

University of Louisville

## ThinkIR: The University of Louisville's Institutional Repository

---

Electronic Theses and Dissertations

---

5-2019

### Segmentation and classification of lung nodules from Thoracic CT scans : methods based on dictionary learning and deep convolutional neural networks.

Mohammad Mehdi Farhangi  
*University of Louisville*

Follow this and additional works at: <https://ir.library.louisville.edu/etd>



Part of the [Computer Engineering Commons](#)

---

#### Recommended Citation

Farhangi, Mohammad Mehdi, "Segmentation and classification of lung nodules from Thoracic CT scans : methods based on dictionary learning and deep convolutional neural networks." (2019). *Electronic Theses and Dissertations*. Paper 3301.  
<https://doi.org/10.18297/etd/3301>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

SEGMENTATION AND CLASSIFICATION OF LUNG NODULES FROM  
THORACIC CT SCANS: METHODS BASED ON DICTIONARY LEARNING  
AND DEEP CONVOLUTIONAL NEURAL NETWORKS

By  
Mohammad Mehdi Farhangi

A Dissertation  
Submitted to the Faculty of the  
J.B. Speed School of Engineering of the University of  
Louisville  
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy  
in Computer Science and Engineering

Department of Computer Engineering and Computer  
Science  
University of Louisville  
Louisville, Kentucky

May 2019

Copyright 2019 by Mohammad Mehdi Farhangi

All rights reserved



SEGMENTATION AND CLASSIFICATION OF LUNG NODULES FROM  
THORACIC CT SCANS: METHODS BASED ON DICTIONARY LEARNING  
AND DEEP CONVOLUTIONAL NEURAL NETWORKS

By

Mohammad Mehdi Farhangi

Dissertation approved on

April 23, 2019

by the following dissertation Committee:

---

Dissertation Adviser  
Dr. Amir A. Amini

---

Dissertation Co-adviser  
Dr. Hichem Frigui

---

Dr. Juw Won Park

---

Dr. Nihat Altiparmak

## ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my primary advisor Dr. Amir A. Amini and co-advisor Dr. Hichem Frigui, for their insightful guidance, motivation, and immense knowledge. They have always been patient and supportive in overcoming obstacles I have been facing through my research and writing of this thesis.

I also would like to thank Dr. Albert Seow and his team for analyzing the NLST data; without their help this would not have been possible. I would also like to express my thanks to members of my committee: Dr. Juw Won Park, Dr. Nihat Altiparmak, and Dr. Neal Dunlap for their valuable time and feedback on my thesis. I am also grateful to Dr. Adel Elmaghraby, Chair of CECS Department, who was always supportive during my entire graduate studies.

Last, I should express my gratitude to all other Medical Imaging Lab members at the University of Louisville for their support and for their friendship. I am especially thankful to Benjamin Veasey who developed the software for perparing and pre-processing NLST data and Jungwon Cha who collaborated and helped in different parts of my research.

## ABSTRACT

### SEGMENTATION AND CLASSIFICATION OF LUNG NODULES FROM THORACIC CT SCANS: METHODS BASED ON DICTIONARY LEARNING AND DEEP CONVOLUTIONAL NEURAL NETWORKS

Mohammad Mehdi Farhangi

April 23, 2019

Lung cancer is a leading cause of cancer death in the world. Key to survival of patients is early diagnosis. Studies have demonstrated that screening high risk patients with Low-dose Computed Tomography (CT) is invaluable for reducing morbidity and mortality. Computer Aided Diagnosis (CADx) systems can assist radiologists and care providers in reading and analyzing lung CT images to segment, classify, and keep track of nodules for signs of cancer.

In this thesis, we propose a CADx system for this purpose. To predict lung nodule malignancy, we propose a new deep learning framework that combines Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to learn best in-plane and inter-slice visual features for diagnostic nodule classification. Since a nodule's volumetric growth and shape variation over a period of time may reveal information regarding the malignancy of nodule, separately, a dictionary learning based approach is proposed to segment the nodule's shape at two time points from two scans, one year apart. The output of a CNN classifier trained to learn visual appearance of malignant nodules is then combined with the derived measures of shape change and volumetric growth in assigning a probability of malignancy to the nodule.

Due to the limited number of available CT scans of benign and malignant nodules in the image database from the National Lung Screening Trial (NLST), we chose to initially train a deep neural network on the larger LUNA16 Challenge database which was built for the purpose of eliminating false positives from detected nodules in thoracic CT scans. Discriminative features that were learned in this application were transferred to predict malignancy. The algorithm for segmenting nodule shapes in serial CT scans utilizes a sparse combination of training shapes (SCoTS). This

algorithm captures a sparse representation of a shape in input data through a linear span of previously delineated shapes in a training repository. The model updates shape prior over level set iterations and captures variabilities in shapes by a sparse combination of the training data. The level set evolution is therefore driven by a data term as well as a term capturing valid prior shapes. During evolution, the shape prior influence is adjusted based on shape reconstruction, with the assigned weight determined from the degree of sparsity of the representation. The discriminative nature of sparse representation, affords us the opportunity to compare nodules' variations in consecutive time points and to predict malignancy. Experimental validations of the proposed segmentation algorithm have been demonstrated on 542 3-D lung nodule data from the LIDC-IDRI database which includes radiologist delineated nodule boundaries.

The effectiveness of the proposed deep learning and dictionary learning architectures for malignancy prediction have been demonstrated on CT data from 370 biopsied subjects collected from the NLST database. Each subject in this database had at least two serial CT scans at two separate time points one year apart. The proposed RNN CAD system achieved an ROC Area Under the Curve (AUC) of 0.87, when validated on CT data from nodules at second sequential time point and 0.83 based on dictionary learning method; however, when nodule shape change and appearance were combined, the classifier performance improved to AUC=0.89.



## TABLE OF CONTENTS

Acknowledgments . . . . .	iii
Abstract . . . . .	iv
List of Tables . . . . .	viii
List of Figures . . . . .	ix
INTRODUCTION . . . . .	1
Motivation . . . . .	1
Lung Anatomy . . . . .	2
Segmentation . . . . .	2
Nodule Detection . . . . .	3
Thesis Objectives . . . . .	4
Thesis Organization . . . . .	5
LITERATURE REVIEW ON LUNG NODULE SEGMENTATION . . . . .	7
Literature Review on Lung Nodule Segmentation . . . . .	7
Active Shape Models . . . . .	13
Sparse Shape Composition . . . . .	15
SCOTS: SPARSE LINEAR COMBINATION OF TRAINING SHAPES . . . . .	16
Level Set Segmentation . . . . .	17
Shape Prior Modeling . . . . .	20
Shape prior weighting . . . . .	22
Segmentation algorithm . . . . .	25
Convergence and Complexity . . . . .	30
SEGMENTATION RESULTS . . . . .	31
Dataset . . . . .	31
Evaluation . . . . .	32

Parameters Settings . . . . .	33
Results . . . . .	34
DEEP LEARNING IN MEDICAL IMAGE ANALYSIS . . . . .	45
Deep Learning . . . . .	46
Deep Learning for Detection . . . . .	48
Deep learning for segmentation . . . . .	49
Deep Learning for Computer-Aided Diagnosis . . . . .	50
LUNG NODULE MALIGNANCY PREDICTION BASED ON SINGLE TIME POINT AND SEQUENTIAL TIME POINT CT SCANS . . . . .	52
Recurrent convolutional networks to detect lung nodules based on single time point CT scans . . . . .	55
Use of shape, and volume growth in two consecutive CT Scans . . . . .	61
EXPERIMENTAL RESULTS FOR MALIGNANCY PREDICTION FROM NLST DATA SETS . . . . .	64
Datasets . . . . .	64
Performance Metrics for Single and Consecutive CT Scan Classifiers . . . . .	65
CONCLUSIONS AND FUTURE WORK . . . . .	76
REFERENCES . . . . .	79
Appendix A: Commonly Used Acronyms . . . . .	91
CURRICULUM VITA . . . . .	93

## LIST OF TABLES

1	The algorithm parameters values used for all experiments . . . . .	34
2	Numerical validation of the algorithm on LIDC-IDRI database . . . . .	38
3	Inter-observer variability among four radiologists . . . . .	41
4	Numerical validation of the algorithm, using STAPLE as the ground truth	41
5	Performance of SCoTS on specific nodule types. . . . .	42
6	Comparison with state of the art algorithms . . . . .	43

## LIST OF FIGURES

1	Anatomy of Lung . . . . .	3
2	Classification of lung nodules based on attachment . . . . .	8
3	Variety of nodule shapes . . . . .	12
4	Landmark based shape representation . . . . .	13
5	Representing a nodule shape by SDF . . . . .	19
6	Encoding two shapes by nodule dictionary . . . . .	23
7	Shape prior weighting function . . . . .	25
8	Level set evolution update . . . . .	28
9	Surface evolution in segmentation process . . . . .	35
10	Evolution of sparse representation and energy . . . . .	37
11	Result of the algorithm on samples of the dataset . . . . .	39
12	Result of segmentation using different dictionaries . . . . .	40
13	Perceptron . . . . .	45
14	Annotation variability among readers . . . . .	53
15	Transfer learning for malignancy prediction . . . . .	54
16	Architecture of the proposed deep learning . . . . .	57
17	Diagram of recurrent units . . . . .	59
18	Malignancy prediction from two time points . . . . .	63
19	Nodule changes in two time points . . . . .	66
20	Segmentation Results on NLST database . . . . .	67
21	Performace on first time point . . . . .	70
22	Performance on second time point . . . . .	71
23	performance based on two time points . . . . .	72
24	Prediction correlation . . . . .	73
25	prediction correlation . . . . .	74
26	Overall perfomance of the CAD system . . . . .	75

# CHAPTER I

## INTRODUCTION

### 1 Motivation

Fatality caused by lung cancer has significant proportions; it is estimated that world-wide 1.1 million people die of lung cancer each year [65]. One reason is because lung cancer is not diagnosed in its early stages; indeed in most cases it is diagnosed after it has already spread through the body. Unfortunately, treatments are not very successful at that point. Screening high risk patients with low-dose Computed Tomography (CT) has shown significant reduction of lung cancer mortality rate [61]. The goal of CT screening is to detect cancer in early stages when there are more options to treat the patient.

As a consequence of the benefits provided by low-dose CT imaging, it is being implemented in large scale in the US. However, this extensive usage results in significant increase of reading efforts for the radiologists. Radiologists mostly read and investigate volumetric CT scans to detect small abnormal lesions including pulmonary nodules. Pulmonary nodules are radiologically visible small structures that are roughly spherical [87]. It turns out that these small structured lesions are crucial indicators of lung cancer, and through their size, appearance, and shape of great use for cancer malignancy diagnosis [56]. Based on detected nodules, radiologists and

surgeons can measure the size and characteristics of nodule, , perform biopsies, and surgical intervention in necessary cases to improve the survival rate.

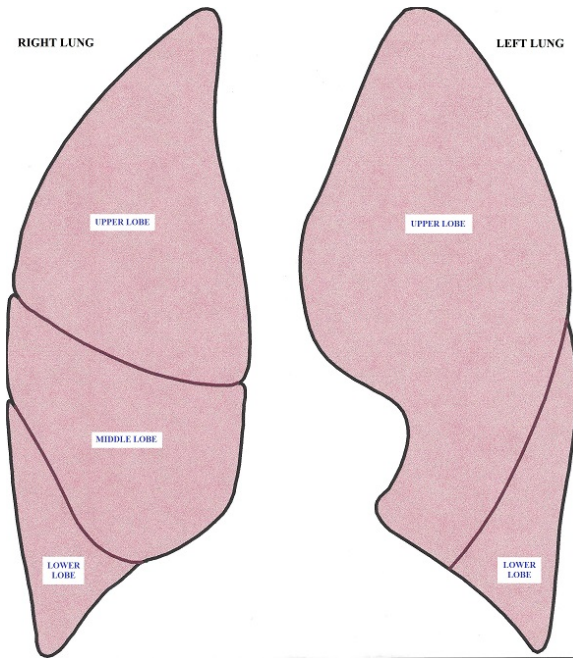
## 2 Lung Anatomy

Lungs are a pair of spongy organs located on both side of the chest. When interpreting the lung CT scans, it is important to have a solid understanding of lung structure. A brief overview of lung anatomy is presented here. Lungs are covered by a tissue layer called pleural, a thin layer of fluid plays as lubricant helping the lung to move smoothly over the exhalation and inhalation.

Each lung can be divided into lobes as shown in Figure 1. Each lobe contains its own separate vascular and lymphatic networks. The right lung is larger because heart is accommodated in the left lung. The left lung divides into upper and lower lobe with a horizontal fissure. The right lung on the other hand, has two fissures which results in three lobes (upper, middle, and lower) lobes.

## 3 Segmentation

Segmentation is a fundamental step of most CAD systems. By segmenting the nodule, the shape characteristics can be extracted, helping the radiologists analyze the malignancy as malignant nodules are more speculated. It also provides information on nodule growth over a period of time. However, segmentation is a challenging problem in lung CT due to the confounding factor that nodules can be attached to the pleural surface with the same Hounsfield Units (HU) or they might have significant overlap with neighboring vessels. From this point of view, Kostis et al. [46] divided



**Figure 1.** Anatomy of Lung [2].

lung nodules into four different categories named well-circumscribed, juxta-vascular, juxta-pleural, and pleural tail nodules. The last three cases cause difficulties for density based segmentation algorithms because for these cases the nodules have HU very similar to those in the vascular and pleural space.

#### 4 Nodule Detection

To diagnose lung cancer, it is important to detect and interpret the lung nodules. Fortunately, Low dose CT affords a significant improvement to lung nodule detection in patients in comparison to chest X-ray so that 20% reduction in mortality is achieved with low-dose CT scans [61, 62]. However, false positives remains high

with this modality and a post processing step is needed to reduce false positives. A nodule detector system typically consists of two steps: 1) candidate detection and 2) false positive reduction.

In candidate detection, a large number of candidates are extracted from the whole volumetric lung CT. The aim of this step to include all the nodules in the large set of candidates using a variety of characteristics like intensity, shape features, morphology, etc. This step detect nodules with a very high sensitivity without forcing the system to keep the false positives at a low rate. Reducing the number of false positives is typically carried out in the second stage where an effective classifier in conjunction with discriminative features serve to remove false positives and provide the output with a high sensitivity at a low false positive per scan.

## 5 Thesis Objectives

The need to go through large number of high resolution CT scans with many slices makes the reading and interpretation of images an arduous task - a tedious process for radiologists. Computer-Aided Detection (CAD) systems have come to assist the radiologists in the reading and interpretation by detecting, segmenting, and keeping track of nodule changes in CT scans and making the process more efficient.

The main objective of this thesis is to develop a CAD system to segment and classify lung nodules in lung CT scans. For nodule segmentation, a general framework is presented which considers shape variability of nodules and brings this information into the segmentation process. We capture the best shape by approximating the evolving surface of a level set model through a linear combination of training shapes



in a subspace, resulting in a sparse representation of nodule shapes. Sparse encoding provides a tool to validate the shape approximation and force the shape from growing once it reaches the boundary of the nodule.

In nodule classification, we propose to classify nodules as benign and malignant from single time point CT scans, from a pair of consecutive scans, and from a combination of the two. With the remarkable successes of convolutional neural networks (CNNs) in computer vision [47, 67], and medical imaging [52], for nodule classification from single time point scans, we represent nodules with high level features learned from training samples by using a deep convolutional neural network. Deep learning approaches afford us to automatically extract the optimal features for the application of lung nodule classification without the need to investigate and extract hand crafted features. A patch surrounding the nodule is separated and multiple layers including convolutional and max pooling layers are applied to the patches to both find and extract the best features to distinguish between benign and malignant nodules. When considering a pair of CT scans acquired over for example a one year period, we would like to determine a quantitative metric of similarity of the two shapes. To this end we represent both nodule through a sparse representation of nodule shapes in two time points and propose a similarity metric from which a malignancy score is derived.

## 6 Thesis Organization

The thesis is organized in the following order:

- Chapter I presents the introduction on the thesis and its objective.

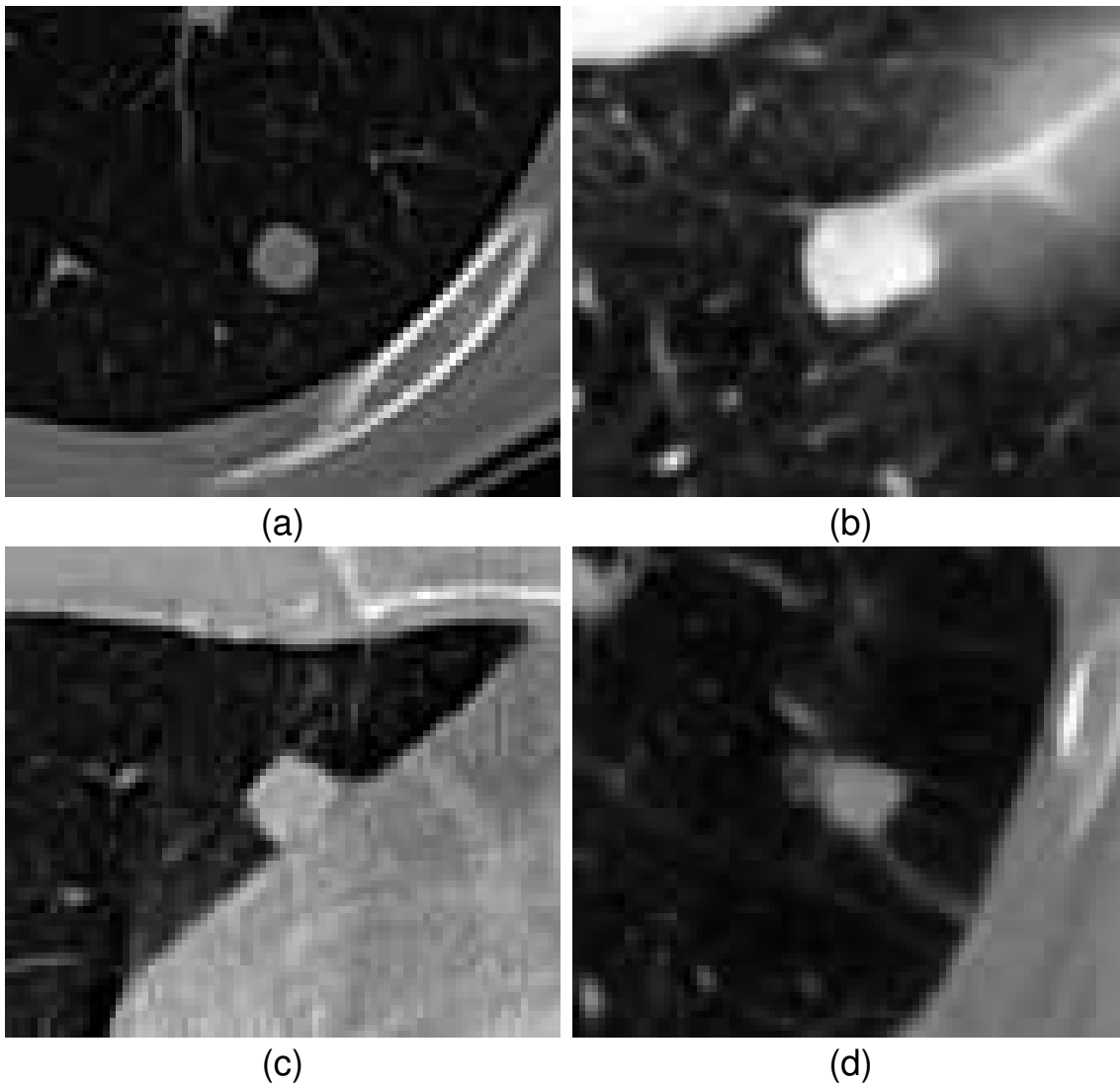
- Chapter II reviews the literature on lung nodule segmentation of CAD systems.
- Chapter III illustrates the proposed method for lung nodule segmentation.
- Chapter IV presents experimental results and evaluation of segmentation on a public database.
- Chapter V reviews the literature on deep learning and its application in medical imaging.
- Chapter VI presents details of our method for malignancy prediction from sequential CT scans.
- Chapter VII reports the experimental results of malignancy prediction evaluated on the NLST database.
- Chapter VIII concludes the thesis and proposes future works.

## CHAPTER II

### LITERATURE REVIEW ON LUNG NODULE SEGMENTATION

#### 1 Literature Review on Lung Nodule Segmentation

Segmentation of nodules is a fundamental step in every CAD system designed to detect lung cancer. Several attempts have been made to automatically segment nodule boundaries from X-ray computed tomography (CT) scans. But, segmentation is challenging. The challenges mostly come from low contrast and noise. Further confounding the task is that nodules can have a variety of shapes and may be attached to the pleural surface with the same Hounsfield Unit (HU). They may also have significant connections to neighboring vessels. From this point of view, Kostis et al. [46] divided lung nodules into four different categories named well-circumscribed, juxta-vascular, juxta-pleural, and pleural tail nodules. The last three cases pose difficulties for intensity based segmentation algorithms because nodules have HU very similar to those in the vascular and pleural space. Examples of these four categories are shown in Figure 2. In the literature, the challenges have been addressed utilizing shape prior information and refinement [68, 60, 31, 95, 22, 101, 97]. The purpose of shape refinement being to keep the segmented area similar to expected shapes of nodules that are available.



**Figure 2.** Classification of lung nodules based on attachment. (a): Well-circumscribed, (b): Juxta-vascular, (c): Juxta-pleural, (d): Pleural tail

To deal with nodules having different shape and HU characteristics, previously proposed methods restrict the segmentation to specific nodule types since shape refinement can then be done more accurately [59]. For example, [46, 25] proposed to segment only small solid nodules with homogeneous and solid texture whereas authors in [11] proposed an algorithm to deal with juxta-vascular nodules. In [58], the authors restricted the method to juxta-pleural nodules and in [49], a method was proposed to segment non-solid and part-solid nodules from the rest of the CT scan.

In addition, some studies proposed preprocessing and post processing techniques to exclude vessels and the pleural surface from the final segmentation result; in [48], the authors proposed to perform a rough segmentation of juxta-pleural nodules, including the pleural surface in the segmented result. Considering the fact that lungs are mostly convex except for the diaphragm and the cardiac region, the only cavity in the rough segmentation is separated using convex hull operation as the nodule boundary. Threshold-based region growing approach was followed by two post processing steps in [49] to detach pleural surface and vessels from non-solid and part-solid nodules. In [5], a multilevel approach consisting of Otsu thresholding, region growing, fuzzy connectivity analysis, morphological operations, and thickening was proposed to segment various types of pulmonary nodules. In [8], another multilevel approach was proposed where segmentation was applied to each nodule slices individually. The authors modeled nodules change in consecutive slices by motion estimation assuming the differences between nodule in slices are small. They subtracted lesion slice from background slice followed by Otsu thresholding and morphological operations to detach connected organs. Finally, to segment juxta-vascular nodules, [59] extracted the

vessels from an initial segmentation by taking advantage of anatomical information where vessels occupy a narrow region of the volume. By defining a threshold on the percentage of segmented voxels in a fixed cubic size, vessels were separated from the nodule.

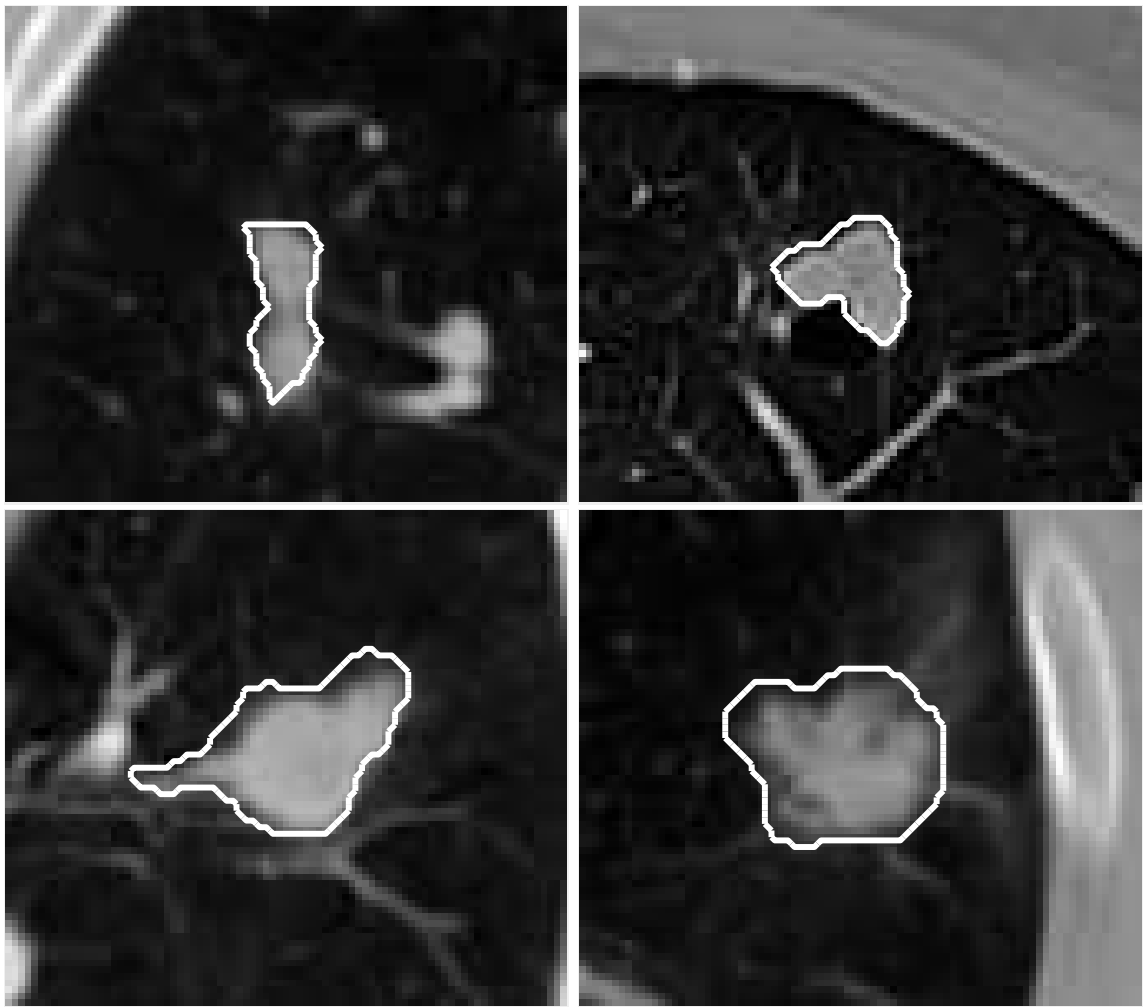
Reeves, *et al.* [68] utilized adaptive thresholding, in which the threshold is computed separately for each scan to compensate for the variations between two consecutive scans. The midpoint of the nodule and parenchyma’s density was chosen as the threshold. They applied geometrical constraints to keep the segmented nodule in a spherical shape while removing vessels from the final segmentation. Dehmeshki, *et al.* [22] proposed sphericity contrast based region growing, in which each pixel of the boundary was added to the segmented region according to their intensity contrast and distance to the center of the region. While the latter metric followed the spherical shape assumption, another shape constraint was applied to stop the segmentation if the size was greater than a predefined threshold. In [91], region growing was also used where the decision of neighbors to be included in the nodule area was carried out by machine learning methods. For each voxel, a feature vector was extracted and fed to a classifier to predict its label based on a linear or non-linear classifier trained over pre-segmented nodules.

Way, *et al.* [95] proposed to segment 3D lung nodules by a snake based algorithm [43]. They extended gradient, energy and curvature to 3D images, and defined a prior mask energy which penalized the curves that grow beyond the pleural surface. In another attempt to incorporate active contours in this application, Farag, *et al.* [31] proposed a 2D variational approach to model density of nodules as nonparametric

Gaussian distributions favoring elliptical shapes. Graph-cut based segmentation is another successful approach to segment nodules in the lung. Ye, *et al.* [101] used density, shape and spatial features to define an energy function for segmentation. The graph cut algorithm was applied to the over-segmented image previously obtained from unsupervised clustering of image features forcing a spherical shape prior. Another attempt in this area was by Cha, *et al.* [97] where a sequence of graph-cuts iterations were performed, each of which with different unary potentials computed from both intensity values and a shape prior. Shape prior in each iteration changes to adapt to the segmented volume in the current iteration.

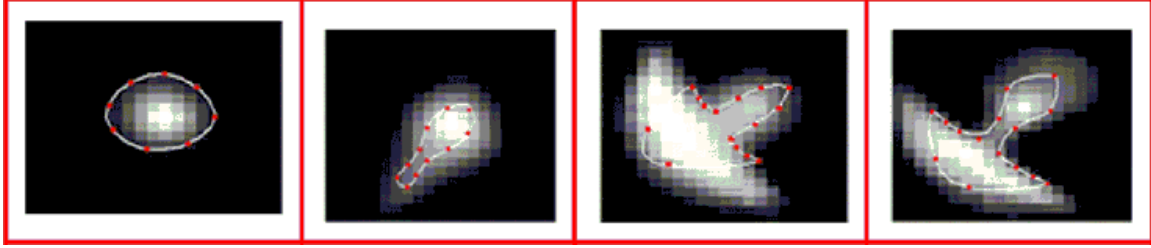
As it can be seen, many of the previously proposed methods assume that the nodule has spherical or ellipsoidal shape that might be attached to vessel or pleura with similar densities. However, as may be seen in Fig. 3, this assumption is unrealistic in general as the nodules can have complicated structures. The ellipsoidal shape assumption is a limitation, resulting in the segmented surface to become overly rounded, preventing capture of fine structures.

There are methods to incorporate shape prior without restricting the shape to predefined ellipsoid or rounded shapes. In the next section we explore these methods which mostly rely on a training sets consisting of shapes that resemble the segmenting target object. Two of the frequently used methods in this area are Active Shape Model (ASM) [19] and Sparse Shape Composition (SSC) [104].



**Figure 3.** Nodules appear with different shapes in the lung





**Figure 4.** Encoding nodule shapes by landmarks. Taken from [30].

## 2 Active Shape Models

Active shape models are popular methods in medical image processing to incorporate shape information into anatomical segmentation of objects from Magnetic Resonance (MR) brain images [29, 92], lung CT [92, 96], and liver CT [96]. In its classical form, ASM finds object boundaries using Principal Component Analysis (PCA) with boundary of training shapes given as a set of landmark coordinates. Shapes are represented by a set of landmarks delineating the boundary of the object. An example can be seen in Figure 4. Coordinates of the landmarks are vectorized and stored in a 1-D matrix. The training set is then comprised of all the landmark coordinates stored in this format. ASM then extracts statistical information from the training set. Mean vector shows the average of landmarks in 2-D space, and eigenvectors computed from PCA analysis shows the direction of significant change in the training set. Having mean vector and most significant eigenvectors the direction of variations are computed and any segmented region can then be refined using these information.

Refining shapes using ASM has found many applications in segmentation algorithms. As some examples, in [12] an iterative graph-cut approach was proposed

where the initial segmentation provided by graph-cut is refined using extracted statistical information by ASM. The authors validated the feasibility of applying their method to segmenting organ in 3-D CT and MR images.

Other variations of ASM have also been used in graph-cut segmentation framework. For example, Grosgeorge et al. [37] used ASM analysis to extract shape information and refine a rough segmented area that was generated by graph-cut segmentation. Shapes in their approach have been represented by the Signed Distance Functions (SDF) rather than landmarks. They managed to accurately segment the right ventricle (RV) in MR images and overcome the challenges with low contrast images and tissues attached to the organ. Cha et al. [97, 96, 9] also utilized ASM analysis to iteratively refine the segmented shapes of lung, liver, and lung nodules that outlined using signed distance function.

Another framework which the active shape model has found success in level set based segmentation. In [50] principal component analysis was performed on a set of signed distance functions (SDF) and increased the accuracy of geodesic active contour [6] in segmenting the corpus callusom from MR images. Delineating shapes by SDF, which is considered as an implicit representation, has advantages over explicit ones like landmark based representation [20]. For example, landmark based representations do not permit topological changes of the evolving boundary because this representation is restricted to the labeling of space and making the splitting and merging of curves very complicated if not impossible. In addition, landmark based representations cannot be extended to 3-D space in a straightforward manner and require tedious human effort to find the correspondences between the landmarks in

training shapes. With the same shape representation, Tsai et al. [89] added scale, rotation and translation to the parameter of curve representation. They tested the method on segmenting left ventricle from cardiac MR images as well as prostate segmentation.

### 3 Sparse Shape Composition

Another approach proposed by Zhang et al. [104], represents new shapes by a sparse combination of training shapes. In their approach, shapes were modeled by landmarks and sparse representation of testing shapes provides them with the means to capture small details even when these variations were present in a small fraction the of training set. These small details are usually ignored and missed when new shapes are modeled by statistical methods like ASM. Sparse representation also provides robustness to false appearance information such as when landmarks are misplaced around the object boundary. Their model was applied to two segmentation problems; liver segmentation in PET-CT images and rodent brain segmentation in MR images [103]. Xing and Yang [100] embedded SSC inside a snake deformable model [43] to segment malignant lung nodules.

## CHAPTER III

### SCOTS: SPARSE LINEAR COMBINATION OF TRAINING SHAPES

In this chapter, a new segmentation algorithm is proposed which can deal with the aforementioned challenges in a unified framework, permitting application of the same method to different nodule categories [32, 33]. We consider shape variability of nodules and bring this information to the segmentation process. In contrast to previously proposed approaches, we have an adaptive model of the shape which dynamically contributes to the segmentation during surface evolution. Nodule shapes in our method are not restricted to a predefined structure; instead, we capture the best shape model by approximating the evolving surface by a linear combination of training shapes in a subspace, resulting in a sparse representation of nodule shapes. Sparse representation provides us with the means to neglect the nodules in the training set that do not contribute to a linear approximation of an input shape. It also affords us the opportunity to recover local details even if these details are only present in a small fraction of training shapes. In addition, sparsity can be viewed as validation for segmentation. In other words, when the segmenting surface can be reconstructed by a sparse linear combination of training shapes, it is likely that the algorithm has reached the true boundary of the nodule. In that case, the contribution of the shape

prior should be increased in order to stop the surface from evolving into surrounding tissues.

Sparse representations have already found success in other applications such as face recognition [98] and snake based segmentation [104]. In this chapter we propose an algorithm that incorporates sparse shape representation into a level set segmentation framework. For this purpose, the region based active contour algorithm was extended to 3D, permitting shape information to be incorporated by 3D Signed Distance Functions (SDF) delineating 3D shapes.

## 1 Level Set Segmentation

Although any curve evolution algorithm could be used within the proposed framework, we have adapted the Chan-Vese algorithm [10] which is a region-based active contour. The advantages of this method over the edge based realization of active contours [6] in lung nodule segmentation application are described here:

*First*, the Chan-Vese segmentation is more robust to noise and blurry edges; a situation that is common in lung nodule CT images. Thus, it can handle non-solid nodules better; these nodules suffer from a lack of sharp boundaries.

*Second*, the Chan-Vese algorithm is less sensitive to contour position compared to edge based active contours. Specifically, edge based active contours in their classical form grow or shrink until they reach the edges, but the contour keeps growing or shrinking if they miss an edge. In contrast, contours that are driven by Chan-Vese energy functional can return back and modify themselves in case they pass through the nodule boundary. This property makes region based active contours

more suitable for the present application since the contribution of shape prior, might slightly mislead the contour during the evolution.

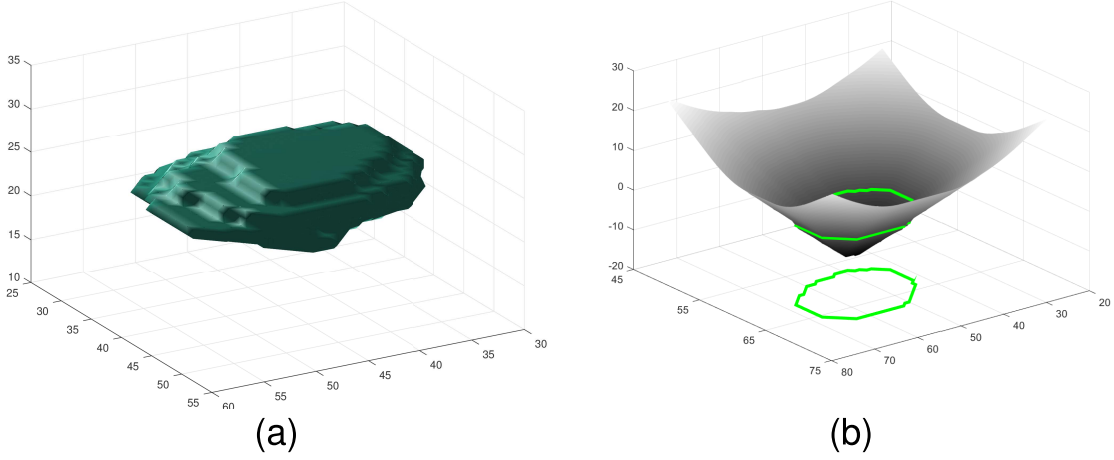
Our method is modeled to deform a segmenting surface by a force based on the Chan-Vese image intensity and shape prior information until the deformable surface stops in a location that separates two homogeneous regions and best approximates shape prior.

Let  $I$  be the volumetric image and  $\Omega$  the domain of  $I$ .  $C \in \Omega$  is a surface that separates image volume into two segmented regions, and  $x \in R^3$  denotes an arbitrary point in the volume  $I$ . To find the desired segmenting surface  $C$ , we minimize the following functional:

$$\begin{aligned}
E_{cv}(C) = & k_1 SurfaceArea(C) + k_2 Volume(inside(C)) \\
& + k_3 \int_{inside(C)} (I - \mu)^2 dx \\
& + k_4 \int_{outside(C)} (I - \nu)^2 dx.
\end{aligned} \tag{1}$$

In this equation,  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$  are weighting parameters and  $\mu$  and  $\nu$  represent the average intensity values inside and outside of the surface  $C$  respectively. The first two terms, defined on the area and volume of the surface, control the smoothness and the volume of the separating surface. The last two terms are external energies which help separate two homogeneous regions in the image.

To utilize the advantages of implicit representation, the surface  $C$  may be represented as the zero level set of Lipschitz function  $\phi$  [64].  $\phi$  is a Signed Distance Function (SDF) which encodes the distance of every point  $x \in R^3$  in the image to the boundary.  $\phi(x)$  is positive if the point  $x$  lies outside the surface, and is negative



**Figure 5.** Representing a nodule shape by a Signed Distance Function (SDF). Left: zero level set of the SDF, Right: the SDF of a mid slice of the nodule shape, superimposed with nodule boundary shown in green on the surface and also projected onto the x-y plane. For any point, SDF encodes the closest distance to the boundary. For points inside the boundary  $SDF < 0$  and for points outside of the boundary  $SDF > 0$ .

if  $x$  is inside the surface. Locations where  $\phi$  crosses zero represent the bounding surface. Fig. 5 shows construction of the SDF for an arbitrary nodule shape. Fig. 5(a) shows the nodule and Fig. 5(b) displays the SDF of mid slice of this nodule.

With embedding the surface  $C$  inside the zero level set of the SDF  $\phi$ , equation (1) may be written as:

$$\begin{aligned}
 E_{cv}(\phi) = & k_1 \int_{\Omega} | \nabla H(\phi(x)) | dx \\
 & + k_2 \int_{\Omega} H(\phi(x)) dx \\
 & + k_3 \int_{\Omega} (I - \mu)^2 H(\phi(x)) dx \\
 & + k_4 \int_{\Omega} (I - \nu)^2 (1 - H(\phi(x))) dx,
 \end{aligned} \tag{2}$$

where  $\nabla$  is the gradient operator and  $H$  is the Heaviside function defined as:

$$H(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases}.$$

To minimize  $E_{cv}$ , the Euler-Lagrange equation is derived and  $\phi$  is iteratively updated:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)[k_1 \text{div}(\frac{\nabla \phi}{|\nabla \phi|}) - k_2 - k_3(I - \mu)^2 + k_4(I - \nu)^2]. \quad (3)$$

In (3), the symbol *div* refers to the divergence operator and  $\text{div}(\frac{\nabla \phi}{|\nabla \phi|})$  computes the curvature of the isosurfaces embedded in  $\phi$ . Further details about these and the numerical implementations can be found in [10]. To simplify subsequent descriptions, (3) is rewritten as:

$$\phi(t+1) = \phi(t) + V_{cv}(t), \quad (4)$$

where  $V_{cv}(t)$  is the product of the right hand side of equation (3) and the time step, and is the amount of update for each iteration.

## 2 Shape Prior Modeling

One challenge in shape based segmentation is the alignment problem; to measure shape variations it is essential to compare like parts of shapes. Leventon, *et al.* [50] showed that SDFs are robust to slight misalignment helping avoid exact registration. Thus, in the proposed algorithm, prior to performing segmentation, shapes are roughly aligned by computing the center of SDFs; subsequently applying an intrinsic alignment method as proposed by Cremer, *et al.* [21]. The center of any input shape



$\phi$  is computed from:

$$\mu_\phi = \int_{\Omega} x h(\phi(x)) dx, \quad h(\phi(x)) = \frac{H(\phi(x))}{\int_{\Omega} H(\phi(x)) dx}. \quad (5)$$

The new shape  $\phi$  is then translated so that its center is aligned with the center of a selected reference shape  $\phi_0$ .

To start building the shape prior, all registered SDFs ( $\varphi_i; i = 1, \dots, n$ ) are vectorized and stored in columns (bases) of a matrix  $D = [\varphi_1 | \varphi_2 | \dots | \varphi_n] \in R^{m \times n}$  referred to as the dictionary with each column called an atom. Here,  $m$  is the number of voxels in each training image and  $n$  is the size of the training shapes. New shapes are modeled by a linear subspace of dictionary  $D$ . Specifically, having the dictionary  $D$ , any new SDF  $\phi \in R^m$  may be approximated as

$$\phi \simeq \tilde{\phi}(\mathbf{x}) = D\mathbf{x}, \quad (6)$$

where  $\mathbf{x} \in R^n$  is a weight vector which determines the contribution of each training shape in modeling the new shape  $\phi$ . Linear combinations of SDFs in (6) do not necessarily result in a valid SDF representation [50]. However, since the linear combination will have positive and negative values, it encodes a shape at its zero level set. Starting from the zero level set, the encoded shape is then re-initialized to generate an SDF. The vector  $\mathbf{x}$  in (6) is chosen to minimize the error between the input shape  $\phi$  and its approximation; i.e.,

$$\underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\phi - D\mathbf{x}\|_2^2. \quad (7)$$

Including all the atoms in the linear approximation in (7) will result in departure from the valid shape space. Thus, we seek to represent the nodule boundary

compactly by an appropriate combination of training shapes, neglecting shapes from irrelevant nodule types. To achieve this, a constraint is included to limit the number of atoms that contribute to the linear approximation. More formally, a constraint is added to (7) to minimize:

$$\begin{aligned} & \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\phi - D\mathbf{x}\|_2^2 \\ & \text{subject to } \|\mathbf{x}\|_0 \leq k. \end{aligned} \tag{8}$$

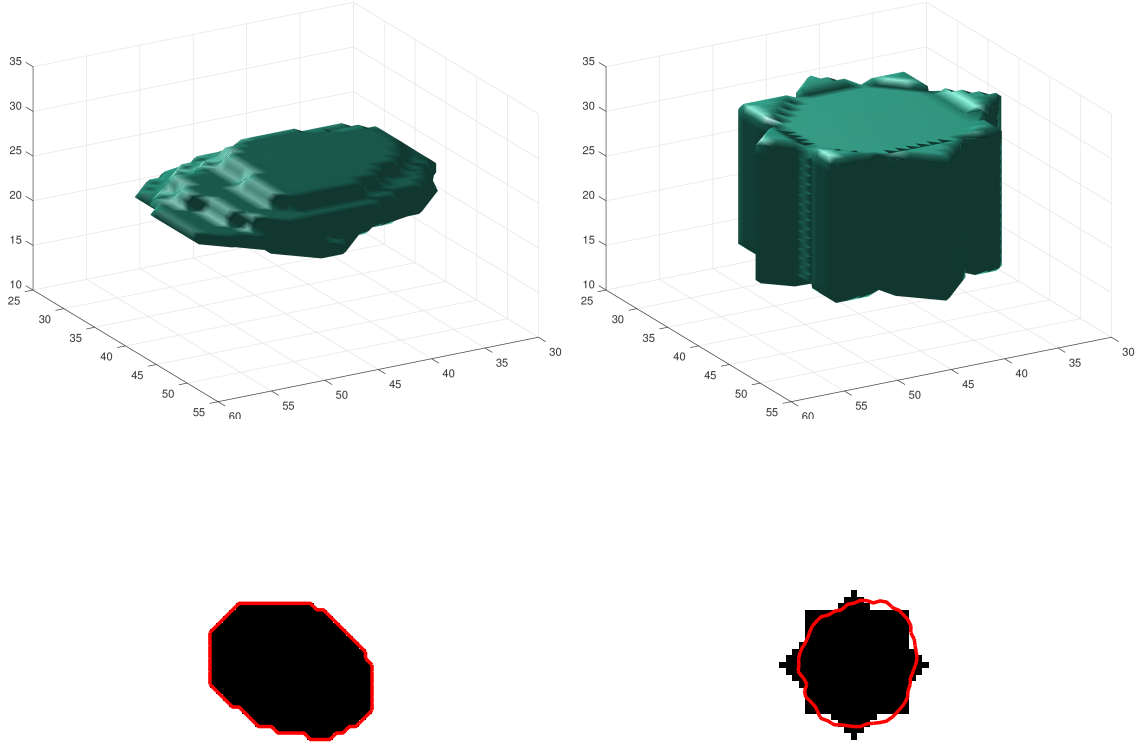
Here,  $\|\mathbf{x}\|_0$  denotes the  $l_0$  norm, which counts the number of non-zero elements in vector  $\mathbf{x}$ , and the parameter  $k$  controls the sparsity of  $\mathbf{x}$ . This problem is NP-hard in nature and cannot be solved efficiently. However, recent papers have shown that if the solution is sparse enough, the  $l_0$  constraint can be replaced by the  $l_1$  norm [28]. Adding the constraint to the objective function with a penalty term  $\lambda$ , equation (8) is rewritten as

$$\underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\phi - D\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1; \lambda \geq 0. \tag{9}$$

Unlike (8), the optimization problem in (9) is convex and can be solved efficiently.

### 3 Shape prior weighting

Another interpretation for (8) is that  $\phi$  is a noisy or irregular shape which needs to be reconstructed from the bases in a shape dictionary. A sparse coding based on the dictionary gives us a low dimensional representation of  $\phi$ , removing irregularities from it. However, if the irregularities are dense, or the input shape is far from the atoms in the dictionary, for a fixed non-zero value of  $\lambda$  in (9),  $\phi$  cannot be sparsely represented. In other words, the sparse weight vector  $\mathbf{x}$  emerging from the dictionary



**Figure 6.** *Top left:* the zero level set of an SDF corresponding to an actual nodule boundary. *Top right:* the zero level set of an SDF with a structure significantly different from nodule boundaries. *Bottom left:* the mid plane of SDF corresponding to the actual nodule boundary with the best  $D\mathbf{x}$  superimposed. *Bottom right:* the mid plane of the arbitrary SDF together with the best  $D\mathbf{x}$  superimposed (see text).

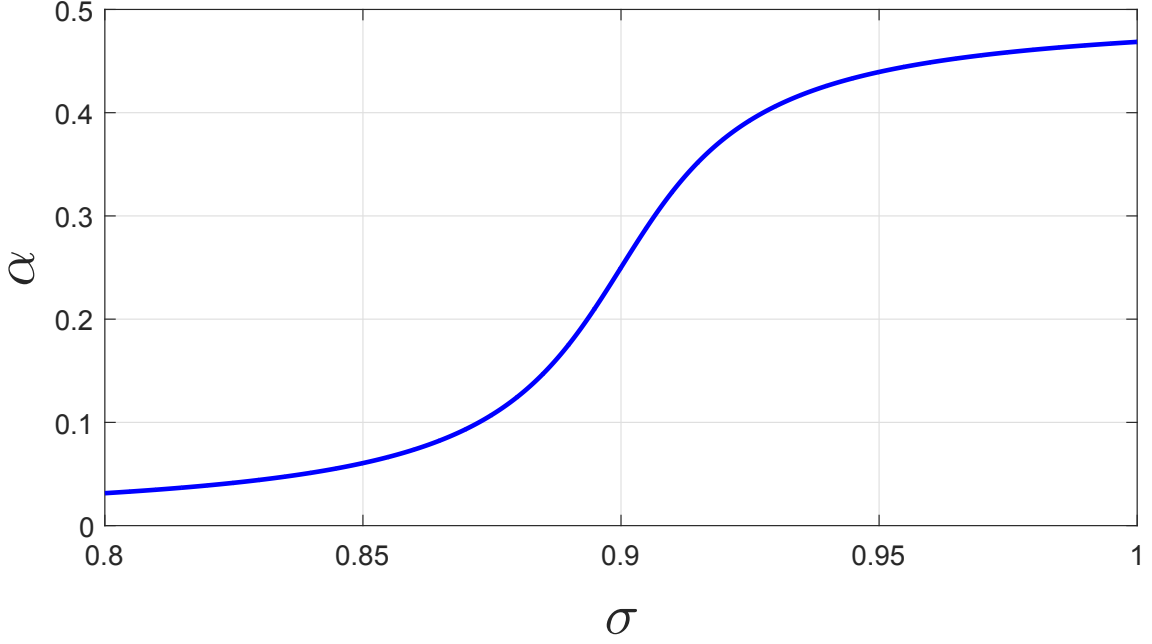
is sufficient to construct the shape prior as well as to decide the level of contribution of shape prior in segmentation.

To better illustrate this, we investigate two cases: the first one has a boundary with an SDF that is already provided in the dictionary (top left shape in Fig. 6).

The second case is when the new shape is significantly different from all atoms in the dictionary (top right shape in Fig. 6). For the first case, by solving (9), the reconstructed shape is highly sparse, resulting from just one non-zero value in vector  $\mathbf{x}$ , an actual shape in the dictionary. The middle slice of the reconstructed shape is shown in the bottom left of Fig. 6 with the contour accurately delineating the input shape. On the other hand, when the input shape is not similar to dictionary atoms as in the top right of Fig. 6, a significant number of shapes in the dictionary need to participate in the linear approximation in order to form a structure similar to the input shape. The reconstructed shape, shown in bottom right of Fig. 6, is not sparse in this case – this reconstruction required more than 80% of the dictionary atoms in the linear approximation for the same value of  $\lambda$  in (9). Despite the large number of atoms used, the input shape cannot be accurately captured by the dictionary.

To take advantage of sparsity, the shape prior weighting is made related to the level of sparsity of vector  $\mathbf{x}$  in (9); when the surface is evolving and is not close to a shape in the dictionary, it is primarily driven by the low level Chan-Vese energy function. Once the surface starts to form a nodule boundary, the sparsity increases and correspondingly, relative to Chan-Vese energy, the weighting for the shape prior is automatically increased. This prevents the evolving surface from leaking inside neighboring organs which have a similar range of HU.

In order to map the sparsity to shape prior influence, a metric is defined so that the more sparse the representation the closer its value is to 1:  $\sigma(t) = e^{-s(t)}$ , where  $s(t)$  is the sparsity ratio of  $\mathbf{x}$  at iteration index  $t$ , i.e, the number of non-zero elements in vector  $\mathbf{x}$  divided by the length of the vector. The shape prior weighting plotted



**Figure 7.** Shape prior weighting,  $\alpha$ , plotted against  $\sigma$  (see equation (10))

in Fig. 7 is as follows,

$$\alpha(\sigma) = \frac{\tan^{-1}(50(\sigma(t) - 0.9)) + \pi/2}{2\pi}. \quad (10)$$

The saddle point of  $\tan^{-1}$  is set to  $\sigma = 0.9$  around which the shape prior weighting changes significantly.

#### 4 Segmentation algorithm

This section provides a framework to guide the segmenting surface to separate two homogeneous regions while forcing the segmentation to be consistent with the training shapes. For this purpose, we add another term to the surface evolution in (3). The new term alongside with original active contour equation helps guide the sur-

face not only by low level intensity statistics, but also by high level shape prior information. Specifically, in each step of surface evolution, it moves in a direction to optimize the Chan-Vese energy function and concurrently projects the segmenting surface into a linear subspace of training shapes. The  $l_1$  constraint in (9) provides us the appropriate subspace.

The first term of (9) is both convex and differentiable. The second term is convex, but not differentiable. Thus, unfortunately its gradient is not well defined. For an optimization problem with this structure, a class of algorithms called coordinate-wise optimization converges to the global optimum [90]. Here, we adapt an algorithm of this class, called shooting algorithm [34], to work in conjunction with implicit surface evolution in our segmentation framework. At an abstract level, the algorithm works by fixing all the entries of solution vector  $\mathbf{x}$  except one, and optimizes the objective function along that dimension. That is, in each step, the algorithm moves along a specific dimension and finds the optimum value for that specific entry.

To illustrate the algorithm in more details, we introduce the following notations:

$$\begin{aligned} D &= [\varphi_1 | \varphi_2 | \dots | \varphi_n], \quad \mathbf{x} = [x_1, x_2, \dots, x_n], \\ D^{(-i)} &= [\varphi_1 | \dots | \varphi_{i-1} | \varphi_{i+1} | \dots | \varphi_n], \\ \mathbf{x}^{(-i)} &= [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]. \end{aligned}$$

The non-differentiable part of (9) can be separated into individual coordinate wise components, each of which can be solved directly by applying Karush-Kuhn-Tucker necessary conditions. By fixing the values of  $\mathbf{x}^{(-i)}$ , the  $i^{th}$  coordinate wise component of (9) is obtained by solving the following:

$$\operatorname{argmin}_{x_i} \frac{1}{2} \|\phi_i - \varphi_i x_i\|_2^2 + \lambda |x_i| + \lambda \|\mathbf{x}^{(-i)}\|_1; \lambda \geq 0, \quad (11)$$

where  $\phi_i = \phi - D^{(-i)}(\mathbf{x})^{(-i)}$  is the error of the approximation computed when including all the atoms of the dictionary except  $\varphi_i$ . To find the best value for the coefficient  $x_i$ , (11) minimizes the error of approximation by adding the contribution of  $\varphi_i$  to the linear approximation. Following the Karush-Kuhn-Tucker necessary conditions, the optimal solution is obtained as follows:

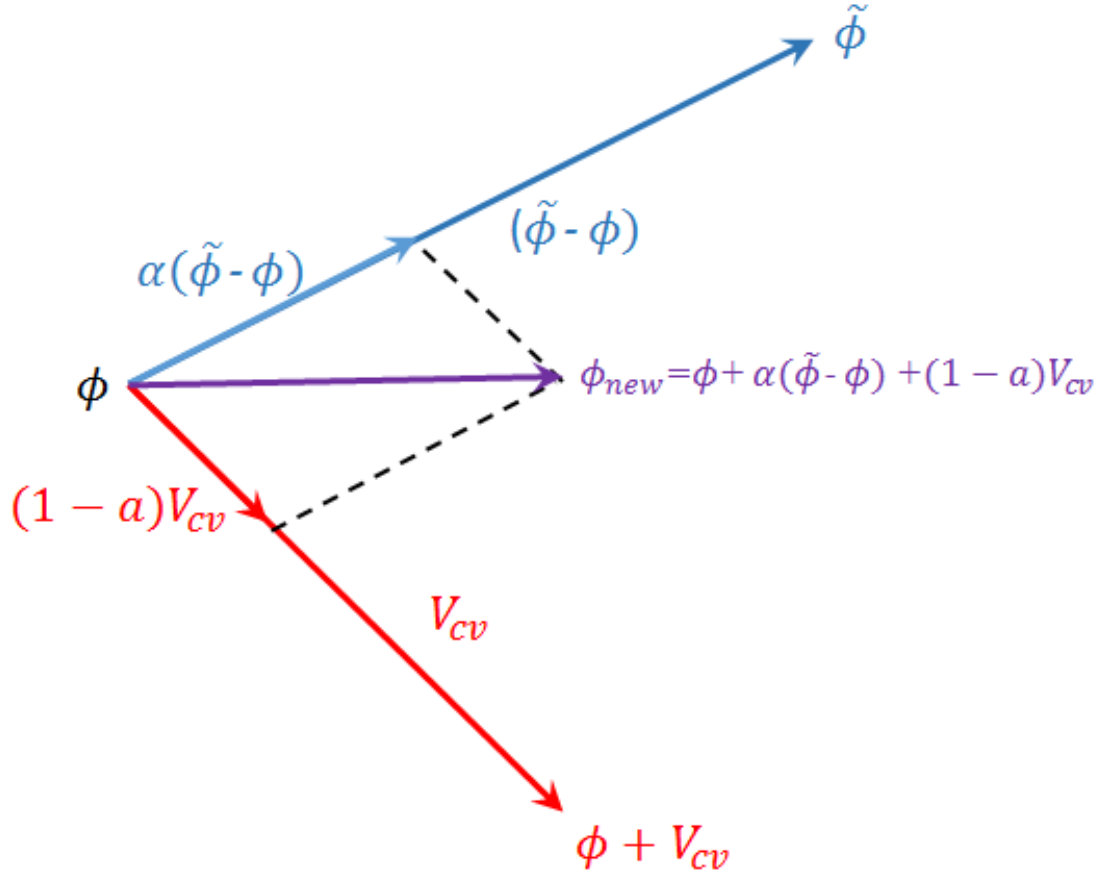
$$x_i^* = \begin{cases} \frac{(\phi_i^T \varphi_i - \lambda)}{\varphi_i^T \varphi_i}, & \text{if } \phi_i^T \varphi_i - \lambda > 0. \\ \frac{(\phi_i^T \varphi_i + \lambda)}{\varphi_i^T \varphi_i}, & \text{if } \phi_i^T \varphi_i + \lambda < 0. \\ 0, & \text{if } -\lambda \leq \phi_i^T \varphi_i \leq \lambda. \end{cases} \quad (12)$$

Equation (12) simply computes the projection of error ( $\phi_i$ ) onto the  $i^{th}$  atom of the dictionary. If the size of this projection is in  $[-\lambda, \lambda]$ ,  $x_i$  is set to zero, otherwise its value is set to  $\frac{(\phi_i^T \varphi_i - \lambda)}{\varphi_i^T \varphi_i}$  or  $\frac{(\phi_i^T \varphi_i + \lambda)}{\varphi_i^T \varphi_i}$  depending on whether the projection is greater than  $\lambda$  or less than  $-\lambda$ . In summary, this algorithm iteratively optimizes the objective function along one coordinate, while keeping other coordinates fixed.

Having the final approximated surface, we now can not only evolve the surface in the direction of Chan-Vese update, but also in a direction that minimizes the distance between the surface and its approximation. This update is illustrated in more details in Figure 8, where a linear combination of the Chan-Vese and shape update is computed and the surface is moved in that direction. To better illustrate equation update shown in Figure 8, we inspect two extreme cases:

$\alpha = \mathbf{0}$ : in this case, the term corresponding to shape update vanishes and the surface is driven by the Chan-Vese equation. The algorithm in this case is the same as classical Chan-Vese algorithm.

$\alpha = \mathbf{1}$ : the vector corresponding to the Chan-Vese update is multiplied by zero and



**Figure 8.** The surface evolution update is computed by a linear combination of updates in the direction of the Chan-Vese and the shape approximation.



does not play any role in this case. Thus, the curve is guiding by the shape energy update and in each iteration the surface is updated rendering the approximated shape prior.

The proposed algorithm for image segmentation is illustrated in Algorithm 1. It incorporates the shooting algorithm into level set evolution and forces the surface to separate homogeneous regions and at the same time keeps it similar to a linear combination of training shapes in the dictionary.

---

**Algorithm 1** Image Segmentation by SCoTS

---

```

1: Initialize  $\mathbf{x}$  randomly and  $\phi$  by the SDF of a  $5 \times 5$  square
2: while  $\|E(t+1) - E(t)\| > Threshold$  do
3:   Compute approximation:  $\tilde{\phi} = D\mathbf{x}$ 
4:   update  $\phi(t+1) = \phi(t) + \alpha(\sigma) (\tilde{\phi}(t) - \phi(t))$ 
       $+ (1 - \alpha(\sigma)) V_{cv}(t)$ 
5:   for  $i = 1 : n$  do
6:     using current value of  $\mathbf{x}^{(-i)}$  solve (11).
7:     Suppose  $x_i^*$  is the solution of (11), update  $i^{th}$ 
      element of  $\mathbf{x}$  to  $x_i^*$ .
8:   end for
9:   Compute number of non-zeros in  $\mathbf{x}$  and update  $\alpha(\sigma)$ 
10: end while

```

---

In this algorithm,  $\tilde{\phi}$  represents sparse approximation of the evolving surface. The shape prior weighting function  $\alpha(\sigma)$  determines the level of trust in the shape prior and is a function of sparsity (10). The algorithm consists of two nested loops. In the outer loop the segmenting surface gets updated in the direction of a linear combination of Chan-Vese and sparse shape prior terms. The inner loop refines the updated shape and brings it into the valid shape space.

## 5 Convergence and Complexity

### Convergence

The stopping criterion for the proposed algorithm is based on the change in the energy of the segmenting surface. If the energy does not change from one iteration to the next, the algorithm stops. We define the energy function as the sum of the Chan-Vese energy function and the computed sparse linear approximation.

$$E(\phi(t), \mathbf{x}(t)) = E_{cv}(\phi(t)) + E_{sp}(\phi(t), \mathbf{x}(t)). \quad (13)$$

$E_{cv}$  is the Chan-Vese energy function and  $E_{sp}$  is the energy term defined in (9). Although convergence of the algorithm cannot be mathematically proven, we have empirically observed that as long as the evolving surface is roughly initialized at the center of the nodule, the algorithm converges.

### Time Complexity

The algorithm complexity is a function of Chan-Vese and shape approximation complexity. In each iteration of the outer loop, one iteration of Chan-Vese, as well as  $n$  iterations of the inner loop are applied. The inner loop solves (11) through conditions provided in (12). However, as pointed out in [72], all the required values of  $\phi_i^T \varphi_i$  and  $\varphi_i^T \varphi_i$  for (12) can be pre-computed before starting the inner loop. Thus, since the complexity of Chan-Vese in each iteration is  $O(N \times M \times P)$ , where  $N \times M \times P$  is the size of the image, the complexity of one iteration of the proposed algorithm is  $O(N \times M \times P + n)$ , where  $n$  is the total number of shapes in the dictionary.

## CHAPTER IV

### SEGMENTATION RESULTS

#### 1 Dataset

The proposed algorithm for nodule segmentation has been validated on a subset of data from the lung image database consortium image collection (LIDC-IDRI) [4, 18]. In LIDC-IDRI, each dataset is a breath-held 3D CT image of the thorax with size  $512 \times 512$ . The number of slices varying between 95 and 672 and the in-plane pixel size varies between 0.5 and 0.8 mm/pixel. The range for the kVp for these data was 120-140 with 120 as the average and 20.99 as the standard deviation. The range for the mA was 30-634 with 215.9 as the average and 145.1 as the standard deviation. LIDC-IDRI contains lung CT scans from 1018 patients with nodule annotations provided by four experienced radiologists. It should be noted that the 4 radiologists who delineated the LIDC data differ between cases so that not the same 4 individuals read and delineated each scan. Therefore, for the rest of the paper, the reader should keep in mind that "radiologist j's delineations" ( $j=1, \dots, 4$ ) can indeed be from a number of individual radiologists.

Since nodules of size less than 3 mm are considered inconsequential, for our validation we only included data with nodule size greater than or equal to 3 mm in

diameter. Furthermore, we selected those nodules that were in common among all four radiologists. This left 542 nodules that fit inside predefined cubic volumes.

## 2 Evaluation

The approach adopted to validation is 10 fold cross validation [45], four times, once for each radiologist’s delineations. That is, 542 nodules as delineated by each radiologist were divided into 10 groups (folds). The SDFs of shapes corresponding to the nodules of 9 folds served as the atoms of the dictionary and the algorithm was tested on nodules in the 10<sup>th</sup> fold. The testing fold was successively changed between the 10 groups, each providing a different segmentation accuracy.

The accuracy of segmentation is measured based on the Dice Similarity Coefficient (DSC), Jaccard index (J), True Positive Rate (TPR), and False Positive Rate (FPR):

$$\begin{aligned} DSC &= \frac{2N(A \cap B)}{N(A) + N(B)}, & TPR &= \frac{TP}{TP + FN}, \\ J &= \frac{N(A \cap B)}{N(A \cup B)} = \frac{DSC}{2 - DSC}, & FPR &= \frac{FP}{FP + TN}, \end{aligned} \quad (14)$$

where in (14),  $A$  denotes the set of voxels classified by the algorithm to belong to the lung nodule,  $B$  is the ground truth from manual delineation,  $N(V)$  is the voxel count in the set  $V$ ,  $TP$  is the number of nodule voxels correctly segmented by the algorithm,  $FN$  represents the number of nodule voxels mistakenly segmented as background,  $FP$  is the number of non-nodule voxels mistakenly segmented as belonging to the nodule, and  $TN$  is the number of non-nodule voxels that are correctly segmented.

For more detailed evaluations, we also classified the nodules into several categories based on texture and attachment. Information about nodules texture may be found

in the LIDC XML Base Schema [1], where each radiologist rated the lesions related to in several categories including texture. The texture was classified from (1-5), where label 5 indicates solid nodules, label 3 corresponds to part-solid, and label 1 represents non-solid nodules. From the selected dataset, we chose nodules for which more than one radiologist’s rating was part-solid or non-solid. The selected subset contained 37 samples which met this criterion. In addition to this classification, a thoracic radiologist (Dr. Seow) further classified the nodules in our LIDC-IDRI datasets. It turned out 209 nodules are well-circumscribed, 178 are juxta-vascular while the total number of juxta-pleural and pleural tail nodules combined is 155. Since only a small fraction of nodules were classified as pleural tail, we combined these with juxta-pleural nodules.

### 3 Parameters Settings

The algorithm works on a region of interest (ROI) with size  $75 \times 85 \times 45$  with the nodule approximately centered. The approach to selection of the ROI was to keep the ROI size (in voxels) large enough to include all the nodules in the dataset. We extracted the size of all nodules as delineated by radiologists and determined the number of voxels they occupy in each dimension for the dataset. The largest number of voxels occupied for all nodules was 71 in the first dimension, 80 in the second dimension and 35 in the third dimension. We picked  $75 \times 85 \times 45$  for the size of ROI to ensure that all nodules delineated by all radiologists would fit inside the ROI.

The algorithm’s parameters were determined through optimization of the Dice Similarity Coefficient (DSC) for a limited number of samples of the three nodule

**Table 1.** The algorithm parameters values used for all experiments

Parameter	$k_1$	$k_2$	$k_3$	$k_4$	$\lambda$
Value	0.2	0	5	5	150

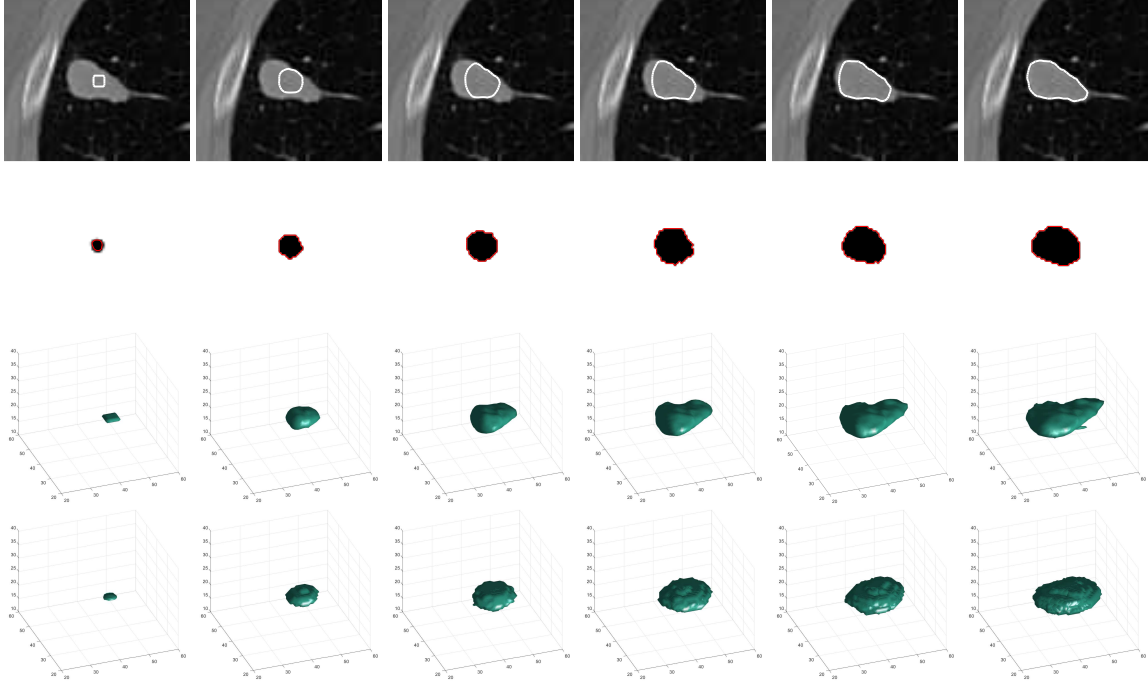
classes: well-circumscribed, juxta-pleural, and juxta-vascular. Following this approach we arrived at parameter values in Table 1.

In applying the segmentation algorithm, a point at the center of nodule is selected manually around which a  $5 \times 5$  square was prescribed as the zero level set of an SDF which iteratively evolved into a 3D shape and into neighboring slices.

#### 4 Results

The first experiment, illustrated in Fig. 9, demonstrates the surface and shape evolution through different steps. The initial and final segmenting surface are shown in the first and last columns of this figure respectively. Also, a few intermediate steps of segmentation are shown in the second to fifth columns of this figure. The zero level set of adaptive shape prior and its evolution has also been shown in the second and fourth rows. As the segmenting surface evolves, the shape prior changes and adapts itself to the segmenting surface. For this experiment, each iteration of the algorithm approximately took 7.3 seconds of CPU time in MATLAB on an AMD 3.9 GHz with 32 GB RAM of memory. The algorithm converged after 50 iterations and segmented the volume in 6 minutes.

A second result is used to demonstrate the evolution and convergence of the sparse representation. Fig. 10(a) shows the mid slice of an input nodule superimposed with



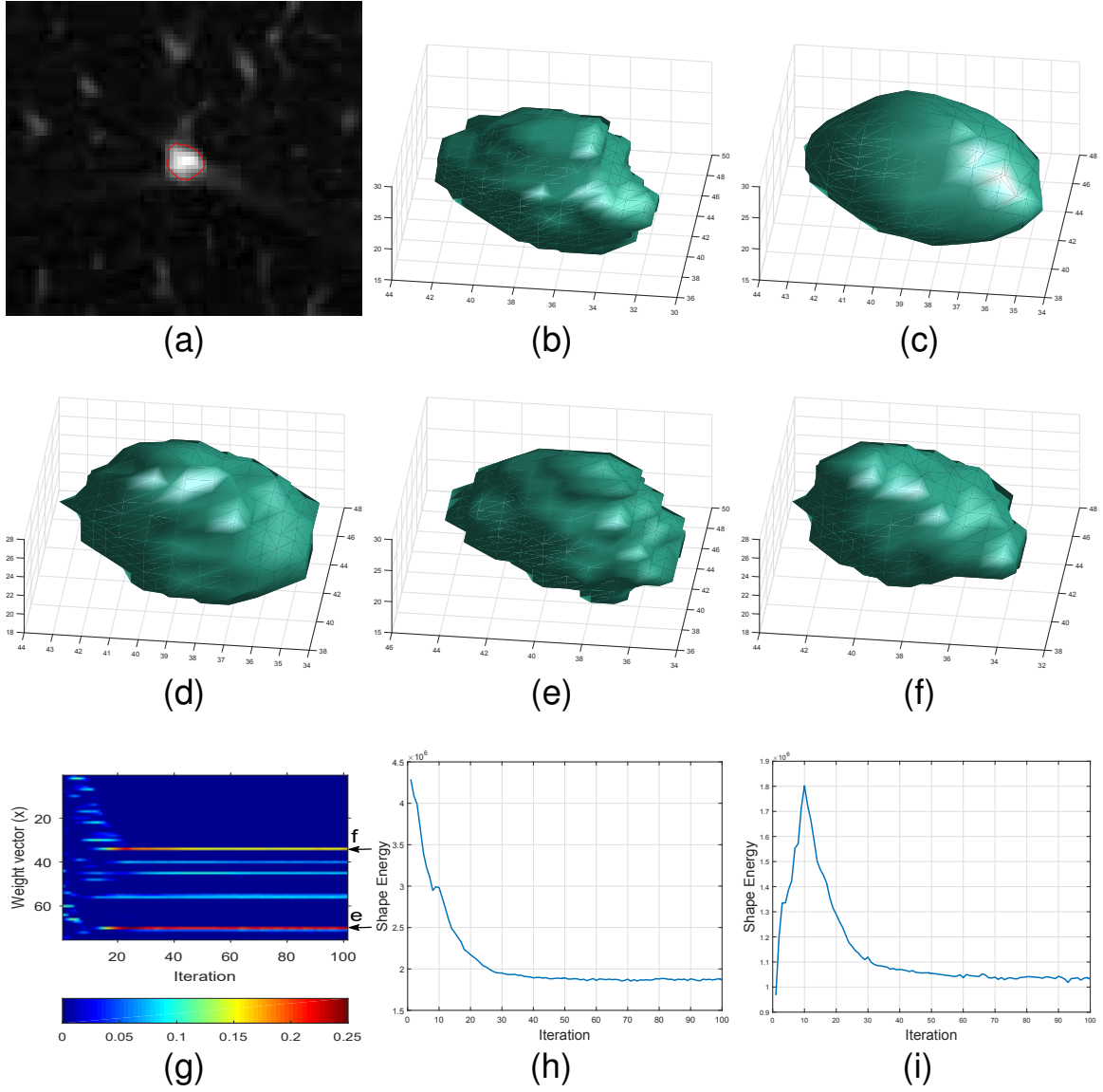
**Figure 9.** Surface and adaptive shape prior evolution. First column corresponds to the initialization. Columns 2-6 show iterations 6, 13, 20, 27, and the final result in iteration 50. First and third row show the evolution of segmentation in 2D and 3D respectively and second and fourth rows show the zero level set of adaptive shape prior in 2D and 3D.

the segmentation surface plotted in red. Fig. 10(b) shows this nodule in three dimensions delineated by the radiologist. Fig. 10(c) shows the final segmenting surface. Shape prior constructed from dictionary atoms is shown in Fig. 10(d). Fig. 10(e) and Fig. 10(f) represent two shapes from the dictionary that had significant contribution in building the shape prior. Fig. 10(g) shows the evolution of vector  $\mathbf{x}$  (equation (6)) over 100 iterations of the algorithm. Every iterative update of vector  $\mathbf{x}$  is stored as a column of a matrix with increasing iterations going from left to right. Small entries are shown in blue and for some the color shifts towards red as

the iterations and values increase. For the sake of simplicity only those entries of vector  $\mathbf{x}$  that changed were shown in this figure. We can see that for the first 20 iterations, selected atoms varying from one iteration to another. Subsequently, the algorithm promotes few atoms, two of which were shown in Fig. 10(e) and 10(f). This result is quite typical for the algorithm in that any valid nodule shape can be represented by 1 – 3% of the shape atoms in the dictionary. In Fig. 10(h) the total energy (equation (13)) is plotted over 100 iterations. The plot shows that the total energy converges and does not change significantly after 30 iterations. Finally, in Fig. 10(i), shape energy is presented. Initially, the shape energy oscillates with no apparent intention to converge. However, once the evolving surface forms a nodule boundary structure, the shape energy starts to decrease, resulting in convergence.

Fig. 11 shows segmentation of some samples of LIDC-IDRI dataset. It demonstrates how the proposed algorithm can distinguish the nodule from the surrounding tissues like vessel or pleura. It can be seen that when the nodule is not spherical, the proposed method can still capture the nodule. The last row in this figure shows samples for which the algorithm did not generate satisfactory results. The reason behind the failures maybe attributed to the fact that the framework is general and tries to segment all types of nodules. Failures mostly happen in situations where the nodules are attached to organs with similar HUs and at the same time the shape prior of the nodule cannot be sparsely reconstructed from the dictionary. As a result, the shape prior energy is dominated by Chan-Vese energy function and the segmenting surface leaks into adjacent regions. Another type of failure is related to very small and non-solid textured nodules which pose difficulty for Chan-Vese algorithm





**Figure 10.** (a) Mid slice of a nodule and its boundary, (b) ground truth nodule boundary in 3D, (c) segmented boundary in 3D, (d) shape prior constructed from dictionary atoms, (e,f) dictionary atoms with highest weights in nodule boundary reconstruction, (g) evolution of column vector  $\mathbf{x}$  (only 75 elements of the vector are shown) over 100 iterations - the weight of atoms corresponding to shapes (e) and (f) are distinct in yellow and red colors with arrows pointing to them. The range of weights during evolution for elements of the vector  $\mathbf{x}$  was  $[0, 0.23]$ . (h,i) total energy and shape energy over 100 iterations of the algorithm.

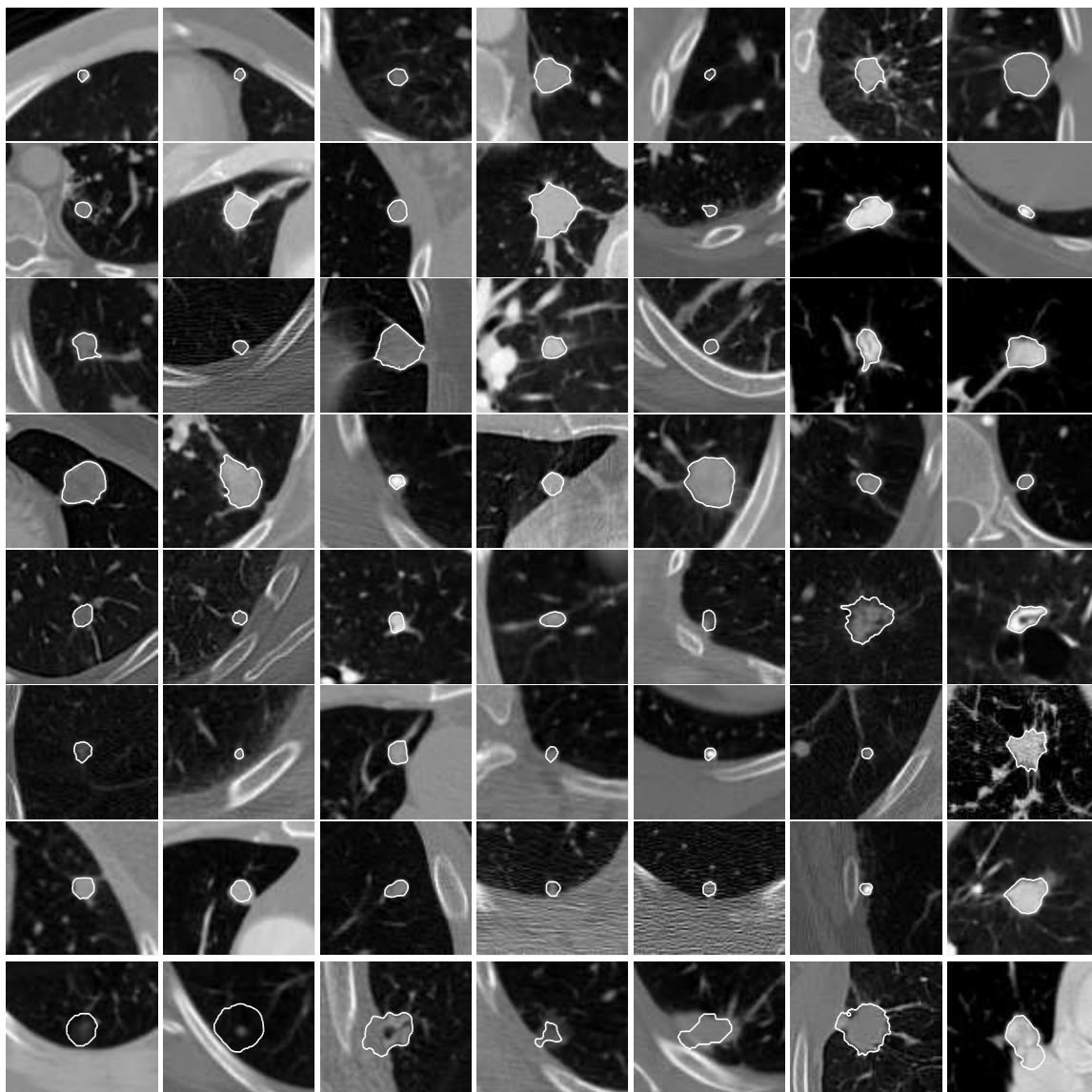
**Table 2.** Numerical validation of the algorithm (DSC, TPR, FPR) on 542 nodule CT dataset using Radiologist  $j$ ’s delineations for both training and testing with 10 fold cross validation. When rounded to the nearest hundredth, for all cases FPR was zero and therefore has not been included in the table. In this approach all delineations from each radiologist is randomly split into 10 groups (folds). Delineations in 9 of the folds are used to construct the dictionary. The image data in the 10<sup>th</sup> fold are used for testing. See text for additional descriptions.

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4
SCoTS	$DSC = 0.72 \pm 0.15$ $TPR = 0.77 \pm 0.16$	$DSC = 0.71 \pm 0.17$ $TPR = 0.8 \pm 0.16$	$DSC = 0.72 \pm 0.16$ $TPR = 0.77 \pm 0.17$	$DSC = 0.71 \pm 0.17$ $TPR = 0.78 \pm 0.16$

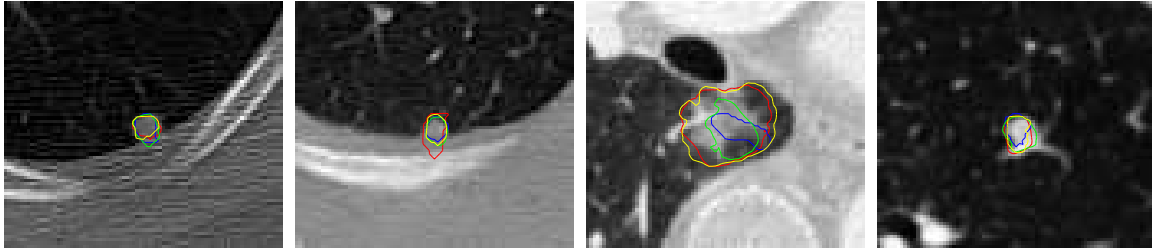
to grow properly. For such cases, the shape prior energy is dominant (see last row of Fig. 11). These nodules have shapes that does not appear in the training set. Such a nodule when attached to a vessel or pleura poses a challenge to the algorithm. In this case, neither the density nor proposed shape prior information stops the surface from growing into surrounding tissues.

In Table 2 we report the segmentation accuracy of SCoTS with respect to each radiologist’s delineations separately. Column  $j$  reports the accuracy when radiologist  $j$ ’s delineations were used to build the dictionary. Subsequently, 10 fold cross validation was performed based on this dictionary. The average DSC and TPR for each of the 10 testing folds were averaged in order to produce the entries in the table.

Although in Table 2 the segmentation accuracies are very close to one another, the segmented nodules are not identical and in some cases even dissimilar. Fig. 12 shows four images in which dictionaries constructed from different experts’ delineations produced significantly different segmented surfaces. In this figure, red, blue, green, and yellow curves identify the final segmentation produced by a dictionary constructed from nodules delineated by radiologist number 1, 2, 3, and 4 respec-



**Figure 11.** Segmentation of some samples of the lung dataset. Segmented boundaries were painted in white color. For these results, the dictionary was built based on radiologists 1's delineations followed by the 10 fold cross validation approach. Top 7 rows represent cases where the algorithm successfully segmented out the nodule from surrounding tissues. The last row shows 7 cases where the algorithm did not generate satisfactory results. These nodules have shapes that may not appear in the training set. We define failure cases as samples which have DCS less than 0.6. By this definition, the average percentage of failure over all samples is 13.5%.



**Figure 12.** Segmentation results based on dictionaries built based on first, second, third, and fourth radiologists’ delineations. For each case, different colors represent different dictionaries. The figure illustrates the resulting variability of the final segmentation.

tively. This behavior suggests the possibility of combining the results in a consensus manner where the segmented surfaces obtained by each dictionary are fused by an aggregation method such as [94] to provide higher accuracies.

To assess inter-observer variability in LIDC-IDRI database, radiologists’ delineations are compared to each other in Table 3. In this table and for column  $j$ , radiologist  $j$ ’s delineations were used as the ground truth for direct comparison with the radiologist in row  $i$ . The table confirms inter-observer variability as the computed DSC and TPR are not exactly equal to 1. Therefore, we provide another validation in Table 4 based on a consensus scheme with four radiologists’ delineations known as simultaneous truth and performance level estimation (STAPLE) [94]. In a capsule, STAPLE works by jointly optimizing the TPR and the FPR in an expectation maximization framework. It generates a probabilistic estimate for each voxel which when thresholded over 0.5 produces a consensus mask as ground truth. In Table 4 this mask has been used as ground truth. More specifically, in producing the entries in column  $j$  of this table, 9 folds of radiologist  $j$ ’s delineations were used as

**Table 3.** Inter-observer variability among four radiologists for delineation of nodules in the LIDC-IDRI database. Reported in rows  $i$  column  $j$  is the DSC and TPR for radiologists  $i$  using radiologist  $j$  delineations as ground truth. When rounded to the nearest hundredth, for all cases, FPR was zero and therefore has not been included in the table.

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4
Radiologist 1	$DSC = 1$ $TPR = 1$	$DSC = 0.79 \pm 0.09$ $TPR = 0.85 \pm 0.13$	$DSC = 0.80 \pm 0.10$ $TPR = 0.84 \pm 0.13$	$DSC = 0.79 \pm 0.10$ $TPR = 0.82 \pm 0.13$
Radiologist 2	$DSC = 0.79 \pm 0.09$ $TPR = 0.78 \pm 0.14$	$DSC = 1$ $TPR = 1$	$DSC = 0.81 \pm 0.10$ $TPR = 0.81 \pm 0.15$	$DSC = 0.79 \pm 0.10$ $TPR = 0.79 \pm 0.14$
Radiologist 3	$DSC = 0.80 \pm 0.10$ $TPR = 0.79 \pm 0.14$	$DSC = 0.81 \pm 0.10$ $TPR = 0.84 \pm 0.14$	$DSC = 1$ $TPR = 1$	$DSC = 0.79 \pm 0.10$ $TPR = 0.80 \pm 0.14$
Radiologist 4	$DSC = 0.79 \pm 0.10$ $TPR = 0.79 \pm 0.16$	$DSC = 0.79 \pm 0.10$ $TPR = 0.83 \pm 0.15$	$DSC = 0.79 \pm 0.10$ $TPR = 0.82 \pm 0.14$	$DSC = 1$ $TPR = 1$

**Table 4.** Numerical validation of the algorithm (DSC, TPR, FPR) on 542 nodule CT dataset using Radiologist  $j$ ’s delineations for training and STAPLE for testing with 10 fold cross validation. When rounded to the nearest hundredth, for all cases, FPR was zero and therefore has not been included in the table.

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4
SCoTS	$DSC = 0.73 \pm 0.16$ $TPR = 0.84 \pm 0.14$	$DSC = 0.74 \pm 0.16$ $TPR = 0.77 \pm 0.15$	$DSC = 0.73 \pm 0.16$ $TPR = 0.75 \pm 0.17$	$DSC = 0.73 \pm 0.16$ $TPR = 0.77 \pm 0.15$

dictionary and the output of SCoTS on test nodules in the 10<sup>th</sup> fold were compared with the ground truth from STAPLE. Comparing tables 2 and 4, the ground truth obtained by STAPLE resulted in higher accuracy compared than the case where one radiologist’s delineation served as the ground truth (Table 2).

To evaluate how SCoTS performs on different nodule classes, the relevant DSC and TPR values were evaluated on different nodule classes – results are reported in Table 5. Similar to Table 2, dictionary construction and ground truth alternates between four radiologists in a round robin fashion. It can be seen that performance is best on well-circumscribed nodules, and lowest on juxta-pleural/pleural tail nodules.

**Table 5.** Performance of SCoTS on specific nodule types.

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4
Well-circumscribed	$DSC = 0.75 \pm 0.14$ $TPR = 0.78 \pm 0.17$	$DSC = 0.75 \pm 0.13$ $TPR = 0.79 \pm 0.15$	$DSC = 0.75 \pm 0.13$ $TPR = 0.77 \pm 0.17$	$DSC = 0.75 \pm 0.13$ $TPR = 0.79 \pm 0.16$
Juxta-vascular	$DSC = 0.73 \pm 0.14$ $TPR = 0.74 \pm 0.16$	$DSC = 0.73 \pm 0.13$ $TPR = 0.73 \pm 0.16$	$DSC = 0.72 \pm 0.16$ $TPR = 0.70 \pm 0.19$	$DSC = 0.73 \pm 0.14$ $TPR = 0.73 \pm 0.16$
Juxta-pleural & plural tail	$DSC = 0.64 \pm 0.20$ $TPR = 0.78 \pm 0.17$	$DSC = 0.62 \pm 0.23$ $TPR = 0.78 \pm 0.18$	$DSC = 0.64 \pm 0.20$ $TPR = 0.77 \pm 0.17$	$DSC = 0.63 \pm 0.22$ $TPR = 0.78 \pm 0.17$

This is not surprising because these are the least and most challenging cases respectively. Another observation is that no matter which radiologist’s delineations were used to construct the dictionary, SCoT’s performance is similar on well-circumscribed nodules. We see reduction in performance on juxta-pleural and pleural tail nodules.

Table 6 compares results of SCoTS to other nodule segmentation methods which have been tested on the LIDC-IDRI database. Although all papers cited in Table 6 make use of the LIDC-IDRI database for validation, their experimental setup are different from our approach. The approach to validation in [11] split samples from the LIDC database into two groups, training the model on one group and testing on the second group without any cross validation. The authors of [49] used around 100 nodule data to build their model, and although they used the LIDC-IDRI data for testing, they did not specify if the training data were from the LIDC-IDRI database or from another database. While the method proposed in [8] used the LIDC-IDRI database for testing, it did not involve any training. The method in [5] also did not involve any training but for testing, the authors of [5] used the same nodule data sets in the LIDC-IDRI database that we have used for our training and testing. To compare our results with [5], a voxel probability map metric used in [5] was

**Table 6.** Comparison of performance of SCoTS with previously proposed nodule segmentation methods applied to LIDC-IDRI data. For this table, the dictionary used was built by radiologist 1’s delineations, followed by 10 fold cross validation. First three rows compares SCoTS with two other methods considering all nodule types. The next two rows compares SCoTS with [11] on juxta-vascular nodules and the last two rows compares SCoTS with [49] considering only non-solid or part-solid nodules. We report TPR with  $M_{pr} = 100\%$  as done for MRFC-OB [5] for SCoTS.

	$TPR(\%)$	$FPR(\%)$	$J$
SCoTS	$90.26 \pm 11.42$	$0.3 \pm 0.5$	$0.57 \pm 0.16$
MRFC-OB [16]	$95.50 \pm 7.86$	N/A	N/A
Cavalcanti, <i>et al.</i> [17]	93.53	0.89	N/A
SCoTS (JV)	$88.91 \pm 11.74$	$0.13 \pm 0.46$	$0.57 \pm 0.14$
Chen, <i>et al.</i> [12]	88.89	10.19	N/A
SCoTS (non/part-solid)	$81.73 \pm 23.57$	$0.9 \pm 1.2$	$0.36 \pm 0.23$
Lassen, <i>et al.</i> [14]	N/A	N/A	$0.50 \pm 0.14$

calculated for SCoTS. The map is computed for each voxel  $c$  of the nodule as follows: if all radiologists classify  $c$  as being part of the nodule, then the  $M_{pr}(c) = 100\%$ . At the other extreme, if  $c$  is not assigned a nodule label by any of the radiologists, then  $M_{PR}(c) = 0\%$ . Based on this definition,  $TPR$  is defined as true positive rate computed with  $M_{pr}(c) = 100\%$  as ground truth. Further details about this metric can be found in [5]. Table 6 compares the segmentation accuracies based on  $TPR$  and Jaccard index ( $J$ ) defined in (14).

Despite the fact that results in [5] are better than SCoTS, we should point out that in [5] each nodule class is treated separately with a different segmentation approach. In [5] after applying Otsu thresholding, a connectivity analysis was done to determine the nodule type, separating juxta-vascular and well-circumscribed nodules from juxta-pleural and pleural tail nodules. Attached tissues were separated from

pleural tail or pleural surface either by morphological operations or a thickening algorithm. However, SCoTS extracts the nodule shape by sparse representation and without any preprocessing or classification step in advance. Also, the method proposed in [8] needs an ROI as the input to the algorithm, used as a reference slice for background without nodule tissue. From this point of view, the proposed method is more general since it does not require the user to provide nearest slice to lesion.

The performance of the SCoTS on juxta-vascular nodules (SCoTS (JV)) is also shown in Table 6. Despite the close accuracy, the method in [11] is applicable to juxta-vascular nodules and in this sense less general than the method proposed in this paper.

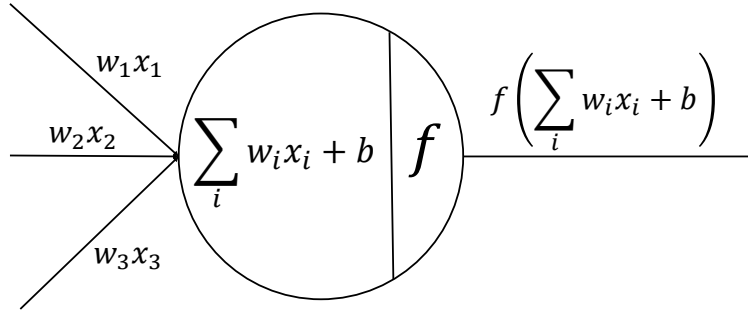
From the last two rows of Table 6, it can be seen that SCoTS does not perform well on non-solid and part-solid nodules. The main reason for this diminished performance lies in the Chan-Vese energy function and not shape prior energy. The Chan-Vese algorithm tries to separate homogeneous regions from ROI whereas the non-solid nodules have inhomogeneous appearance and pose difficulty for SCoTS. However, it should be noted that the algorithm in [49] requires additional manual steps as the user is required to draw the largest diameter of the nodule.



## CHAPTER V

### DEEP LEARNING IN MEDICAL IMAGE ANALYSIS

Traditionally Computer Aided Diagnosis (CAD) systems have relied on handcrafted features and classifier systems to distinguish between benign and malignant pathologies. With the advent of deep learning systems, optimum features are learned for the classification task at hand. This chapter reviews fundamentals of deep learning and its success in medical image detection, segmentation, and diagnosis of anatomical objects.



**Figure 13.** Diagram of a perceptron. A weighted averaging of inputs is added with bias and activated by activation function  $f$  generating one output.

## 1 Deep Learning

In this section and using artificial neural networks, we focus on learning hierarchical feature representations to effectively represent the data for the task of classification and segmentation in medical images. The perceptron [70] is the fundamental building blocks of artificial networks (Fig. 13). It receives a set of inputs  $x_i$ 's and generates an output by computing a weighted average of inputs and passing it through an activation function. The weights  $w_i$ 's are the parameters of the perceptron and are learned during the training phase.

Recently, Convolutional Neural Networks (CNNs) that are a special case of neural networks were found to achieve state-of-the-art performance in image classification related applications. In CNNs the parameters of the network are shared among multiple perceptrons implying that features are relevant in different locations of an image. In convolutional layers multiple input matrices (channels) are fed to a convolutional layer and generate multiple output matrices:

$$O_j = f\left(\sum_{i=1}^N I_i * K_{ij} + B_j\right), \quad (15)$$

where  $I_i$ 's are input matrices that are convolved with corresponding kernel matrix  $K_j$ , added with bias  $B_j$  to produce an output matrix  $O_j$ . The output can be generated for a total of  $j = 1, \dots, M$  matrices. Each kernel extracts a set of local features and swipes over the whole image. The activation function  $f$  is a function that determines whether a neuron should fire or not depending on the inputs to the neuron. The popular activation function used in deep learning is Rectified Linear Unit (ReLU):

$$\max(0, x)$$

The minimum value of the activation function is set to 0 and the maximum can be infinity. This essentially means fewer nodes are firing, giving the benefit of a lighter model with sparsity.

For dimensionality reduction in CNN architectures, pooling layers are embedded in which the neighboring neurons in the input are summarized in a single output. A typical choice for pooling is max-pooling with  $2 \times 2$  kernels. In this layer, from the 4 elements in a  $2 \times 2$  neighborhood of input matrix, the maximum value is selected as the output. This operation effectively reduces the size of the input matrix by half in each direction.

The kernel weights (feature extractors) are jointly learned with the classification layer using stochastic gradient descent and updated in each iteration to minimize the error of class prediction  $E = (y \log(p) + (1 - y) \log(1 - p))$ , where  $y$  is the true label of the sample and  $p$  is the predicted probability:

$$w_i(t + 1) = w_i(t) - \eta \frac{\partial E}{\partial w_i}. \quad (16)$$

In Eq. (16),  $w_i(t)$  is the current weight,  $\frac{\partial E}{\partial w_i}$  is the error gradient with respect to the weight  $w_i$ ,  $t$  is the iteration number, and  $\eta$  is the learning rate that controls the speed of learning. Instead of updating the kernels after every single samples, they are updated after a number of samples passes through the network. This number is referred to as batch size and is one of the hyper parameters of the network which should be set in the training phase.

The rest of this chapter reviews applications of deep learning in medical image analysis.

## 2 Deep Learning for Detection

Location and interpretation of anatomical structures are critical in radiologists' workflow. This is achieved by identifying object of interests signatures that can distinguish them from the rest of the image. These anatomical signatures can be extracted automatically by computational methods. Early computational algorithms were designed specifically for each task by creating specific image filters to extract image features. With availability of more and more medical image data sets, deep learning technologies are now more effective and are being applied in this domain.

With early successes of deep learning and convolutional neural networks in 2-D computer vision applications [47, 84, 39, 82, 73] some studies have attempted to adapt these to medical imaging applications by extracting 2-D planes from 3-D volumes. Authors in [93] extracted 2-D mid-axial, mid-coronal, and mid-sagittal planes from 3-D thoracic CT scans and building an RGB-like image with the planes and feed it to OverFeat network [73] to detect lung nodules from other thoracic lesions. To illustrate the effectiveness of 2-D CNNs, a similar framework was used in [71] for two other CT imaging applications: locating sclerotic metastases in spine imaging and lymph nodes and colonic polyps in cancer monitoring and screening. To increase the robustness, these authors augmented their data during training by randomly scaling, shifting and rotating the images. To better utilize volumetric information in images, authors in [74] increased the number of 2-D views to 9 and used ensemble methods to fuse information from different views for detection of pulmonary nodules. In [88], the pulmonary nodule detection was performed using multiple CNNs parsing both

PET and CT data. Multi-stream CNNs typically used in these studies are often referred to by 2.5 dimensional representation of volumetric 3-D images [78].

To take advantage of information in all directions of volumetric image, some studies have proposed use of 3-D coevolution kernels. In these systems, kernels sweep the volume not in x and y directions alone but in the z direction as well, capturing more contextual information [66, 35, 38]. In [35], a 3-D CNN was used to find micro-bleeds in brain MRI. Authors have used the same CNN architecture for the application of false positive reduction in thoracic CT scans.

### 3 Deep learning for segmentation

Long et al. [54] revised and adapted convolutional neural networks in order to make them capable of semantically segmenting natural scene images. The key insight in their work was to replace fully connected layers with convolutional kernels permitting the network to receive inputs with various sizes rather than a fixed size as it is forced in CNNs; fully connected layers can be viewed as convolutions with kernels that cover their entire input regions. Moreover, their modification permitted making dense prediction for per-pixel tasks such as semantic segmentation and with the transposed convolution layer the size of output was matched with the input image. Subsequently, the authors in [102] used the similar 2-D FCN with a new Jaccard distance loss function to segment skin lesions in dermoscopic images. In order to incorporate more contextual information to segment subcortical regions in MR images, Dolz et al. [26] used a 3-D FCN.

Recently, FCNs were modified in the form of a new architecture called U-Net [69]

where the up-sampling part of the FCN was enriched with more channels of features, allowing the network to propagate contextual information to higher resolution layers; down-sampling and up-sampling parts of the network are symmetric and form a U shaped architecture with cross links from the down-sampling part of the network to the up-sampling part. The authors segmented neuronal structures in electron microscopic and cell segmentation in light microscopic images. The 2-D operators in [9] were replaced with their 3-D counterparts in [16] to provide segmentation in 3-D volumetric images and the proposed algorithm evaluated on xenopus kidney segmentation in confocal microscopic data. Authors in [57] developed a 3-D U-Net architecture called V-Net with a new loss function directly optimizing the Dice Similarity Coefficient and applied to segmenting prostate in MR volumes.

#### 4 Deep Learning for Computer-Aided Diagnosis

The major task of CADx systems is to diagnose disease by discriminating between benign and malignant image regions or volumes. Similar to the CAdE systems which detect disease and have utilized deep learning based methods, CADx systems have also seen application of deep learning architectures and obtained state-of-the-art performance. In early attempts, Cheng et al. [13] used a deep learning diagnosis technique for breast ultrasound lesions and lung CT nodules. Authors in [53] used a 2-D CNN for the both applications of prostate cancer diagnosis in biopsy specimens and breast cancer metastasis detection for sentinel lymph nodes. In [51], diagnosis of Alzheimer disease was performed based on MRI and PET modalities. 3-D CNNs were used to predict disease but also were used to estimate a missed modality. For the

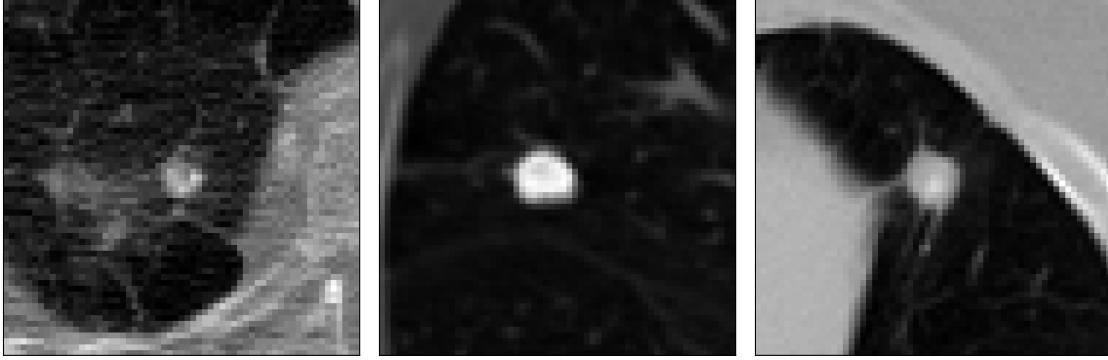
same application, authors in [80] used a more recent CNN architecture called deep polynomial networks which calculate a linear or quadratic function of input feature maps. In order to compensate for the small number of training samples available, in [55] two CNNs were initially trained on ImageNet [23] and learned features were subsequently transferred to diagnose thyroid nodule in ultrasound images. 2-D CNNs [7, 79], multi-view 2-D CNNs [63], and 3-D CNNs [24, 41] were the most common deep learning methodologies that have been applied for diagnosing pulmonary nodules in CT scans.

## CHAPTER VI

### LUNG NODULE MALIGNANCY PREDICTION BASED ON SINGLE TIME POINT AND SEQUENTIAL TIME POINT CT SCANS

Cancer diagnosis from low dose CT images still remains a challenging task due to the subtle visual appearance differences between benign and malignant nodules, even to the trained radiologists. As an example, mid-level slices of three nodules from LIDC- IDRI database are depicted in Fig. 14. Nodules shown in this figure belong to patients 658, 766, and 974 of LIDC-IDRI database. For all three nodules, two radiologists classified them as malignant (assigning a malignancy score of 4 or 5), and two other radiologists assigned a malignancy score of 1 out of 5 pointing to a benign diagnosis. Such inter-variability reaffirms the previously reported finding that lung cancer is the third most frequently missed diagnosis based on expert readers' visual assessment [15]. Computer-aided Diagnosis (CAD) systems may therefore prove useful in assisting radiologists in the malignancy prediction task. As performed in the National Lung Screening Trials, it is also typical that subjects with increased risk of developing lung cancer are scanned on an annual basis. Therefore, provided two consecutive CT scans from the same nodule, it was hypothesized that a CAD system can extract more information and confidence about the diagnosis of the nodule. In this chapter, we propose two individual methods for classification of lung nodules.



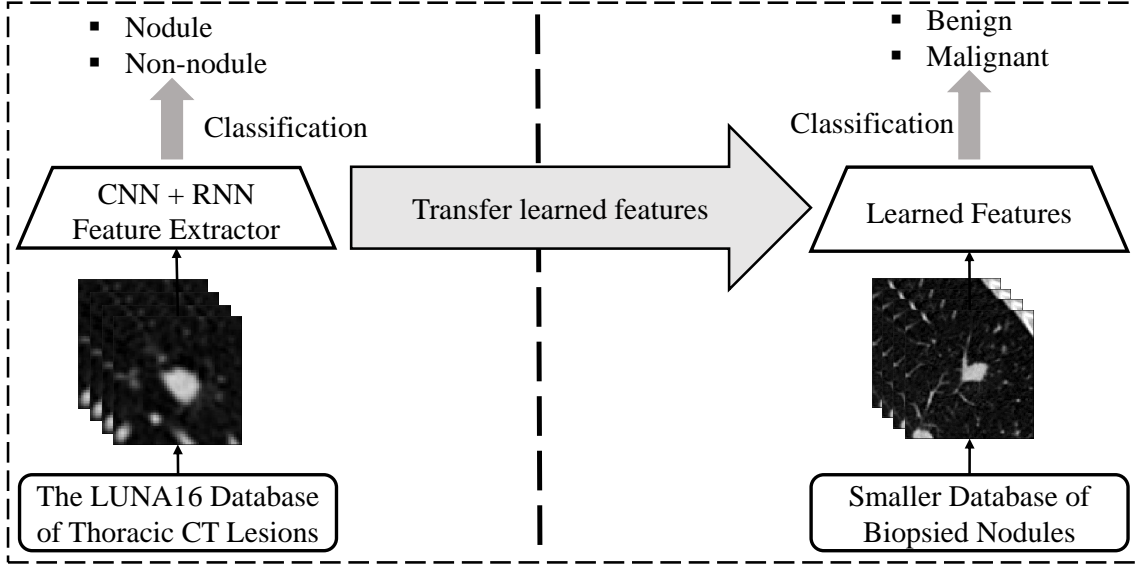


**Figure 14.** From left to right: mid-slice of nodule data from patient numbers 658, 766, and 974 of LIDC-IDRI database. For all three nodules, two radiologists labeled them as benign and two others diagnosed them as malignant. The figure illustrates the inter-variability in radiologists’ annotations.

The first method’s prediction is based on the visual appearance in a single scan and the second method looks at two scans together and assigns a malignancy score based on the changes in nodule shape. The two methods were also combined using a simple fusion approach and as expected the combination of the two methods improved the classifier’s performance. The chapter describes a method to fuse the two results.

### Transfer Learning

For image and object classification, convolutional neural networks (CNN) have achieved significant success in a variety of applications [47, 67, 52]. However, an obstacle with deep learning methods is the fact that a huge database consisting of thousands of samples is necessary for proper training. Unfortunately, a comprehensive dataset on lung nodule samples is not available to efficiently apply a deep learning method. A major technique to overcome the challenge is to use pre-trained CNN features followed with a supervised tuning of a classifier. A special case of this approach is called



**Figure 15.** Diagram of the proposed neural network for nodule classification illustrated for a 16 slice volume of interest input. A deep 2-D CNN extracts in-plane features followed by GRU units to extract inter-slice dependencies. The learned features are finally passed to classification layer to classify lesions.

transfer learning in which final classification layers are re-trained to adapt to the proposed application [81]. Promising results reported in [86], offer this approach to help avoid large amount of labeled training data and to save time to train the CNN weights.

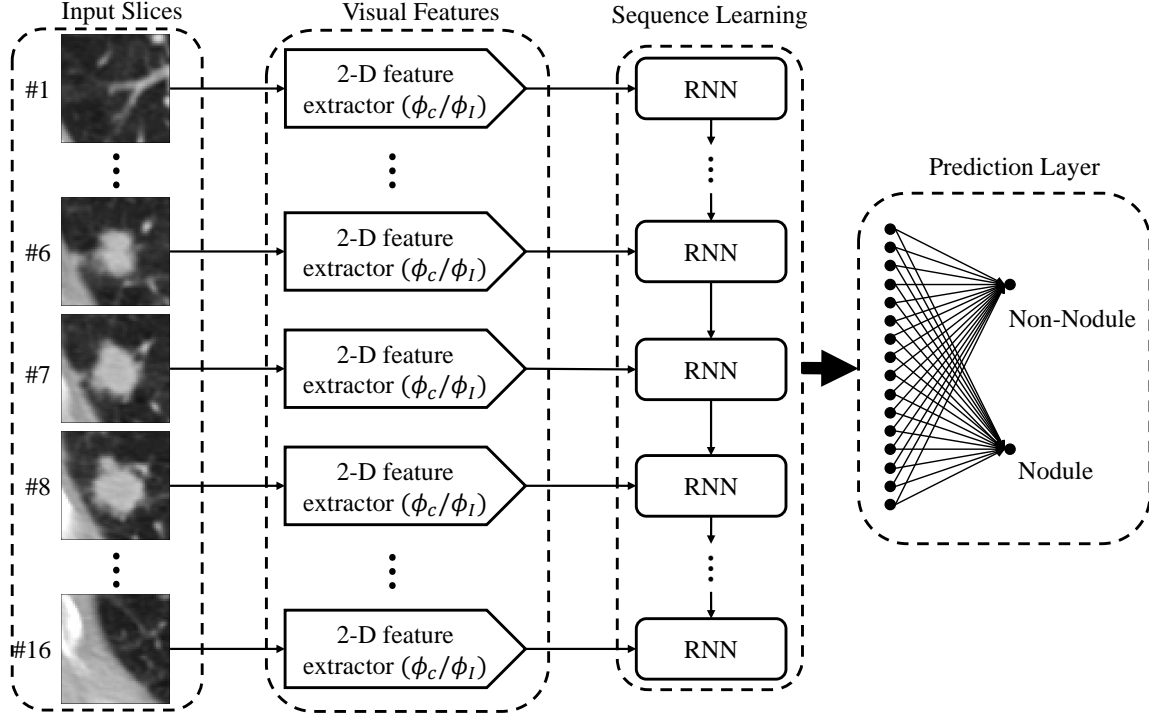
Fig. 15 summarizes the training process of the proposed CADx system. A deep network is first trained on samples available in the LUNA16 database with the aim to learn discriminative pulmonary nodule features in CT data. The weights of this trained network are then transferred into the new application of benign/malignant classification of nodules. The top classification layers are then trained from scratch to discriminate benign vs malignant nodules with the feature extractor layer’s weights frozen.

## 1 Recurrent convolutional networks to detect lung nodules based on single time point CT scans

In order to compensate for the insufficient number of malignant and benign nodules available to us, we chose to use transfer learning from data provided by the LUNA16 Challenge for detection of nodules in chest CT [75]. The task in the LUNA16 challenge was to reduce the number of false positives (e.g. blood vessels, thoracic airways, etc.) from lung nodule candidates generated by a candidate generator software. The data used in this challenge are taken from Lung Image Database Consortium Image Collection (LIDC-IDRI) database [4]. LIDC-IDRI is a publicly available database containing thoracic CT scans from 1018 patients. It contains nodule annotations provided by four experienced radiologists; of these, nodules with diameters  $\geq 3$  mm that were annotated by at least 3 radiologists were included in the training set. This left us with 1186 nodules. In order to learn nodule features, more than 750,000 regions were extracted from the aforementioned CT scans as negative samples using the algorithms described in [74] as with the LUNA16 Challenge data. The actual number of true nodules in the LUNA16 database is 1186 (selected by at least three radiologists) and is very small compared to the total number of candidate locations in this dataset (more than 750,000 candidate locations). Thus, in order to compensate for this heavy imbalance, we artificially augmented the nodule data as follows. We translated the position of each candidate by 2 voxels along each dimension. They were also rotated in the interval of (0,360) degrees by 10 degrees increments; in total, 47 samples artificially generated for each available training nodule. Due to their

high count, no augmentation was performed on the non-nodule candidate locations. Detecting and classifying nodules in volumetric CT scans is inherently a 3-D classification problem. However, to take advantage of the architectures proposed in 2-D computer vision applications, early deep learning based systems proposed to treat the problem in 2-D [74, 93, 42, 83, 17, 99]. These 2-D architectures however, partially ignores the data in 3-D volumes as the information is collapsed into 2-D planes. Here, we propose an alternative architecture presented in Fig. 16, to take advantage of all the information in 3-D volumes and computationally less expensive in compare to their 3-D counterparts.

The architecture of the feature extractor is summarized in Fig. 16. In order to capture the information in all dimensions of the Volume of Interest (VOI), the proposed network was designed to capture in-plane features using 2-D CNNs while modeling the contextual information across axial slices with RNNs, i.e., this architecture essentially treats the slices in a volume as temporal information as commonly performed in RNNs. Recurrent networks are the state-of-the-art frameworks in sequence learning [27, 3] and here were adapted for volumetric analysis. Each slice in a volume is processed with 2-D CNNs whose weight are shared among all CNNs. The first convolutional layer consists of 64 kernels of size  $5 \times 5$ , producing a feature map with dimension  $64 \times 48 \times 48$ . The second, third, and fourth convolutional layers have 64 kernels of size  $3 \times 3$ . Also, the convolutions were zero-padded so that any reduction in size would be due to the sub-sampling; sub-sampling was performed by a factor of 2 using max-pooling layers after each of the last three convolutional layers. In-plane 2-D features were passed to 128 bi-directional Gated Recurrent Unit (GRU)



**Figure 16.** The general framework of the proposed approach to classify volumetric CT scans inspired by the radiologists reading process. Features extracted by individual 2-D CNNs from each slice are passed into RNN units to capture volumetric information. The output of the RNNs is passed to a fully connected layer followed by a softmax layer to classify each sequence of 2-D slices as nodule or not.

units [15] to capture inter-slice dependencies and to generate 256 output for each volume in the image. Subsequently, 256 outputs generated from 128 bidirectional GRU units were passed to a fully connected layer with 256 nodes, followed by a soft-max classification layer. The whole network was trained end-to-end to minimize the cross entropy error.

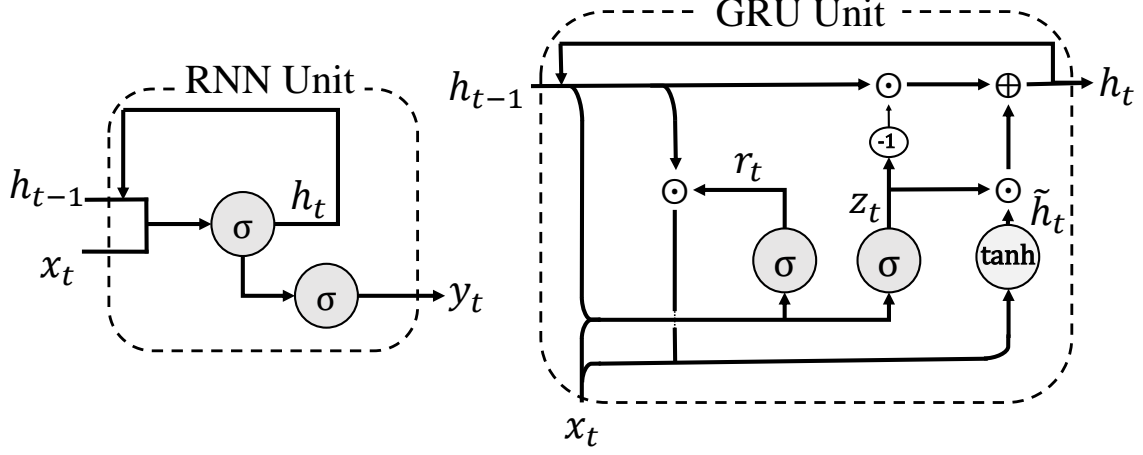
The classification layers of the network trained on the LUNA16 database were subsequently removed and replaced by new fully connected layers (the remaining

sections of the network were left untouched with their parameters frozen). The parameters for the fully connected classification layers were then learned from scratch on the benign and malignant nodules obtained from the NLST database after augmentation. The initial convolutional layers on the other hand were kept fixed. In order to account for nodule size variations, the network generated confidences for input volumes of size  $20 \text{ mm}^3$  and  $30 \text{ mm}^3$  which were then simply averaged and reported as final malignancy prediction score.

### Inter-slice learning

Recurrent neural networks are the natural architectures to analyze time series data such as speech, text, and video. Unlike the CNN architecture, an RNN utilizes loops and internal states that get updated as new inputs are processed, allowing it to develop an internal memory in order to make predictions based on serial data. This advantage has resulted in state-of-the-art in time-series learning related applications such as speech [3] and activity recognition [27].

For the analysis of CT scans, consecutive slices can be treated as temporal data. Thus, through connecting the visual features extracted from consecutive slices to time-series learning units such as RNN, our network learns the interrelationship between the slices using the RNNs. This process is analogous to how radiologists locate nodules in a stack of 2-D CT slices. Rather than using the information in a single slice, they scroll back and forth through the CT slices and make their decision regarding the presence or absence of a nodule based on patterns observed in a stack of slices.



**Figure 17.** L

eft: diagram of RNN units, right: diagram of GRU unit used to model CT slice dependencies.

In their classical form (Fig. 17, left), RNNs capture temporal information by mapping the stack of input slices to hidden states and outputs:

$$\begin{aligned} h_t &= \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \\ y_t &= \sigma(W_{hy}h_t + b_y). \end{aligned} \tag{17}$$

In Eq. 17,  $\sigma$  is an activation function,  $x_t$  is the input to the RNN units generated by the feature extractor block, i.e.,  $x_t = \phi_{\mathcal{C}}(S_i)$ ,  $h_t \in R^N$  is the hidden state with  $N$  hidden units,  $z_t$  is the output corresponding to the  $t^{th}$  slice in the sequence, and  $W$ 's and  $b$ 's are the parameters of the RNN that are learned during the training phase. For a volume that consists of  $T$  slices ( $t = 1, 2, \dots, T$ ), the hidden states and outputs for each slice are computed sequentially; i.e.  $h_1, z_1, h_2, z_2, \dots, h_T, z_T$ .

The classical RNNs suffer from vanishing and exploding gradients [40] resulting in low performance due to their inability to maintain memory over long series. In our

case, this could lead to a learned model where the contribution of the late slices of the volume may not be fully exploited. To overcome this limitation, several RNN architectures that are capable of learning long-term dependencies have been proposed. These architectures include Long Short Term Memory (LSTM) [40] and more recently Gated Recurrent Unit (GRU) [14]. In our proposed architecture, we model the slice dependency using the GRU units (Fig 17). This is because GRUs are simpler with fewer training parameters but still has been found to provide performance comparable to LSTMs [14]. GRUs learn long-term dependencies by introducing update and reset gating units. The gates find a way to optionally let the information flow through the network using sigmoid and point-wise multiplication operation.

More specifically, having  $\sigma(x) = (1 + e^{-x})^{-1}$  and  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  as the activation functions, a GRU updates its hidden states and outputs as follows:

$$\begin{aligned}
r_t &= \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r), \\
z_t &= \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z), \\
\tilde{h}_t &= \tanh(W_{xo}x_t + W_{ho}(r_t \odot h_{t-1}) + b_o), \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t.
\end{aligned} \tag{18}$$

In Eq. 18,  $\odot$  represents the element wise multiplication and  $\tilde{h}$  is the unit's new memory content. The reset gate  $r_t \in R^N$  determines how the input and the current memory are combined to generate the new memory content. For example, when it is *off*, the unit's new memory is only obtained based on the input data. The update gate  $z_t \in R^N$  determines how the unit updates its content by partially forgetting the existing memory and adding the new memory content, i.e.,  $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$ . In the extreme cases, when  $z_t = 0$  the unit is not updated and when  $z_t = 1$  the unit



is updated by the new memory.

Once the CNN and RNN weights are trained on the LUNA16 database, fully connected layers of the trained network is replaced by new fully connected layers and are trained from scratch based on single scans contain benign or malignant nodules.

## 2 Use of shape, and volume growth in two consecutive CT Scans

Based on SCoTS, discussed in Chapter III, in this section we develop machinery for comparing a nodule's shape and volume change in two consecutive CT scan.

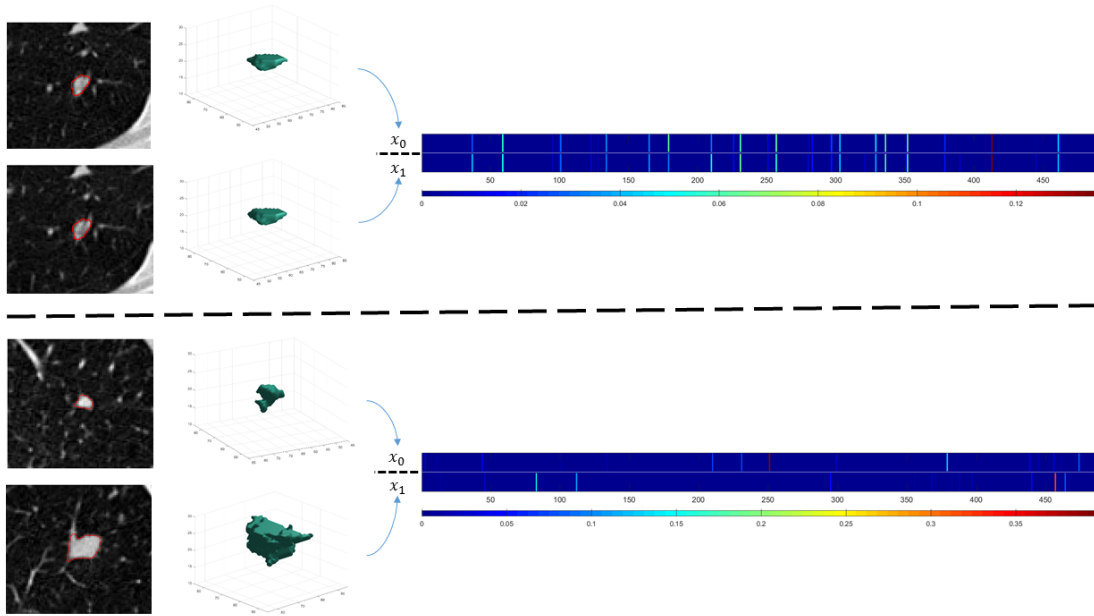
### Methodology

In order to compare nodule shapes and volume in two time consecutive time points, we apply the SCoTS algorithm to the two time points. This provides us with two weight vectors  $x_0$  and  $x_1$  (see Eq. (9), Chapter III). When the nodule does not grow or goes through any type of a shape change dictionary atoms needed to reconstruct the nodule shape in two time points would be identical and moreover the weight vectors  $x_0$  and  $x_1$  would be very similar. On the other hand, when the nodule's shape varies significantly, the atoms and the weight vectors that emerge from the dictionary are different, and therefore there is a higher probability for malignancy. Fig. 18 illustrates the classification performed based on shape variation extracted with SCoTS from consecutive scans. Based on the number of shared atoms in two representation a malignancy confidence is assigned to the nodule:

$$p = 1 - \frac{\|\mathbf{x}_0 \wedge \mathbf{x}_1\|_0}{\|\mathbf{x}_0 \vee \mathbf{x}_1\|_0}, \quad (19)$$

where  $|||_0$  is the  $l_0$  norm counting number of non-zero values,  $\wedge$  and  $\vee$  are logical *and* and *or*, and  $p$  is the malignancy score. If the dictionary response to the nodule shape at time point 0 is close to the dictionary response at time point 1, the logical *and* and *or* would be similar, and the value of fraction in equation 19 gets close to one and the malignancy score assigned to the nodule approaches zero. On the other hand, when the nodule in the two time points are decoded by different atoms in the dictionary, the fraction gets close to zero and  $p$  approaches 1. In summary, more elements shared in  $x_0$  and  $x_1$  result in lower malignancy confidence based on equation 19.

To better visualize the dictionary response to benign and malignant nodules, in Fig. 18 a pair of corresponding benign and malignant nodules are illustrated. The nodules were segmented at time 0 and time 1 and their SDF was computed. The shape of each nodule was approximated by the atoms of the dictionary. The figure visualizes the vector  $\mathbf{x}$  that was obtained from the approximation. For better visibility, the entries of vector are color coded with zero values in dark blue and shifting to red by increasing the entries. It was observed that vector  $x_0$  the produced weighting vector is sparse with zero values in both the benign and the malignant cases. For the benign nodule the vector  $x_1$  activated very similar atoms in the dictionary as  $x_0$ . This was not the case for the case of the malignant nodule. It is also observed that for the benign sample almost the same atoms of the dictionary activated in two time points as the nodule shape does not change significantly in the baseline and follow up scan. Specifically, for the benign nodule  $p = 0.13$  and for the malignant nodule,  $p = 0.99$ .



**Figure 18.** Dictionary response to cases of benign and malignant nodules. Top row shows a benign nodule segmented in two successive scans. The vector  $\mathbf{x}$  was color coded in time points 0 and 1 with lower values in blue and higher values in red. For the top nodule in this example,  $p=0.14$ , and for the bottom nodule,  $p=0.99$ . Please see text for description.

In this Chapter we have discussed CNN and RNN approaches to learning the visual appearance of nodules in single time point CT scans and proposed machinery based on dictionary learning to learn change in nodule shape from consecutive scans. In the next Chapter we will discuss experimental results for the proposed methods.

## CHAPTER VII

### EXPERIMENTAL RESULTS FOR MALIGNANCY PREDICTION FROM NLST DATA SETS

This Chapter reports the performance of methodologies for malignancy prediction which were proposed in Chapter VI. First, the database that was used for training and evaluation was described and next each model's performance is reported. The performance of the fusion of the RNN model for single time point nodule appearance with shape and growth dictionary learning model as the final performance of the CAD system is finally reported.

#### 1 Datasets

The data for training and evaluation was collected from the The National Lung Screening Trial (NLST) [61]. NLST was a multi-year study confirmed that screening high-risk individuals with low dose CT scanning can reduce morbidity and mortality due to lung cancer. Approximately, 54,000 participants enrolled and screened annually for 2 year span. Of these, CT scans of 15,000 subjects were obtained through the Cancer Data Access System (CDAS) administered by the National Cancer Institute/National Institutes of Health. Majority of these scans are normal with no pathologies. To develop the database for this study we collected data from partici-

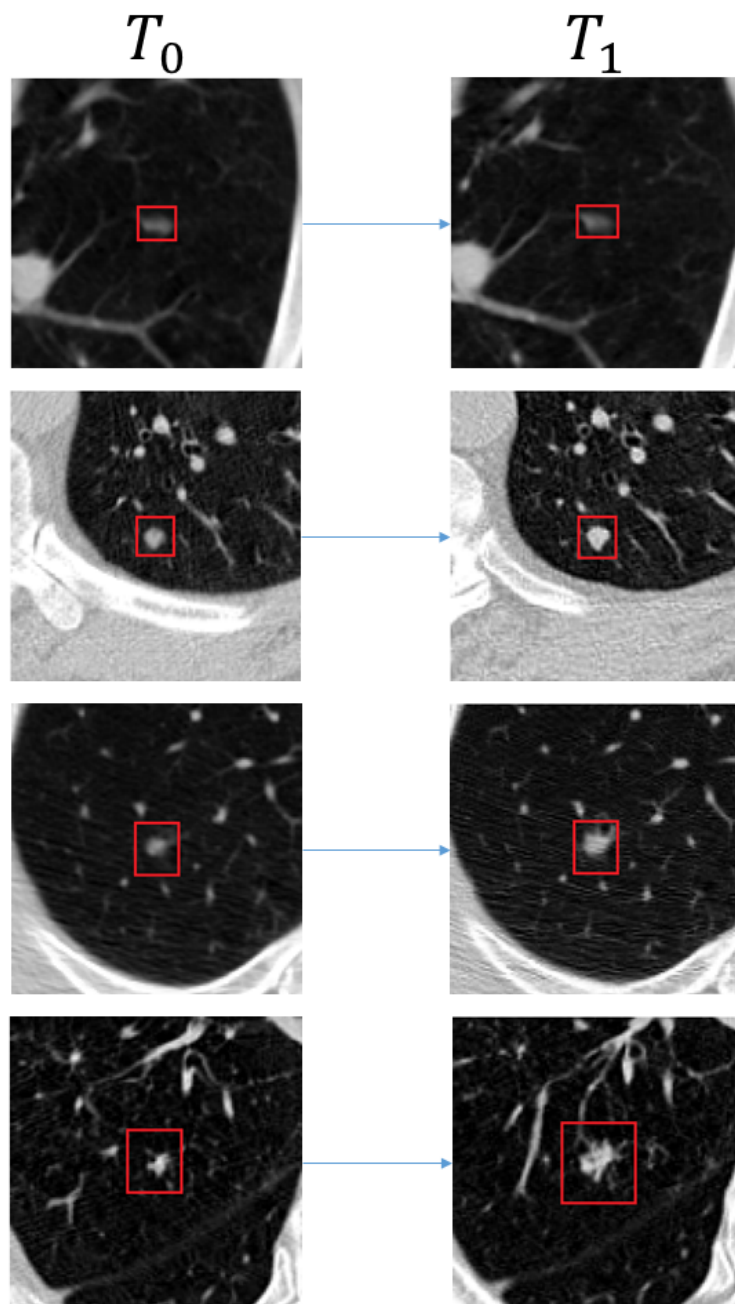
pants only if a pathologically proven malignant/benign lesion diagnosis and at least one follow up scan was available. From these data, we selected 221 benign and 149 malignant biopsied proven samples. All scans were isotropically re-sampled to have a uniform resolution of  $0.625 \times 0.625 \times 2 \text{ mm}^3$ . The HUs were clipped to the range of (-1000,400) and mapped to the range of (0,1).

Although NLST CT scans are read by a radiologist, the exact location of biopsied lesions are only qualitatively provided and accurate coordinate of nodules are not available. Therefore, a collaboration was initiated with the Department of Radiology at the University of Louisville to discern and extract the exact physical coordinate of nodules. Fig. 19 illustrates the axial view of four sample in the NLST database in two consecutive time points. The nodules are superimposed with the annotation provided by a thoracic radiologists who marked the boundary of the nodule in the single mid-axial slice of the nodule.

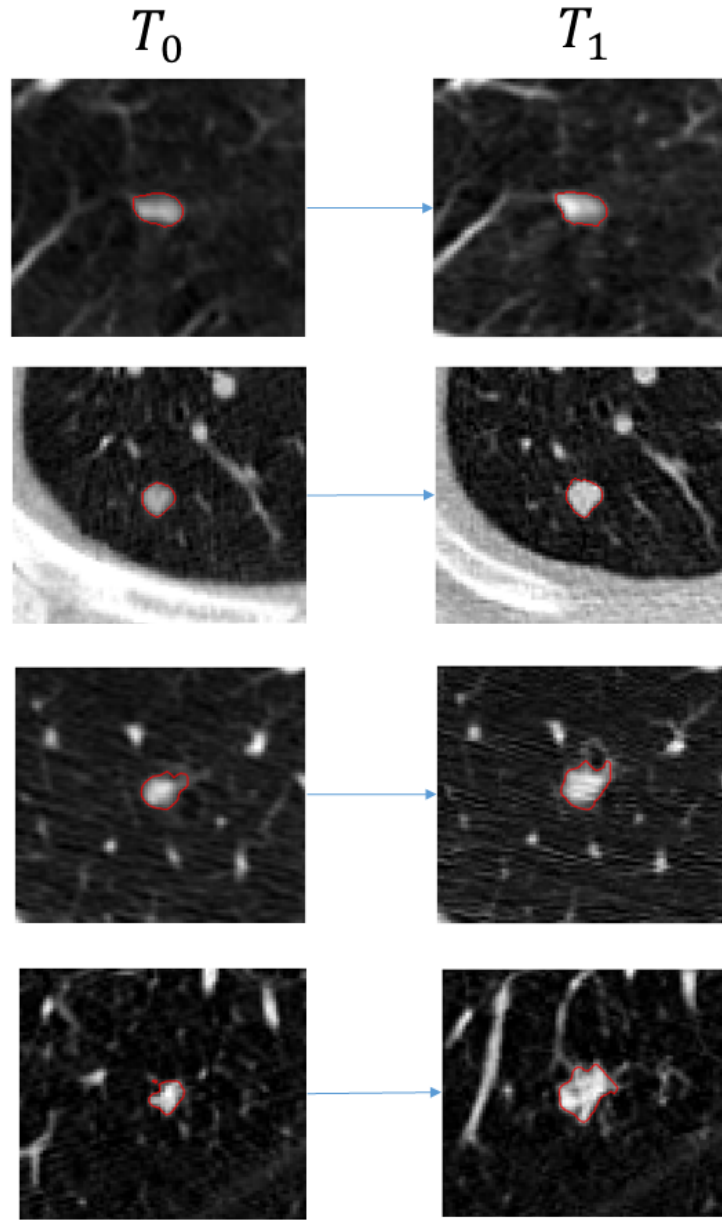
Starting from the center of the ROI, SCoTS [32] was used to segment the nodule from the rest of scan. In total, 740 samples related to two time points of 370 subjects, accounting for 221 benign and 149 malignant nodules were segmented in 3-D. As some examples, Fig. 20 shows segmented nodules (benign and malignant) in two time points. In this figure, the top two rows show samples of benign nodules and the bottom two show samples of malignant nodules.

## 2 Performance Metrics for Single and Consecutive CT Scan Classifiers

The performance of the approach to classify nodules based on visual appearances in single scan or from consecutive scans is described in this section.



**Figure 19.** Nodule ROIs marked in two time points. Top two rows show benign nodules and the next two show malignant ones.



**Figure 20.** Nodules segmented using SCoTS algorithm [32]. Top two rows show benign nodules and the next two show malignant ones.

We extracted patches of size  $20\text{ mm}^3$ ,  $30\text{ mm}^3$ , and  $40\text{ mm}^3$  from isotropically re-sampled scans centered around each nodule. To increase the size of the training data, data augmentation was performed. In details, each nodule was translated by 2 voxels along each coordinate axis and rotated in the transverse plane between  $(0,360)$  degrees interval with 10 degrees incremental step. The augmentation provided 47 additional samples for each available training nodule.

To evaluate the approach on all available samples we used 5 folds cross validation where the folds created in patient level, i.e., nodules of same patients in different time points were kept in same folds. The model was trained to learn discriminative features based on data in 4 folds and was subsequently evaluated on the nodule data in the 5<sup>th</sup> fold. The testing fold rotated 5 times in a round robin fashion while training was performed on the other folds. This approach which is called 5 fold cross validation permits the opportunity for testing the model’s accuracy over all samples. The parameters of the network were randomly initialized using Glorot initialization [36] and updated using a batch size of 32 samples. Adam optimizer [44] was used to minimize binary cross entropy loss function with a learning rate set to 0.001. The network trained end-to-end for 20 epochs.

Training was performed on NVIDIA 1080 Ti, with 12 Giga Byte memory, and 1.6 GHz GPU clock rate on a machine with 32 GB RAM memory and Radeon R7 CPU with 3.9 GHz clock rate. The algorithm developed in Python version 3.5.2 with Keras version 2.2.4 and tensorflow version 1.8.0 as deep learning libraries.

The evaluation here is in terms of Receiver Operating Characteristic (ROC) that measures the true positive rate (sensitivity) for different false positive rates. Another



metric that is used is the classification accuracy reporting the percentage of correctly diagnosed nodules for a specific threshold:

$$\frac{True\ Positives + True\ Negatives}{Total\ nubmer\ of\ samples}$$

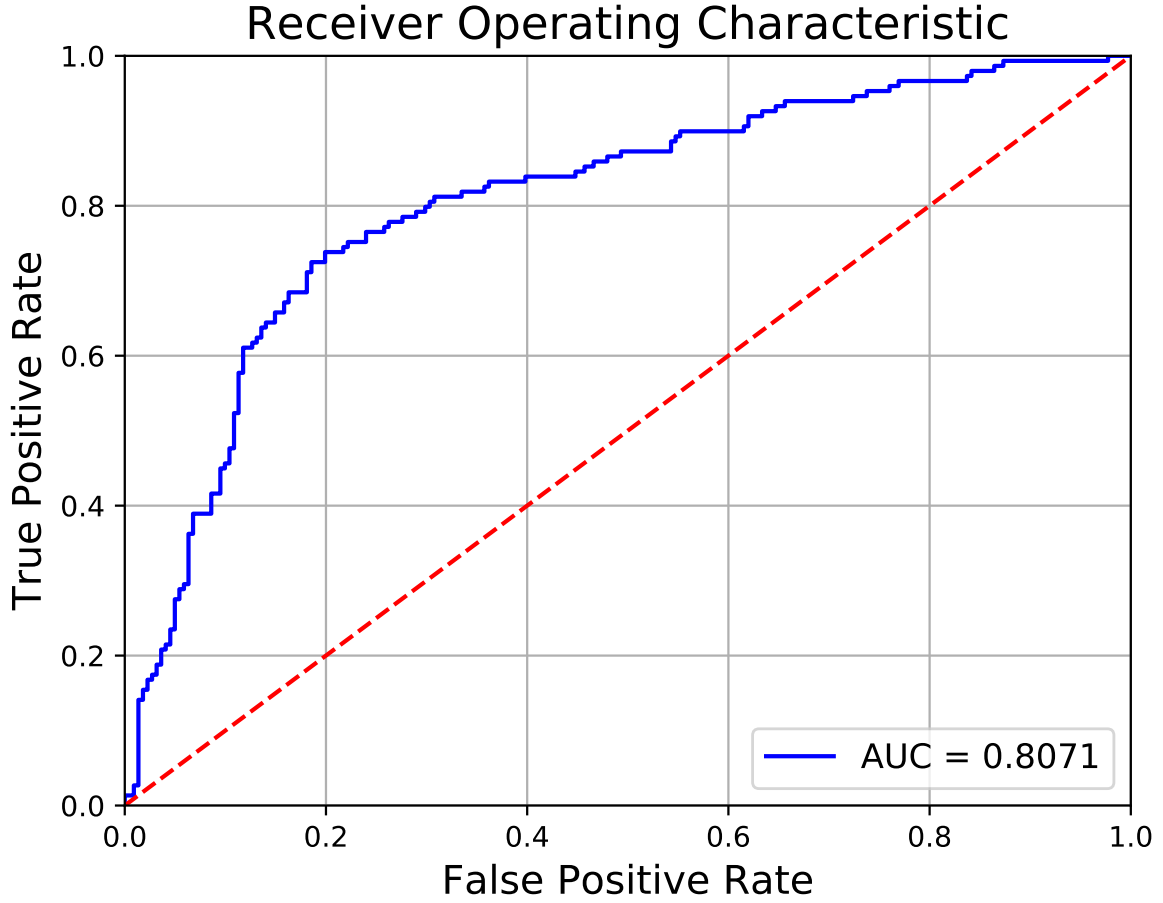
The system's performance in terms of ROC when it is trained on data from both time points and evaluated on the nodule data in time 0 is plotted in Fig. 21. The system achieved an area under the ROC curve (AUC) equal to 0.8071 and an accuracy of 76.21% with a classification threshold = 0.5.

When tested on data from time point 1 the performance improved mostly due to the fact that the appearance of nodules at time point 1 has more details. The ROC curve for this evaluation is plotted in Fig. 22. In this case the AUC boosts to 0.8673 with an accuracy of 80.27% with a classification threshold = 0.5.

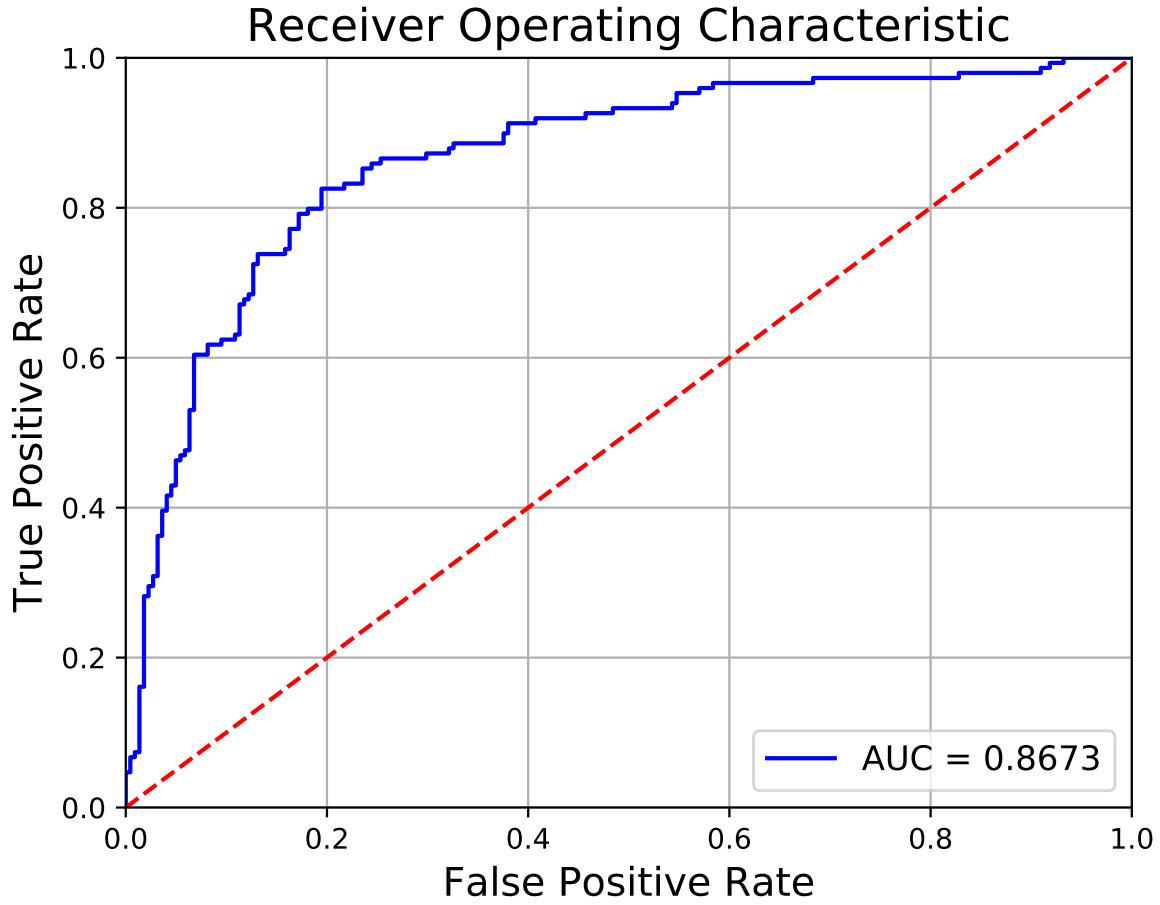
In order to capture shape variation in two time points the model needs a set of delineated nodules to construct the nodule shape dictionary. The result of evaluation of the classifier which uses malignancy score (equation 19 of chapter VI) is plotted as an ROC in Fig. 23. The area under the curve for the ROC; i.e, AUC=0.8324 with accuracy=77.30%.

The correlation among different methodologies prediction were plotted in Fig. 24 and Fig. 25 benign samples shown in dark blue and malignant ones shown in yellow. In Fig. 24 the scatter plot of scores from the RNN on time point 1 versus the RNN on time 0 is shown; the correlation coefficient is equal to 0.89 implying the two time points are highly correlated.

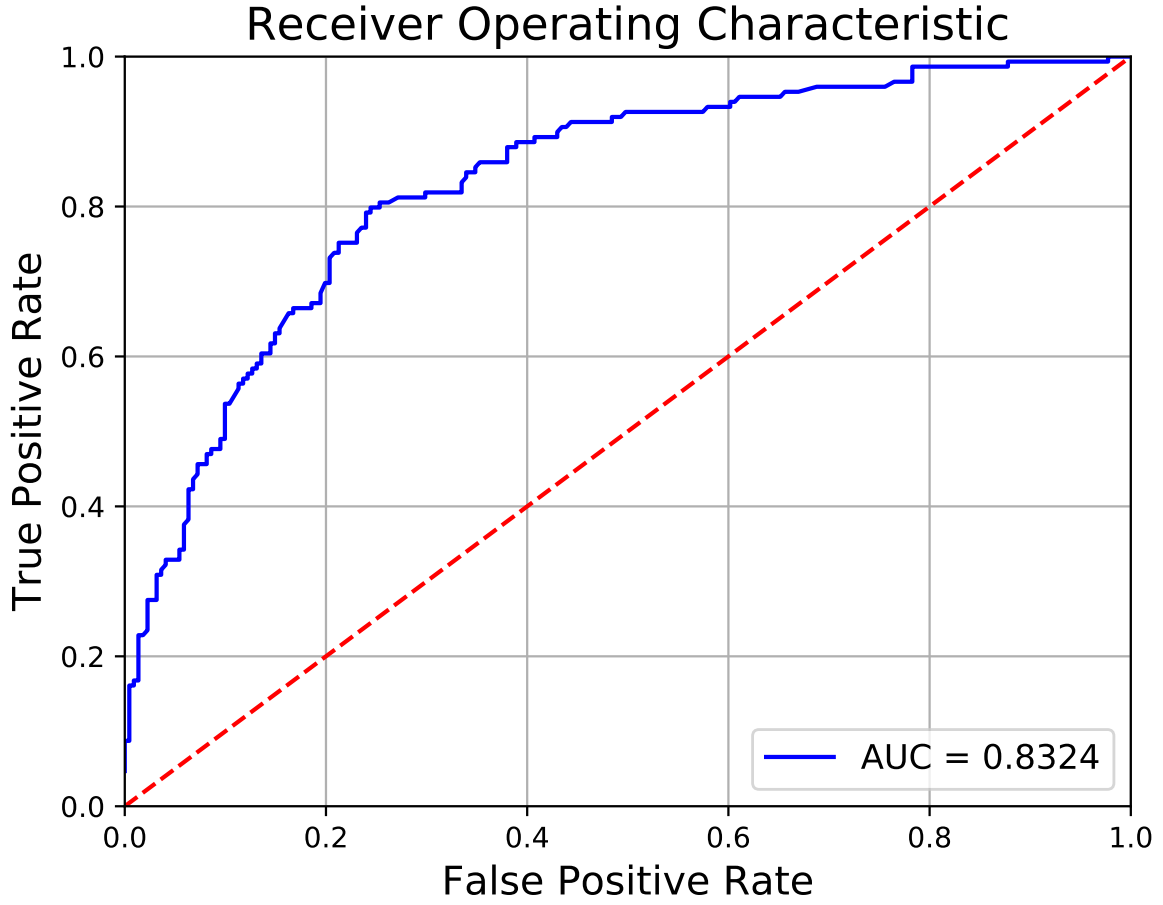
$$\rho = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 (y - m_y)^2}}$$



**Figure 21.** ROC plot for nodule classification with the proposed recurrent neural network (RNN) of Chapter VI when using visual appearance alone. Training is on CT patches from time 0 and time 1 but with testing only on time 0 with 5 fold cross validation. Testing was on 370 biopsied samples of NLST database.

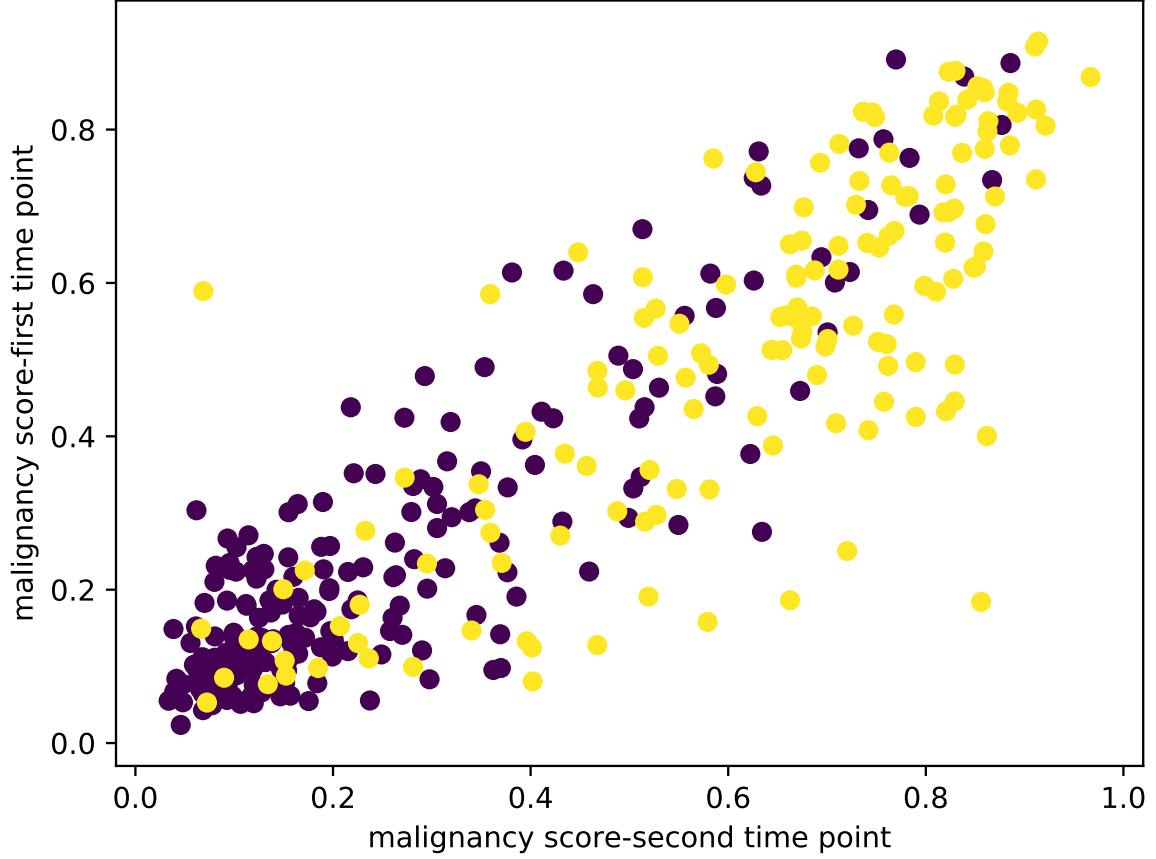


**Figure 22.** ROC plot for nodule classification with the proposed recurrent neural network (RNN) of Chapter VI when using visual appearance alone. Training is on CT patches from time 0 and time 1 but with testing only on time 1 with 5 fold cross validation. Testing was on 370 biopsied samples of NLST database.



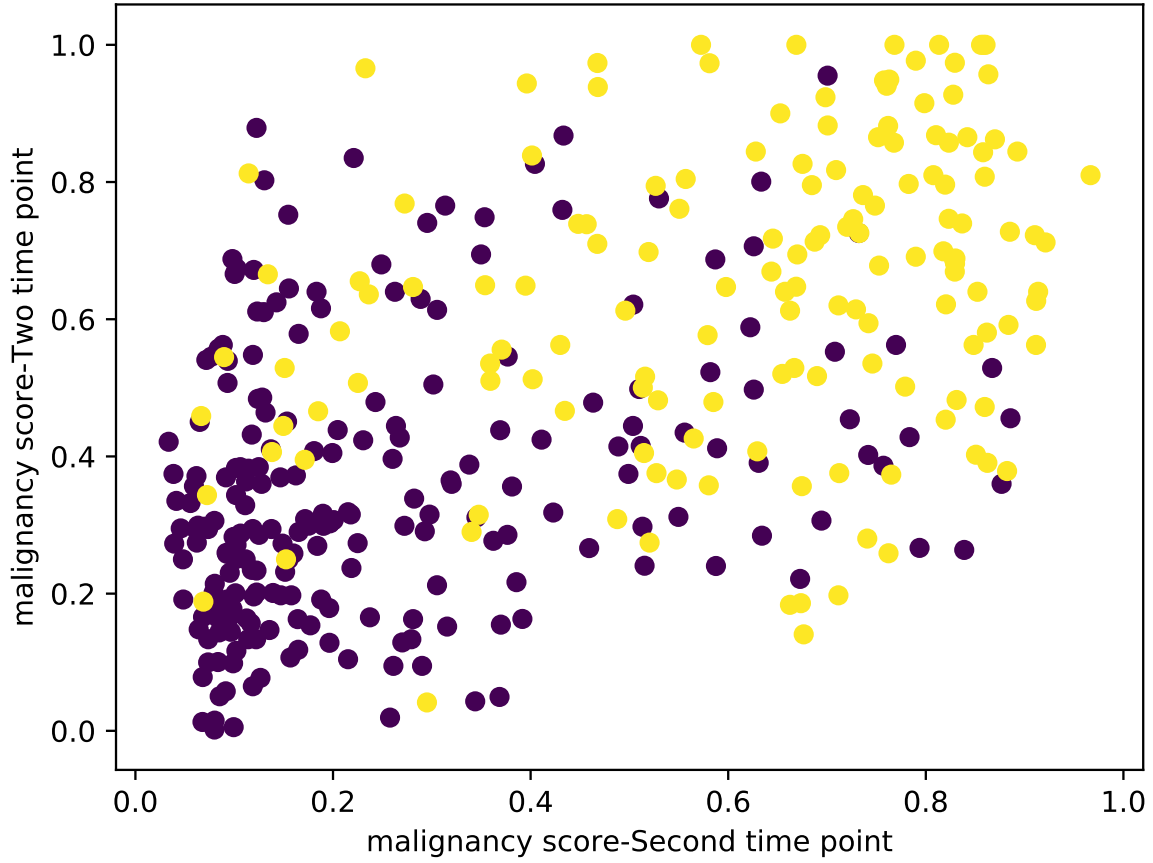
**Figure 23.** ROC for a classifier which uses equation (19) for Chapter VI to assign malignancy score utilizing corresponding nodule data from two time points.

, where  $x$  and  $y$  refer to the probability vector for the samples, and  $m_x$  and  $m_y$  are the mean of the respective probabilities for each model. On the other hand considering the correlation between RNN derived malignancy scores based on time point 1 nodule data (Fig. 25) and malignancy score based on dictionary learning utilizing CT data at two time points (equation (19) of Chapter VI) , it is observed that samples are separated in this plot with a lower correlation coefficient (equal to 0.56), suggesting an ensemble of these two classifiers can improve performance. Predictions can be

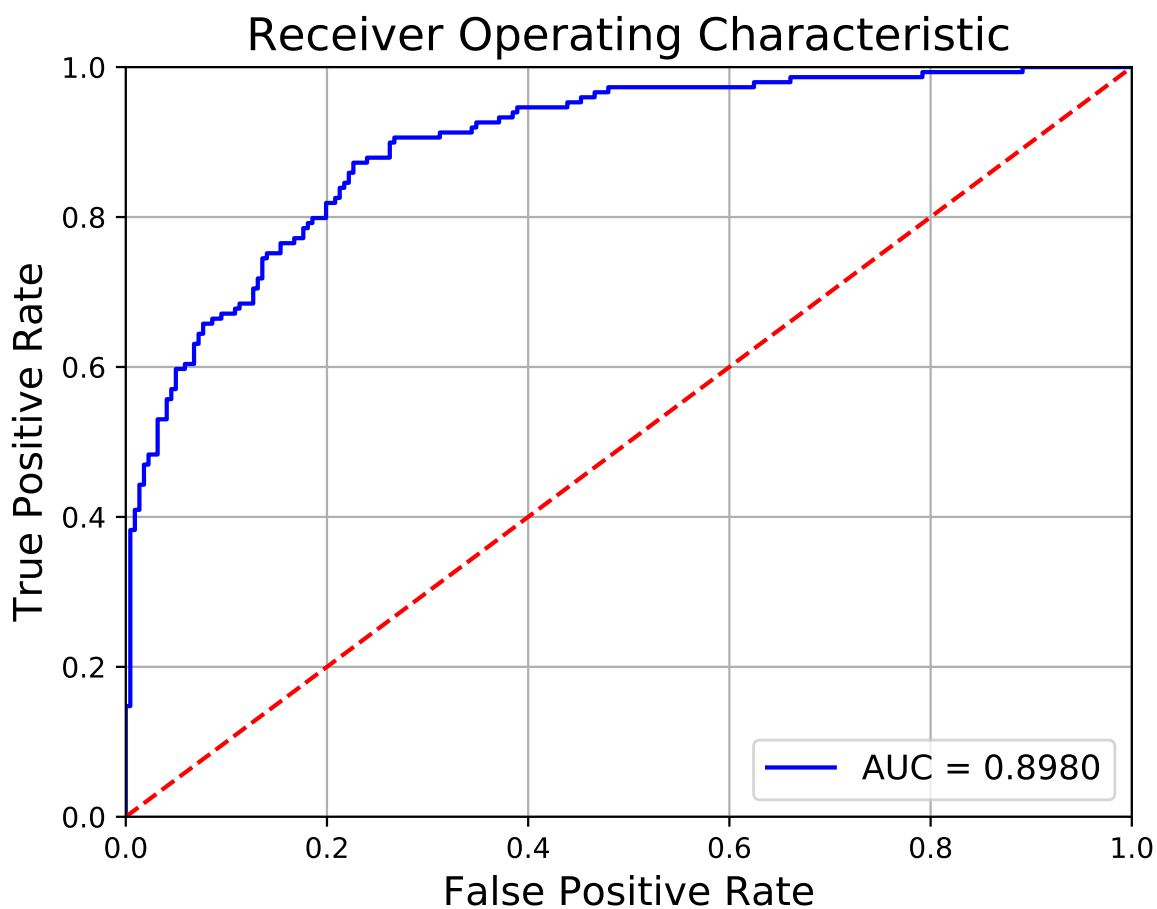


**Figure 24.** Scatter plot of malignancy prediction scores from the RNN model based on time 1 (horizontal axis) and time 0 (vertical axis) data, each calculated independently. Two methods are highly correlated with the correlation coefficient  $= 0.89$ . Each circle shows one sample out of 370 subjects benign nodules shown in dark blue and malignant samples shown in yellow.

fused with different variety of methods [77, 76]. To fuse two classifiers we simply average the malignancy scores provided by the two methods. Performance of the resulting method in terms of ROC is shown in Fig. 26 -  $AUC = 0.898$  and accuracy  $= 80.81\%$ .



**Figure 25.** Scatter plot of predicted malignancy probabilities obtained based on time 1 nodule CT derived by the RNN model versus shape change malignancy score making use of nodule CT data in time point 0 and time point 1. The scores exhibit significant scatter with a correlation coefficient = 0.56. Each circle represents one subject out of 370 with benign nodules shown in dark blue and malignant ones shown in yellow.



**Figure 26.** ROC plot for the method which fuses the RNN and Dictionary learning malignancy predictions through simple averaging of malignancy scores predicted by each.

## CHAPTER VIII

### CONCLUSIONS AND FUTURE WORK

In this thesis, a CAD system has been developed which diagnoses nodule malignancy from consecutive thoracic scans acquired one year apart. To capture nodule shape variations over time, a general framework for segmentation of anatomical shapes with application to different classes of lung nodules was presented. The method incorporates shape prior information within a sparse representation framework for active contour based segmentation of lung nodules. The discriminative nature of sparse representation, which was used to model new shapes, helped limit shape variations to a valid shape space prescribed by the training data.

The segmentation results showed that despite the generality of the algorithm, the method can work for different types of nodules; well-circumscribed, juxta-vascular, juxta-pleural, and pleural tail, achieving performance metrics close to previously published methods which were developed for segmentation of specific nodule types or required classification of nodules in advance. In comparison to prior publications, performance of SCoTS on non-solid and part-solid nodules was less competitive mainly due to the inhomogeneous texture of these nodules. We believe that although SCoTS was applied to nodule shape segmentation, it should prove general and applicable to a wide variety of medical image segmentation problems.



The proposed segmentation algorithm was used as one approach to predict malignancy of nodules from two consecutive scans. Corresponding nodules in two time points were segmented and their shape was sparsely represented by shapes in a training set. Malignancy score was then computed by comparison of the weight vectors of the sparse representation of the nodules in the two time points - as similarity between the representations determines the malignancy of the nodule. In addition to shape variation over time, a separate model based on recurrent neural networks was also developed which predicts malignancy based on visual appearance. Due to the limitation over the number of available samples in this application, we followed transfer learning, with features extracted from LUNA16 thoracic CT challenge transferred and adapted for nodule malignancy prediction. Subsequent to this, fully connected layers were trained on biopsied samples from the NLST database to learn features which distinguish malignant from benign nodules.

Future work can be pursued in three different directions. *First*, The boundary extraction of nodules in first and second time points can be improved by advanced state-of-the-art segmentations specifically developed for nodule segmentation in thoracic CT scans. Examples of these methods is U-Net [69, 16] which emerged more recently as part of the substantial success of deep learning based methods for image segmentation. In some extreme cases where nodule has significant attachment to the pleural surface and the nodule’s texture is non-solid, SCoTS fails to properly segment the nodule which is; mostly due to the generality of the algorithm. *Second*, the proposed sparse coding approach takes into account only nodule’s shape in consecutive time points. Ignoring the textural and visual information in the proposed

framework, is the main disadvantage which resulted in lower performance compared to the case where malignancy was predicted based on a single scan. As illustrated, even though shape information could still boost performance when fused with the deep learning methodology, we did not make use of all of the visual information in consecutive scans. We hypothesize that application of deep learning techniques to model sequential information with networks such as RNNs can extract more visual and temporal information in consecutive scans and further boost the performance. Another approach that can extract information effectively in two time points is the Siamese network [85]. In this framework two CNNs can work in tandem on a VOI around a nodule in both time points and extract effective features that discriminate benign from malignant nodules. *Third*, like any other deep learning application, having more data available for training can improve performance of the CAD system. By delineating more biopsied malignant and benign samples from the NLST database, and feed more data to the CAD system during the training performance can be improved.

## REFERENCES

- [1] LIDC XML Base Schema 2011.
- [2] Lung Anatomy Description. <http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm>.
- [3] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Qiang Cheng, Guoliang Chen, et al. Deep speech 2: End-to-end speech recognition in english and mandarin. In *International conference on machine learning*, pages 173–182, 2016.
- [4] Samuel G Armato, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, 38(2):915–931, 2011.
- [5] Pawel Badura and Ewa Pietka. Soft computing approach to 3D lung nodule segmentation in CT. *Computers in biology and medicine*, 53:230–243, 2014.
- [6] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997.
- [7] Jason L Causey, Junyu Zhang, Shiqian Ma, Bo Jiang, Jake A Qualls, David G Politte, Fred Prior, Shuzhong Zhang, and Xiuzhen Huang. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Scientific reports*, 8(1):9286, 2018.
- [8] Pablo G Cavalcanti, Shahram Shirani, Jacob Scharcanski, Crystal Fong, Jane Meng, Jane Castelli, and David Koff. Lung nodule segmentation in chest computed tomography using a novel background estimation method. *Quantitative imaging in medicine and surgery*, 6(1):16, 2016.

- [9] Jungwon Cha, Mohammad Mehdi Farhangi, Neal Dunlap, and Amir A Amini. Segmentation and tracking of lung nodules via graph-cuts incorporating shape prior and motion from 4D CT. *Medical physics*, 45(1):297–306, 2018.
- [10] Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- [11] Kan Chen, Bin Li, Lian-fang Tian, Wen-bo Zhu, and Ying-han Bao. Vessel attachment nodule segmentation using integrated active contour model based on fuzzy speed function and shape–intensity joint Bhattacharya distance. *Signal Processing*, 103:273–284, 2014.
- [12] Xinjian Chen and Ulas Bagci. 3D automatic anatomy segmentation based on iterative graph-cut-ASM. *Medical physics*, 38(8):4610–4622, 2011.
- [13] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Scientific reports*, 6:24454, 2016.
- [14] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [15] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [16] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432, 2016.
- [17] Francesco Ciompi, Kaman Chung, Sarah J Van Riel, Arnaud Arindra Adiyoso Setio, Paul K Gerke, Colin Jacobs, Ernst Th Scholten, Cornelia Schaefer-Prokop, Mathilde MW Wille, Alfonso Marchiano, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Scientific reports*, 7:46479, 2017.

- [18] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.
- [19] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [20] Daniel Cremers. Image segmentation with shape priors: Explicit versus implicit representations. *Handbook of Mathematical Methods in Imaging*, pages 1909–1944, 2015.
- [21] Daniel Cremers, Stanley J Osher, and Stefano Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International journal of computer vision*, 69(3):335–351, 2006.
- [22] Jamshid Dehmehski, Hamdan Amin, Manlio Valdivieso, and Xujiong Ye. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. *IEEE Transactions on Medical Imaging*, 27(4):467–480, 2008.
- [23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [24] Raunak Dey, Zhongjie Lu, and Yi Hong. Diagnostic classification of lung nodules using 3D neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 774–778. IEEE, 2018.
- [25] Stefano Diciotti, Simone Lombardo, Massimo Falchini, Giulia Picozzi, and Mario Mascalchi. Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. *IEEE Transactions on Biomedical Engineering*, 58(12):3418–3428, 2011.
- [26] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*, 170:456–470, 2018.
- [27] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceed-*

ings of the *IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

- [28] David L Donoho. For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.
- [29] Nicolae Duta and Milan Sonka. Segmentation and interpretation of mr brain images. an improved active shape model. *IEEE Transactions on Medical Imaging*, 17(6):1049–1062, 1998.
- [30] Amal Farag, Shireen Elhabian, James Graham, Aly Farag, Salwa Elshazly, Robert Falk, Hani Mahdi, Hossam Abdelmunim, and Sahar Al-Ghaafary. Modeling of the lung nodules for detection in LDCT scans. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 3618–3621. IEEE, 2010.
- [31] Amal A Farag, Hossam E Abd El Munim, James H Graham, and Aly A Farag. A novel approach for lung nodules segmentation in chest CT using level sets. *IEEE Transactions on Image Processing*, 22(12):5202–5213, 2013.
- [32] M Mehdi Farhangi, Hichem Frigui, Albert Seow, and Amir A Amini. 3-D active contour segmentation based on sparse linear combination of training shapes (SCoTS). *IEEE Transactions on Medical Imaging*, 36(11):2239–2249, 2017.
- [33] Mohammad M Farhangi, Hichem Frigui, Robert Bert, and Amir A Amini. Incorporating shape prior into active contours with a sparse linear combination of training shapes: Application to corpus callosum segmentation. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6449–6452. IEEE, 2016.
- [34] Wenjiang J Fu. Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics*, 7(3):397–416, 1998.
- [35] Mohsen Ghafoorian, Nico Karssemeijer, Tom Heskes, Mayra Bergkamp, Joost Wissink, Jiri Obels, Karlijn Keizer, Frank-Erik de Leeuw, Bram van Ginneken, Elena Marchiori, et al. Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. *NeuroImage: Clinical*, 14:391–399, 2017.

- [36] 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.
- [37] Damien Grosgeorge, Caroline Petitjean, J-N Dacher, and Su Ruan. Graph cut segmentation with a statistical shape model in cardiac MRI. *Computer Vision and Image Understanding*, 117(9):1027–1035, 2013.
- [38] Sardar Hamidian, Berkman Sahiner, Nicholas Petrick, and Aria Pezeshk. 3D convolutional neural network for automatic detection of lung nodules in chest CT. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 1013409. International Society for Optics and Photonics, 2017.
- [39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [40] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [41] Sarfaraz Hussein, Kunlin Cao, Qi Song, and Ulas Bagci. Risk stratification of lung nodules using 3d cnn-based multi-task learning. In *International conference on information processing in medical imaging*, pages 249–260. Springer, 2017.
- [42] Hongyang Jiang, He Ma, Wei Qian, Mengdi Gao, and Yan Li. An automatic detection system of lung nodule based on multigroup patch-based deep learning network. *IEEE journal of biomedical and health informatics*, 22(4):1227–1237, 2018.
- [43] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [44] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [45] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.
- [46] William J Kostis, Anthony P Reeves, David F Yankelevitz, Claudia I Henschke, et al. Three-dimensional segmentation and growth-rate estimation of

- small pulmonary nodules in helical CT images. *IEEE Trans. Med. Imaging*, 22(10):1259–1274, 2003.
- [47] 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.
  - [48] J-M Kuhnigk, Volker Dicken, Lars Bornemann, Annemarie Bakai, Dag Wormanns, Stefan Krass, and H-O Peitgen. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Transactions on Medical Imaging*, 25(4):417–434, 2006.
  - [49] BC Lassen, C Jacobs, JM Kuhnigk, B van Ginneken, and EM van Rikxoort. Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans. *Physics in Medicine & Biology*, 60(3):1307, 2015.
  - [50] Michael E Leventon, W Eric L Grimson, and Olivier Faugeras. Statistical shape influence in geodesic active contours. In *Computer vision and pattern recognition, 2000. Proceedings. IEEE conference on*, volume 1, pages 316–323. IEEE, 2000.
  - [51] Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 305–312, 2014.
  - [52] 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.
  - [53] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermesen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, 6:26286, 2016.
  - [54] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.



- [55] Jinlian Ma, Fa Wu, Jiang Zhu, Dong Xu, and Dexing Kong. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics*, 73:221–230, 2017.
- [56] Heber MacMahon, John HM Austin, Gordon Gamsu, Christian J Herold, James R Jett, David P Naidich, Edward F Patz Jr, and Stephen J Swensen. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*, 237(2):395–400, 2005.
- [57] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Fourth IEEE International Conference on 3D Vision (3DV)*, pages 565–571, 2016.
- [58] Jan Hendrik Moltz, Jan-Martin Kuhnigk, Lars Bornemann, and H Peitgen. Segmentation of juxtapleural lung nodules in ct scan based on ellipsoid approximation. In *Proceedings of First International Workshop on Pulmonary Image Processing. New York*, pages 25–32, 2008.
- [59] Sudipta Mukhopadhyay. A segmentation framework of pulmonary nodules in lung CT images. *Journal of digital imaging*, 29(1):86–103, 2016.
- [60] Keelin Murphy, Bram van Ginneken, Arnold MR Schilham, BJ De Hoop, HA Gietema, and Mathias Prokop. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Medical image analysis*, 13(5):757–770, 2009.
- [61] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.
- [62] National Lung Screening Trial Research Team. The national lung screening trial: overview and study design. *Radiology*, 258(1):243–253, 2011.
- [63] Aiden Nibali, Zhen He, and Dennis Wollersheim. Pulmonary nodule classification with deep residual networks. *International journal of computer assisted radiology and surgery*, 12(10):1799–1808, 2017.
- [64] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

- [65] D Maxwell Parkin. Global cancer statistics in the year 2000. *The lancet oncology*, 2(9):533–543, 2001.
- [66] Aria Pezeshk, Sardar Hamidian, Nicholas Petrick, and Berkman Sahiner. 3D convolutional neural networks for automatic detection of pulmonary nodules in chest CT. *IEEE journal of biomedical and health informatics*, 2018.
- [67] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [68] Anthony P Reeves, Antoni B Chan, David F Yankelevitz, Claudia I Henschke, Bryan Kressler, and William J Kostis. On measuring the change in size of pulmonary nodules. *IEEE Transactions on Medical imaging*, 25(4):435–450, 2006.
- [69] 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, 2015.
- [70] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [71] Holger R Roth, Le Lu, Jiamin Liu, Jianhua Yao, Ari Seff, Kevin Cherry, Lauren Kim, and Ronald M Summers. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Transactions on Medical Imaging*, 35(5):1170–1181, 2016.
- [72] Mark Schmidt. Least squares optimization with L1-norm regularization. *CS542B Project Report*, 504:195–221, 2005.
- [73] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.
- [74] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J Van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sánchez, and Bram van Ginneken. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169, 2016.

- [75] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas van den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Medical image analysis*, 42:1–13, 2017.
- [76] AliAsghar ShahrjooiHaghighi, Hichem Frigui, Xiang Zhang, Xiaoli Wei, Biyun Shi, and Craig J. McClain. Ensemble Feature Selection for Biomarker Discovery in Mass Spectrometry-based Metabolomics. In *Proceedings of the 34th Annual ACM Symposium on Applied Computing*. ACM, 2019.
- [77] Aliasghar Shahrjooihaghighi, Hichem Frigui, Xiang Zhang, Xiaoli Wei, Biyun Shi, and Ameni Trabelsi. An ensemble feature selection method for biomarker discovery. In *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 416–421. IEEE, 2017.
- [78] Dinggang Shen, Guorong Wu, and Heung-II Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.
- [79] 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.
- [80] Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer’s disease. *IEEE journal of biomedical and health informatics*, 22(1):173–183, 2018.
- [81] 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 Transactions on Medical Imaging*, 35(5):1285–1298, 2016.
- [82] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [83] Nannan Sun, Dongbao Yang, Shancheng Fang, and Hongtao Xie. Deep Convolutional Nets for Pulmonary Nodule Detection and Classification. In *Interna-*

*tional Conference on Knowledge Science, Engineering and Management*, pages 197–208. Springer, 2018.

- [84] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [85] Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.
- [86] 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.
- [87] Maxine Tan, Rudi Deklerck, Bart Jansen, Michel Bister, and Jan Cornelis. A novel computer-aided lung nodule detection system for CT images. *Medical physics*, 38(10):5630–5645, 2011.
- [88] Atsushi Teramoto, Hiroshi Fujita, Osamu Yamamuro, and Tsuneo Tamaki. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Medical physics*, 43(6Part1):2821–2827, 2016.
- [89] Andy Tsai, Anthony Yezzi, William Wells III, Clare Tempany, Dewey Tucker, Ayres Fan, W Eric Grimson, and Alan S Willsky. A shape-based approach to the segmentation of medical imagery using level sets. 2003.
- [90] Paul Tseng et al. Coordinate ascent for maximizing nondifferentiable concave functions. 1988.
- [91] Bram Van Ginneken. Supervised probabilistic segmentation of pulmonary nodules in CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 912–919. Springer, 2006.
- [92] Bram Van Ginneken, Alejandro F Frangi, Joes J Staal, Bart M ter Haar Romeny, and Max A Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, 2002.

- [93] Bram Van Ginneken, Arnaud AA Setio, Colin Jacobs, and Francesco Ciompi. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In *2015 IEEE 12th International symposium on biomedical imaging (ISBI)*, pages 286–289. IEEE, 2015.
- [94] Simon K Warfield, Kelly H Zou, and William M Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903, 2004.
- [95] Ted W Way, Lubomir M Hadjiiski, Berkman Sahiner, Heang-Ping Chan, Philip N Cascade, Ella A Kazerooni, Naama Bogot, and Chuan Zhou. Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours. *Medical physics*, 33(7Part1):2323–2337, 2006.
- [96] Jung won Cha, M Mehdi Farhangi, Neal Dunlap, and Amir Amini. Volumetric analysis of respiratory gated whole lung and liver CT data with motion-constrained graph cuts segmentation. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pages 3405–3408. IEEE, 2017.
- [97] Jung won Cha, Mohammad M Farhangi, Neal Dunlap, and Amir Amini. 4D lung tumor segmentation via shape prior and motion cues. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 1284–1287. IEEE, 2016.
- [98] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.
- [99] Hongtao Xie, Dongbao Yang, Nannan Sun, Zhineng Chen, and Yongdong Zhang. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition*, 85:109–119, 2019.
- [100] Fuyong Xing and Lin Yang. Robust selection-based sparse shape model for lung cancer image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 404–412. Springer, 2013.
- [101] Xujiang Ye, Gareth Beddoe, and Greg Slabaugh. Automatic graph cut segmentation of lesions in CT using mean shift superpixels. *Journal of Biomedical Imaging*, 2010:19, 2010.

- [102] Yading Yuan, Ming Chao, and Yeh-Chi Lo. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Transactions on Medical Imaging*, 36(9):1876–1886, 2017.
- [103] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris N Metaxas, and Xiang Sean Zhou. Deformable segmentation via sparse shape representation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 451–458. Springer, 2011.
- [104] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris N Metaxas, and Xiang Sean Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical image analysis*, 16(1):265–277, 2012.

## Appendix A: Commonly Used Acronyms

ASM - Active Shape Model

AUC - Area Under Curve

CNN - Convolutional Neural Network

CADe - Computer Aided Detection

CADx - Computer Aided Diagnosis

CT - Computed Tomography

DSC - Dice Similarity Coefficient

FPR - False Positive Rate

GRU - Gated Recurrent Unit

HU - Hounsfield Unit

LDCT - Low Dose Computed Tomography

LSTM - Long Short Term Memory

MRI - Magnetic Resonance Imaging

NLST - National Lung Screening Trial

PCA - Principal Component Analysis

RNN - Recurrent Neural Networks

ROC - Receiver operating characteristic

ROI - Region of Interest

SCoTS - Sparse Combination of Training Shapes

SDF - Signed Distance Function

SSC - Sparse Shape Composition

STAPLE - Simultaneous Truth and Performance Level Estimation

TPR - True Positive Rate

VOI - Volume of Interest



## CURRICULUM VITA

Mohammad Mehdi Farhangi  
Medical Imaging Lab, Lutz Hall, Room 410  
University of Louisville, Louisville, KY 40292, USA  
Email: m0farh03@louisville.edu

### Education

Ph.D., University of Louisville, Louisville, KY, April 2019

M.S., Iran University of Science and Technology, Terhan, Iran, January 2012

B.S., Ferdowsi University of Mashhad, Mashhad, Iran, September 2009

### Research Experience

- Summer Intern: Division of Imaging, Diagnostics, and Software Reliability (DIDSR), Food and Drug Administration, May 2018 - December 2019
- Research Assistant: Medical Imaging Lab, University of Louisville, KY, 2014 - 2019
- Research Assistant: Computer Vision Lab, Iran University of Science and Technology, 2009 - 2012

### Publications and Presentations

1. Cha, J., **Farhangi, M. M.**, Dunlap, N., & Amini, A. A. (2018). Segmentation and tracking of lung nodules via graphcuts incorporating shape prior and motion from 4D CT. *Medical physics*, 45(1), 297-306.

2. **Farhangi, M. M.**, Frigui, H., Seow, A., & Amini, A. A. (2017). 3-D active contour segmentation based on sparse linear combination of training shapes (scots). *IEEE Transactions on Medical Imaging*, 36(11), 2239-2249.
3. won Cha, J., **Farhangi, M. M.**, Dunlap, N., & Amini, A. (2017, July). Volumetric analysis of respiratory gated whole lung and liver CT data with motion-constrained graph cuts segmentation. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 3405-3408). IEEE.
4. won Cha, J., **Farhangi, M. M.**, Dunlap, N., & Amini, A. (2016, August). 4D lung tumor segmentation via shape prior and motion cues. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1284-1287). IEEE.
5. **Farhangi, M. M.**, Frigui, H., Bert, R., & Amini, A. A. (2016, August). Incorporating shape prior into active contours with a sparse linear combination of training shapes: Application to corpus callosum segmentation. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 6449-6452). IEEE.