the input image size, so the network can be trained and deployed on images of different sizes. Our network was trained on fixed-size image crops of 64x200x200 due to limited GPU memory, however, in some cases the entire lung field does not fit in this crop. At test time there is more memory available as mini-batches are not used and gradients do not need to be stored for backpropagation. As a result, at test time much larger inputs can be used. In fact, the entire lung region, regardless of size, can be used as input and inference can be done in one forward pass per image. This is extremely efficient compared to patchwise approaches. In addition, Seg3DNet can accommodate different input image sizes, avoiding aggressive rescaling and interpolation that might degrade the fissure signal.

This is the first study to evaluate a fissure detection method on a dataset of this size and diversity: 3706 COPDGene subjects with TLC and FRC scans and 20 lung cancer subjects with 4DCT scans. The COPDGene data used for training and evaluation came from 21 different institutions. Different scanner makes and models were used, as well as different reconstruction algorithms. The diversity of the evaluation set was further enriched with a lung cancer dataset of 4DCT

scans. These scans were acquired at a lower dose during breathing, resulting in poorer image quality, motion blurring, and/or artifacts which were not present in the training dataset. Robustness to such diversity is generally a challenge when designing rule-based algorithms for image segmentation: it can be difficult to achieve similar performance across different scanning protocols and diseases.

Fissure detection performance was evaluated on four methods: Hessian-based, DoS, a deep learning approach using the U-Net architecture, and the proposed FissureNet. FissureNet and U-Net both greatly outperformed the Hessian and DoS methods on all datasets. Hessian and DoS methods were not able to detect weak fissures and produced many false positives at blood vessels and diseased regions. FissureNet consistently outperformed U-Net; while both methods demonstrated high sensitivity for fissure detection, FissureNet predicted fewer false positives. This can be attributed to the larger input patches and coarse-to-fine cascade, allowing the network to use more global context to differentiate true fissures from disease that resembles fissures.

On the COPDGene evaluation dataset, all methods performed better on TLC scans compared to FRC scans in terms of PR-AUC. However, in the COPDGene trial the TLC scans were acquired at a higher dose and thus the image quality was better, so better performance was expected. In the future, comparing images of the lung at different inspiration levels acquired using the same dose would help determine which inspiration level is best for fissure detection. Although the performance on FRC images was worse, the FissureNet results are nonetheless impressive for lower dose scans. This demonstrates the ability of FissureNet to generalize across different

scanning protocols. The COPDGene dataset consisted of subjects with a wide range of disease severity, encompassing all GOLD stages. It is more challenging to detect fissures in heavily diseased cases as alterations in the underlying tissue can resemble the fissure and/or result in abnormal tissue appearance. Performance of FissureNet was robust to these challenges.

Training a multi-class network for the right lung results in the ability to distinguish between oblique and horizontal fissures. This is the first fissure detection method to make this distinction. Since the ultimate goal is to divide the lungs into lobes, unique predictions for different fissures facilitates straightforward post-processing. A limitation of training FissureNet using a ground truth containing only oblique and horizontal fissures is an inability to detect accessory fissures. While accessory fissures have exhibit similar local appearance compared to the major fissures, the proposed FissureNet learns high level information encoded in the particular shapes and orientations of the oblique and horizontal fissures. However, introducing an accessory fissure class and providing additional annotation in the training data could extend the network's capability.

Detection of the right horizontal fissure was consistently worse than the oblique fissures for the COPDGene dataset. The orientation of the horizontal fissure is often parallel with the axial imaging plane, potentially obscuring the fissure in CT images. It is not uncommon for horizontal fissures to be radiographically incomplete or missing, hindering identification even by human analysts. Interestingly, on the 4DCT dataset the ROF has a higher ASD compared to the RF.

The COPDGene ground truth fissures used for evaluation have several limitations. The fissures were automatically extracted from lobar segmentations resulting in complete fissure boundaries for all cases even those with radiographically incomplete or missing fissures. In such cases, the extrapolated or interpolated fissure location is highly subjective and evaluating the performance of any automated method using such a ground truth is limited in these regions. Furthermore, the ground truth fissures in the COPDGene evaluation dataset were generated using the same method as the training dataset (Apollo software followed by manual correction). This introduces a bias for learning-based methods to identify complete fissures in unseen subjects regardless of actual fissure integrity. An additional possible bias may be attributed to the FissureNet and U-Net methods being trained on the COPDGene dataset, while the Hessian and DoS methods were developed on an independent dataset.

To address these limitations, evaluation was performed on a dataset of lung cancer subjects with 4DCT scans. The ground truth fissure segmentations for this dataset were generated manually. Additionally, both complete and visible-only fissures were annotated. All fissure detection methods performed worse on the 4DCT dataset compared to the COPDGene dataset. The 4DCT scans use a lower dose and commonly have motion artifacts and blurring, resulting in decreased fissure visibility. All methods performed better using the visible-only fissure ground truth.

A drawback of our method, and of deep learning in general, is the requirement of a large training dataset with ground truth segmentations. Manual segmentation is tedious, time-consuming, and typically performed by a medical imaging expert

analyst. Additionally, a high-end GPU card was required for training the network and such a card may not be available on a standard workstation. However, once the network is trained, it can be deployed on a low-end consumer GPU. Although there is a large overhead in training time (48 hours), processing time is only 20 seconds per image.

Tajbakhsh et al. [138] analyzed how well networks trained on natural images transferred to medical images and found pretraining resulted in improved or equal performance compared to random initialization. No transfer learning was used in this study due to limited availability of pretrained weights for 3D architectures. This is an area for potential further development.

The proposed method is designed exclusively for fissure detection and does not provide a complete lobar segmentation. However, the high specificity of our method facilitates lobar segmentation with simple post-processing (i.e. thresholding, morphological operations, and connected component analysis). For challenging cases with incomplete fissures, a more sophisticated surface-fitting technique might be used for post-processing. For example, optimal surface finding graph search could be used to divide the lung into lobes, defining the graph costs by fissure probabilities.

6.6 Conclusion

We have proposed a method for automatic detection of pulmonary fissures in CT images using a deep learning framework. We presented a novel coarse-to-fine cascade of ConvNets called FissureNet, and a novel 3D segmentation architecture called Seg3DNet. Fissure detection was evaluated with two rule-based methods

(Hessian and DoS) and two learning-based methods (FissureNet and U-Net). The learning-based methods outperformed the rule-based methods. Furthermore, FissureNet outperformed U-Net as it was capable of learning larger-scale global features. FissureNet achieves high sensitivity for fissure detection while producing very few false positives, allowing for straightforward post-processing to obtain a final lobar segmentation. The results show that FissureNet is robust to different CT scanners, scanning protocols (low-dose and normal-dose), inspiration levels (TLC and FRC), imaging modalities (breath-hold vs. 4DCT), and severities of pulmonary disease.

CHAPTER 7

PULMONARY LOBE SEGMENTATION

Submitted as: SE Gerard and JM Reinhardt: Pulmonary Lobe Segmentation Using a Sequence of Convolutional Neural Networks for Marginal Learning. International Symposium on Biomedical Imaging, 2019.

Segmentation of the pulmonary lobes in computed tomography images is an important precursor for characterizing and quantifying disease patterns, regional functional analysis, and determining treatment interventions. With the increasing resolution and quantity of scans produced in the clinic automatic and reliable lobar segmentation methods are essential for efficient workflows. In this work, a deep learning framework is proposed that utilizes convolutional neural networks for segmentation of fissures and lobes in computed tomography images. A novel pipeline is proposed that consists of a series of 3D convolutional neural networks to marginally learn the lobe segmentation. The method was evaluated extensively on a dataset of 1076 CT images from the COPDGene clinical trial, consisting of scans acquired multiple institutions using various scanners. Overall the method achieved median Dice coefficient of 0.993 and a median average symmetric surface distance of 0.138 mm across all lobes. The results show the method is robust to different inspiration levels, pathologies, and image quality.

7.1 Introduction

The human lungs are anatomically divided into five compartments called lobes. Infolding of the outer pleural membrane forms the boundary between adjacent lobes,

called fissures. The left lung has two lobes separated by the left oblique fissure. The right lung has three lobes separated by the right oblique fissure and the right horizontal fissure. In cross sections of computed tomography (CT) images the fissures appear as thin curvilinear surfaces that are slightly brighter in intensity than the surrounding lung parenchyma. The lobes are supplied by separate airways and blood vessels and are therefore generally functionally independent. Furthermore, diseases are often constrained by the fissure boundary and do not spread across lobes. Identification of the lobar regions on CT images is an important precursor that allows for quantitative characterization of disease and regional functional analysis.

The increasing resolution and quantity of scans have placed a demand on developing automatic and robust lobar segmentation methods. Assuming a lung segmentation is available, accurate identification of the fissure could presumably provide a lobar segmentation. However, it is generally not this straightforward as fissures are often radiographically incomplete or missing and do not completely divide the lung into lobes. Additionally, pathology such as fibrosis or emphysema can locally resemble fissures or obscure the shape and appearance of the fissure. Current fissure detector methods are sensitive to these factors and do not produce high quality fissure segmentations to allow straightforward division of lungs to lobes.

To overcome these challenges, many works have proposed sophisticated pipelines consisting of a sequence of modules to achieve lobe segmentation: lung segmentation, fissure detection, removal of falsely detected fissures, and surface interpolation and/or extrapolation [161, 146, 74, 163, 141, 75, 114]. Many of these methods are dependent

on local image appearance filters which can be sensitive to weak fissures and prone to false positives due to the limited locality of the filters. To mitigate this, several works incorporate anatomical priors in the form of blood and airway vasculature maps [141, 146, 75] or atlas-based initialization [161, 146]. Doel et al. [30] present an extensive review of pulmonary lobar segmentation methods.

Deep learning using convolutional neural networks (ConvNets) have recently been successful across many applications in computer vision, including medical imaging. However, few works have attempted to address pulmonary lobe segmentation. A progressive holistically nested neural networks followed by a random walker post-processing step for lobe segmentation was proposed in [46]. A limitation of this method is that training was performed on 2D slices, removing global contextual cues and 3D smoothness. In [150], lobe segmentation was directly learned on downsampled CT images using a 3D DenseNet.

For accurate lobe segmentation, it is important to learn both global and local features. Local features at high-resolution are important for detecting the exact fissure location which forms the lobar boundary. Global features are important for learning the the lungs are divided into five connected components and learning the relative position of lung tissue with respect to the fissure. Global features are especially important when the fissure is radiographically weak or incomplete, as information from the vasculature, airways, and lobe shape can provide can guide the segmentation.

In this work, a deep learning approach for lobe segmentation is proposed which utilizes a sequence of 3D ConvNet models to allow both local and global features to

be learned. The lobe segmentation is incrementally learned by training the networks to learn the following sequence of segmentation tasks: fissure region of interest (ROI), fissure, low-resolution lobe, and finally high-resolution lobe. Knowledge is propagated between networks in a feed-forward manner by using the output of a network as an input to the next network(s) in the series.

7.2 Methods

Given a CT image I and a lung segmentation M, the proposed method assigns each voxel a label from the discrete class set Y: the left lower lobe (LLL), left upper lobe (LUL), right lower lobe (RLL), right middle lobe (RML), right upper lobe (RUL), or non-lung (BKG). The learned model pipeline is illustrated in Figure 7.1. Separate models are trained for the left lung and the right lung and the predictions are merged in a post-processing step.

7.2.1 Image Datasets

The dataset of CT images used in this study consists of scans acquired from the COPDGene study, a large-scale multi-center clinical trial studying genetic and imaging biomarkers of chronic obstructive pulmonary disease (COPD) [115]. For each subject a pair of 3D breath-hold CT scans are acquired, one scan at total lung capacity (TLC) and one scan at functional residual volume (FRC). The TLC scans were acquired with a dose of 200 mAs and the FRC scans were acquired with a lower dose of 50 mAs.

A subset of 1601 subjects was used for training the fissure and lobe segmentation models, and a disjoint subset of 576 subjects was used for evaluating the model.

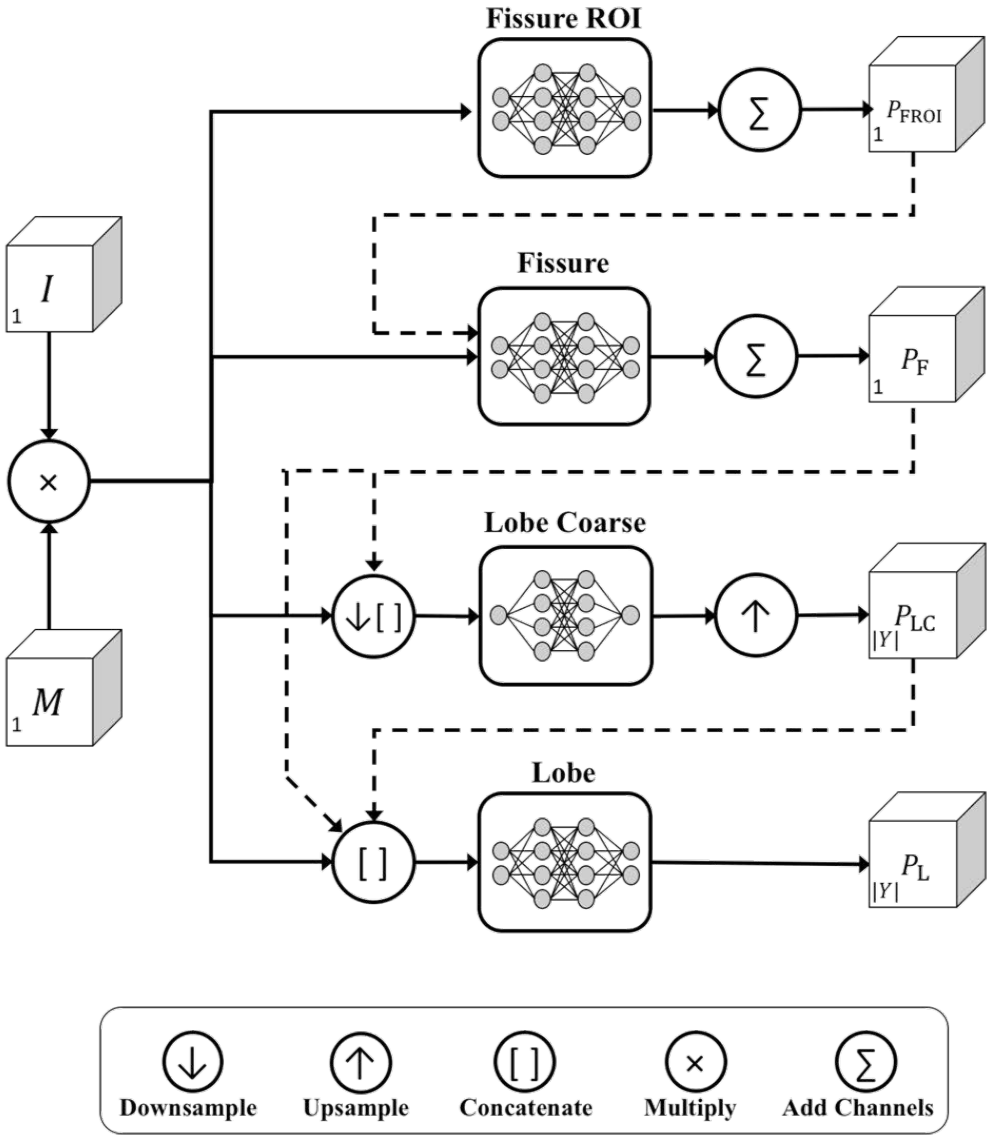


Figure 7.1. Lobe segmentation pipeline.

For each subject, both TLC and FRC scans were used for training and evaluation. Ground truth lung and lobe segmentations were obtained using a commercial software package (Apollo, VIDA Diagnostics, Coralville, IA). The automatic segmentations were manually inspected and edited if necessary. Ground truth fissure segmentations were automatically extracted from the lobar segmentations by identifying adjacent voxels with different lobe labels.

7.2.2 Preprocessing

The proposed method assumes each CT image has a corresponding lung mask which labels every voxel as left lung, right lung, or background. For consistency, all CT images and lung masks are resampled to have isotropic voxels with size 1 mm^3. To reduce the image size, crops with dimensions $256 \times 256 \times 256$ voxels are extracted at two locations, one centered around the left lung and one centered around the right lung. The crops extracted from the CT image and lung mask are be denoted I and M, respectively. The lung mask is used to mask out the background of the CT image, i.e., for every voxel that is labeled background in M the corresponding voxel in I is set to -1024. The masked image, $I \times M$, serves as the input to all networks presented in the following sections.

7.2.3 Seg3DNet

The lobe segmentation pipeline consists of a sequence of four ConvNets; each of these uses the Seg3DNet architecture [49]. Seg3DNet is a 3D, fully-convolutional network (FCN) [125] with an encoder and decoder module; see Figure 7.2. The encoder is similar to a 3D extension of U-net [119], however, the decoder module is

modified to consume less GPU memory. The input and output images of Seg3DNet both have three spatial dimensions with the same size. Additionally, the input and output images have a channel dimension which can be different sizes; the number of input channels C corresponds to the number of input feature maps and the number of output channels $|Y|$ corresponds to the number of classes.

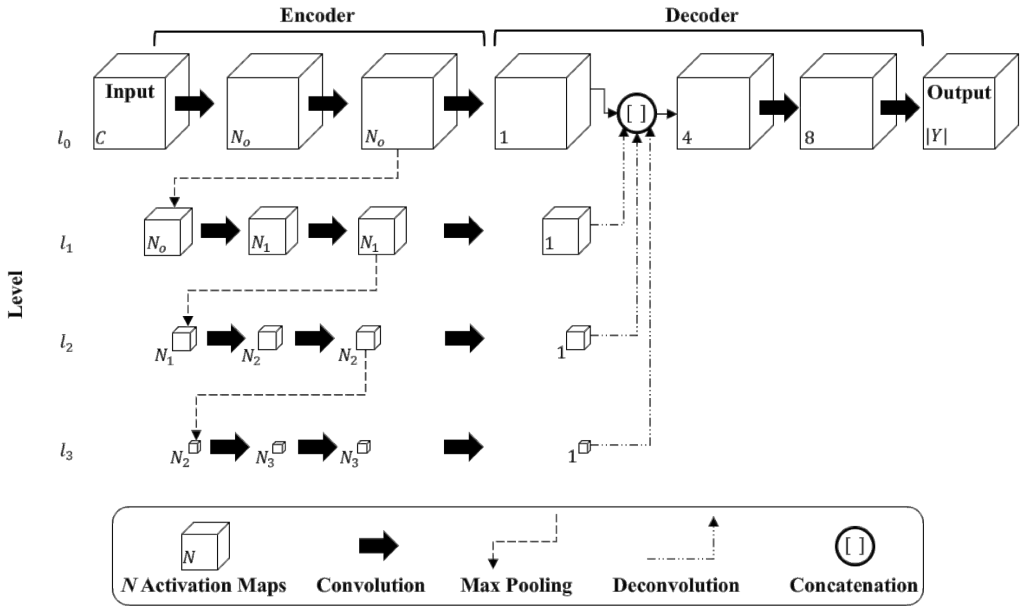


Figure 7.2. Seg3DNet architecture. The number of channels for each image representation is denoted in the lower left corner of each cube. For the encoder module, the number of channels for all image representations in level l_i is $N_i = 2^{i+5}$ for $i = 0, .., L$. The relative spatial size of the image representations are drawn to scale. At each level the spatial dimensions of the image representation gets downsampled by a factor of two.

7.2.4 FissureNet

The first network and the second network predict a fissure ROI P_{FROI} and a precise fissure P_F, respectively. This is the same as the FissureNet method, see [49] for detailed description. Briefly, the first network predicts a fissure ROI, where all voxels within 5 mm of the fissure are defined as the fissure ROI. The second network predicts the precise fissure location, with the prediction weighted by the fissure ROI prediction. This alleviates the class imbalance problem between fissure and non-fissure voxels. FissureNet predicts a vector-valued probability image where each voxel represents a probability distribution over the different major fissures (left oblique, right oblique, and right horizontal) and non-fissure. The probability of each fissure is accumulated to produce a single channel fissureness P_F image that is utilized in the lobe networks.

7.2.5 Coarse Lobe Network

The coarse lobe network learns a low-resolution lobe segmentation. This network is trained with aggressively downsampled images, allowing the network to be trained with images of the entire lung field. This makes it possible to learn global features and contextual information. The masked CT image $I \times M$ and the fissureness image P_F, and the target lobe segmentation L are downsampled to $64 \times 64 \times 64$, a downsampling factor of four along each dimension. Linear interpolation is used for downsampling $I \times M$ and P_F and nearest neighbor interpolation is used for downsampling L. Prior to downsampling, Gaussian smoothing is performed. The input to the network is the downsampled $I \times M$ and P_F images concatenated along the channel dimension. The output of the network is a vector-valued image where

each voxel represents the probability distribution over each of the lobes and the background. This output is upsampled to the $256 \times 256 \times 256$ voxel resolution to produce P_{LC}, which is input to the final lobe network.

7.2.6 Lobe Network

The final network in the pipeline learns a high resolution lobe segmentation. The input to this network is the 1 mm resolution $I \times M$, P_F, and P_{LC} images concatenated along the channel dimension. At this resolution, the network cannot be trained on the entire $256 \times 256 \times 256$ image due to GPU memory constraints. Sagittal slabs with size $64 \times 256 \times 256$ voxels are extracted at random locations along the left-right axis, and these crops are used for training. The output of the refine lobe network, P_L, is a vector-valued probability image, i.e., each voxel is a 6-element vector representing the probability of the voxel belonging to each of the lobes and background.

7.2.7 Post-processing

A simple post-processing is used to obtain a final discrete segmentation from P_L. First, all voxels with a high confidence prediction for one lobe are assigned to that lobe, using a high confidence prediction threshold of $p_i > 0.70$. A distance transform is applied to each of the resulting high confidence lobe regions and the remaining low confidence voxels are assigned to the lobe to which they are nearest.

7.2.8 Implementation

The lobe segmentation pipeline is implemented with open source frameworks Theano [140] and Lasagne [28]. The networks are trained using NVIDIA GPU cards:

P100 with 16 GB RAM for low resolution networks and P40 with 24 GB RAM for the high-resolution networks. Adam optimization [67] is used for training with an initial learning rate of 5^{-4} and Xavier normal initialization [50] is used.

7.3 Results

The performance of the proposed method was evaluated on 1076 scans of subjects with COPD. Figure 7.3 qualitatively compares the automated lobe segmentation results to the manual segmentation. Quantitative comparison was performed using the average symmetric surface distance and the Dice coefficient to measure volume overlap distance between surfaces, respectively. Overall, the proposed method achieved a median Dice coefficient of 0.993 and a median ASSD of 0.138 mm. Table 7.1 shows the a summary of the medians and Figure 7.5 shows the quartiles of the ASSD and DICE distributions stratified by lobe and lung volume, Figure 7.4 displays histograms of ASSD and DICE for all evaluation cases. The results show TLC performance was consistently better than FRC performance. Furthermore, the right lobes performed slightly worse compared to the left lobes. Figure 7.6 shows the method was robust to different levels of disease severity as assessed by GOLD level.

7.4 Discussion

The proposed method was evaluated on scans acquired at TLC and FRC. Performance on FRC images was slightly worse compared to TLC images, however, the performance on FRC images was still very high. In general, it is more difficult to segment lobes in scans acquired at lower lung volumes, such as FRC and residual volume (RV), compare to TLC. At lower lung volumes there is less air in the lungs,

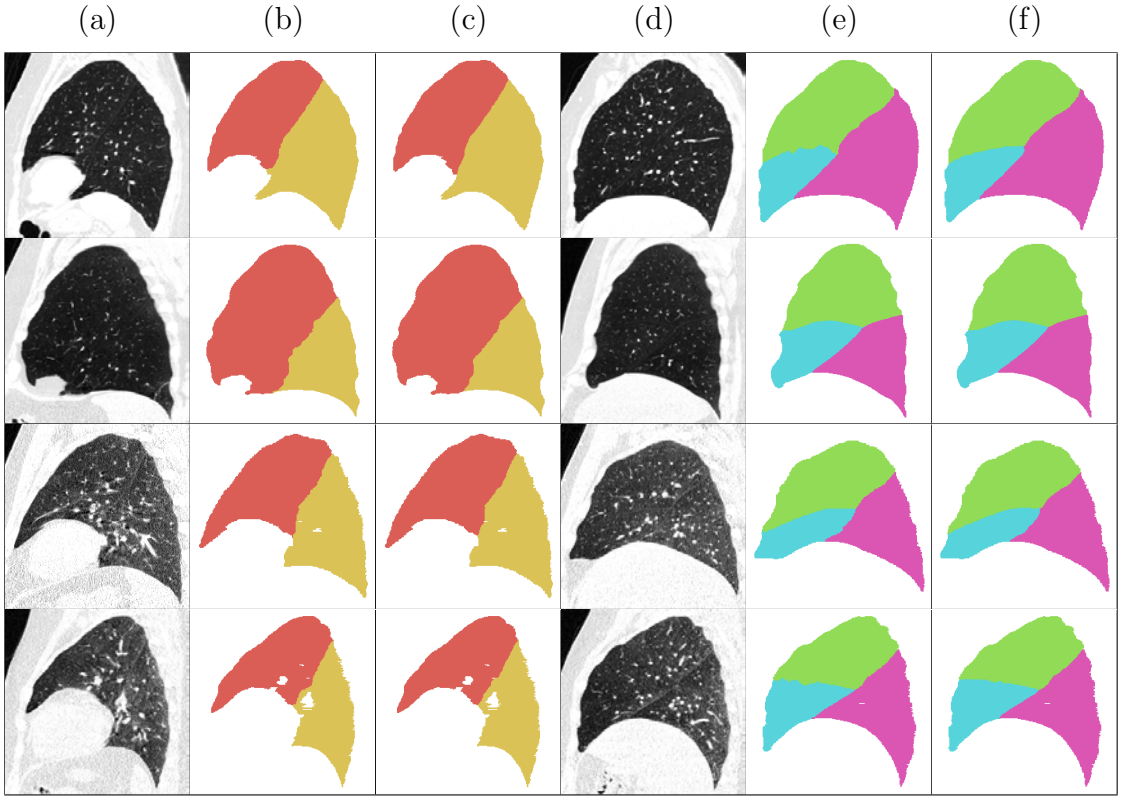


Figure 7.3. Representative results for four subjects in rows one to four, with TLC scans in rows 1-2 and FRC scans in rows 3-4. Left lungs in columns (a)-(c) and right lungs in columns (d)-(f). (a) and (d) CT slice, (b) and (e) ground truth lobe segmentation, (c) and (f) predicted lobe segmentaiton.

Table 7.1. Median values across all evaluation scans, stratified by lung volume and lobe. ASSD units are in mm

Volume	Lobe	Dice	ASSD
TLC	LLL	0.997	0.074
	LUL	0.997	0.071
	RLL	0.996	0.093
	RML	0.983	0.302
	RUL	0.993	0.185
FRC	LLL	0.994	0.100
	LUL	0.995	0.089
	RLL	0.993	0.112
	RML	0.975	0.371
	RUL	0.989	0.227

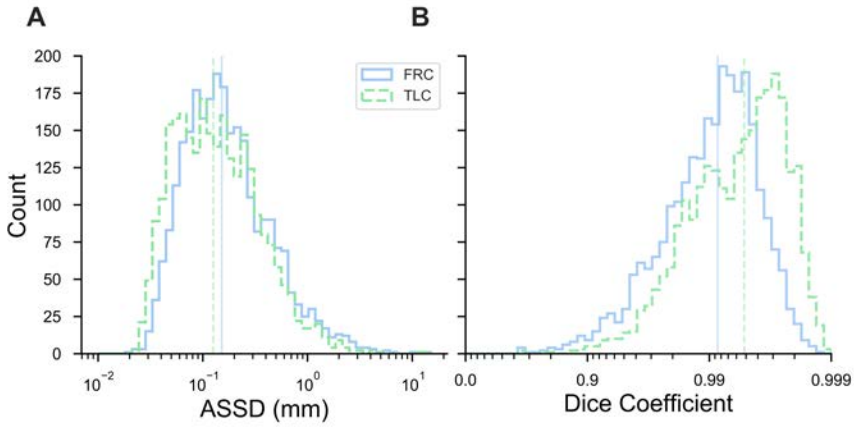


Figure 7.4. Histogram of ASSD (A) and Dice (B) results stratified by TLC scans (green) and FRC scans (blue). Vertical lines represent medians of respective distributions.

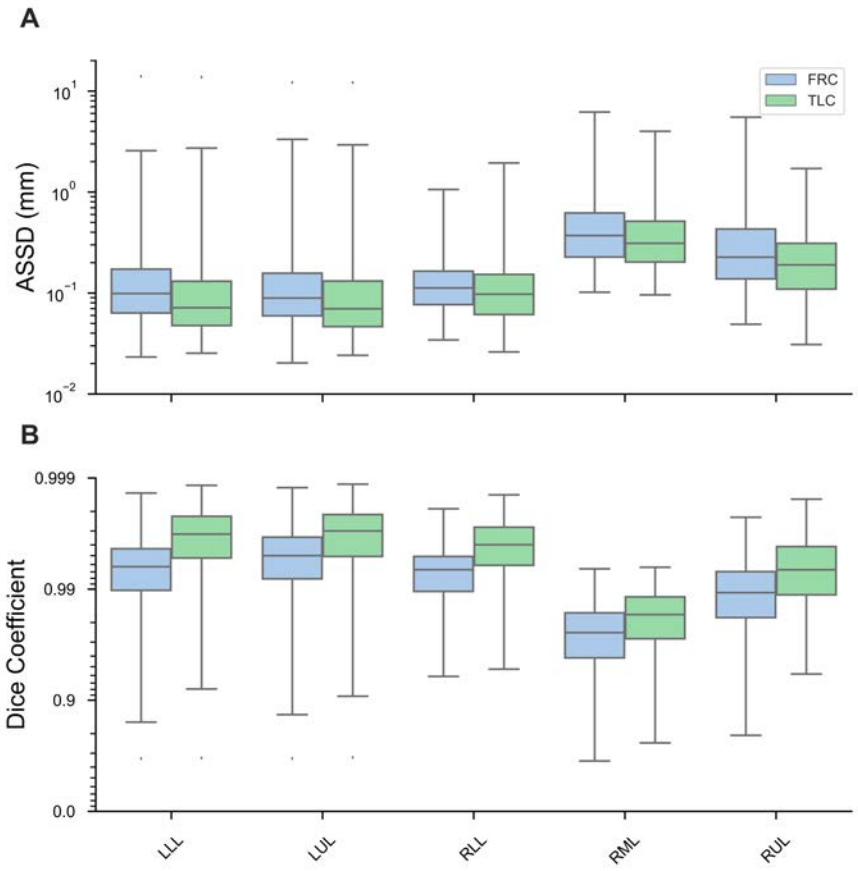


Figure 7.5. ASSD (A) and Dice (B) results stratified by lobe and lung volume: TLC (green) vs FRC (blue).

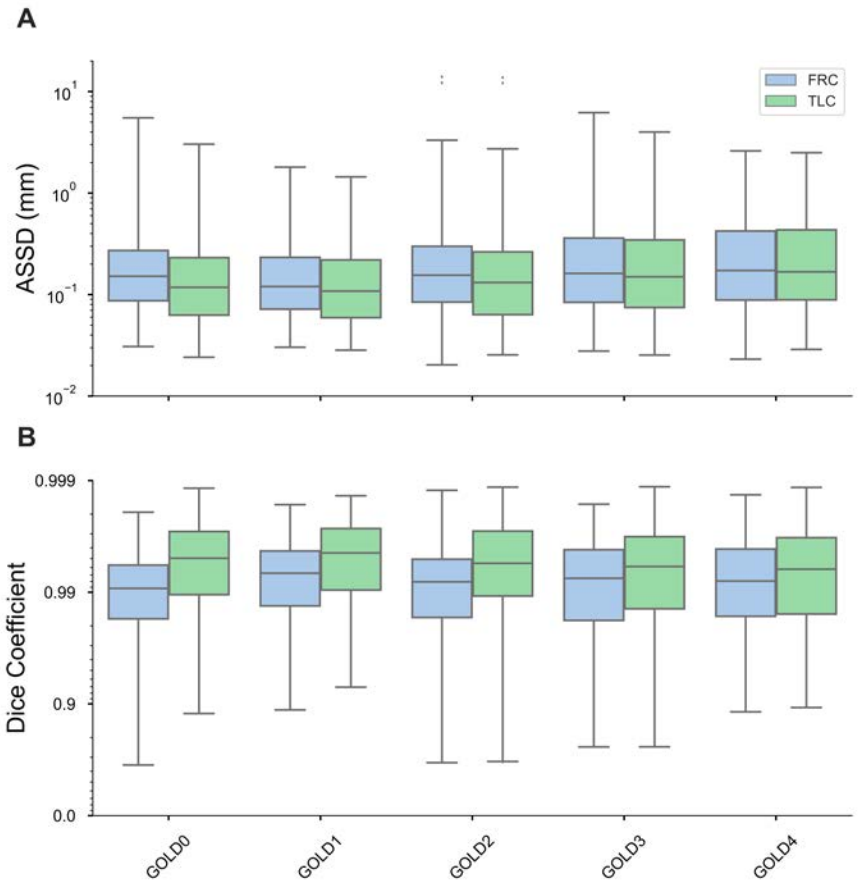


Figure 7.6. ASSD (A) and Dice (B) results stratified by GOLD level and lung volume: TLC (green) vs FRC (blue).

resulting in lower contrast between the fissure and lung parenchyma. The FRC scans in this study were acquired at a lower dose compared to the TLC scans, resulting in poorer image quality for the FRC scans. This is another contributing factor for the drop in performance. However, the absolute performance metrics were still good, demonstrating the proposed method is robust challenges imposed by low imaging dose and small lung volumes.

Segmentation performance on the left lobes was consistently better than the right lobes, with RML and RUL having the worst performance. This is consistent with previous works, and can be explained by poor visibility of the right horizontal fissure, which separates the parenchyma above the RLL into RML and RUL. Nonetheless, the proposed method was robust to poor visibility of the horizontal fissure and was able to reliably segment the RML and RUL.

The evaluation dataset consisted of subjects from all GOLD stages. More diseased subjects are generally more difficult to segment, pulmonary disease can locally resemble the fissure or distort the true fissure appearance and shape. Although there was a small decrease in performance with increasing GOLD level, the proposed method still showed high performance for all GOLD stages.

Automated runtime for the proposed method was less than three minutes when run on a NVIDIA P100 GPU card, this includes the inference times for eight networks. Overhead of training all networks was approximately 96 hours.

7.5 Conclusion

A novel pipeline was proposed which uses a series of interconnected ConvNets to incrementally learn lobe segmentation in CT images. The method achieved high performance on a dataset of 1076 CT scans, one of the most extensive evaluations that has been performed for lobar segmentation. The method was robust to different inspiration levels, poor image quality, and disease severities. This allows for efficient and reliable regional analysis on large-scale datasets both in the clinic and in research settings.

CHAPTER 8

DISCUSSION AND CONCLUSION

Deep learning using convolutional neural networks (ConvNets) are successful at performing many tasks in computer vision, including image classification, segmentation, object recognition, and captioning. Since the breakthrough of AlexNet [73] in the 2012 ImageNet competition [27], there has been an explosion of research related to deep learning and ConvNets. The majority of this research is applied to natural 2-dimensional (2D) images due to the availability of large public datasets such as ImageNet. In recent years, the medical imaging community has taken notice and applied this powerful technique to medical image analysis [53, 82, 126, 68].

In this work, a pipeline for pulmonary segmentation was proposed that uses a series of inter-connected 3D ConvNets. The pipeline sequentially performs segmentation of lungs, fissures, and lobes in computed tomography (CT) images. A multi-resolution ConvNet model was proposed, which is capable of learning from large volumetric images without loss of global context and spatial smoothness. Similar multi-resolution regimes have shown to be beneficial in image processing applications including statistical shape models [20] and image registration [18, 124]. In this work, the multi-resolution model consisted of two resolution levels, however, this could be extended to more levels. First, a low-resolution model is learned by training a ConvNet on aggressively downsampled images. At low resolutions the network can be trained on entire images; this gives the network the capacity to learn global patterns. A high-resolution

model is subsequently learned which uses both the high-resolution CT images and the low-resolution model prediction. This allows for precise edge information to be learned and global information to be incorporated.

The proposed pipeline was evaluated extensively to ensure robustness. Multiple datasets were used for evaluation, these datasets incorporate a variety of imaging protocols, scanners, lung volumes, imaging modalities, pathologies, and species. The datasets included scans of human subjects with chronic obstructive pulmonary disorder (COPD), idiopathic pulmonary fibrosis (IPF), and lung cancer. In addition to the human scans, an animal dataset was utilized which consisted of three animal species with models of acute respiratory disorder (ARDS). The human scans included multiple lung volumes: total lung capacity (TLC), functional residual volume (FRC), residual volume (RV), and intermediate volumes from 4DCT scans. Different imaging doses were used resulting in different image qualities. The modules in the pipeline proved to be robust to these challenges, demonstrating its utility in high-throughput processing.

In medical imaging applications, the majority of methods train ConvNets using a subset of the image, e.g., 2D slices or small 3D patches cropped from the image. While these methods show promising results, training on small subsets of the image greatly reduces the extent of global context that can be learned by the network. For segmentation of pulmonary structures, global context from the surrounding anatomy is important for discrimination of different structures. For example. in injured lungs the local intensity and texture features can be indistinguishable from the surround soft tissue. For accurate inclusion of injury in lung segmentation, the rib cage provides

perhaps the most valuable information for approximating the location of the lung boundary. The multi-resolution model proposed in this work, allows for learning global features without sacrificing high-resolution information.

The proposed pipeline was developed on 3D images; for processing 4DCT images each 3D phases is independently processed. Ideally, the temporal patterns in the datasets would be exploited and temporal consistency would be ensured through 4D processing. In future work, recurrent neural networks (RNNs) could be explored; RNNs are capable of learning temporal patterns in sequential data. This would require training data with all phases segmented which is currently not available for these datasets. Alternatively, a hybrid algorithm using ConvNets with optimal surface finding [79] graph optimization would allow temporal constraints to be incorporated. This would be less computationally efficient compared to a purely ConvNet-based pipeline. Furthermore, defining temporal constraints for the graph optimization is challenging. There are large variations in lung motion regionally, between subjects, and even within the same subject. While 4D processing is ideal for 4DCT images, applying the proposed pipeline to each phase independently produced visually temporal consistent results and it may be possible that 4D processing will add complexity without significant improvements.

Data augmentation is widely used to artificially generate more training data. This technique involves applying simple affine transformations to images, e.g., rotations, translations, and scaling, as well as intensity transformations. This type of data augmentation is appropriate for natural images, however, it is not realistic for some

types of medical images. CT images have a fixed orientation and standardized intensity units. Applying such transformations introduces variations in the data that do not naturally exist. The natural variations in anatomical structures are informative features that can be learned by the network. This is especially true for global feature learning or the low-resolution model. For this reason, we did not use data augmentation in this work.

A drawback of deep learning is the dependency on a large dataset of annotated images. The performance of deep learning based methods lies not only in the architecture and algorithm developed, but also in the diversity of the dataset. Therefore, fair comparison of novel algorithms should be done on a common dataset to eliminate the contribution of the data itself. While this is done in the computer vision field on the ImageNet dataset, much of the development and evaluation in medical imaging domains uses independent datasets.

A common question when apply deep learning to a new application is how much training data is needed. Unfortunately, there is not a universal answer to this question; the amount of data required is dependent on the specific problem that is being explored. Some problems are more complex and require larger models and thus more training data. Another factor to consider is the variability in the training data. If more variations exist, more training data is required to represent all the variations in the learned model. Several works have explored methods for estimating sample size for required for classification performance [39, 90]. This technique involves creating a learning curve using a small dataset and fitting an inverse power law. While there

is no universal answer to the amount of training data required for deep learning, a commonality for all applications of deep learning has been the more training data available, the better the performance.

Another drawback of deep learning is the computational overhead in training. For large 3D datasets, training a single model can take several days. Proper hyperparameter optimization involves retraining the same model many times until the best combination of hyperparamters is found. Furthermore, a high-end GPU card with sufficient RAM is required for training. While training ConvNets is computationally expensive and time-consuming, after training is completed these models are very computationally efficient and can even be executed on a CPU.

The "black-box" nature of deep learning deters its use in medical applications. Rule-based methods or feature engineering may be favored since the model can be explained and if something fails it can be understood. For medical image segmentation, the results can be reviewed and one does not need to understand what the model learned. Medical image segmentation is a time-consuming and tedious task, but can be done by human analyst, making it an ideal application for deep learning.

The scope of applications that could benefit from multi-resolution models exceeds image segmentation, which is just an initial step for image analysis. For example, detecting image biomarkers for various pulmonary diseases is a large area of research. Multi-scale processing for learning biomarkers provides more context for learning compared to 2D slices. Furthermore, extracting a subset of 2D slices could omit the pathological region altogether. Another application of multi-resolution

ConvNets is image registration. Conventional image registration algorithms use multi-resolution pyramids, however, this has not been explored in deep learning based image registration methods.

In this dissertation, a pipeline was developed for segmentation of pulmonary structures in CT images. The pipeline consists of modules for segmentation of lungs, fissures, and lobes. Multi-scale ConvNets were proposed to enable learning features of multiple scale range in large volumetric images. Multi-scale processing is similar to the way human image analyst read medical images; first the entire lung is viewed and subsequently abnormal regions are viewed more closely. The proposed pipeline was extensively evaluated on multiple datasets to ensure robustness. The high performance and computational efficiency of the pipeline is a promising step towards high-throughput medical image analysis in clinical and research settings. CT images provide a rich source of information regarding pulmonary diseases. Unfortunately, much of this information not fully utilized due to the number of slices in volumetric images and the high volume of scans produced. Multi-scale ConvNets have great potential of providing an extra pair of eyes to extract and understand information in medical images.

Appendices

APPENDIX A

4D LUNG SEGMENTATION

Published as: SG Yeary, GE Christensen, JE Bayouth, S Bodduluri, Y Pan, J Guo, K Du, JH Song, B Zhao, I Oguz, JM Reinhardt: 4D Lung CT Segmentation for Radiation Therapy Applications. ICART: Imaging and Computer Assistance in Radiation Therapy, 2015.

Radiation therapy protocols for lung cancer treatment planning commonly use a 3D CT image to delineate critical regions of interest and a 4D respiratory-gated CT to track tumor respiratory motion. 4D CT data sets contain a great deal of information and can be analyzed to obtain quantitative data on respiratory system dynamics, mechanics, and function. We describe a method that transfers segmentation contours from the 3D planning CT to the entire 4D data set, even though the lung changes size and shape during image acquisition due to respiration. Our proposed method uses deformable image registration to align all phases of the 4D CT into the coordinate system of the 3D planning CT. A graph optimization algorithm is used to segment the lung boundaries of the aligned phases of the 4D data set. The algorithm is initialized with the 3D segmentation of the planning CT. The segmentations of the individual phases in the planning CT coordinate system are then transformed back into each of the original phase coordinate systems providing 3D segmented volumes at each phase of the 4D CT. We tested the method on six data sets from subjects about to undergo radiation therapy for lung cancer. For each subject a 3D planning CT was segmented as part of the treatment planning process. This planning CT was

used to segment the 4D CT at ten phase points across the respiratory cycle. The results of our method were compared against manual segmentations of four phases giving an average surface distance of 0.417 mm and average Dice coefficient of 0.983. The 4D segmentations appear to be more consistent across phases than the manual segmentations, especially near the mediastinum. The results show that this method can provide accurate 4D segmentations from a single segmented 3D CT image.

A.1 Introduction

Respiratory-gated 4D CT imaging is used in radiation therapy planning for lung cancer treatment to measure tumor position at different points in the respiratory cycle. These data sets can be further analyzed to obtain quantitative measurements of lung dynamics [155], lung mechanics [3], and regional lung ventilation [31]. Accurate lung segmentation of the constituent 3D volumes of a 4D data set is required before quantitative analysis can be performed.

An initial 3D CT scan (referred to here as the "planning" CT) is conducted on all patients that undergo a 4D CT scan. The planning scan is performed in breath-hold and typically provides higher signal-to-noise and superior spatial resolution than a 4D CT acquisition. As a result, the planning CT image provides a high quality image of the lung anatomy and — due to acquisition during a single breath hold — has fewer motion artifacts than a 4D CT data set. The planning CT is commonly used for radiation therapy planning and includes lung and tumor segmentations generated by an image analyst.

Several different approaches for lung tissue segmentation in 3D CT images have

been proposed, including optimal thresholding [60], adaptive border tracking [111], graph search algorithms [61], and segmentation-by-registration with the use of a lung atlas [145]. These algorithms vary in segmentation accuracy, computational complexity, and robustness to abnormalities and disease, but it is unlikely any 3D algorithm could accurately and consistently segment all volumes of a 4D data set independently. It seems likely that a direct 4D segmentation may give better results than 3D volume-by-volume segmentation as all of the available image information is used and temporal coherence across respiratory phases can be ensured. Additionally, since 4D CT data sets typically have lower image quality and more motion artifacts than the typical planning CT, conventional segmentation algorithms may fail or perform poorly when applied to the individual 3D volumes of the 4D CT.

Lung segmentation in a 4D CT data set can be viewed as a multiple surface segmentation problem, where prior known interactions between surfaces can be used to improve segmentation robustness. Optimal surface finding is a graph search method that has been used in many applications to segment multiple surfaces in n-D medical images [79]. In [97], optimal surface finding is used to segment multiple objects and surfaces in the brain. In [153] and [137], an active shape model is combined with optimal surface finding to simultaneously segment multiple lung volumes. In this work, we propose a similar method for lung segmentation in 4D CT data sets consisting of ten volumes imaged at different phases during the breathing cycle.

Our 4D segmentation algorithm is based on a combination of deformable image registration and 4D optimal surface finding. First, image registration is used to align

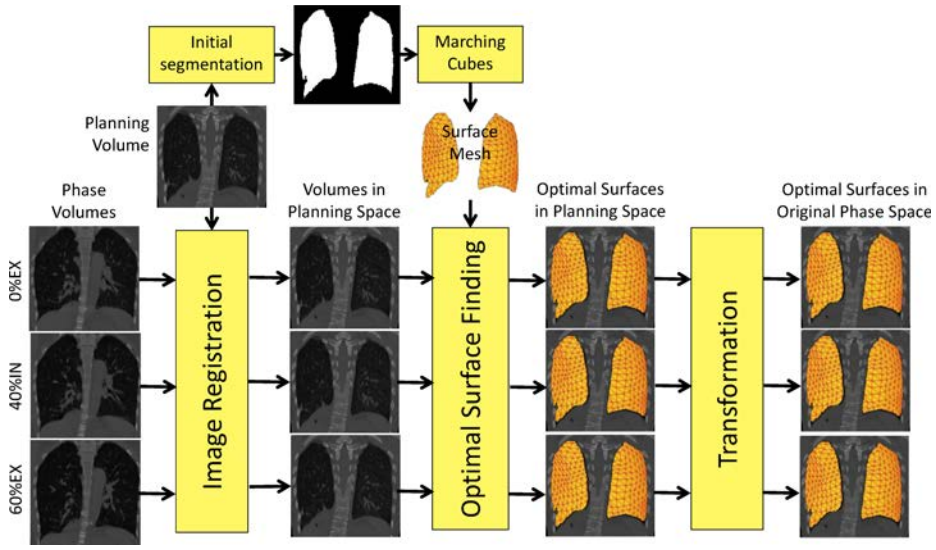


Figure A.1. Processing flowchart. The process starts with the planning CT segmentation and produces N segmented 3D phase volumes.

all volumes of the 4D data set to the coordinates of the planning CT so that the planning CT segmentation can be used as a shape prior for all 4D CT lung volumes. A 4D geometric graph is constructed using shape priors and the optimal surface is found using an efficient maximum flow algorithm.

A.2 Methods

Our proposed 4D segmentation method consists of two stages: image registration and optimal surface finding. The processing pipeline is shown in Figure A.1. We assume that the 4D CT consists of N individual 3D phase volumes sampled across the respiratory cycle. We use the notation 20%IN to refer to the 3D image reconstructed at 20% of tidal lung volume during inspiration, and 80%EX to refer to the 3D image reconstructed at 80% tidal lung volume during expiration.

A.2.1 Registration

In the first stage, each 3D volume of the 4D data set is roughly aligned to the planning CT volume using deformable image registration. Our method does not require an accurate registration, so a low resolution registration framework can be used to reduce computation time. In this work, we used the Elastix image registration software [69] to perform the alignment of the planning CT to each of the 3D volumes. A b-spline transform was used with a grid spacing of 20 mm. Normalized correlation was used for the similarity metric and a gradient descent optimization algorithm was used with a maximum of 1000 iterations. Note that it is possible to perform all registrations simultaneously since each planning CT to 3D phase volume registration is performed independently. Thus, performing the registrations in parallel can reduce the overall computation time.

After performing the image registration, we obtain a set of deformed 3D images in the same coordinate space as the planning CT image, allowing the existing planning CT segmentation to be used as a subject specific shape prior for the next stage.

A.2.2 Optimal Surface Finding

Optimal surface finding is a graph search framework used for simultaneous segmentation of multiple interacting surfaces in images [97]. The lungs at each phase image are treated as surfaces and the interactions between different phases represent temporal constraints. A shape prior similar to the true surface is used to initialize the graph search space. Graph $G(V, E)$ consisting of node set V and edge set E is constructed in a region around the shape prior. In our work, the planning CT

segmentation is utilized as the shape prior for all phases. Since the same shape prior was used for all phases, all sub-graphs G^t had the same geometric structure in space. This greatly simplified the 4D graph construction.

For multiple surface segmentation the following graph construction is done for each surface, using similar graph construction parameters to those used in [137]. First, the marching cubes algorithm is used to transform the planning CT segmentation into a mesh representation with approximately 10,000 vertices. A search profile, or graph column V_i, is constructed at each mesh vertex and nodes $v_{i,j}$ are sampled at equal distances along the profile. For the experiments presented in this paper, the graph column length was 60 nodes sampled at a distance of 0.35 mm. The search profiles were defined using the electric field lines generated by treating each mesh vertex as a charged point source as described in [160].

Graph edges are introduced to define feasible surfaces as those which maintain the topology of the shape prior and have a certain degree of surface smoothness. This is achieved by introducing intracolumn edges that force the optimal surface to intersect each graph column V_i at exactly one node $v_{i,j}$ and intercolumn edges that restrict the surface height from changing more than Δ nodes between neighboring columns. The intracolumn edges $< v_{i,j}, v_{i,j-1} >$ were introduced within each column of the graph and the intercolumn edges $< v_{i,j}, v_{k,j-\Delta} >$ were introduced between all neighboring columns V_i and V_k. For experiments used in this paper, we used $\Delta = 12$ nodes. Each node $v \epsilon V$ was assigned a cost inversely proportional to the likelihood the surface contains $v_{i,j}$. The gradient magnitude of the deformed phase images was

used as the cost function in this work. The cost of a feasible surface is the summation of all node costs on the surface. The optimal surface corresponds to the surface with the minimum cost among all feasible surfaces.

Interactions between surfaces were enforced by introducing temporal edges $< v_{i,j}^{t}, v_{i,j-\delta}^{t+1} >$ between corresponding columns of the surfaces. In our 4D lung segmentation, the deformed lung of each image phase $t = 1 \ldots N$ was treated as a surface and the temporal constraint δ enforced surface consistency between phases. Here we used $\delta = 20$ nodes.

A maximum flow algorithm was used to find the globally optimal solution of the cost function as described in [11]. This resulted in N surfaces in the deformed image space. The surfaces were transformed back to the coordinates of each original 3D phase volume using the transformations obtained from the registration.

A.2.3 Data Sets and Experimental Setup

4D CT data sets and a planning CT image from six lung cancer subjects about to undergo radiation therapy were used for this study. All data were gathered under a protocol approved by the University of Iowa Institutional Review Board (IRB 200905703). Each 4D CT contained ten 3D volumes retrospectively reconstructed in 20% steps from inspiration to exhalation. All images were resampled to obtain $1 \times 1 \times 1$ mm^3 voxels. Ground truth segmentations were generated for four respiratory phases (0%EX, 40%IN, 100%IN, 60%EX) by a trained image analyst (Y.P.) using the Analyze^{TM} software package. The same image analyst segmented the planning CT image for each subject to provide initial shape priors.

Table A.1. Dice coefficients for the 4D segmentation compared to the ground truth segmentations for each of the four 3D phase volumes used for evaluation. Dice coefficients are calculated separately for left and right lungs. Mean is calculated across four phases and both lungs for each subject.

Subject	0%EX		40%IN		100%IN		60%EX		Mean
ID	Left	Right	Left	Right	Left	Right	Left	Right	
A	0.985	0.986	0.988	0.988	0.985	0.986	0.985	0.986	0.986
B	0.976	0.983	0.976	0.986	0.973	0.984	0.974	0.984	0.980
C	0.978	0.978	0.977	0.985	0.984	0.982	0.984	0.977	0.981
D	0.988	0.988	0.981	0.982	0.983	0.978	0.982	0.979	0.983
E	0.981	0.983	0.982	0.981	0.982	0.983	0.983	0.981	0.982
F	0.985	0.983	0.984	0.984	0.984	0.985	0.985	0.985	0.985

To test the sensitivity of the method to the segmentation shape prior, we ran an additional experiment for one data set where different 3D phase volumes were used in place of the planning CT image. This was done by using the phase volumes at 20%IN, 60%IN, 40%EX, and 20%EX to establish the initial shape priors for the optimal surface finding.

A.3 Results

A.3.1 4D Segmentation with planning CT as Initial Segmentation

Tables A.1 and A.2 show the Dice coefficients and mean absolute surface distance for each of the four evaluation phases. Figure A.2 shows the automatic 4D segmentation, displayed at two of the individual phase reconstructions, for one subject.

Table A.2. Average symmetric absolute surface distance ± standard deviation (mm) for the 4D segmentation compared to the ground truth segmentations for each of the four 3D phase volumes used for evaluation. Mean is calculated across four phases and both lungs for each subject.

Subject ID	0%EX		40%IN	
	Left	Right	Left	Right
A	0.385 ± 0.807	0.437 ± 1.203	0.315 ± 0.860	0.353 ± 0.790
B	0.428 ± 0.890	0.422 ± 0.766	0.410 ± 0.715	0.394 ± 0.743
C	0.431 ± 0.755	0.536 ± 1.375	0.501 ± 1.126	0.320 ± 0.682
D	0.242 ± 0.573	0.297 ± 0.784	0.435 ± 0.959	0.491 ± 1.194
E	0.499 ± 1.315	0.536 ± 1.555	0.455 ± 1.086	0.498 ± 0.959
F	0.307 ± 0.566	0.443 ± 1.135	0.342 ± 0.655	0.376 ± 0.836

Subject ID	100%IN		60%EX		Mean
	Left	Right	Left	Right	
A	0.407 ± 0.841	0.459 ± 0.976	0.417 ± 0.854	0.429 ± 0.853	0.400
B	0.464 ± 0.950	0.434 ± 0.773	0.441 ± 0.778	0.407 ± 0.717	0.425
C	0.365 ± 0.896	0.423 ± 0.909	0.349 ± 0.839	0.520 ± 0.871	0.431
D	0.380 ± 0.694	0.542 ± 0.911	0.399 ± 0.795	0.547 ± 1.197	0.417
E	0.422 ± 0.747	0.468 ± 0.986	0.424 ± 1.009	0.495 ± 0.983	0.475
F	0.347 ± 0.662	0.338 ± 0.625	0.340 ± 0.716	0.353 ± 0.716	0.356

Figure A.2. Automatic 4D segmentation results for subject A. From left to right: 0%EX coronal plane, 0%EX transverse plane, 100%IN coronal plane, 100%IN transverse plane.

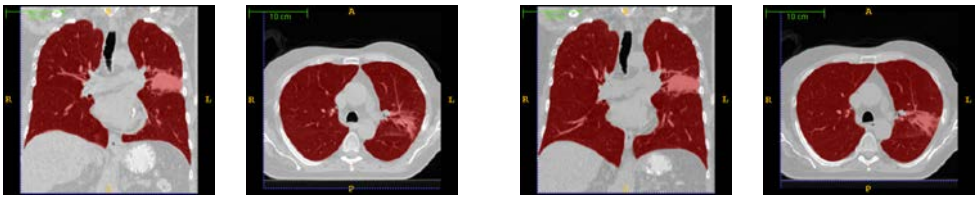


Table A.3. Dice coefficients for the 4D segmentation of subject A using four different initial segmentations as the shape prior. Mean is calculated across four phases and both lungs for each subject.

Initialization	0%EX		40%IN		100%IN		60%EX		Mean
Phase Volume	Left	Right	Left	Right	Left	Right	Left	Right	
20%IN	0.982	0.985	0.985	0.986	0.984	0.987	0.982	0.985	0.984
60%IN	0.981	0.984	0.985	0.987	0.983	0.987	0.981	0.986	0.984
40%EX	0.982	0.985	0.985	0.987	0.982	0.986	0.982	0.986	0.984
20%EX	0.981	0.985	0.985	0.986	0.983	0.987	0.981	0.985	0.984

A.3.2 Sensitivity to Initial Segmentation

The sensitivity to the choice of initial segmentation was tested by segmenting one 4D data set (subject A) with four different phase volumes used as the initial segmentation. Table A.3 gives the Dice coefficients for the four evaluation volumes with four different initializations.

A.4 Discussion

The results show that image registration and optimal surface finding can be used to produce an accurate 4D segmentation from one initial 3D segmentation. In our experiments an average symmetric absolute surface distance of 0.417 mm

and a mean Dice coefficient of 0.983 were obtained. This mean surface error is less than half of a voxel. We observe that most errors were where the airways enter the lungs at the mediastinum. This is a subjective area to manually segment, as different boundaries may be chosen by different analysts. For comparison, a 3D thresholding based algorithm similar to [60] was applied to each volume and an average symmetric absolute surface distance of 0.5644 mm and a mean Dice coefficient of 0.9790 were obtained. Although the simple thresholding gave good results, tumors were not included in the segmentation thus it is not suitable for this application. The same mean Dice coefficient of 0.984 was obtained for all initial segmentations showing the algorithm it robust to the initial segmentation. Since all phase volumes were segmented simultaneously by the optimal surface finding, our final segmentation contours are consistent across all phases. The segmentations appear to be more consistent across phases than the manual segmentations, especially near the mediastinum. Additionally, the method was able to include large chest wall tumors in the segmentation, which is a difficult problem due to the similarity in intensity with the surrounding tissue.

We were able to use a computationally inexpensive registration by using the optimal surface finding to refine the initial segmentation result. In many cases it is even possible to use an affine registration when the lung deformation across respiration is not too large. The optimal surface finding incorporates shape prior information, surface smoothness constraints, and temporal surface constraints making it ideal for 4D lung segmentation. The optimal surface finding guarantees a globally optimal

minimization of the selected cost function.

The experiments were run on a Linux machine with an Intel Xeon 2.27 GHz CPU and 48 GB of RAM. The deformable image registration takes 2 minutes per phase, which can be done in parallel. The optimal surface finding takes approximately 4.5 minutes with our parameters. This yields a total time of 6.5 minutes to produce segmentations of all 10 volumes of a 4D data set. The manual segmentation used for evaluation took approximately 30 minutes per 3D volume.

The proposed framework is extensible to segmentation of other surfaces. For example, sublobar segmentations can be used to study the mechanical properties and sliding of the lung lobes. Our framework can be easily extended to handle lobar segmentations.

A.5 Summary

We proposed a method for 4D segmentation of lung tissue in respiratory-gated data sets. The method utilizes the planning CT segmentation to obtain accurate segmentations for all phase volumes of a 4D data set. An average surface distance of 0.417 mm and an average Dice coefficient of 0.983 were achieved. Based on preliminary results we showed that the method is robust to the initial planning CT image segmentation. The 4D segmentation is valuable for further quantitative analysis of the data sets.

A.6 Acknowledgments

This work was supported in part by NIH grant CA166703.

APPENDIX B

ALPHA SHAPES FOR LUNG TUMOR INCLUSION

Published as: SE Gerard, HJ Johnson, JE Bayouth, GE Christensen, K Du, J Guo, JM Reinhardt: Alpha shapes for lung segmentation in the presence of Large Tumors. 6th International Workshop on Pulmonary Image Analysis, 2016.

Lung segmentation is a critical initial step in planning radiation therapy interventions for lung cancer patients. Achieving robust automatic segmentation of lungs with large tumors is challenging due to large variations in lung morphology, tumor location, and tumor shape between subjects. We present an automatic method to segment lungs with large tumors in CT images using an initial intensity based segmentation followed by alpha shape construction and graph search. We evaluated our method by comparing automated segmentations to manual segmentations on twelve subjects. Computed metrics for segmentation quality include average surface distance of 0.727 mm and average DICE coefficient of 0.970. These results demonstrate that the proposed method accurately segments the entire lung regions both free of and in the presence of large tumors.

B.1 Introduction

Radiation therapy interventions of lung cancer patients involve acquiring a thoracic CT scan to assess the disease state and to design a patient specific treatment plan. Typically a respiratory correlated scan, or 4D CT, is also acquired prior to treatment. This produces a huge amount of data, requiring automatic computer-aided methods for further analysis. Accurate delineation of the lungs is a critical initial first

step for treatment planning and quantitative analysis.

Normal lung tissue has high contrast with the surrounding anatomy, making lung segmentation relatively robust using simple thresholding based methods [54, 60]. However, in the presence of large tumors, conventional threshold based methods fail because the radiodense tumors are excluded from the segmentation. An accurate lung tissue segmentation including the tumor is required during radiation therapy planning, e.g., to predict the dose delivered to the normal lung tissue and the tumor or to perform tissue motion or biomechanical analyses.

Several approaches have been used to segment pathological lungs. Statistical shape models (SSM) use a training set of lung shapes with annotated corresponding landmarks to capture the major modes of lung shape variation in a model that can be fit to new lung cases [131, 136]. The SSM training set is limited to a subset of possible lung shape variations and thus it often fails on cases that are not represented in the training set. Segmentation by registration methods [130, 143] make use an atlas or multiple atlases with ground truth segmentations. The atlas segmentation is mapped to a test case using image registration. Atlas based segmentations require a very accurate registration to get adequate results. Recently, machine learning based methods have been proposed to classify pathological lung tissue [88, 132, 131]. The LObe and Lung Analysis 2011 (LOLA11) challenge evaluated different algorithms on a publicly available dataset of 55 chest CT scans with varying degrees of pathology.

In this paper we propose an alpha shape approach for inclusion of large tumors in lung segmentations. The concept of using alpha shapes to represent a point set

was originally presented in [33]. In [135], alpha shapes were used for boundary reconstruction of shapes in 2D images. In [17] alpha shapes were used to generate patient specific maps for renal segmentation in 3D ultrasound images. To the best of our knowledge, this is the first time that alpha shapes have been applied to the problem of lung segmentation.

B.2 Methods

The proposed method has three main steps: an initial intensity based segmentation, an alpha shape computation of the initial mask, and an optimal graph search for final refinement. Figure B.1 shows a flowchart of the method. Each of these steps will be described next.

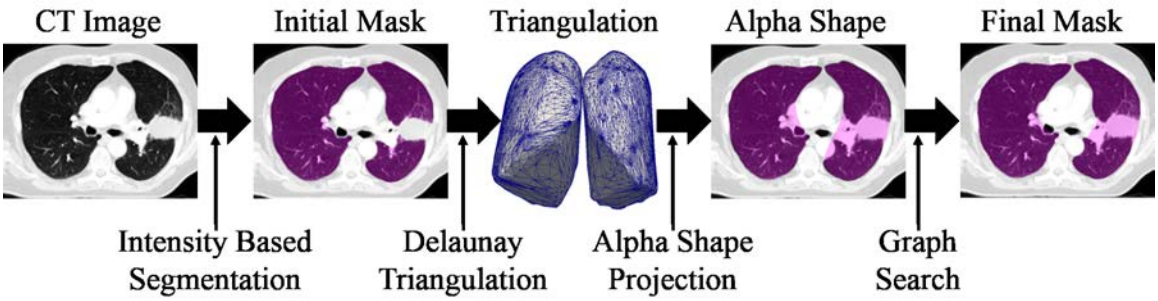


Figure B.1. Flowchart of the proposed segmentation method.

B.2.1 Initial Segmentation

First, we obtain an initial segmentation of the lungs using an intensity-based approach [54]. The method from [54] consists of three main steps: extraction of lungs using optimal thresholding, separation of the right and left lungs, and optimal

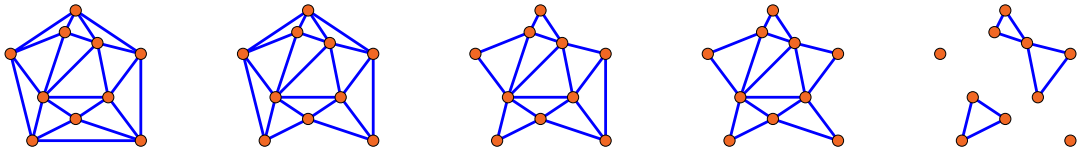
smoothing to smooth the lung boundaries. This gives separate masks for both left and right normal lung tissue lung, however large dense tumors are not included in the segmentation. We will refer to this result as the initial mask. The next steps will process initial masks for the left and right lung separately.

B.2.2 Alpha Shapes

Given a finite set of points $P \in \mathbb{R}^n$ and $0 \leq \alpha \leq \infty$, an α-shape is one way to represent the shape of P. If $\alpha = \infty$, the α-shape is equivalent to the convex hull of P. A family of shapes can be obtained by varying α; as α decreases the α-shape gradually becomes more constricted to a tighter fit around P.

The Delaunay triangulation of a point set can be used to calculate the α -shape. The Delaunay triangulation of a point set $DT(P)$ is the triangulation that maximizes the minimum angle of each d-simplex $t \in DT(P)$. Note: in 2D and 3D the d-simplices t are triangles and tetrahedrons, respectively. $DT(P)$ is the dual graph of the Voronoi diagram of P. The α-shape consists of a subset of d-simplices $t \in DT(P)$; it is the union of d-simplices $t \in DT(P)$ with circumsphere radius $R_c(t) \leq \alpha$.The surface mesh representation of the α-shape consists of all boundary facets of these d-simplices. The Delaunay triangulation and α-shape calculation is illustrated in Figure B.2 using a toy example and Figure B.3 using a 2D lung example.

The lung shape is approximately convex along the pleural surface, so the convex hull of the initial mask gives a good estimate of the boundary for this surface, even in presence of large chest wall tumors. However, the mediastinum and diaphragm regions are not convex so the convex hull oversegments these regions. α-shapes allow



(a) $\alpha = \infty$ (b) $\alpha = 33$ (c) $\alpha = 25$ (d) $\alpha = 20$ (e) $\alpha = 15$

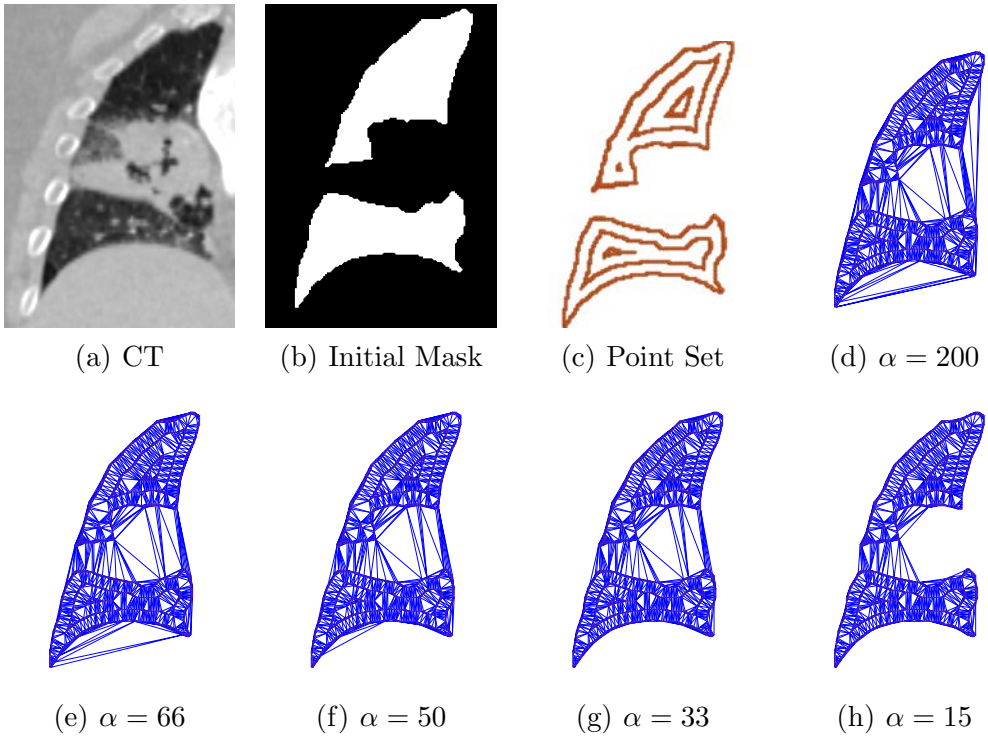
Figure B.2. Simple example showing Delaunay triangulation and α-shape calculation. The red circles are the point set. B.2(a)- B.2(e) show the Delaunay triangles that make up the α-shape for decreasing values of α. B.2(a) $\alpha = \infty$ is equivalent to the Delaunay triangulation and convex hull. As α decreases, triangles are removed that have a circumcircle with radius greater than α. For each example, the α-shape is the union of all triangles.

us to elegantly obtain a family of shapes that smoothly interpolate between the convex hull and initial mask to represent the lung shape.

As the α-shape is defined on a point set, rather than an image, first we obtain a set of points $p \in P$ representing the initial mask. We could use the physical locations of all image voxels in the mask as our point set, however for time efficiency we chose to only take a sample of these voxels. We performed image erosion and subtraction operations to generate a set of concentric contours of the initial mask and used these voxel locations for P. To ensure all of the initial mask is included in the α-shape, the separation between concentric contours needs to be small enough such that the d-simplices formed between contours have a circumradius smaller than α.

There is a trade off between oversegmentation and undersegmentation when choosing α. If α is too large, the mediastinum is oversegmented, however decreasing α too much can remove tumors (Figure B.4). We experimented with different values of α to obtain lung shapes that included tumor regions and minimized oversegmentation

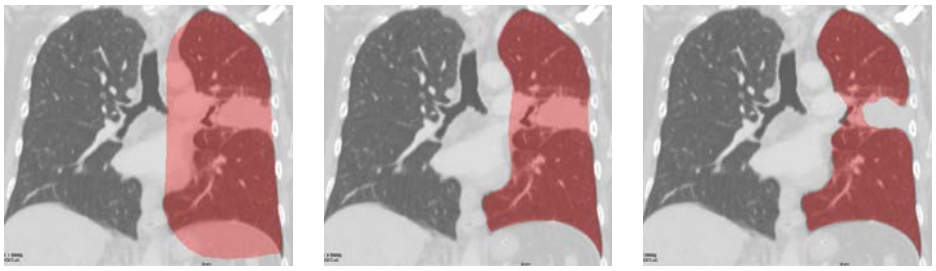
near the diaphragm and mediastinum. The diaphragm is easily removed due to the broad concave structure, however the mediastinum has many very small concavities resulting in minor oversegmentation that cannot be removed without removing the tumor. We empirically determined $\alpha = 25$ to give good results for all subjects in this study. Prior to the next step, we performed Laplacian smoothing [59] on the alpha shape surface to remove sharp edges.



(a) CT (b) Initial Mask (c) Point Set (d) $\alpha = 200$

(e) $\alpha = 66$ (f) $\alpha = 50$ (g) $\alpha = 33$ (h) $\alpha = 15$

Figure B.3. B.3(a) CT image showing large radiodense tumor in right lung, B.3(b) initial mask produced by thresholding, B.3(c) point set used to represent initial mask, B.3(d)- B.3(h) Delaunay triangles that form the alpha shape for decreasing alpha values. Note: this is a 2D example for illustrative purposes; the actual method is performed in 3D with tetrahedrons instead of triangles.



(a) $\alpha = 200$ (b) $\alpha = 25$ (c) $\alpha = 9$

Figure B.4. Alpha shape of initial mask for different values of α. B.4(a) is approximately equal to the convex hull of the initial mask, B.4(c) is approximately equal to the initial mask.

B.2.3 Graph Search

To reduce minor mediastinum oversegmentation resulting from small concavities, we use a graph search framework to find the optimal lung surface. The method will only briefly be described here; see [79] for a detailed description of the method.

The graph search method requires an initial surface to initialize the graph construction, here we use the alpha shape from the previous step. A graph $G(E, V)$ consisting of node set V and edge set E is built in a margin around the initial surface. The nodes have an associated cost reflecting the unlikeliness that it belongs to the lung surface. We use the inverse gradient of the CT image for the cost, as the lung surface has a high image gradient. Edges are used to enforce smoothness constraints, or how much the topology of the resulting segmentation can deviate from the initial surface. The globally optimal surface is obtained by using a maximum flow algorithm on a closely derived graph.

As a final step we performed a logical OR operation between the graph search

result and the initial mask. This adds back in sharp edges, such as those at the costaphrenic angle, that are lost when the smoothing is performed.

B.3 Data Sets and Experimental Setup

In this study we used breath-hold thoracic CT scans from twelve lung cancer subjects about to undergo radiation therapy. All scans were gathered under a protocol approved by the University of Iowa Institutional Review Board (IRB 200905703). All images were resampled from $0.98 \times 0.98 \times 2$ mm^3 to obtain $1 \times 1 \times 1$ mm^3 isotropic voxels. We chose 11 subjects that had a large lung tumor in either the left or right lung, and one subject with no lung tumors.

Our method was evaluated by comparing to manual segmentations generated by a radiation therapy physicist using the MimVista 6.4.7 software (MIM Software, Cleveland, OH) following a protocol similar to that used for radiation therapy planning. The DICE coefficient was used for a metric of volume overlap, and the average unsigned symmetric surface distance was used to measure the distance between the lung boundaries. The analysis was performed on both the left and right lungs of all subjects. Additionally, we compared the results of lungs with and without tumors.

The proposed method was implemented using the Jupyter Notebook [105] rapid prototyping environment, algorithms from ITK [64] python wrapped in SimpleITK [85] in coordination with vtkDelaunay3D from VTK (www.vtk.org).

B.4 Results

Figure B.5 shows the average unsigned surface distances for the initial mask and our final segmentation for both tumor and non-tumor groups. Figure B.6 shows

the result at each step of the proposed method and the manual segmentation.

B.5 Discussion

The method shows good results with an average unsigned surface distance of 0.727 mm and an average DICE coefficient of 0.970 for all subjects. This surface distance is on the sub-voxel level. Additionally, we split the lungs into tumor and non-tumor groups. The DICE coefficient was 0.967 and 0.971 for tumor and non-tumor groups, respectively. The average unsigned surface distance was 0.758 mm and 0.702 mm for tumor and non-tumor groups, respectively. A t-test was performed to compare the surface distance error for lungs with tumors and lungs without tumors and no significant difference was found (p-value = 0.314). The surface distance for the initial mask was 1.368 mm and 0.710 mm for tumor and non-tumor groups, respectively. These results show that compared to the initial mask, our method significantly improves the average surface distance error for lungs with tumors (p-value 0.0008) and does not significantly change the average surface distance error for lungs without tumors (p-value 0.867).

By visual inspection, it is clear that the majority of the discrepancy between our results and the manual segmentations occur at the mediastinum. This is a subjective region to segment, as can be observed on coronal slices of the manual segmentations. The manual segmentations were done on transverse slices and variation on where the mediastinal boundary was define between slices resulted in the boundary not being smooth in 3D. The proposed method produces a smooth mediastinum boundary that is consistent between subjects.

It is important to note that this method is limited by the convex hull of the initial segmentation. If the tumor is outside the convex hull, it will fail to be included in the final result. For example, tumors that span the entire apex region of the lung will not be included. This is a current limitation to our method, but could possibly be overcome by including points from the rib cage when calculating the convex hull.

In future work we plan on identifying the optimal α for each subject rather than using the same α for all subjects. Additionally, we plan to experiment with automatically identifying a spatially varying alpha to overcome over segmentation near mediastinum.

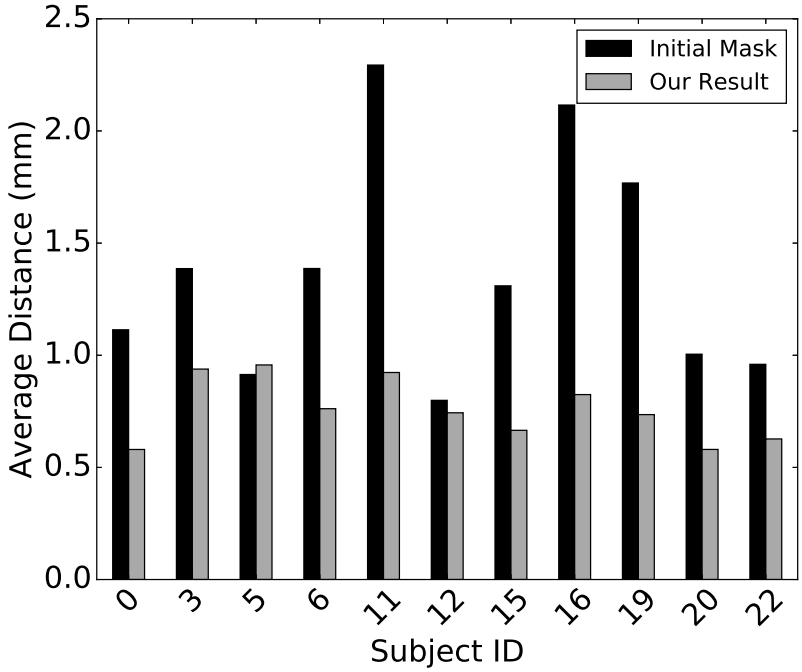
The experiments were run on a Linux machine with an Intel Xeon 2.27 GHz CPU and 48 GB of RAM. Generating the alpha shape and the graph search take approximately 13 and 1 minute(s) of computer time, respectively. The manual segmentation took on average 53 minutes per subject.

B.6 Summary

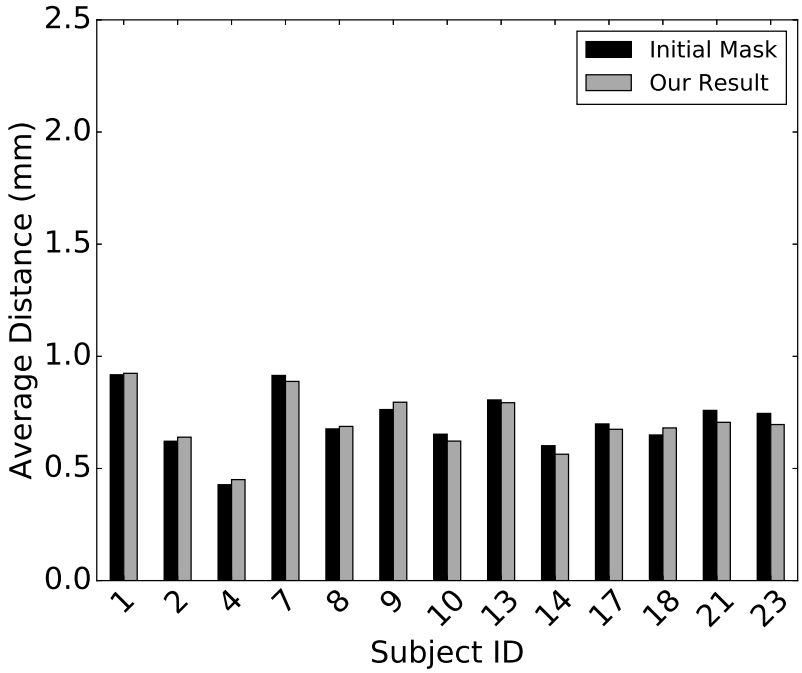
We proposed a method for segmentation of lungs in the presence of large tumors. The method utilizes an intensity based segmentation, alpha shapes correction, and a graph search framework final refinement. An average surface error distance of 0.727 mm and DICE coefficient of 0.970 and were achieved. The accurate lung segmentation with inclusion of tumor regions is valuable for radiation therapy treatment planning and further quantitative analysis.

B.7 Acknowledgments

This work was supported in part by NIH grant CA166703.



(a)



(b)

Figure B.5. Average unsigned surface distances for both the initial mask and our final result compared to the manual segmentation. B.5(a) and B.5(b) show results for lungs with and without tumors, respectively.

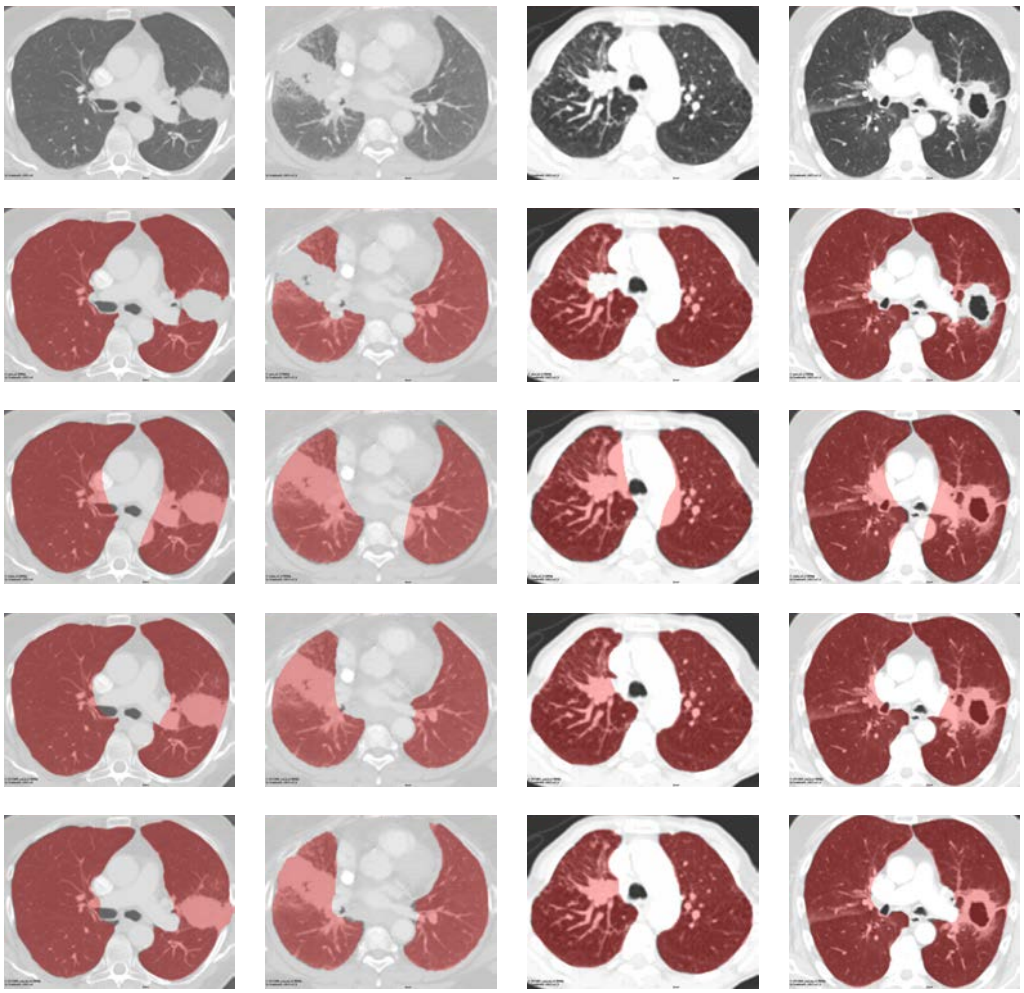


Figure B.6. Segmentation at each step of the proposed method for four different data sets displayed in columns one to four. Row 1: CT image. Row 2: initial mask. Row 3: alpha shape of the initial mask. Row 4: final segmentation using the proposed automatic method. Row 5: manual segmentation.

REFERENCES

[1] Cancer statistics center. https://cancerstatisticscenter.cancer.org/#/. Accessed: 2016-06-20.

[2] Lung cancer screening - patient version. http://www.cancer.gov/types/lung/patient/lung-screening-pdq#section/all. Accessed: 2016-06-20.

[3] Ryan E. Amelon, Kunlin. Cao, Kai Ding, Gary E. Christensen, Joseph M. Reinhardt, and Madhavan L. Raghavan. Three-dimensional characterization of regional lung deformation. *J Biomechanics*, 44:2489–2495, Sep 2011.

[4] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Medical Imaging*, 35(5):1207–1216, 2016.

[5] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Thomas Geiser, Andreas Christe, and Stavroula Mougiakakou. Semantic segmentation of pathological lung tissue with dilated fully convolutional networks. *arXiv preprint arXiv:1803.06167*, 2018.

[6] SG Armato III and Heber MacMahon. Automated lung segmentation and computer-aided diagnosis for thoracic CT scans. In *International Congress Series*, volume 1256, pages 977–982. Elsevier, 2003.

[7] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.

[8] Carolyn E Behrendt. Mild and moderate-to-severe COPD in nonsmokers: distinct demographic profiles. *Chest*, 128(3):1239–1244, 2005.

[9] CL Bellardine Black, AM Hoffman, LW Tsai, EP Ingenito, B Suki, DW Kaczka, BA Simon, and KR Lutchen. Impact of positive end-expiratory pressure during heterogeneous lung injury: insights from computed tomographic image functional modeling. *Annals of biomedical engineering*, 36(6):980–991, 2008.

[10] Adel Boueiz, Yale Chang, Michael H Cho, George R Washko, Raul San José Estépar, Russell P Bowler, James D Crapo, Dawn L DeMeo, Jennifer G Dy, Edwin K Silverman, P J Castaldi, and COPDGene Investigators. Lobar emphysema distribution is associated with 5-year radiological disease progression. *Chest*, 2017.

[11] Y. Boykov and M.P. Jolly. An experimental comparison of min-cut/max-flow algorithms for energy minimizations in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1124–1137, 2004.

[12] Matthew S Brown, Jonathan G Goldin, Michael F McNitt-Gray, Lloyd E Greaser, Amita Sapra, Kuo-Tung Li, James W Sayre, Katherine Martin, and Denise R Aberle. Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function. *Medical physics*, 27(3):592–598, 2000.

[13] A. S. Buist, W. M. Vollmer, and M. A. McBurnie. Worldwide burden of COPD in high- and low-income countries. Part I. The burden of obstructive lung disease (BOLD) initiative. *Int. J. Tuberc. Lung Dis.*, 12(7):703–708, Jul 2008.

[14] Alysson R Carvalho, Peter M Spieth, Paolo Pelosi, Marcos F Vidal Melo, Thea Koch, Frederico C Jandre, Antonio Giannella-Neto, and Marcelo Gama de Abreu. Ability of dynamic airway pressure curve profile and elastance for positive end-expiratory pressure titration. *Intensive care medicine*, 34(12):2291, 2008.

[15] Maurizio Cereda, Yi Xin, Hooman Hamedani, Giacomo Bellani, Stephen Kadlecek, Justin Clapp, Luca Guerra, Natalie Meeder, Jennia Rajaei, Nicholas J Tustison, et al. Tidal changes on CT and progression of ARDS. *Thorax*, pages thoraxjnl–2016, 2017.

[16] Maurizio Cereda, Yi Xin, Natalie Meeder, Johnathan Zeng, YunQing Jiang, Hooman Hamedani, Harrilla Profka, Stephen Kadlecek, Justin Clapp, Charuhas G Deshpande, et al. Visualizing the propagation of acute lung injury. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 124(1):121–131, 2016.

[17] Juan J Cerrolaza, Christopher Meyer, James Jago, Craig Peters, and Marius George Linguraru. Positive delta detection for alpha shape segmentation of 3D ultrasound images of pathologic kidneys. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 711–718. Springer, 2015.

[18] Prachya Chalermwat and Tarek El-Ghazawi. Multi-resolution image registration using genetics. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 2, pages 452–456. IEEE, 1999.

[19] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):681–685, 2001.

[20] Timothy F Cootes, Christopher J Taylor, and Andreas Lanitis. Active shape models: Evaluation of a multi-resolution method for improving image search. In *BMVC*, volume 1, pages 327–336. Citeseer, 1994.

[21] Silvia Coppola, Sara Froio, and Davide Chiumello. Lung imaging in ARDS. In *Acute Respiratory Distress Syndrome*, pages 155–171. Springer, 2017.

[22] David Couper, Jeffrey L Curtis, Natalia A Gouskova, Nadia N Hansel, Eric A Hoffman, Richard E Kanner, Eric Kleerup, Stephen C Lazarus, Fernando J Martinez, Robert Paine, et al. Clinical significance of symptoms in smokers with preserved pulmonary function. 2016.

[23] David Couper, Lisa M LaVange, MeiLan Han, R Graham Barr, Eugene Bleecker, Eric A Hoffman, Richard Kanner, Eric Kleerup, Fernando J Martinez, Prescott G Woodruff, et al. Design of the subpopulations and intermediate outcomes in copd study (spiromics). *Thorax*, 69(5):492–495, 2014.

[24] Harvey O Coxson. Quantitative computed tomography assessment of airway wall dimensions: Current status and potential applications for phenotyping chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 5(9):940–945, 2008.

[25] LM Cuevas, PM Spieth, AR Carvalho, MG de Abreu, and E Koch. Automatic lung segmentation of helical-ct scans in experimental induced lung injury. In *4th European Conference of the International Federation for Medical and Biological Engineering*, pages 764–767. Springer, 2009.

[26] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[27] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.

[28] Sander Dieleman, Jan Schlüter, Colin Raffel, Eben Olson, Søren Kaae Sønderby, Daniel Nouri, et al. Lasagne: First release., August 2015.

[29] Kai Ding, John E. Bayouth, John M. Buatti, Gary E. Christensen, and Joseph M. Reinhardt. 4DCT-based measurement of changes in pulmonary function following a course of radiation therapy. *Medical Physics*, 37(3):1261–1272, March 2010.

[30] Tom Doel, David J Gavaghan, and Vicente Grau. Review of automatic pulmonary lobe segmentation methods from CT. *Computerized Medical Imaging and Graphics*, 40:13–29, 2015.

[31] Kaifang Du, John E Bayouth, Kunlin Cao, Gary E Christensen, Kai Ding, and Joseph M Reinhardt. Reproducibility of registration-based measures of lung tissue expansion. *Medical Physics*, 39(3):1595–1608, 2012.

[32] R. Eberhardt, D. Gompelmann, F. J. Herth, and M. Schuhmann. Endoscopic bronchial valve treatment: Patient selection and special considerations. *Int J Chron Obstruct Pulmon Dis*, 10:2147–2157, 2015.

[33] Herbert Edelsbrunner, David G. Kirkpatrick, and Raimund Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 1983.

[34] Jan Ehrhardt, Cristian Lorenz, et al. *4D Modeling and estimation of respiratory motion for radiation therapy*. Springer, 2013.

[35] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[36] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[37] A Fernandez-Bustamante, RB Easley, M Fuld, D Mulreany, Deokiee Chon, JF Lewis, and BA Simon. Regional pulmonary inflammation in an endotoxemic ovine acute lung injury model. *Respiratory physiology & neurobiology*, 183(2):149–158, 2012.

[38] Ana Fernandez-Bustamante, R Blaine Easley, Matthew Fuld, Daniel Mulreany, Eric A Hoffman, and Brett A Simon. Regional aeration and perfusion distribution in a sheep model of endotoxemic acute lung injury characterized by functional computed tomography imaging. *Critical care medicine*, 37(8):2402–2411, 2009.

[39] Rosa L Figueroa, Qing Zeng-Treitler, Sasikiran Kandula, and Long H Ngo. Predicting sample size required for classification performance. *BMC medical informatics and decision making*, 12(1):8, 2012.

[40] Centers for Disease Control, Prevention (CDC, and others). Chronic obstructive pulmonary disease among adults–united states, 2011. *MMWR. Morbidity and mortality weekly report*, 61(46):938, 2012.

[41] ARDS Definition Task Force, VM Ranieri, GD Rubenfeld, et al. Acute respiratory distress syndrome. *Jama*, 307(23):2526–2533, 2012.

[42] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998.

[43] Craig J Galbán, Meilan K Han, Jennifer L Boes, Komal A Chughtai, Charles R Meyer, Timothy D Johnson, Stefanie Galbán, Alnawaz Rehemtulla, Ella A Kazerooni, Fernando J Martinez, et al. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nature medicine*, 18(11):1711, 2012.

[44] Luciano Gattinoni, Pietro Caironi, Massimo Cressoni, Davide Chiumello, V Marco Ranieri, Michael Quintel, Sebastiano Russo, Nicolò Patroniti, Rodrigo Cornejo, and Guillermo Bugedo. Lung recruitment in patients with the acute respiratory distress syndrome. *New England Journal of Medicine*, 354(17):1775–1786, 2006.

[45] Luciano Gattinoni, Pietro Caironi, Paolo Pelosi, and Lawrence R Goodman. What has computed tomography taught us about the acute respiratory distress syndrome? *American journal of respiratory and critical care medicine*, 164(9):1701–1711, 2001.

[46] Kevin George, Adam P Harrison, Dakai Jin, Ziyue Xu, and Daniel J Mollura. Pathological pulmonary lobe segmentation from CT images using progressive holistically nested neural networks and random walker. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 195–203. Springer, 2017.

[47] Sarah E Gerard, Jacob Herrmann, David W Kaczka, and Joseph M Reinhardt. Transfer learning for segmentation of injured lungs using coarse-to-fine convolutional neural networks. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 191–201. Springer, 2018.

[48] Sarah E. Gerard, Hans J. Johnson, John E. Bayouth, Gary E. Christensen, Kaifang Du, Junfeng Guo, and Joseph M. Reinhardt. Alpha shapes for lung segmentation in the presence of large tumors. In *6^{th} International Workshop on Pulmonary Image Analysis*, pages 9–17, 2016.

[49] Sarah E Gerard, Taylor J Patton, Gary E Christensen, John E Bayouth, and Joseph M Reinhardt. FissureNet: A deep learning approach for pulmonary fissure detection in CT images. *IEEE Trans. Medical Imaging*, 2018. PMID: 30106711.

[50] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[51] Thomas Godet, Matthieu Jabaudon, Raïko Blondonnet, Aymeric Tremblay, Jules Audard, Benjamin Rieu, Bruno Pereira, Jean-Marc Garcier, Emmanuel Futier, and Jean-Michel Constantin. High frequency percussive ventilation increases alveolar recruitment in early acute respiratory distress syndrome: an experimental, physiological and CT scan study. *Critical Care*, 22(1):3, 2018.

[52] D. Gompelmann, M. Schuhmann, K. Kontogianni, C. Heussel, J. Tschirren, R. Eberhardt, and F. J. F. Herth. Fissure integrity and emphysema distribution in patients with severe chronic obstructive pulmonary disease. *Amer. J. Respiratory and Critical Care Medicine*, page A4985, 2017. (abstract).

[53] Hayit Greenspan, Bram Van Ginneken, and Ronald M Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.

[54] J Guo, MK Fuld, SK Alford, JM Reinhardt, and EA Hoffman. Pulmonary analysis software suite 9.0: Integrating quantitative measures of function with structural analyses. In *First International Workshop on Pulmonary Image Analysis*, pages 283–292, 2008.

[55] Junfeng Guo, Matthew K. Fuld, Sara K. Alford, Joseph M. Reinhardt, and Eric A. Hoffman. Pulmonary Analysis Software Suite 9.0: Integrating quantitative measures of function with structural analyses. In Matthew Brown, Marleen de Bruijne, Bram van Ginneken, Atilla Kiraly, Jan Martin Kuhnigk, Cristian Lorenz, Kensaku Mori, and Joseph Reinhardt, editors, *First International Workshop on Pulmonary Image Analysis*, pages 283–292, 2008.

[56] Adam P Harrison, Ziyue Xu, Kevin George, Le Lu, Ronald M Summers, and Daniel J Mollura. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from ct images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 621–629. Springer, 2017.

[57] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563, 2009.

[58] Jacob Herrmann, Eric A Hoffman, and David W Kaczka. Frequency-selective computed tomography: Applications during periodic thoracic motion. *IEEE transactions on medical imaging*, 36(8):1722–1732, 2017.

[59] Leonard R Herrmann. Laplacian-isoparametric grid generation scheme. *Journal of the Engineering Mechanics Division*, 102(5):749–907, 1976.

[60] Shiying Hu, Eric A Hoffman, and Joseph M Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-Ray CT images. *IEEE Transactions on Medical Imaging*, 20(6):490–498, 2001.

[61] Panfang Hua, Qi Song, Milan Sonka, Eric A Hoffman, and Joseph M Reinhardt. Segmentation of pathological and diseased lung tissue in CT images using a graph-search algorithm. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 2072–2075. IEEE, 2011.

[62] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

[63] Carlos A Jiménez-Ruiz, Fernando Masa, Marc Miravitlles, Rafael Gabriel, José Luis Viejo, Carlos Villasante, Victor Sobradillo, IBERPOC Study Investigators, et al. Smoking characteristics: differences in attitudes and dependence between healthy smokers and smokers with COPD. *Chest*, 119(5):1365–1370, 2001.

[64] Hans J. Johnson, Matthew M McCormick, and Luis Ibanez. The ITK Software Guide Book 2: Design and Functionality-Volume 2. 2015.

[65] David W Kaczka, Kunlin Cao, Gary E Christensen, Jason HT Bates, and Brett A Simon. Analysis of regional mechanics in canine lung injury using forced oscillations and 3D image registration. *Annals of biomedical engineering*, 39(3):1112–1124, 2011.

[66] Paul Keall. 4-dimensional computed tomography imaging and treatment planning. In *Seminars in Radiation Oncology*, volume 14, pages 81–90. Elsevier, 2004.

[67] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[68] Eyal Klang. Deep learning and medical imaging. *Journal of thoracic disease*, 10(3):1325, 2018.

[69] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim. Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29:196–205, Jan 2010.

[70] Tobias Klinder, Hannes Wendland, and Rafael Wiemker. Lobar fissure detection using line enhancing filters. In *Medical Imaging: Image Processing*, page 86693C, 2013.

[71] T. D. Koster and D. J. Slebos. The fissure: Interlobar collateral ventilation and implications for endoscopic therapy in emphysema. *Int J Chron Obstruct Pulmon Dis*, 11:765–773, 2016.

[72] T. D. Koster, E. M. van Rikxoort, R. H. Huebner, F. Doellinger, K. Klooster, J. P. Charbonnier, S. Radhakrishnan, F. J. Herth, and D. J. Slebos. Predicting Lung Volume Reduction after Endobronchial Valve Therapy Is Maximized Using a Combination of Diagnostic Tools. *Respiration*, 92(3):150–157, 2016.

[73] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[74] Jan-Martin Kuhnigk, Volker Dicken, Stephan Zidowitz, Lars Bornemann, Bernd Kuemmerlen, Stefan Krass, Heinz-Otto Peitgen, Silja Yuval, Hans-Holger Jend, Wigbert S Rau, et al. New tools for computer assistance in thoracic CT. Part 1. Functional analysis of lungs, lung lobes, and bronchopulmonary segments. *Radiographics*, 25(2):525–536, 2005.

[75] Bianca Lassen, Eva M van Rikxoort, Michael Schmidt, Sjoerd Kerkstra, Bram van Ginneken, and Jan-Martin Kuhnigk. Automatic segmentation of the pulmonary lobes from chest CT scans based on fissures, vessels, and bronchi. *IEEE Transactions on Medical Imaging*, 32(2):210–222, 2013.

[76] Joseph K Leader, Bin Zheng, Robert M Rogers, Frank C Sciurba, Andrew Perez, Brian E Chapman, Sanjay Patel, Carl R Fuhrman, and David Gur. Automated lung segmentation in X-Ray computed tomography: development and evaluation of a heuristic threshold-based scheme1. *Academic radiology*, 10(11):1224–1236, 2003.

[77] Yann LeCun, LD Jackel, Leon Bottou, A Brunot, Corinna Cortes, JS Denker, Harris Drucker, I Guyon, UA Muller, Eduard Sackinger, et al. Comparison of learning algorithms for handwritten digit recognition. In *International conference on artificial neural networks*, volume 60, pages 53–60. Perth, Australia, 1995.

[78] Baojun Li, Gary E Christensen, Eric A Hoffman, Geoffrey McLennan, and Joseph M Reinhardt. Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. *Academic radiology*, 10(3):255–265, 2003.

[79] Kang Li, Xiaodong Wu, Danny Z Chen, and Milan Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):119–134, 2006.

[80] Qiang Li, Shusuke Sone, and K Doi. Selective enhancement filters for nodules, vessels, and airway walls in two-and three-dimensional CT scans. *Medical Physics*, 30(8):2040–2051, 2003.

[81] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, pages 844–848. IEEE, 2014.

[82] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[83] Antonio M López, David Lloret, Joan Serrat, and Juan J Villanueva. Multilocal creaseness based on the level-set extrinsic curvature. *Computer Vision and Image Understanding*, 77(2):111–144, 2000.

[84] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[85] Bradley Christopher Lowekamp, David T Chen, Luis Ibanez, and Daniel Blezek. The design of SimpleITK. *Frontiers in Neuroinformatics*, 7(45), 2013.

[86] Afarine Madani, Caroline Keyzer, and Pierre-Alain Gevenois. Quantitative computed tomography assessment of lung structure and function in pulmonary emphysema. *European Respiratory Journal*, 18(4):720–730, 2001.

[87] Bangalore S Manjunath and Wei-Ying Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, 1996.

[88] Awais Mansoor, Ulas Bagci, Ziyue Xu, Brent Foster, Kenneth N Olivier, Jason M Elinoff, Anthony F Suffredini, Jayaram K Udupa, and Daniel J Mollura. A generic approach to pathological lung segmentation. *IEEE Transactions on Medical Imaging*, 33(12):2293–2310, 2014.

[89] John E McDonough, Ren Yuan, Masaru Suzuki, Nazgol Seyednejad, W Mark Elliott, Pablo G Sanchez, Alexander C Wright, Warren B Gefter, Leslie Litzky, Harvey O Coxson, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *New England Journal of Medicine*, 365(17):1567–1575, 2011.

[90] Sayan Mukherjee, Pablo Tamayo, Simon Rogers, Ryan Rifkin, Anna Engle, Colin Campbell, Todd R Golub, and Jill P Mesirov. Estimating dataset size requirements for classifying dna microarray data. *Journal of computational biology*, 10(2):119–142, 2003.

[91] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[92] National Emphysema Treatment Trial Research Group. Patients at high risk of death after lung-volume–reduction surgery. *N Engl J Med*, 2001(345):1075–1083, 2001.

[93] Idiopathic Pulmonary Fibrosis Clinical Research Network. Prednisone, azathioprine, and n-acetylcysteine for pulmonary fibrosis. *New England Journal of Medicine*, 366(21):1968–1977, 2012.

[94] Idiopathic Pulmonary Fibrosis Clinical Research Network. Randomized trial of acetylcysteine in idiopathic pulmonary fibrosis. *New England Journal of Medicine*, 370(22):2093–2101, 2014.

[95] Kenneth B Newman, David A Lynch, Lee S Newman, Denise Ellegood, and John D Newell Jr. Quantitative computed tomography detects air trapping due to asthma. *Chest*, 106(1):105–109, 1994.

[96] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1520–1528, 2015.

[97] I. Oguz and M. Sonka. LOGISMOS_B: Layered optimal graph image segmentation of multiple objects and surfaces for the brain. *IEEE TMI*, 33:1220–1235, 2014.

[98] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[99] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[100] Yue Pan, Gary E. Christensen, Oguz C. Durumeric, Sarah E. Gerard, Joseph M. Reinhardt, and Geoffrey D. Hugo. Current- and varifold-based registration of lung vessel and airway trees. In *7th International Workshop on Biomedical Image Registration*, June 2016.

[101] Taylor J Patton, Sarah E Gerard, Wei Shao, Gary E Christensen, Joseph M Reinhardt, and John E Bayouth. Quantifying ventilation change due to radiation therapy using 4DCT jacobian calculations. *Medical physics*, 45(10):4483–4492, 2018.

[102] Luis Felipe Paula, Tyler J Wellman, Tilo Winkler, Peter M Spieth, Andreas Güldner, Jose G Venegas, Marcelo Gama de Abreu, Alysson R Carvalho, and Marcos F Vidal Melo. Regional tidal lung strain in mechanically ventilated normal lungs. *Journal of Applied Physiology*, 121(6):1335, 2016.

[103] Gaetano Perchiazzi, Christian Rylander, Savino Derosa, Mariangela Pellegrini, Loredana Pitagora, Debora Polieri, Antonio Vena, Angela Tannoia, Tommaso Fiore, and Göran Hedenstierna. Regional distribution of lung compliance by image analysis of computed tomograms. *Respiratory physiology & neurobiology*, 201:60–70, 2014.

[104] Gaetano Perchiazzi, Christian Rylander, Antonio Vena, Savino Derosa, Debora Polieri, Tommaso Fiore, Rocco Giuliani, and Göran Hedenstierna. Lung regional stress and strain as a function of posture and ventilatory mode. *Journal of applied physiology*, 110(5):1374–1383, 2011.

[105] Fernando Pérez and Brian E. Granger. IPython: A system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

[106] Victor M Pinto-Plata, Hana Müllerova, John F Toso, Maurille Feudjo-Tepie, Joan B Soriano, Rupert S Vessey, and Bartolome R Celli. C-reactive protein in patients with COPD, control smokers and non-smokers. *Thorax*, 61(1):23–28, 2006.

[107] Alfredo Morales Pinzon. *Lung segmentation and airway tree matching: application to aeration quantification in CT images of subjects with ARDS*. PhD thesis, Université de Lyon, 2016.

[108] Alfredo Morales Pinzón, Maciej Orkisz, Jean-Christophe Richard, and Marcela Hernández Hoyos. Lung segmentation in 3D CT images from induced acute respiratory distress syndrome.

[109] Mithun N Prasad, Matthew S Brown, Shama Ahmad, Fereidoun Abtin, Jared Allen, Irene da Costa, Hyun J Kim, Michael F McNitt-Gray, and Jonathan G Goldin. Automatic segmentation of lung parenchyma in the presence of diseases based on curvature of ribs. *Academic radiology*, 15(9):1173–1180, 2008.

[110] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 246–253. Springer, 2013.

[111] J. Pu, J. Roos, C. A. Yi, S. Napel, G. D. Rubin, and D. S. Paik. Adaptive border marching algorithm: Automatic lung segmentation on chest CT images. *Comput Med Imaging Graph*, 32(6):452–462, Sep 2008.

[112] J. Pu, Z. Wang, S. Gu, C. Fuhrman, J. K. Leader, X. Meng, J. Tedrow, and F. C. Sciurba. Pulmonary fissure integrity and collateral ventilation in COPD patients. *PLoS ONE*, 9(5):e96631, 2014.

[113] Jiantao Pu, Justus Roos, A Yi Chin, Sandy Napel, Geoffrey D Rubin, and David S Paik. Adaptive border marching algorithm: automatic lung segmentation on chest CT images. *Computerized Medical Imaging and Graphics*, 32(6):452–462, 2008.

[114] Jiantao Pu, Bin Zheng, Joseph K Leader, Carl Fuhrman, Friedrich Knollmann, Amy Klym, and David Gur. Pulmonary lobe segmentation in CT examinations using implicit surface fitting. *IEEE Transactions on Medical Imaging*, 28(12):1986–1996, 2009.

[115] Elizabeth A Regan, John E Hokanson, James R Murphy, Barry Make, David A Lynch, Terri H Beaty, Douglas Curran-Everett, Edwin K Silverman, and James D Crapo. Genetic epidemiology of COPD (COPDGene) study design. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 7(1):32–43, 2011.

[116] Joseph M. Reinhardt, Gary E. Christensen, Eric A. Hoffman, Kai Ding, and Kunlin Cao. Registration-derived estimates of local lung expansion as surrogates for regional ventilation. In Nico Karssemeijer and Boudewijn Lelieveldt, editors, *Information Processing in Medical Imaging*, volume 4584 of *Lecture Notes in Computer Science*, pages 763–774, Utrecht, July 2007. Springer-Verlag.

[117] Joseph M. Reinhardt, Kai Ding, Kunlin Cao, Gary E. Christensen, Eric A. Hoffman, and Shalmali V. Bodas. Registration-based estimates of local lung tissue expansion compared to xenon-CT measures of specific ventilation. *Med. Imag. Analysis*, 12(6):752–763, Dec. 2008.

[118] Martine Remy-Jardin, Jacques Remy, Franck Deschildre, Dominique Artaud, Jean Paul Beregi, Claude Hossein-Foucher, Xavier Marchandise, and Alain Duhamel. Diagnosis of pulmonary embolism with spiral CT: Comparison with pulmonary angiography and scintigraphy. *Radiology*, 200(3):699–706, 1996.

[119] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[120] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533, 1986.

[121] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

[122] Margaret L Salisbury, David A Lynch, Edwin JR Van Beek, Ella A Kazerooni, Junfeng Guo, Meng Xia, Susan Murray, Kevin J Anstrom, Eric Yow, Fernando J Martinez, et al. Idiopathic pulmonary fibrosis: the association between the adaptive multiple features method and fibrosis outcomes. *American journal of respiratory and critical care medicine*, 195(7):921–929, 2017.

[123] Yoshinobu Sato, Shin Nakajima, Nobuyuki Shiraga, Hideki Atsumi, Shigeyuki Yoshida, Thomas Koller, Guido Gerig, and Ron Kikinis. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis*, 2(2):143–168, 1998.

[124] Julia A Schnabel, Daniel Rueckert, Marcel Quist, Jane M Blackall, Andy D Castellano-Smith, Thomas Hartkens, Graeme P Penney, Walter A Hall, Haiying Liu, Charles L Truwit, et al. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–581. Springer, 2001.

[125] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017.

[126] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[127] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, and Jie Tian. Multi-scale convolutional neural networks for lung nodule classification. In *International Conference on Information Processing in Medical Imaging*, pages 588–599. Springer, 2015.

[128] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Medical Imaging*, 35(5):1285–1298, 2016.

[129] Jered P Sieren, John D Newell Jr, R Graham Barr, Eugene R Bleecker, Nathan Burnette, Elizabeth E Carretta, David Couper, Jonathan Goldin, Junfeng Guo, MeiLan K Han, et al. Spiromics protocol for multicenter quantitative computed tomography to phenotype the lungs. *American journal of respiratory and critical care medicine*, 194(7):794–806, 2016.

[130] Ingrid Sluimer, Mathias Prokop, and Bram Van Ginneken. Toward automated segmentation of the pathological lung in CT. *IEEE Transactions on Medical Imaging*, 24(8):1025–1038, 2005.

[131] Michal Sofka, Jens Wetzl, Neil Birkbeck, Jingdan Zhang, Timo Kohlberger, Jens Kaftan, Jérôme Declerck, and S Kevin Zhou. Multi-stage learning for robust lung segmentation in challenging CT volumes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 667–674. Springer, 2011.

[132] Yang Song, Weidong Cai, Jinman Kim, and David Dagan Feng. A multistage discriminative model for tumor and lymph node detection in thoracic images. *IEEE Transactions on Medical Imaging*, 31(5):1061–1075, 2012.

[133] Lauge Sorensen, Saher B Shaker, and Marleen de Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559–569, 2010.

[134] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[135] Peer Stelldinger, Ullrich Köthe, and Hans Meine. Topologically correct image segmentation using alpha shapes. In *Discrete Geometry for Computer Imagery*, pages 542–554. Springer, 2006.

[136] Shanhui Sun, Christian Bauer, and Reinhard Beichel. Automated 3-d segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach. *IEEE Transactions on Medical Imaging*, 31(2):449–460, 2012.

[137] Shanhui Sun, Milan Sonka, and Reinhard R. Beichel. Graph-based 4D lung segmentation in CT images with expert-guided computer-aided refinement. *IEEE 10th International Symposium on Biomedical Imaging*, 2013.

[138] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.

[139] Omid Talakoub, Emma Helm, Javad Alirezaie, Paul Babyn, Brian Kavanagh, Francesco Grasso, and Doreen Engelberts. An automatic wavelet-based approach for lung segmentation and density analysis in dynamic CT. In *Computational Intelligence in Image and Signal Processing, 2007. CIISP 2007. IEEE Symposium on*, pages 369–374. IEEE, 2007.

[140] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016.

[141] Soumik Ukil and Joseph M. Reinhardt. Anatomy-guided lung lobar surface detection in X-ray CT images. *IEEE Trans. Medical Imaging*, 28(2):202–214, 2009. PMID: 19188109.

[142] Igor Rafael S Valente, Paulo César Cortez, Edson Cavalcanti Neto, José Marques Soares, Victor Hugo C de Albuquerque, and João Manuel RS Tavares. Automatic 3D pulmonary nodule detection in CT images: A survey. *Computer methods and programs in biomedicine*, 124:91–107, 2016.

[143] Eva M Van Rikxoort, Bartjan De Hoop, Saskia Van De Vorst, Mathias Prokop, and Bram Van Ginneken. Automatic segmentation of pulmonary segments from volumetric chest CT scans. *IEEE Transactions on Medical Imaging*, 28(4):621–630, 2009.

[144] Eva M van Rikxoort, Bartjan de Hoop, Max A Viergever, Mathias Prokop, and Bram van Ginneken. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*, 36(7):2934–2947, 2009.

[145] Eva M. van Rikxoort, Bartjan de Hoop, Max A. Viergever, Mathias Prokop, and Bram van Ginneken. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*, 36, 2009.

[146] Eva M van Rikxoort, Mathias Prokop, Bartjan de Hoop, Max A Viergever, Josien PW Pluim, and Bram van Ginneken. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE Transactions on Medical Imaging*, 29(6):1286–1296, 2010.

[147] Eva M Van Rikxoort and Bram Van Ginneken. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. *Physics in Medicine and Biology*, 58(17):R187, 2013.

[148] Eva M van Rikxoort, Bram van Ginneken, Mark Klik, and Mathias Prokop. Supervised enhancement filters: Application to fissure detection in chest CT scans. *IEEE Transactions on Medical Imaging*, 27(1):1–10, 2008.

[149] Emily S Wan, Peter J Castaldi, Michael H Cho, John E Hokanson, Elizabeth A Regan, Barry J Make, Terri H Beaty, MeiLan K Han, Jeffrey L Curtis, Douglas Curran-Everett, David A Lynch, Dawn L DeMeo, James D Crapo, and Edwin K Silverman. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respiratory research*, 15(1):89, 2014.

[150] Xiaoyong Wang, Pangyu Teng, Pechin Lo, Ashley Banola, Grace Kim, Fereidoun Abtin, Jonathan Goldin, and Matthew Brown. High throughput lung and lobar segmentation by 2D and 3D CNN on chest CT with diffuse lung disease. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 202–214. Springer, 2018.

[151] Qiao Wei, Yaoping Hu, John H MacGregor, and Gary Gelfand. Automatic recognition of major fissures in human lungs. *International Journal of Computer Assisted Radiology and Surgery*, 7(1):111–123, 2012.

[152] Rafael Wiemker, Thomas Bülow, and Thomas Blaffert. Unsupervised extraction of the pulmonary interlobar fissures from high resolution thoracic CT data. In *International Congress Series*, volume 1281, pages 1121–1126. Elsevier, 2005.

[153] Matthias Wilms, Jan Ehrhardt, and Heinz Handels. A 4D statistical shape model for automated segmentation of lungs with large tumors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 347–354. Springer, 2012.

[154] Nicole M Wink, Christoph Panknin, and Timothy D Solberg. Phase versus amplitude sorting of 4D-CT data. *Journal of Applied Clinical Medical Physics*, 7(1):77–85, 2006.

[155] Guorong Wu, Qian Wang, Jun Lian, and Dinggang Shen. Estimating the 4D respiratory lung motion by spatiotemporal registration and super-resolution image reconstruction. *Medical Physics*, 40(3):031710, 2013.

[156] Changyan Xiao, Marius Staring, Juan Wang, Denis P Shamonin, and Berend C Stoel. A derivative of stick filter for pulmonary fissure detection in CT images. In *SPIE Medical Imaging*, pages 86690V–86690V. International Society for Optics and Photonics, 2013.

[157] Changyan Xiao, Berend C Stoel, M Els Bakker, Yuanyuan Peng, Jan Stolk, and Marius Staring. Pulmonary fissure detection in CT images using a derivative of stick filter. *IEEE Transactions on Medical Imaging*, 35(6):1488–1500, 2016.

[158] Yi Xin, Gang Song, Maurizio Cereda, Stephen Kadlecek, Hooman Hamedani, Yunqing Jiang, Jennia Rajaei, Justin Clapp, Harrilla Profka, Natalie Meeder, et al. Semiautomatic segmentation of longitudinal computed tomography images in a rat model of lung injury by surfactant depletion. *Journal of Applied Physiology*, 118(3):377–385, 2014.

[159] Sarah Gerard Yeary, Gary E. Christensen, John E. Bayouth, Sandeep Bodduluri, Yue Pan, Junfeng Guo, Kaifang Du, Joo H. Song, Bowen Zhao, Ipek Oguz, and Joseph M. Reinhardt. 4D lung CT segmentation for radiation therapy applications. In *ICART: Imaging and Computer Assistance in Radiation Therapy*, pages 50–57, 2015.

[160] Yin Yin, Xiangmin Zhang, Rachel Williams, Xiaodong Wu, Donald D. Anderson, and Milan Sonka. LOGISMOS-layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint. *IEEE Transactions on Medical Imaging*, 29, 2010.

[161] Li Zhang, Eric Hoffman, and Joseph M. Reinhardt. Lung lobe segmentation in volumetric X-ray CT images. *IEEE Trans. Medical Imaging*, 25(1):1–16, 2006.

[162] Li Zhang and Joseph M Reinhardt. 3D pulmonary CT image registration with a standard lung atlas. In *Medical Imaging 2000: Physiology and Function from Multidimensional Images*, volume 3978, pages 67–78. International Society for Optics and Photonics, 2000.

[163] Xiangrong Zhou, Tatsuro Hayashi, Takeshi Hara, Hiroshi Fujita, Ryujiro Yokoyama, Takuji Kiryu, and Hiroaki Hoshi. Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images. *Computerized Medical Imaging and Graphics*, 30(5):299–313, 2006.