Machine Learning Driven Active Surfaces for 3D Segmentation of Tumour Lesions in PET Images

by

Payam Ahmadvand

B.Eng., Shahid Rajaee University, 2012

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Summer 2017

APPROVAL

Name: Payam Ahmadvand

Degree: Master of Science

Title of Thesis: Machine Learning Driven Active Surfaces for 3D Segmentation of Tumour Lesions in PET Images

Examining Committee:

Dr. Ze-Nian Li
Professor
Computing Science, Simon Fraser University
Chair

Dr. Ghassan Hamarneh,
Senior Supervisor
Professor
Computing Science, Simon Fraser University

Dr. Ping Tan,
Supervisor
Associate Professor
Computing Science, Simon Fraser University

Dr. Mark S. Drew,
SFU Examiner
Professor
Computing Science, Simon Fraser University

Date Approved: May 9th, 2017

Abstract

One of the key challenges facing wider adoption of positron emission tomography (PET) as an imaging biomarker of disease is the development of reproducible quantitative image interpretation tools. Quantifying changes in tumor tissue, due to disease progression or treatment regimen, often requires accurate and reproducible delineation of lesions. Lesion segmentation is necessary for measuring tumor proliferation/shrinkage and radiotracer-uptake to quantify tumor metabolism. In this thesis, we develop an active surface model for segmenting lesions from PET images. We first implemented a non-convex level set active surface method with likelihood terms trained on manually-collected seeds points. We evaluated this approach on the following datasets: Images of phantoms collected by our collaborators at UBC; Quantitative Imaging Network (QIN) phantom images; and images of real patients from the QIN Head and Neck challenge. Secondly, to avoid user interaction, we developed an improved version of our method by training a machine learning system on anatomically and physiologically meaningful imaging cues to distinguish normal organ activity from tumorous lesion activity. Then, the inferred lesion likelihoods are used to guide a convex active surface segmentation model. The result is a lesion segmentation method that does not require user-initialization, manual seeding, or parameter-tweaking and, thus, guaranteeing reproducible results. We tested this enhanced method on data from the Cancer Imaging Archive. Our method not only produces more accurate segmentation than state-of-the-art segmentation results, but also does not need any user interaction.

Keywords:
Machine Learning, Segmentation, Active Contour Model, Functional Imaging, Positron Emission Tomography (PET), Head and Neck Cancer.

To the memory of my father, Mozafar Ahmadvand

“You were born with wings, why prefer to crawl through life?”

— Molana Jalaluddin (Rumi)

Acknowledgments

I am very grateful for the support of many wonderful people.

I would like to thank Dr. Ghassan Hamarneh, my senior supervisor who supported, taught, and encouraged me in many different ways leading to the accomplishments of my thesis. I deeply appreciate Dr. Ping Tan, my supervisor, Dr. Mark S. Drew, the internal examiner, and Dr. Ze-Nian Li, the defence chair for their valuable time as members of my defence committee.

I want to thank my collaborators in the Vancouver Quantitative Imaging Network (QIN) team: Dr. Francois Benard, Dr. Anna Celler, Dr. Noirin Duggan, Dr. Jesse Tanguay and Hillgan Ma for their great discussions, and fruitful collaboration. I have been fortunate to work in an amazing group with wonderful friends in the Medical Image Analysis Lab (MIAL) at SFU. Especially, I wish to express my gratitude to Dr. Seyed Masoud Nosrati for his full support and kindness in sharing his experiences. Many thanks go to Aïcha BenTaieb, Colin Brown, Jeremy Kawahara, and Saeedeh Afshari who generously took time out of their busy schedules to provide helpful tips and many inspirational discussions. I would also like to thank Dr. Lisa Tang for her wonderful friendship, support, and great advice.

Finally, I would like to thank my mother, Giti Hassanpour and my brother, Pouya Ahmadvand for their support, sacrifices, and unconditional love.

Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Fundamentals

1.1.1 Position Emission Tomography (PET)

PET is an imaging modality that captures functional processes inside the body. It is commonly used in cardiology and neurology for detection and treatment of inflammation and infections, as well as in oncology for diagnosis, staging, and treatment of cancer. PET images show the distribution and concentration of radiotracer uptake, with regions of high uptake typically visualized as brighter pixels compared to normal tissue. One of the key challenges in PET image analysis is producing quantitative measures of tumor characteristics, such as size, shape, and location of the tumor, which are needed in order to precisely localize radiation doses administered in radiation therapy or for evaluating treatment efficacy. A critical challenge towards quantitative imaging is the ability to distinguish normal activity, e.g. in the heart, brain, bladder, and kidneys, from abnormal activity due to the presence of lesions (see Fig. 1.1). The aim of this thesis is to outline a method for automatic localization and segmentation of lesions from PET.

Radioactive tracing

Acquiring a PET scan is typically accompanied by the administration of a radioactive tracer with a short half-life. Such tracer is produced in a cyclotron. The most common radiotracer used in current clinical routine is a fluorine-18 (18F) isotope labeled glucose (18F-FDG –Fluorodeoxyglucose). The radioactive tracer is applied to the patient by intravenous injection. Before the image acquisition, an appropriate resting time for the patient is required. This can guarantee a sufficient distribution of the radioactive tracer inside the body where the tracer 18F-FDG is taken up in the same way as glucose without radioactive tracer. Since malignant tumor cells are fast growing cells, higher uptake

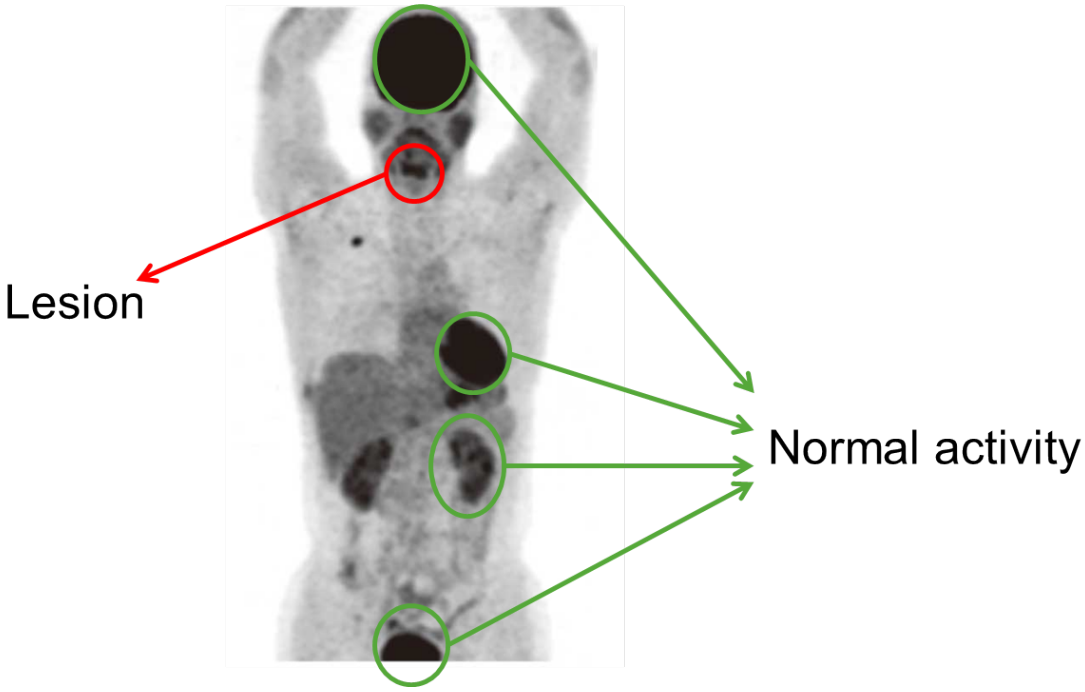


Figure 1.1: A PET image showing normal activity in (from top to bottom, or superior to inferior) the brain, heart, kidneys, and bladder (green) vs lesions located in the head and neck area (red).

of 18F-FDG would take place in lesions. Thus, the location of the lesions will exhibit stronger signal (or brighter pixels) compared to normal cells in PET images [46].

Coincidence detection and imaging

Positron-emitting isotopes are used in PET. Positrons are positively charged anti-electrons produced by unstable radioisotopes. As positrons travel through the tissue, annihilation occurs when positrons interact with electrons, lose kinetic energy, and result in gamma photons moving in opposite directions (anti-parallel). This emission of gamma rays plays a key role in PET imaging. A ring detector surrounds the subject being imaged. This detector is called a coincidence detector since it is designed to detected these anti-parallel gamma rays. If two detected rays are in coincidence, the annihilation of the positron is taking place somewhere between the detectors [32]. All registered coincidences from each detector pair are counted to form a sinogram. Finally, an image reconstruction algorithm is used to transform the sinogram into a 3D PET image (see Fig. 1.2) [19].

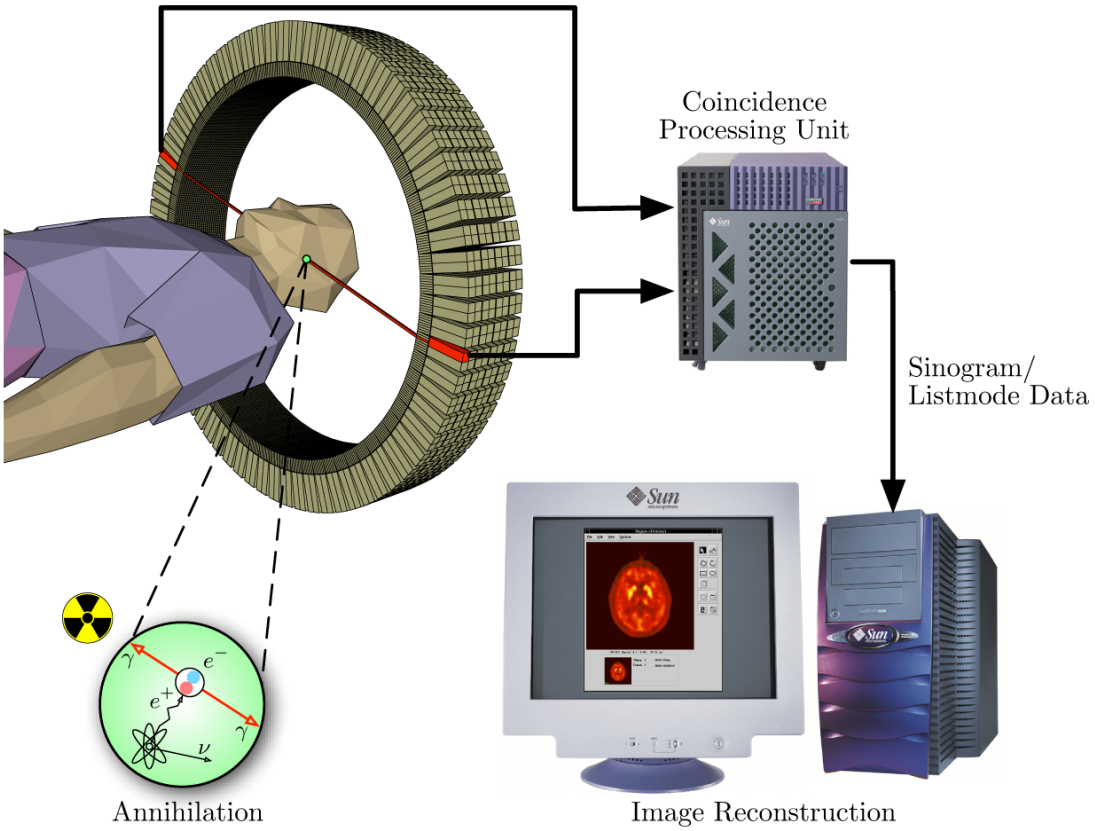


Figure 1.2: The PET scanner structure. Two different detectors get photons emitted during annihilation. The sum of all coincidence events passes into the coincidence processing unit. Finally, a 3D volume is reconstructed from this sinogram [40].

1.1.2 Quantification of PET images

Diagnosis and staging of lesions can be achieved by an oncologist through visual inspection. However, this qualitative analysis for identification of tumour brings limitations to the prediction and assessment of treatment. Quantitative analysis is more suitable because it provides the objectivity that is required for these applications.

Over the past two decades, several quantitative methods have been presented. The most accepted of these indices is the Standardized Uptake Value (SUV). SUV describes the ratio between the concentration of the radiotracer in a region of interest (ROI) and the injected activity (from physical decay) which is normalized by the normalization factor:

$$SUV = \frac{radiotracer\ concentration(\frac{kBq}{ml})}{\frac{injected\ activity(MBq)}{normalization\ factor}} \tag{1.1}$$

The normalization factor takes into account that the radioactive tracer distribution in the patient's body is affected by physique. Body weight (bw), body surface area (bsm), and lean body mass (lbm) are a common normalization factor.

1.1.3 Validation of tumor segmentation

In order to validate a proposed segmentation method, ideally one needs access to ground truth segmentations. However, the absolute truth is not available for real patient data. Generally, expert segmentation is considered as ground truth in the validation of PET segmentation methods. The Dice similarity coefficient is one well known metric for evaluating segmentation, which measures the size of the overlap of the two different segmentations divided by the total size of the two contours or objects. Usually, tumor segmentation results are validated against contours drawn on real images by experienced oncologist.

Physical and digital phantoms are constructed to create a known truth to validate segmentation methods. Therefore, by using these phantom or synthetic images, in which ground truth is known, a segmentation method can be validated.

In order to simulate tumors inside the body, a physical phantom may contain different sized compartments (e.g. spheres) with radiotracers injected into them. The ground truth in physical and digital phantoms is calculated as that total radioactivity in these compartments.

1.2 PET image segmentation techniques

Among the methods proposed for PET segmentation, thresholding methods continue to be popular. In [59, 49, 10, 15, 21] various fixed thresholding methods were proposed. A common approach is to threshold the volume based on the SUV. In [49, 10] it was proposed to use 40-43% of the maximum SUV value as a threshold, however no consensus exists concerning a fixed thresholding level. Depending on the scanner type, reconstruction algorithm and noise, the level may need to be adjusted significantly. Overestimation of lesion boundaries (a larger contour abound the object) is also a problem associated with fixed thresholding methods [23]. Methods for adjusting the threshold level based on the relationship between the true lesion volume and estimated volume with respect to the quality characteristics of the scanner output have also been proposed, however these require an initial estimate of the tumor which is not always available. The choice of parameters used to estimate the threshold range from the estimated lesion volume [21] to more complex analytic expressions

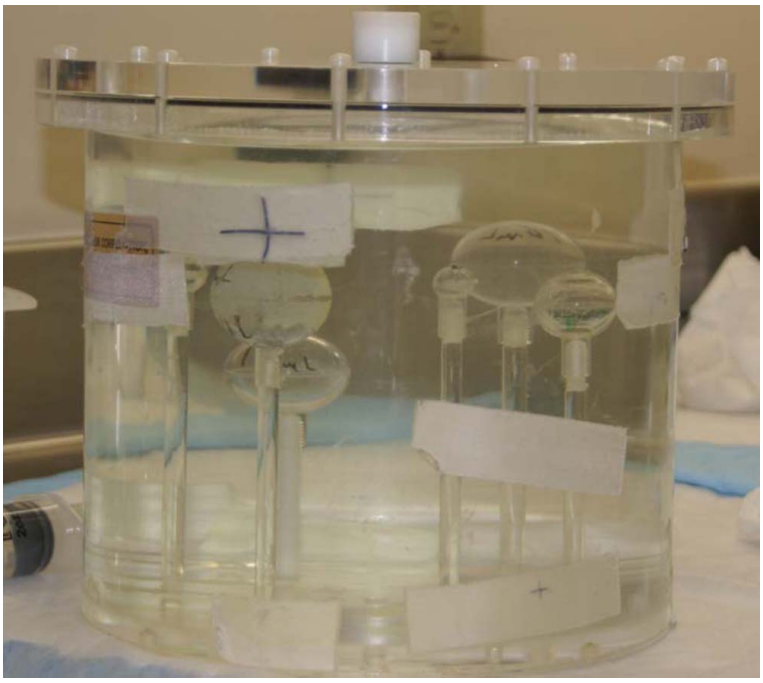


Figure 1.3: A sample of phantom with six spheres in a cylindrical container [44].

involving the source-to-background ratio and full width half maximum (FWHM) [1] of the scanner [48]. Adaptive thresholding methods suffer from some of same limitations as fixed threshold methods, *i.e.* thresholds are specific to different scanners.

More sophisticated thresholding based approaches have been proposed based on stochastic and learning-based methods. In [24] a thresholding based segmentation method was proposed for PET segmentation based on Affinity Propagation (AP) clustering algorithm. First, to create a region of interest for analysis, the lung region is segmented from CT images via region growing algorithm, which is then used to mask the relevant portion of the PET volume. They next estimate the probability density function of the pulmonary region using a kernel density estimation method [13], which is then smoothed. An AP clustering algorithm is then used to approximate clusters and thereby estimate optimal thresholds for segmenting uptake regions. They define the similarity metric of the AP algorithm to be a function of the pairwise probability difference between voxels which is obtained using the estimated pdf (probability density function) as well as the intensity difference. When validated against expert delineations of PET images of rabbits infected with tuberculosis, the method obtained a dice similarity coefficient of $91.25 \pm 8.01\%$. Although initially proposed for

[1] FWHM is a function that gets two independent variables and returns the difference between the two values at which the dependent variable is half of its maximum value

the segmentation of pulmonary infections which often results in spatially diffuse uptake in PET images, in [60] the method was combined with a denoising method and proposed for general PET segmentation.

Extending an earlier work [30], in [29] an algorithm was proposed for PET volume estimation by incorporating fuzzy measures into a Bayesian-based classification. Their approach is called fuzzy locally adaptive Bayesian (FLAB). Previously using one fuzzy measure, in the extended work [29] the proposed algorithm uses 3 'hard' classes and 3 fuzzy levels to define the tumor region. The fuzzy levels are used to describe the membership of a given voxel when its posterior distribution, estimated from maximum posterior likelihood, belongs to a fuzzy domain. The method was not proposed for automatic localization of lesions and requires initialization in the form of a bounding box around the lesion. In [41], a fuzzy C-means (FCM) clustering based approach was proposed for PET segmentation. The method proposes a generalization of the FCM where the Euclidean norm is replaced by kernel-induced distance measure based on the L-p norm, with the value of p automatically estimated based on the data. Validation was carried out on phantom images as well as 9 non-small cell lung cancer (NSCLC) tumors. Higher accuracy compared to FLAB [29] was reported for tumors with heterogeneous uptake, and geometrically complex shapes. Since the method relies on gradient descent algorithm for optimization, the solution is only guaranteed to be locally optimal. The method also requires a volume of interest (VOI) to be defined. In [43], a method based on belief function theory was proposed. They model the segmentation problem as a binary assignment of either low take region (lu) or high uptake region (hu), where the frame of discernment is defined as $\Omega = \{lu, hu\}$, and each voxel in a defined volume of interest represents the information source. They use the FCM algorithm to determine belief masses for each voxel, where assignment for a specific voxel is made based on regional statistics around that voxel. To refine assignments, they use Dempster's rule of the combination on regions of neighboring voxels. Finally, each voxel is assigned to the class either $\{lu\}$ or $\{hu\}$ with the largest belief mass. The method was tested on anthropomorphic phantoms and patient data.

In [20], a method for lesion segmentation in PET based on a combination of the maximum of intensity projections (MIP) and possibility theory was proposed. The method of MIPs was used to achieve better contrast, while the use of possibility theory provided a means to represent uncertainty in the transition between healthy and lesion tissue. The authors report that the method does not globally achieve superior results over some adaptive thresholding methods [49]. For initialization, the method requires the selection of a 3-D region of interest from the user. In another probabilistic approach, Layer et al [42] used a combination of Gaussian mixture model (GMM) and a Markov random field (MRF) model to estimate tumor volume. The approach consists of a coarse estimation step in which an EM-based Gaussian mixture model is used to estimate voxel labels and a refinement step using a Gaussian MRF model with a Gibbs distribution to take account of neighbour dependencies. For initialization, the method requires a volume of interest to be defined

as well as user defined seed points to distinguish foreground and background regions.

In [1], an active contour based solution was proposed for PET segmentation. To enforce data fidelity, the authors used terms similar to the CV model, where instead of the original volume, the similarity is measured against a smoothed version of the data computed using anisotropic diffusion filter. A second regional term was also included using contourlet transform of the image. For length regularization, they used curvature of the evolving curve. The method was solved by means of the level set method. An important drawback of the level set method is that unless the resulting energy is convex, the method is only guaranteed to find locally optimal solutions, in addition, compared to global optimization methods, there is enhanced sensitivity to parameter choice with the active contour method. The authors findings reflect this.

In [9], an approach to distinguish areas of normal and pathology-related uptake in PET-CT scans was proposed. First the PET and CT outputs were filtered separately with PERCIST thresholding [58] applied to the PET volumes and a filter to reveal skeletal regions applied to the CT data. For the PERCIST step, initialization in the form of a sphere of fixed diameter was placed in the liver. Using the output of the skeletal filter on the CT data, matching regions were removed from the PET data. In the next step, regional statistics were computed from both the CT and PET data, and density-based spatial clustering was applied to the PET data to cluster regions fragmented in the earlier thresholding process. Finally, a feature vector was constructed using regional statistics of the clusters, the scale invariant features transform, the histogram of oriented gradients. A radial basis function kernel was used for training, while support vector machine (SVM) was then used to classify the data.The method obtained over 90% prediction score on each of the classes: Brain, Bladder, Heart, L/R Kidney, other. A potential limitation of the method is the implicit assumption that a one-to-one correspondence exists between PET and CT data. Also, the method is not fully automatic but requires a VOI to be defined for the PERCIST step.

In [65, 64], an automated segmentation method for PET-CT was proposed which is based on a combination of decision-tree and K-nearest-neighbor (kNN) classification. The authors describe their method as co-registered multimodality pattern analysis segmentation system (COMPASS). In [65] the method was tested on PET-CT head and neck datasets. Textural features were extracted from the PET and CT images independently, and then ranked using the sequential forward selection method. kNN classification is then applied iteratively to subsets of voxels. In each iteration a different subset of features are considered, with the classifier subdividing the voxels into ‘abnormal’ and ’normal’ subsets, abnormal subsets are classified as such, while normal subsets are subject to further classification considering a further 3 features. A potential drawback of the method is the assumption of a one-to-one correspondence between PET and CT, as well as a dependence on scanner-specific parameters required for a preprocessing thresholding step. The method obtained a mean sensitivity of 0.9 ± 0.12 and mean specificity of 0.95 ± 0.01. In addition to the reviewed works, several supervised and unsupervised methods have been proposed for PET segmentation including

those based on SVMs [36, 62], artificial neural networks (ANN) [55] and spectral clustering [61]. The reader is referred to [23] for a full review of recent methods proposed for PET segmentation.

Graph based methods have also proven popular in the literature. In [5] a co-segmentation method for PET-CT data is proposed using the random walker algorithm [26]. The method automatically finds foreground by first setting a threshold based on the maximum SUV value and then searches neighboring region for voxels with lower SUVs to set as background seeds. The method was tested on 15 clinical PET-CT images and achieved a DSC of 91.44%. In another approach for PET-CT segmentation also within a Random Walker framework, [18] proposed to include a prior based on spatial-topological information extracted from PET images. First, a topology region is generated based on the properties of iso-contours of the image. They then define a function to define the inter-region similarity which is incorporated as a prior in the random walker algorithm. For initialization, the method requires a region of interest to be defined.

In [57] a PET-CT segmentation approach is described in which the segmentation problem is modeled as a MRF and solved in discrete space using the maximum flow algorithm. The method involves the construction of two sub-graphs for the segmentation of the PET and CT data, and the introduction of a consistency constraint modeled as weighted edges connecting the sub-graphs. The consistency constraint introduces a penalty when there is disagreement between labels computed by the PET and CT segmentations. For initialization, one seed point and 2 radii are provided by the user for each target tumor. Parameters for energy terms in the final objective function are determined empirically.

In another recent approach designed for PET-CT, the authors in [35] proposed an approach which uses random walker algorithm to localize object seed points. For tumor delineation, they use a graph-cut framework similar to [57]. The proposed energy function includes region terms based on SUV distribution, the hessian matrix of the volume, and what they describe as a 'downhill' energy, which is designed to quantify the transition between normal and lesion regions in order to delineate these regions more accurately. The term is based on the decreasing rate of SUV uptake as the distance is increased from the maximum intensity point within a tumor. Similar to [57], they also include a penalty term to enforce consistency between segmentations produced from both modalities. The algorithm was validated on 18 PET-CT images and obtained a mean DSC of 0.84 ± 0.058. For initialization the method requires foreground and background seed points to be provided by the user. Compared to variational approaches, graph methods are guaranteed to find global solutions; however, they are unable to compute isotropic energy minimizers and incur a large memory overhead when segmenting fine features.

In recent work [8] a method based on graph-based optimization is used to solve the segmentation problem. This method requires an approximate lesion center point provided by a user to construct a graph structure around the center point, then local image statistics is used to derived the cost function. All hyper-parameter is learnt using training images from QIN HeadNeck dataset,

therefore the method is free from parameters tuning. Although this work provides accurate segmentation, it requires a user interaction per lesion and is not reproducible.

While popular methods for PET segmentation continue to include variants of thresholding [49], more recently sophisticated approaches based on Bayesian-based classification [29], belief function theory [43] and possibility theory [20] have been proposed. Graph based methods based on the random walker algorithm as well as the maximum-flow method have also been reported [18, 57, 35]. Table 1.1 summarizes recent methods for PET and PET-CT segmentation according to whether they handle normal activity, are reproducible[2], the level of user interaction required, as well as the number of parameters which need to be set and whether they utilize CT. The reader is also referred to [23] and [8] for a recent survey.

In this thesis, a fully automated method for PET tumor segmentation is proposed which uses classification to infer a data term which is then included into a level set segmentation. In [1] the level set method is used to minimize the resulting objective function, however the energy model is not convex, which means in practice that the solution is dependent on a good initial placement of the curve. In the proposed method the resulting energy formulation is convex which means that globally optimal solutions are guaranteed and independence to initialization is achieved. Several of the reviewed methods require either a VOI to be defined [41, 29, 20, 42, 43] or background and foreground seed points to be identified [35, 57], in contrast, the proposed method is fully automatic and therefore allows fully reproducible results.

1.3 Head and neck cancer

1.3.1 Anatomy

Typically, head and neck cancers are categorized based on the sites of specific tumors, such as nasopharynx, oral cavity, oropharynx, hypopharynx, larynx, and sinonasal cavity (Fig. 1.4) [11].

Nasopharynx

The topmost part of the pharynx is the nasopharynx, extending from the base of the skull to the top exterior of the soft palate. Usually beginning in the nasopharynx's superior and laternal walls, Nasopharyngeal cancer often obstructs the Eustachian tube's opening; through the pharyngeal walls and tonsillar pillars, these cancers can spread to the nasal cavity and inferiorly to the oropharynx.

[2] A method is described as non-reproducible when its results are dependent on the image-specific parameter tuning/initialization or other user interaction, since any user interaction is subjective and would produce different results.

Table 1.1: Methods proposed for PET/PET-CT segmentation

Method	**Technique**	**Handles Normal Activity** [*]	**Reproducible** [∓]	**User Interaction**	**Parameters** [†]	**Modality**
Song [57] (2013)	Graph based	✗	✗	Seed points and radii for each tumor	7	PET/CT
Ju [35] (2015)	Graph based	✗	✗	Seed points	15	PET/CT
Foster [24] (2014)	Affinity Propagation clust.	✗	◇	Manual correction of registration	7	PET/CT
Lelandais [43] (2012)	Belief-theory with FCM	✗	○	User defined ROI	1	PET/CT
Hatt [29] (2010)	Bayesian based class.	✗	○	User defined ROI	1	PET
Dewalle-Vignion [20] (2011)	Maximum of Intensity Propagations	✗	○	ROI*N [‡]	1	PET
Abdoli [1] (2013)	Active contours	**	✗	Initial curve	3	PET
Layer [42] (2015)	EM-based GMM	✗	✗	ROI & Seed points	5	PET
Bagci [5] (2013)	Graph based	✗	✗	-	5	PET/CT
Bi [9] (2014)	SVM class.	✓	✗	-	5	PET/CT
Cui [18] (2015)	Graph based	✓	✗	-	3	PET/CT
Lapuyade-Lahorgue [41] (2015)	Fuzzy C-means clust.	✓	✗	-	5	PET
Zeng [66] (2013)	Active surface modeling	✓	✗	-	8	PET
Yu [65] (2009)	kNN class.	✓	✓	-	3	PET/CT
Beichel [8] (2016)	Graph modeling	✗	✗	Seed points	0	PET
Yu [65] (2009)	kNN class.	✓	✓	-	3	PET/CT
Our Method	Machine Learning & Convex seg.	✓	✓	-	0	PET

[*]Handles normal activity automatically *i.e.* without user seed points [∓]Accuracy dependent on parameter choice, which includes size of ROI. A method is described as non-reproducible when its results are dependent on the image-specific parameter tuning/initialization or other user interaction, since any user interaction is subjective and would produce different results. [†]Parameters which were empirically set rather than learnt. [○]Sensitivity to ROI selection not tested. [‡]N is the number of projection directions which was set at 3 in [20]. [**]Discussion of initial curve placement not included. [◇]Method is reproducible on the condition that a registered CT is provided and that the method is applied to lung data only.

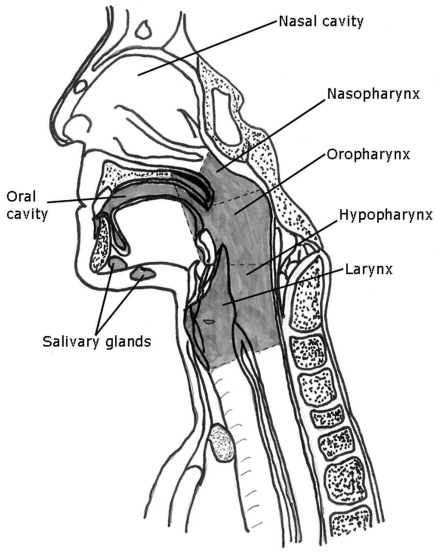


Figure 1.4: Head and Neck Anatomy. Nasopharynx, oral cavity, oropharynx, hypopharynx, larynx, and sinonasal cavity are the most common sites in which tumorous lesions starts developing [11].

Oral cavity

Included in the oral cavity are the gingival, gingivobuccal and buccomasseteric regions, hard palate, retromolar trigone, alveolar ridge, oral tongue, floor of the mouth, lip, and mandible. The oropharynx and the oral cavity are separated by a plane formed by the soft palate, tonsilar pillars and circumvallate papilla. 90 percent of all malignant tumors involving the oral cavity can be accounted for by Squamous Cell Carcinoma (SCC).

Oropharynx

The oropharynx includes the base of the tongue, the tonsillar region, the soft palate, and the pharyngeal wall between the nasopharynx and the pharyngoepiglottic fold. More than 90 percent of oropharyngeal malignancies can be accounted for by SCC and its variants.

Hypopharynx

The section of the upper aerodigestive tract that stretches from the hyoid bone and to the cricoid cartilage is the hypopharynx. The oropharynx is above the hyoid; the hypopharynx becomes the cervical esophagus under the cricoid cartilage. The hypopharynx is usually sectioned off into three regions: the postcricoid region, the lateral and posterior hypopharyngeal walls, and the pyriform

fossa. In excess of 95 percent of all hypopharyngeal tumors are SCCs. Hypopharyngeal tumors can continue being asymptomatic for a substantial amount time. Up to 75 percent of patients have cervical lymph node metastases by the time their cancers are diagnosed.

Larynx

The larynx may be divided into three regions: the supraglottic, glottic, and subglottic. The supraglottic larynx can be found above the true vocal cord (TVC), stretching from the tongue base and valleculae to the laryngeal ventricle. The different parts of the supraglottic larynx are the arytenoid processes of the arytenoid cartilages, the laryngeal ventricle, false vocal cords, the aryepiglottic folds, and the epiglottis. The glottic larynx is made up of the TVC and commissure. The subglottic larynx stretches from the cricoid cartilage to the TVC [63].

1.3.2 Diagnosis

Managing patients with head and neck cancer (HNC) begins from diagnosis and includes processes of checking cancer's clinical presentation, perform a biopsy, and then conducting diagnostic imaging.

1.3.3 Staging

Cancer must be accurately classified, especially because it is such a key factor in influencing treatment decisions. The tumor-node-metastasis (TNM) system is the standard staging system currently being used. It comprises of three parts, each being represented by a letter in the acronym: T for the primary or direct extension, N for secondary or lymphatic involvement, and M for vascular dissemination or distant metastasis [50]. Usually, T is divided into sub-categories, which range from T1 to T4; a higher number indicates the increasing extent of the primary tumor. The advancement of nodal involvement is classified starting from N0 (which indicates no evidence of nodal involvement), and then goes on from N1 to N3, indicating the state of affected lymph nodes (such as the location, size, and the number of lymph nodes involved). The metastasis is described by M0 or M+. A combination of T, N, and M define various stages of HNC, based on the site of the tumor. The head and neck have a complicated lymphatic system; this system has hundreds of lymph nodes, all of which are categorized into 7 levels or 13 sublevels, supraclavicular nodes, and retropharyngeal nodes [37]. Due to the complex lymphatic system and anatomy of the head and neck, HNC staging is also complicated, which has resulted in the use of multiple TNM staging systems for HNC; the staging system used is determined by the anatomical location of a tumor [11].

1.3.4 Treatment

The methods of treatment for HNC are chosen based on the site and stage of cancer, as well as the likelihood of remission and the physical condition of the patient. The aim of treatment is to cure the patient, while inducing the least amount of structural and functional impairment. Surgery and radiation therapy (RT) are the main modalities used for treating HNC. Historically, surgery has been the main treatment for most HNCs in the larynx, oral cavity, and oropharynx. In these instances, postoperative RT is often used in order to control microscope extent. Definitive RT is the main treatment option for nasopharyngeal tumors, as well as certain locally advanced unresectable tumors located in other areas. Surgery is considered a last resort if RT fails. Tumors in the hypopharynx can be treated either by definitive RT, or by surgery. That being said, cosmesis and organ function are commonly compromised by surgery because partial or total resection procedures such as glossectomy, mandibulectomy, and laryngectomy are frequently needed [37, 50]. Locally advanced cancer management has shifted significantly within the last 20 years.

Some of the reasons for developing precision radiotherapy techniques such as intensity modulated radiation therapy (IMRT) and image guided RT (IGRT) include poor outcomes for treating locally advanced HNC, wanting to preserve the patient's organs, as well as wanting better local therapy for unresectable cancers. The goals of the precision radiotherapy techniques that came from this development were to achieve better control of both local and regional disease, to increase survival rates, and to decrease normal tissue toxicity. As of today, precision radiation therapy is usually offered as a means of treating locally advanced resectable tumors in the larynx and oral cavity [12, 27, 37].

1.4 Thesis motivation

Cancer is a leading cause of death in most developed countries and it is responsible for nearly 30% of all deaths in Canada and 25% all deaths in the United States [31]. Head and Neck cancer is the fifth most common type of cancer that causes more than 11,000 deaths in North America and more than 300,000 deaths worldwide annually [33].

Imaging techniques have developed rapidly over the past decades, and indeed it plays an important role in staging and subsequent clinical management of cancer. For diagnoses and treatment, information from different modalities such as computed tomography(CT), magnetic resonance imaging (MRI), and positron emission tomography(PET) is used [51]. CT or MRI scans provide anatomic information, while Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) are nuclear medicine imaging techniques which provide metabolic and functional information. PET uses positron emitting radioisotope as a tracer which provides a better contrast and spatial resolution, while SPECT uses gamma emitting radioisotope as a tracer. SPECT has

less contrast and spatial resolution as compared to PET, but it is cheaper.

Although CT information can be used for quantification of lesions, the development of PET quantification techniques which rely PET information remains relevant, as it simplifies the task of automated contouring, and avoids the issue of misregistration between the CT and PET data sets. Furthermore, CT scans can lead to inaccurate contours due to adjacent atelectasis, or fibrotic tissues post treatment.

In this thesis, we focus on analyzing PET-18F-FDG images for Head and Neck Cancer (HNC) given its importance in cancer care. PET 18F-FDG is the primary modality used to assist oncologists in the localization and segmentation of the target volume of HNC during radiation therapy session. An important use of FDG-PET is a predictor of treatment outcome, for example, when it is used to assess response to chemotherapy and radiation. Manual segmentation can be very time consuming, especially if multiple target lesions need to be contoured. An automated approach would greatly accelerate the workflow of physicians contouring these lesions.

18F-FDG is the dominant radiotracer used in oncological imaging studies, particularly for treatment response assessment. This radiotracer also provides unique challenges due to physiological uptake in structures such as bladder, brain, kidneys, and other normal organs. The bladder is not always in the field of view. For example, in some head and neck protocols, and activity in the bladder is variable. The brain consistently uses glucose, and therefore accumulates FDG. Kidney retention is variable, and depends on the state of hydration of the patient. Cardiac uptake is variable, depending on fatty acid utilization. These are all normal variations which are expected in 18F-FDG studies, which create additional challenges for automated segmentation. Although the proposed method has focused on FDG, which by far is the most common tracer used in oncological PET imaging, the principles and approach outlined in this thesis will be applicable and extended to other radiotracers of interest in oncology (such as 68Ga-DOTATATE, 18F-FDOPA, PSMA imaging agents) as part of future work. Currently, well over 95% of oncological PET/CT studies are performed with 18F-FDG.

1.5 Thesis contribution

In this thesis, we evaluate 3D level set segmentation method in different phantom images provided by our collaborator team at The University of British Columbia. Then, the designed model competes with six other methods in QIN PET segmentation challenge on real data. The result of the QIN challenge reveals none of the proposed methods are reproducible, and all of them require user interactions. As the main contribution of this thesis, we propose a new fully automatic approach to lesion delineation in PET that, in contrast with previous works in Table 1.1 and challenge participants in Table 4.3, is the only method that is (i) fully automatic (i.e. does not require user-initialization or parameter-tweaking when segmenting novel images); (ii) does not require registered CT scans; and

(iii) is able to distinguish between radiotracer activity levels in normal tissue vs. the target tumorous lesions. We focus on 18F-FDG, which is by far the dominant radiotracer used in oncological PET imaging studies. This radiotracer provides unique challenges due to physiological uptake in structures such as the brain, kidneys, and other normal organs. Our method is a hybrid machine learning - active surface formulation, which utilizes physiologically meaningful cues (metabolic activity, anatomical position, and regional appearance statistics) and a convex optimization guaranteeing reproducible results. Beyond the specific application for PET images, our method using seed and machine learning techniques, can provide accurate segmentation in the absence of precise manual segmentation or when a few number of images with manual segmentation are available.

From a technical point of view, we used level set, convex formulation of active surface, random forest, and image feature that were introduced by other works. Then, we (i) extend the convex formulation of the active surface to 3D; (ii) combined machine learning with the active surface; and (iii) designed anatomical position as a novel feature in PET images.

1.6 Auto-bibliography

This thesis is based on the following published works:

- Reinhard Beichel, Brian Smith, Christian Bauer, Ethan Ulrich, Payam Ahmadvand, Mikalai Budzevich, Robert Gillies, Dmitry Goldgof, Milan Grkovski, Ghassan Hamarneh, Qiao Huang, Paul Kinahan, Charles Laymon, James Mountz, John Muzi, M Muzi, Sadek Nehmeh, Matthew Oborski, Yongqiang Tan, Binsheng Zhao, John Sunderland, and John Buatti. Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data, Medical Physics, volume 44, pages 479-496, 2016 [7].

- Payam Ahmadvand, Noirin Duggan, Francois Benard, and Ghassan Hamarneh, Tumour Lesion Segmentation from 3D PET using a Machine Learning driven Active Surface, In Medical Image Computing and Computer-Assisted Intervention Workshop on Machine Learning in Medical Imaging (MICCAI MLMI), volume 10019, pages 271-278, 2016 [2].

- Payam Ahmadvand, Hillgan Ma, Jesse Tanguay, Anna Celler, Francois Benard, and Ghassan Hamarneh. A Level-Set based Segmentation of tumour lesions in PET images. In Quantitative Imaging Network (QIN) Annual Meeting, 2016 [3].

This thesis won the following award:

- The People's Choice Award at the Faculty of Applied Sciences Heat of the Three Minute Thesis, Simon Fraser University, February 2017.

Chapter 2

Methods

2.1 Level set

A large number of problems in computer vision reduce to the study of the evolution of curves by which the boundaries between different objects are defined. These curves can evolve and move according to their own geometries or other cues that the users are provided according to the problem. They are initialized near the object, and then they might disappear, break up, or merge during the evolution time to match the cues (see Fig. 2.1).

Active contour methods have been used in medical image segmentation in the past decades, and have been successful in solving segmentation problems for different modalities. We implement a 2D stochastic active contour scheme (STACS) [52], which is originally introduced for 2D MRI. Then, we extend it to a 3D version in order to segment lesions in 3D PET images.

2.1.1 Energy function

Given a PET image, we aim to develop a method that locates a contour that automatically separates the voxel into two groups: lesion and background. This method encodes three types of prior knowledge into a function with three terms J_1, J_2, and J_3, each of which encodes a different prior. Contour C is embedded as zero level of function ϕ, and the function minimizes:

$$J(\phi) = \lambda_1 J_1(\phi) + \lambda_2 J_2(\phi) + \lambda_3 J_3(\phi) \tag{2.1}$$

2.1.2 Priors (cues)

We utilized the following priors in our formulation.

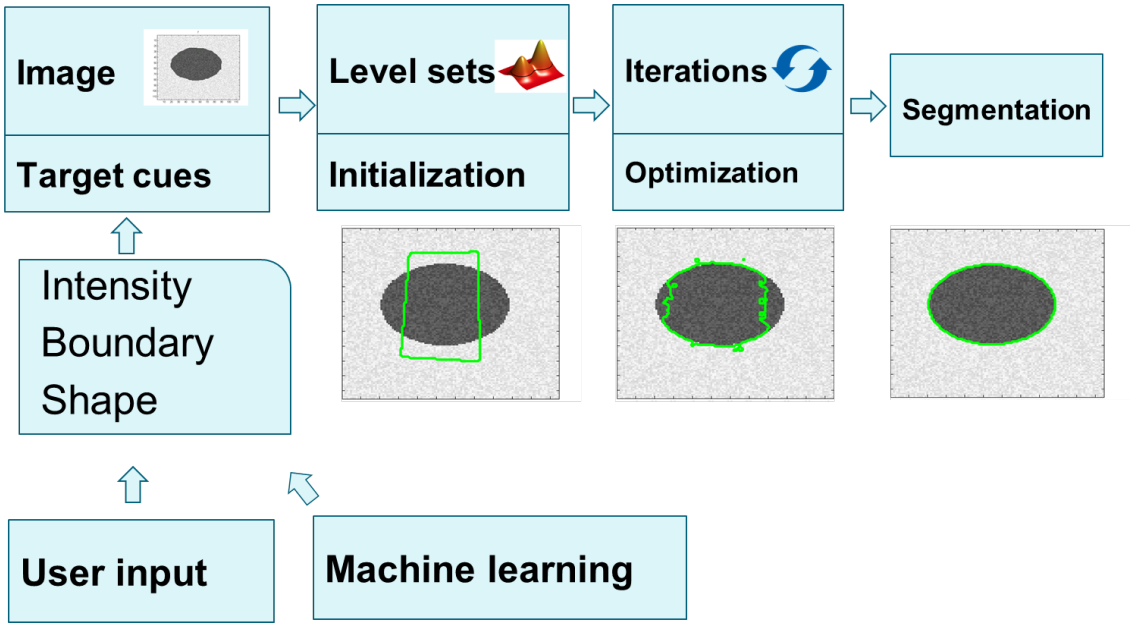


Figure 2.1: An overview of the segmentation system using level set. A zero level set surface is initialized, and then the surface is deformed to match the object based on target cues.

Image intensity

The first term encodes data likelihood based on the intensities of foreground and background voxels.

$$J_1(\phi) = \int_\Omega M_1 H_\varepsilon(\phi(x,y,z)) + M_2[1 - H_\varepsilon(\phi(x,y,z))]dxdydz \tag{2.2}$$

$$M_k = \frac{1}{2}\ln(2\pi\sigma_k^2) + \frac{(u(x,y) - m_k)^2}{2\sigma_k^2} \text{ for } k = 1, 2 \tag{2.3}$$

where M_1 and M_2 are the negative log-likelihoods of the probability density functions of foreground and background, respectively, and

$$H_\varepsilon(\phi) = \frac{1}{2}[1 + \frac{2}{\pi}\arctan(\frac{\phi}{\varepsilon})] \tag{2.4}$$

is a Heaviside function representing the pixels inside the contour, and $1 - H_\varepsilon(\phi)$ is the function which represents the pixels outside the contour.

The intensity priors from background and foreground are collected by seeds or boxes from a region of interest.

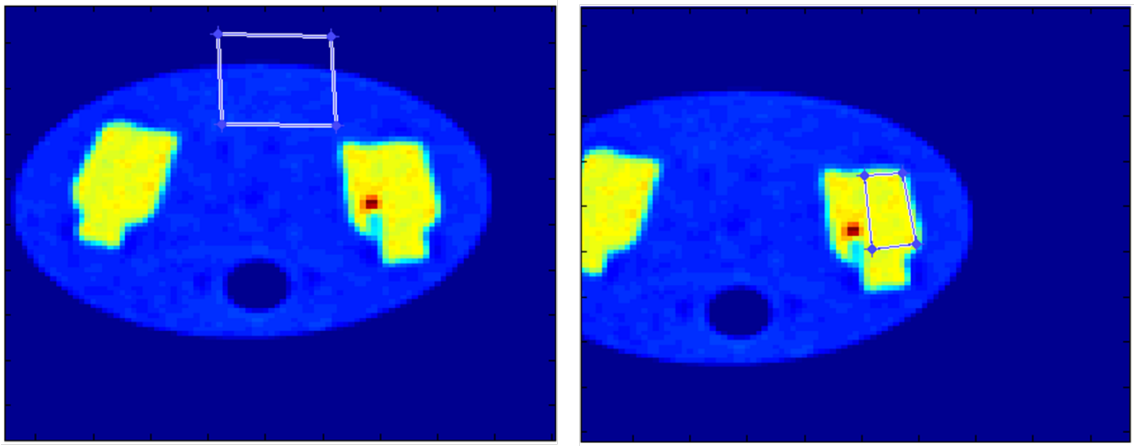


Figure 2.2: Collecting the intensity prior from background. In the left image, air and hot water inside the phantom (section 3.1) is background, and the user is selecting a box that covers both. In the right image, the bottle is considered as background.

Object edge

The points at which radioactivity changes sharply are of interest as they represent the borders of lesions. A simple edge map is the magnitude of the gradient of the volume $v(x, y, z)$, which is commonly used in edge-based active contour models [52]. This may be encoded as follows:

$$\Upsilon(x, y, z) = -|\nabla G_\sigma * v(x, y, z)| \tag{2.5}$$

where ∇ is the gradient operator, G_σ is a 3-D Gaussian kernel with variance σ^2, and $*$ is the 3-D convolution operator. The 3-D Gaussian kernel smoothes out the spurious edges that cause by the gradient operator, also it removes random noise caused by gradient operator (Fig. 2.4). If C is considered as the zero level set of ϕ, (2.5) becomes:

$$J_2(\phi) = \int_\Omega \Upsilon(x, y, z)|\nabla H_\varepsilon(\phi(x, y, z))]dxdydz \tag{2.6}$$

Boundary smoothness

The contour around the segmented lesions should be smooth, not too noisy, or jagged. This can be done by minimizing the entire Euclidean arc length of the contour C:

$$J_3(C) = \int_C ds \tag{2.7}$$

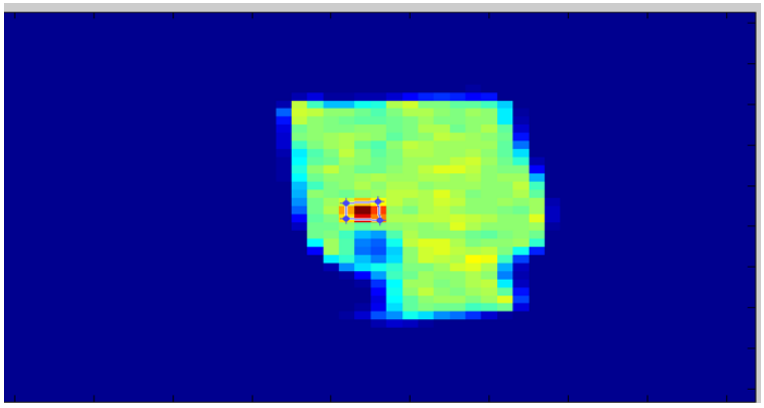


Figure 2.3: Collecting the intensity prior from foreground. A user is drawing a box inside the sphere.

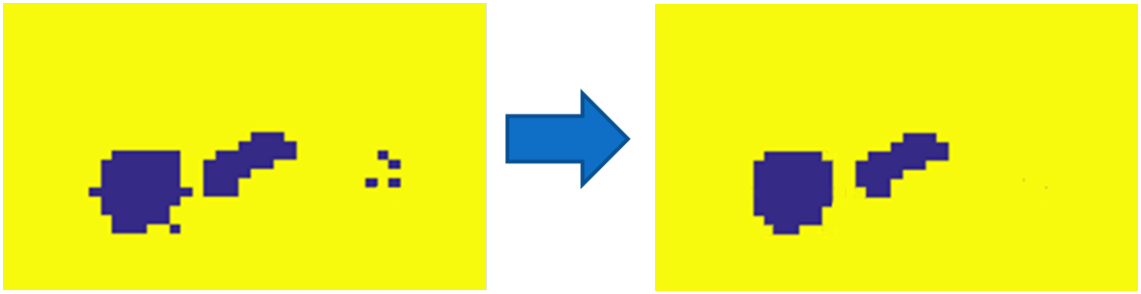


Figure 2.4: The 3-D Gaussian kernel or the regularization term is smoothing out the boundary, removing small, unnecessary components.

where ds is the infinitesimal Euclidean arc length of the contour C. When we minimize $J_3(c)$ alone, the contour C would evolve to become a circle, and finally it shrinks to be disappeared. This smoothness term encodes the prior knowledge that lesion boundaries should be smooth. It also discourages the generation of small noisy contours.

2.1.3 Overview of implementation

To segment a lesion, a zero level set surface is initialized using one or more points (seeds), and then the surface is deformed to match the lesions in the 3D PET image. In particular, the deformation updates are a result of the optimization process described earlier that ensures the resulting segmentation surface: (i) forms a smooth lesion boundary (J_3 in equation 2.1) (by penalizing large surface areas); (ii) passes through voxels of high image intensity gradient magnitude (J_2 in equation 2.1); (iii) whose interior and exterior regions (inside and outside the surface) follow learnt intensity priors

(J_1 in equation 2.1). These three criteria are encoded as energy terms in an energy function.

2.1.4 Implementation details

The intensity priors are Gaussian distributions whose mean and variance are estimated from sample pixels that are drawn inside and outside the lesions. For initialization of the lesion surface and for collecting the intensity priors, the user selected a single seed in the interior of a lesion in a 2D slice and selected one or more regions outside the lesion to collect intensity priors. Five scalar parameters are needed to be set by the user to control the behavior of the level set: One parameter controls the influence of the smoothness term and four parameters(two pairs of starting and ending values) dictate how the influence of the gradient and intensity priors change over iterations. A trial-and-error approach was utilized to adjust the parameters such that the method performs well for the majority of dataset to be segmented. The parameters for the method were fixed across all dataset with the exception for two out of 47 lesions. Also, for the cases where a specific lesion in close proximity should not be included in the segmentation, contours for the touching lesions were initialized separately. Two hundred (200) segmentation update iterations were enough for convergence in all our experiments (see Fig. 2.5 and Fig. 2.6).

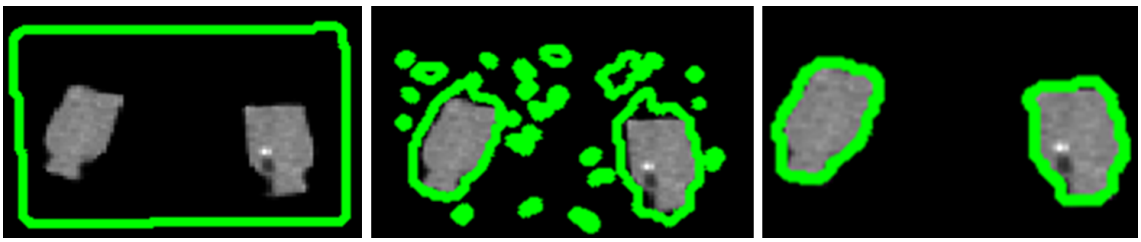


Figure 2.5: Segmenting two bottles: The left image shows the contour initialization. The middle image represents segmentation update iterations. The right image illustrates final convergence.

2.2 Problems with level set

Although level set methodology is one of the most versatile methods in medical image segmentation, it suffers from four major limitations:

1. Manual seeding and initialization requirement: Classical level set requires three types of input from a user for a binary segmentation problem:

   (a) A contour needs to be initialized.

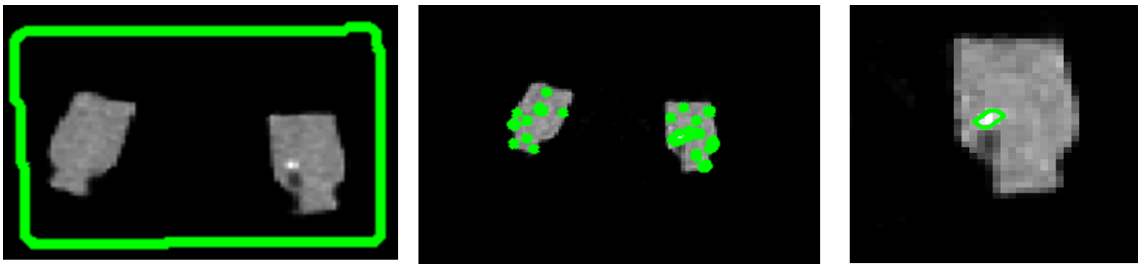


Figure 2.6: Segmenting the sphere inside the bottles: The left image shows the contour initialization. The middle image represents segmentation update iterations. The right image illustrates final convergence.

(b) Seeds from the background is acquired to estimate the probability density function of background's intensity.

(c) Seeds from the foreground is acquired to shape probability density function of the objects.

2. Dependence on intensity information only: The probability density function is usually defined base on intensity or radioactivity in the application of PET images. However, other important information (e.g texture, anatomical position) are ignored, while they very useful for more accurate segmentation.

3. Sensitivity to seeding: Since the intensity prior used for the probability density function is collected from the given image, segmentation is sensitive to the collected seeds. Thus, collecting seeds is subjective to the user interaction, and this is an obstacle to a reproducible segmentation.

4. Sensitivity to initialization (non-convexity): Methods based on the parametric active contour tend to be sensitive on initialization, therefore different initialization can yield worse or better segmentation results.

2.3 Level set with learnt likelihood

At a high level, this method consists of the following steps: Firstly, given our emphasis on head-and-neck (H&N) cancer, the bladder is detected and cropped out by removing all transverse slices inferior to the bladder. In a one-time training stage, seeds from lesions, background, and regions of normal activity (selected manually), along with the associated class labels, are used to train a random forest classifier. The random forest is trained on features extracted from these labeled seed points. For each novel image, a probability map is then generated based on the output of the

classifier, which predicts the label likelihood of each voxel. Finally, an active surface is initialized automatically and is optimized to evolve a regularized volume segmentation delineating only the boundaries of lesions.

2.3.1 Features

Automatic feature learning methods ignore expert domain knowledge and require a large number of datasets in order to alleviate problems like overfitting. However, large datasets are difficult to acquire in many medical imaging domains. We therefore resorted to designing specific features based on the following problem-specific cues. We tried different possible features that are commonly being used in medical image analysis to find the best combination, by which the best classification accuracy can be achieved. While the other features have been introduced by other works, anatomical position as a feature is a new contribution in this thesis.

Radiotracer uptake (5 features)

The first group of features is based on the standardized uptake value (SUV), which plays an important role in locating tumors [38]. SUV is computed as the ratio of the image-derived radioactivity concentration to the whole body concentration of the injected radioactivity.

T-Test reveals activity concentration in different organs has a different distribution in each image (Fig. 2.7). However, the p-value is very close to the cutoff (0.05) when seeds come from different images.

The activity value of each voxel along with max, min, mean and standard deviation of SUV in a window size $3\times3\times3$ around each voxel is added to the features vector to encode activity information. It should be noted SVU max value is frequently used for tumour response assessment and useful to recover multiple target lesions [47].

Anatomical position (1 feature)

Using the approximate position of tumors as a feature is useful since in each type of nonmetastatic cancer, lesions are located in a specific organ (e.g. H&N in our dataset). However, position coordinates must first be described within a standardized frame-of-reference over all the images. The field-of-view of PET images usually spans from the brain to the middle of the femur as the default protocol; however, this is not always the case. To deal with this variability in scans and obtain a common frame of reference across all images, it is necessary to first ensure anatomical correspondence. The most superior point of the bladder, which is a high-uptake organ, is used as the first anatomical landmark. The second landmark used is the most superior part of the image (i.e. top of the brain). To find the bladder as a high-uptake organ, we plot all training images according to the

transverse plane (Fig. 2.8). As can be seen from the plot, two images do not have bladder, and all images with bladder, the bladder is the highest uptake, and a peak can be seen in slices containing bladder (Fig. 2.9). Fig. 2.10 shows the standard deviation in slices containing bladder, has a sharp peak. The highest value of standard deviation can clearly indicate the bladder. If the highest pick is not in the first third of slices from inferior of the body, the image does not have bladder. In this scenario, the second landmark would be the most inferior part of the image.

Each image is then spatially normalized along the axial direction based on these two landmarks to obtain a new normalized axial position feature with values ranging between 0 (most inferior) to 1 (most superior) (Fig. 2.11). No normalization was needed within the transverse plane itself.

PET image texture (8 features)

First, four standard 3-D Haar-like features (edge, line, rectangle, and center-surround) are used to capture the general texture pattern of a $10 \times 10 \times 10$ region around each voxel in an image [28]. Fig. 2.12 show the four type of Haar-like features on a lesion and heart.

Second, the following four texture statistics are calculated for each transverse plane: cluster prominence [56], homogeneity [56], difference variance [28], and inverse difference normalized (INN) [17]. These were found to be particularly useful for distinguishing normal brain activity from H&N lesion activity along the anteroposterior dimension. The four feature values for a given plane were assigned to each voxel in that plane.

Homogeneity (2 features)

Tumor homogeneity is a measure of the uniformity of tumor pixel intensities [39]. The tumor homogeneity is calculated by

$$hmg(P) = \frac{1}{n}\sum_{i=1}^{n}\left(1+\sqrt{\frac{1}{n}\sum_{k=1}^{n}(P_k - P)^2}\right)^{-1} \tag{2.8}$$

where P is a vector of all activity values of 3D neighbors centered around a given voxel, and P_k refers to k th voxel of vector P. We used two homogeneity features at two scales, using window sizes $3 \times 3 \times 3$ and $4 \times 4 \times 4$, i.e. n=27 and 64, respectively.

2.3.2 Normalization

Normalization plays an important role in this problem as each PET images have a different distribution.

Image normalization

To reduce inter-subject image variability, the intensity of all voxels each image is normalized to zero mean and unit variance. The normalization is a technique to reduce noise in PET images, also has the potential for improving the results of the classifier. Applying this normalization is particularly useful for radiotracer uptake features according to T-Test.

Feature normalization

The values of each feature are normalized to zero mean and unit variance over the training samples. Then the mean and variance of training feature vector are used to normalized test feature vector. Cross validation reveals that normalization on feature vector improves the classification accuracy by 3 percent.

2.3.3 Classification

In total, the feature vector assigned to each voxel was of length 16. For training, labeled samples (i.e. feature vectors at seed pixels) were collected from the lesion, body, air background, kidneys, heart, and brain regions. These can be considered as six distinguishable classes; however, we consider the heart and kidneys as one class since they are close to each other in terms of anatomical location and far from H&N lesions. Having the brain as a separate class is helpful to facilitate discrimination between H&N lesions since the brain and lesions are in close proximity.

A random forest classifier is used to predict the label probability on a voxel-by-voxel basis. The parameters of the classifier were trained and applied to new images using leave-one-image-out cross-validation. Having more than 50 decision trees did not increase the classification accuracy, and the value for the number of variables randomly sampled at each split was set to the square root of the number of features, as it was automatically set and found to be robust. Fig. 2.13 shows sample probability maps generated by our trained random forest.

2.3.4 Convex segmentation with learnt likelihoods

To produce a final segmentation of the lesion tissue, the posterior model produced by the random forest classifier is included as a data (likelihood) term into the convex segmentation formulation of Bresson et al. [14]:

$$E(u) = \int_{\Omega} |\nabla u(x)|\, dx + \int_{\Omega} u(x)(P_{obj}(x) - P_{bg}(x) + C_A)dx \tag{2.9}$$

where $u(x)$ in [0,1] is the segmentation label field, Ω is the image domain, the first term is the boundary regularization term, and the second is the data term. P_{obj} is the probability of a lesion

and $P_{bg} = 1 - P_{obj}$ (i.e. P_{bg} groups together the likelihood probabilities of all non-lesion classes), and C_A is a constant that penalizes the area of the segmentation surface and is used to constrain the size of small unconnected components (e.g. small lymph nodes). The contour is automatically initialized as a rectangular prism around the border of the 3D volume. Given the convex formulation, the algorithm converges to the same global solution with any initialization. One hundred (100) segmentation update iterations were enough for convergence in all our experiments [1].

2.3.5 Training stage

Seed selection

For training a model that can classify different region of a PET image, we need samples from different regions in forms of voxel coordinate with corresponding labels. For this purpose, a small tool is developed that automatically travels trough the slices, and meanwhile users can click on the region of interest in the 2-D plane. Then, 3-D coordinate along with corresponding labels (e.g tumor, bladder, background) is stored in a matrix. 1080 seeds have been selected from 10 3D PET training images. Fig. 2.14 shows some seeds collected from the different part of a PET image (coronal plane). Fig. 2.15 illustrates a 3D plot of collocated seeds from one training image. Most of the seeds are collected from the superior part of the body (near head and neck). The ratio between the size of the training set and the dimensionality of the feature vector is a problem studied extensively in the machine learning literature. For a 16-D feature vector, 500-2000 samples set is suggested. We started with 500 samples and increased the number of samples to 1080 in order to reach a reasonable classification accuracy.

Feature selection and setting segmentation parameters

Leave-one-image-out cross-validation technique is used to tune the entire model:

Feature selection There are many possible features that can be used in medical images analysis. However, firstly useful features should be leveraged to improve the accuracy of training model, and speed up the process by removing useless features. Secondly, each feature has its own parameters and the optimal parameters is required to be found. In addition, leave-one-image-out cross-validation confirms all selected features has potential to improve the classifier accuracy.

Segmentation parameters When the highest accuracy based on seeds is achieved, we need to find the optimal value for a single constant of our active surface model. For learning this constant,

[1] A video demo can be found online at https://youtu.be/vuEwCGti_JM

we used some training image with complete and correct segmentation. The optimal value is found by maximizing Dice between the manual segmentation and the automatic segmentation.

2.3.6 Testing stage

Similar to the training phase, each PET image is normalized to zero mean and unit variance. Also, the mean and variance of training feature vector are used to normalized test feature vector. All learnt parameters are fixed and the method ran on test images without any user interactions.

Implementation details

Although training on thousands of samples using random forests is fast (0.15 sec), predicting the labels of millions of voxels using a MATLAB implementation was time-consuming (up to 20 hours per volume). Instead, we used a fast C implementation of random forests [45], with parallel processing distributed on 4 CPU cores (one slice per core). This resulted in a reduction of prediction time by a factor of 50 to reach the average running time of ˜15 minutes per image.

2.4 Addressing the level set limitations

By leveraging machine learning and convex formulation of the active surface, we were able to handle the limitations of level set discussed in section 2.2. We were able to replace the user interaction with machine learning in the first proposed method (see Fig. 2.1).

1. Manual seeding and initialization requirement:
   (a) The contour initialization is done automatically by fixing a rectangular prism at the border of the image. Therefore, the user does not need to initialize the contour manually.
   (b) The seeds are collected from training data *once* during the training phase, and given a new test image, the user interaction for collecting seeds is not required anymore.
2. Dependence on intensity information only: By designing a 16 dimension feature vector, we were able to add more useful and relevant cues such as texture, anatomical position, and homogeneity.
3. Sensitivity to seeding: The seeds are collected during the training phase only, and the subjective manual seeding process does not exist anymore at test time.
4. Sensitivity to initialization (non-convexity): The convex formulation of active surface is not sensitive to initialization, therefore different initializations yield the same segmentation results.

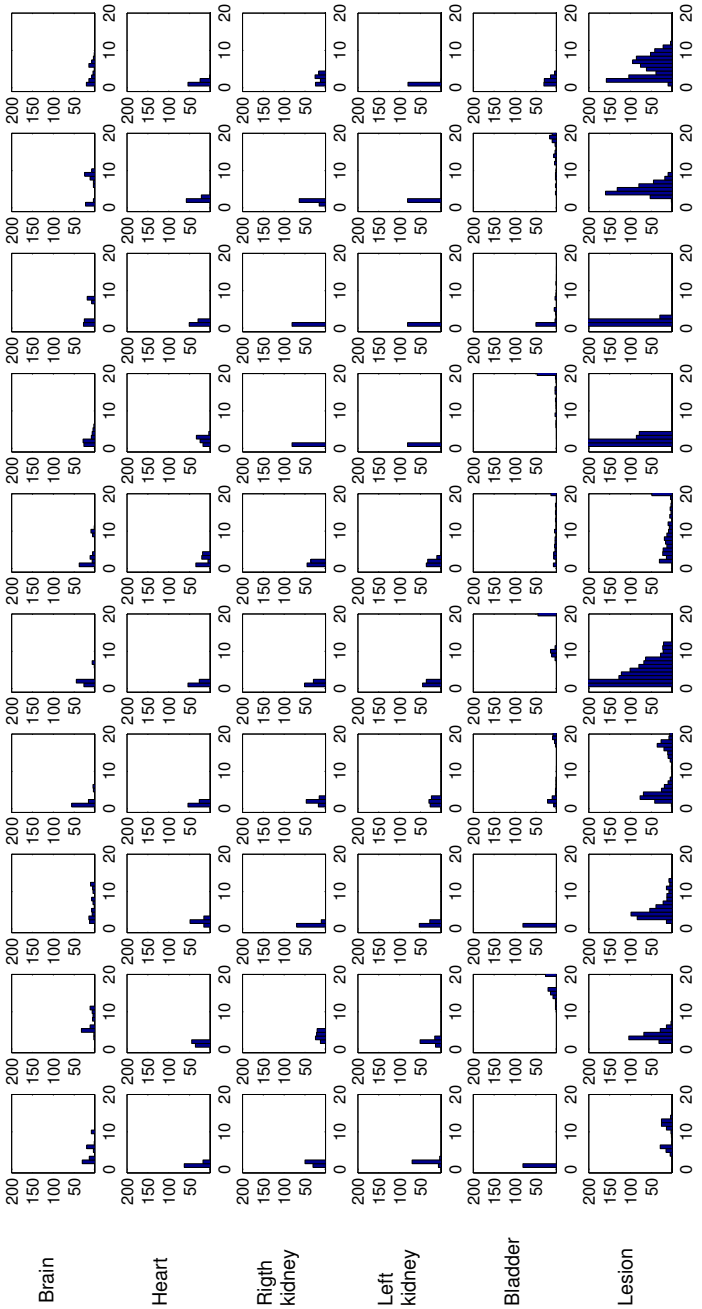


Figure 2.7: The histograms of seeds from a PET images showing the distribution of active organs: brain, heart, left kidney, right kidney, bladder, and lesion.

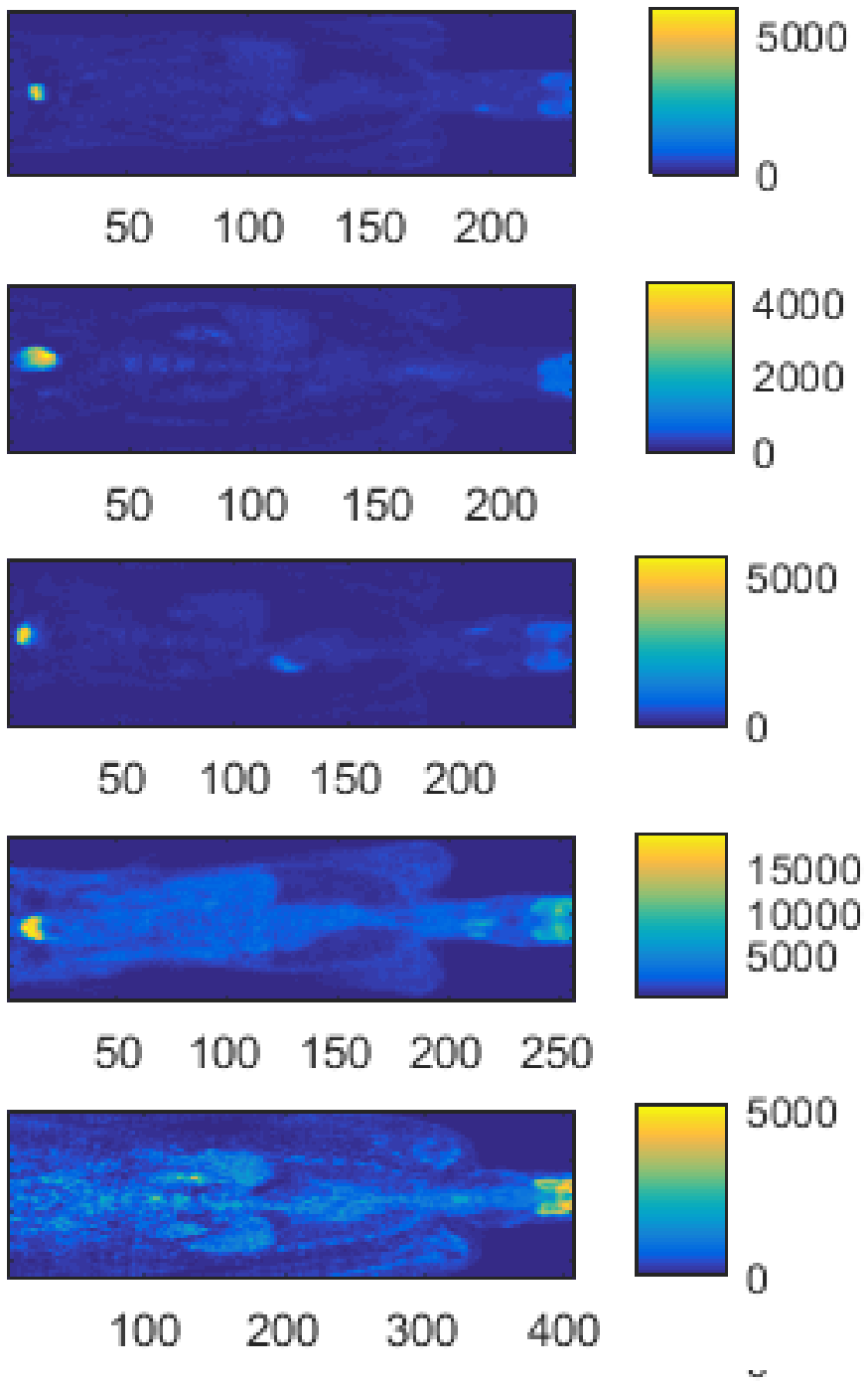


Figure 2.8: Transverse plane view of five training images. All images with bladder, the bladder is the highest uptake.

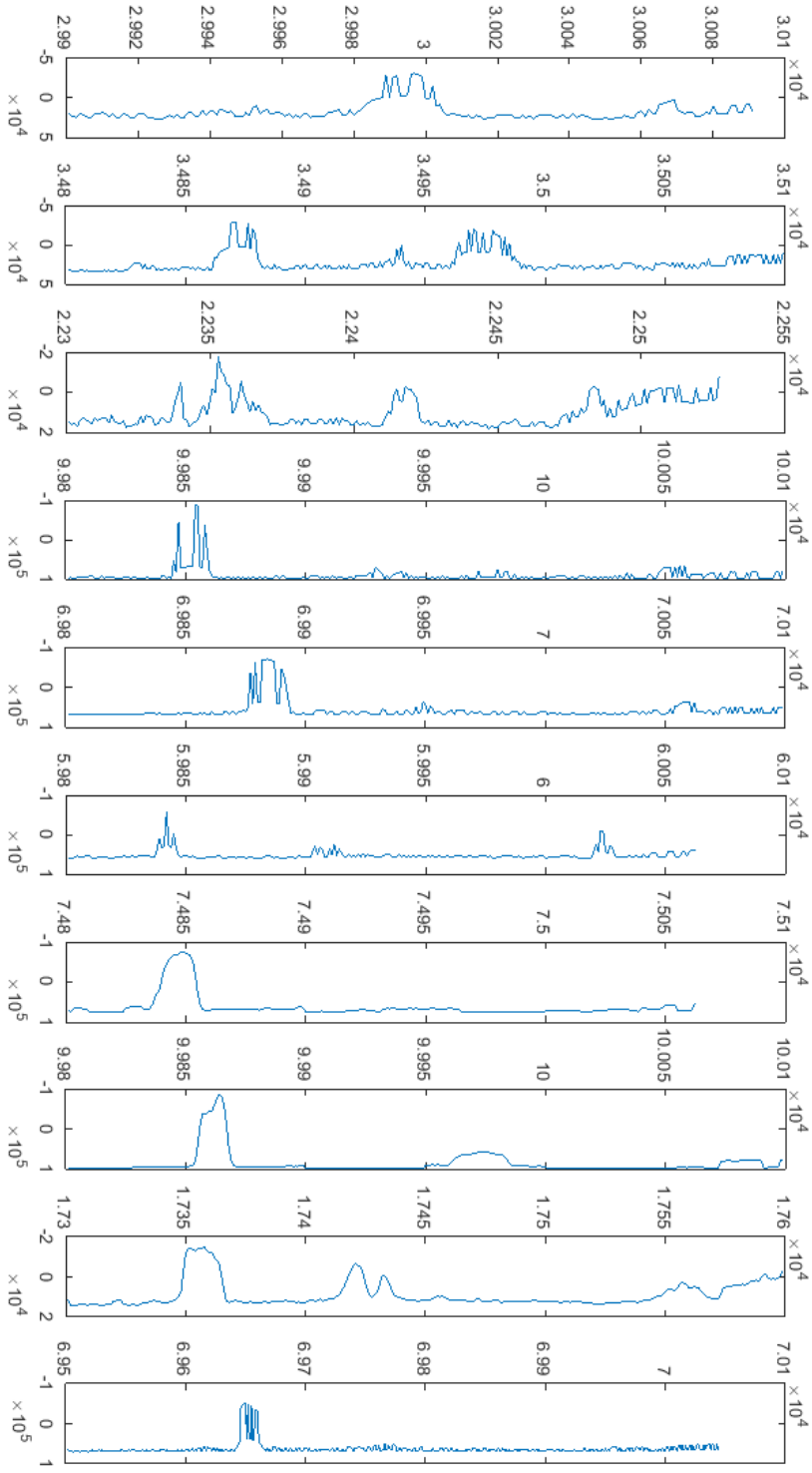


Figure 2.9: Variance plot from transverse plane of ten training images. The variance in slices containing bladder, has a sharp peak.

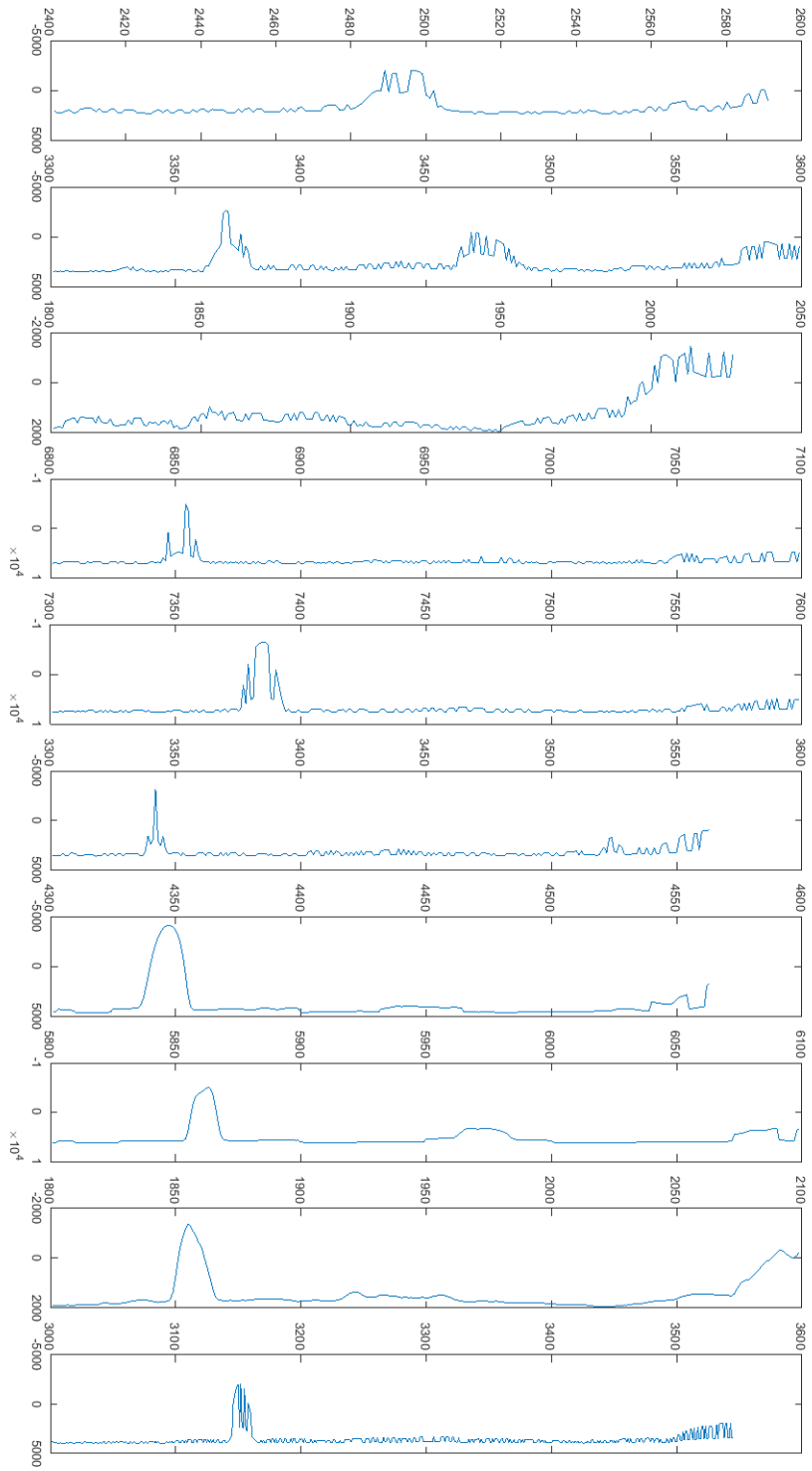


Figure 2.10: Standard deviation plot from transverse plane of ten training images. The standard deviation in slices containing bladder, has a sharp peak.

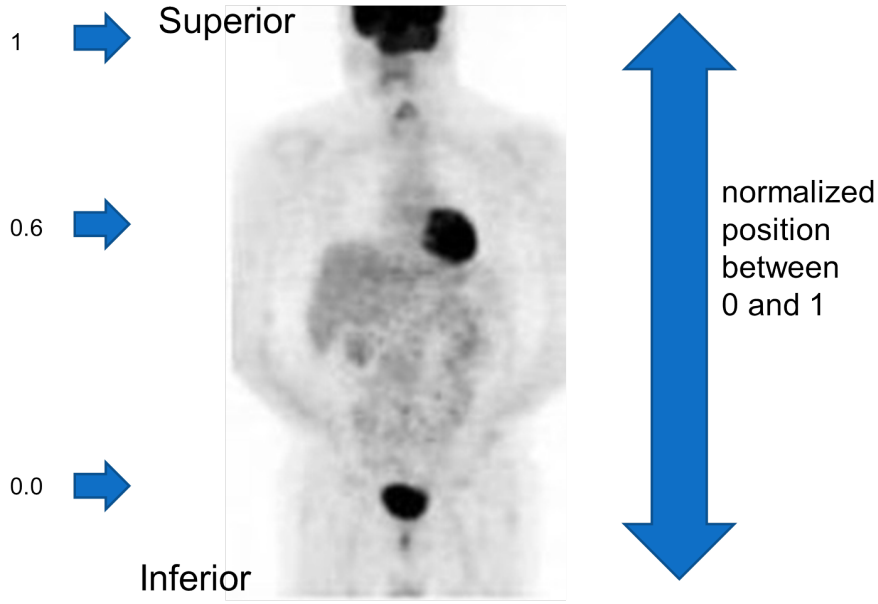


Figure 2.11: An image is normalized along the axial direction based on these two landmarks to obtain a new normalized axial position feature with values ranging between 0 (most inferior) to 1 (most superior).

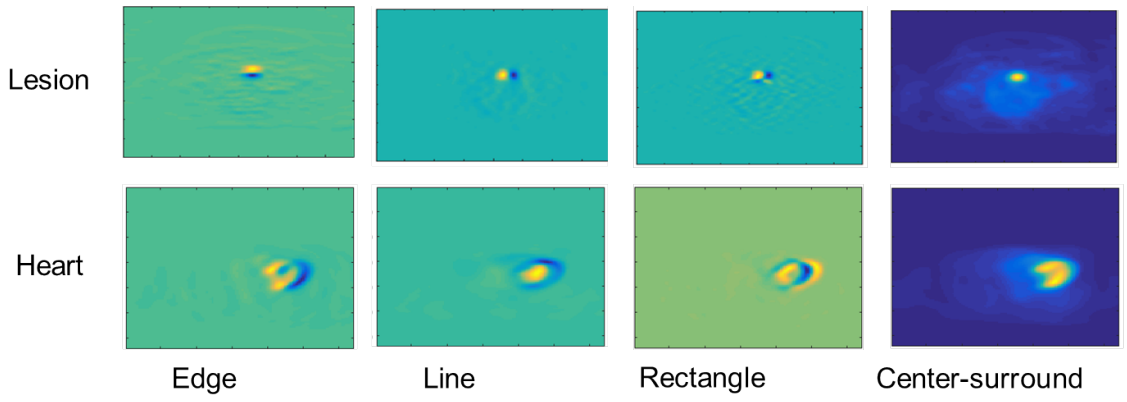


Figure 2.12: Four standard 3-D Haar-like features in transverse plane that shows lesions and heart.

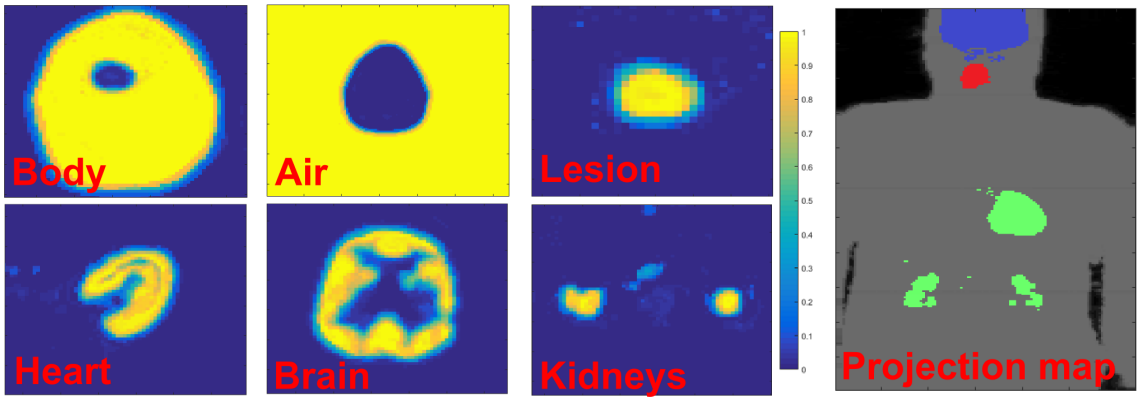


Figure 2.13: Example of class probabilities. Images on the left: Probability maps over different classes in the transverse plane (different scales used for clarity). Right: Maximum probability projection.

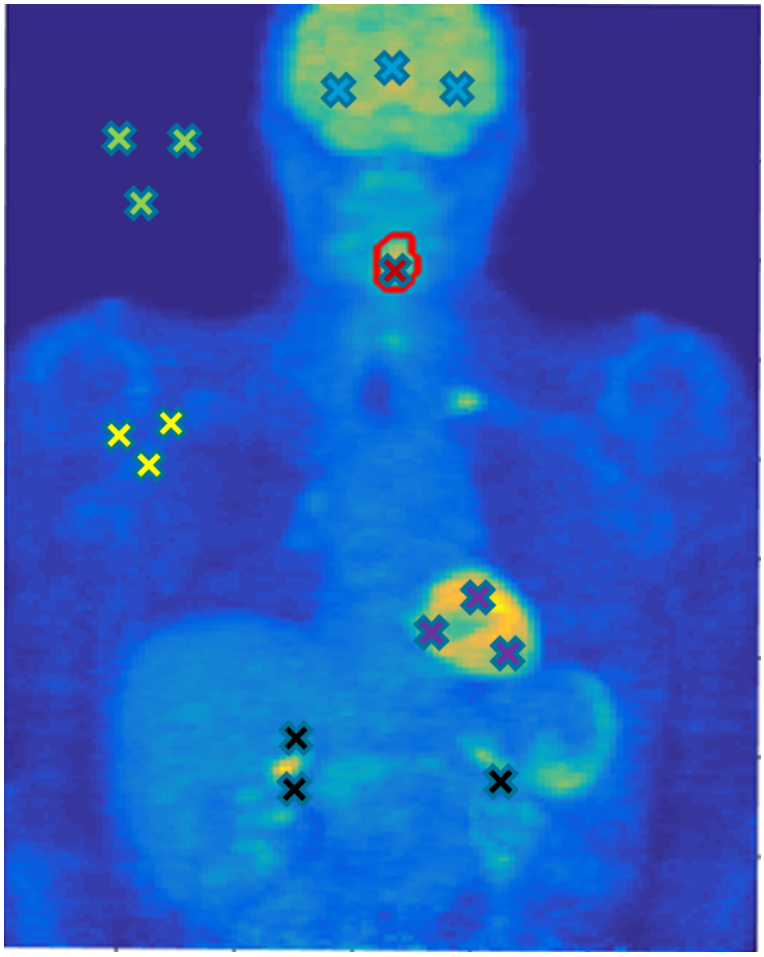


Figure 2.14: Selected seeds from a training image. The seeds are selected from lesion, body, air background, kidneys, heart, and brain regions.

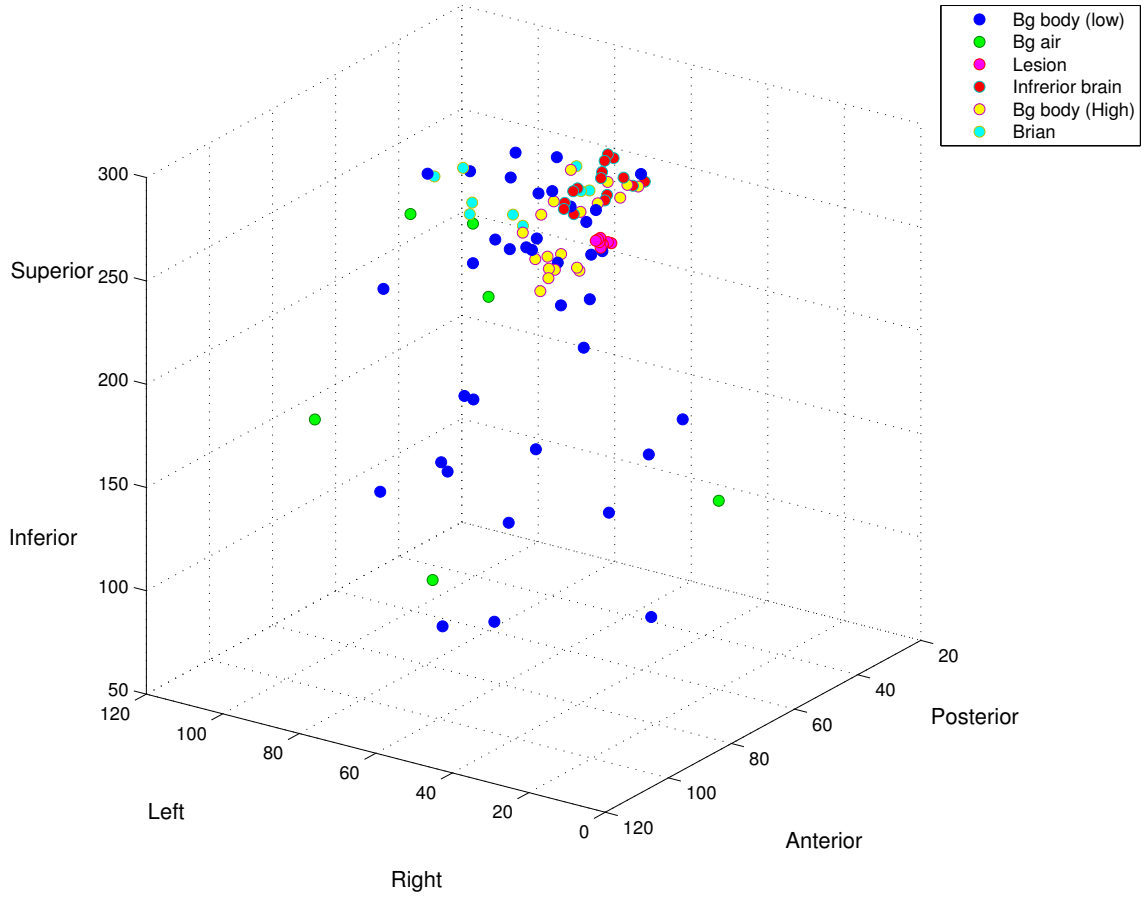


Figure 2.15: 3D Plot of collected seeds from a training image. The seeds corresponds to different regions: lesion, body, air background, kidneys, heart, and brain.

Chapter 3

Data

3.1 UBC phantom

Two independent phantom scans were carted by Medical Imaging Research Group, Department of Radiology, University of British Columbia. The inserts in the phantom were filled with activity. The sphere-to-background ratio (SBR) of the activity was aimed to be approximately 4 or 8, following the NEMA recommendations for calibration standards. This ratio approximates the uptake ratio of a typical patient. However, in order to simplify the complexity of the scans, air and water were first used in the background. Next, hot water (water with radioactivity) with different concentration is used as background to generate more challenging images. Following Poisson nature of photon emissions, scans with shorter duration have increased noise due to fewer detected events.

3.1.1 NEMA-IEC phantom

A phantom from National Electrical Manufactures Association (NEMA) International Electrotechnical Commission (IEC) PET body was used for this scan. The images from these experiments were reconstructed using two different types of algorithms: 1) the ordered subset expectation maximization (OSEM) [53] algorithm with a penalized likelihood reconstruction using relative difference penalty (RPD); and 2) the block sequential regularized expectation maximization (BSREM) [4]. Both algorithms were reconstructed with and without time-of-flight (TOF) information. In order to simulate scans with increased noise levels, the raw data of the phantom experiments resorted into new data sets with scan durations corresponding to 30 seconds, 1 minute, 2 minutes, 3 minutes, 5 minutes and 8 minutes. Table 3.1 shows the volumes and activity levels of the six spheres used in the NEMA-IEC phantom (see Fig. 3.1).

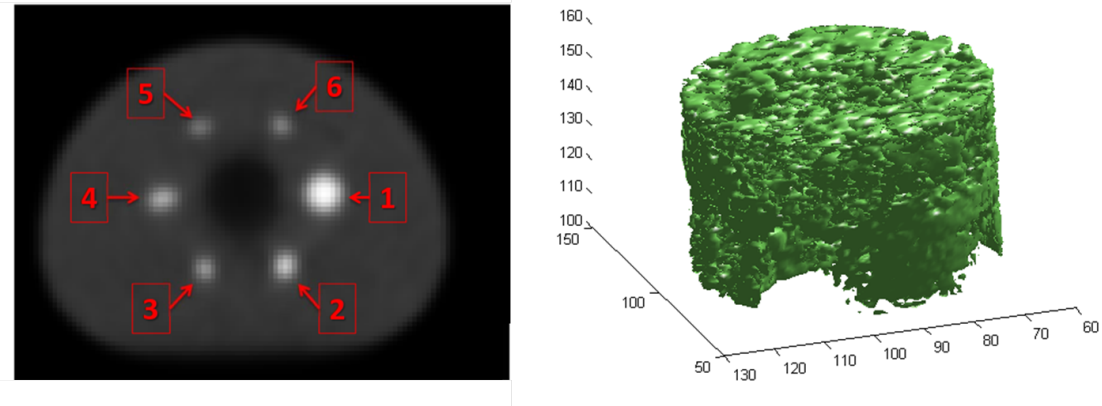


Figure 3.1: NEMA-IEC Phantom with six spheres in different sizes. The left image is axial view, and the right one is the 3D view.

3.1.2 Elliptical lung-spine body

Elliptical Lung-Spine Body Phantom is designed to create new phantom data that simulates a body with long cancer (see Fig. 3.2). The raw data were reconstructed by block sequential regularized expectation maximization (BSREM)[4] with scan durations corresponding to 30 seconds, 1 minute, 2 minutes, 3 minutes, and 5 minutes. F-18 PET/CT image of an anthropomorphic torso phantom scanned with: six spheres (0.5 - 16 mL, SBR* = 8), two bottles (200mL, SBR = 4), and the second smallest sphere placed inside bottle (1 mL, SBR = 2). Table 3.1 represents the volumes and activity levels of the six spheres and two bottles used in the Elliptical Lung-Spine Body phantom.

3.2 The Cancer Imaging Archive (TCIA)

TCIA[1] is a project funded by the Cancer Imaging Program of the National Cancer Institute that identifies and provides a large archive of medical images from patients with cancer. The data is accessible for downloading as a service and is organized as “Collections”, typically named by a common disease (e.g. Head and Neck, Lung Cancer), image modality (PET, CT, MRI, etc) or research focus. TCIA uses DICOM as the primary file format for image storage. More details and supporting data related to the images such as patient outcomes, treatment details, genomics, pathology, expert analyses, expert manual segmentation are also provided when available [16].

The Quantitative Imaging Network (QIN) grows from the National Cancer Institute (NCI) program announcement “Quantitative Imaging for Evaluation of Responses to Cancer Therapies”. The

[1] http://www.cancerimagingarchive.net/

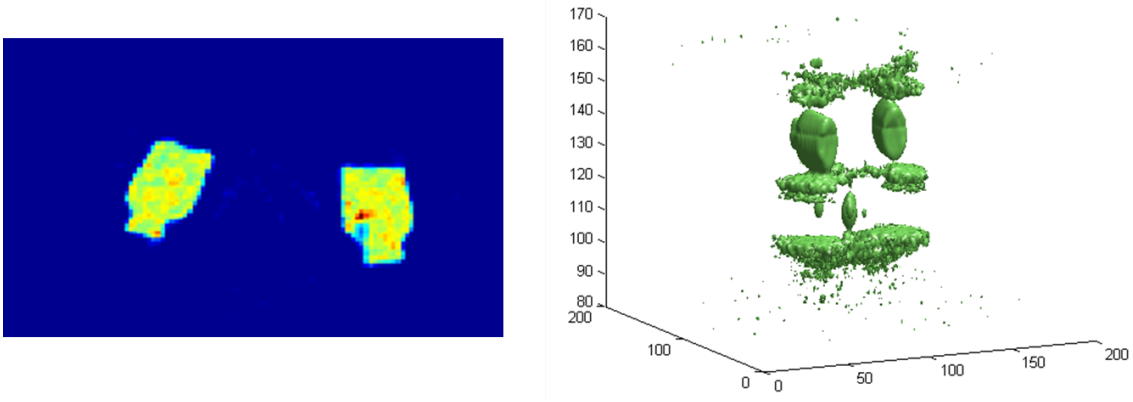


Figure 3.2: Elliptical Lung-Spine Body Phantom with six spheres and two bottles. One of the sphere is located inside a bottle to simulates a lesion inside a organ. The left image is axial view, and the right one is the 3D view.

network is designed to promote research and development of quantitative imaging methods for the measurement of tumor response to therapies in clinical trial settings, with the overall goal of facilitating clinical decision making (Fig. 3.3). Quantitative Imaging Network, with the mission of improving the role of quantitative imaging for clinical decision making in oncology by the developing and validating data acquisition, analysis methods, and tools to tailor treatment for individual patients and predict or monitor the response to drug or radiation therapy introduce, has introduced several collections in TCIA.

3.2.1 QIN-HEADNECK

We used QIN-HEADNECK collection which is a set of head and neck cancer patients' multiple positron emission tomography/computed tomography (PET/CT), with 18F-FDG scans-before and after therapy-with follow up scans where clinically indicated [16, 22, 54]. This collection consists of 156 patients with at least one PET/CT scan and related clinical data available for study (mean 3.05 studies/patient collected during a total of 472 visits) and 60 images have manual segmentation[22]. They are stored in forms of binary segmentations that are encoded in the PixelData attribute of the SEG object, and are considered as an array of bits in which voxel at each frame has a bit. Separate frames are considered for each slice of the volume, though all are encoded in a single multi-frame object. When multiple segments (i.e., primary tumor and lymph nodes) are produced by the operator during a single session using a single segmentation tool, they are stored in a single SEG object, with each segment for each slice stored in a separate frame. Since all tumor and lymph nodes are in head and neck area, the segmentation of volume would have many empty

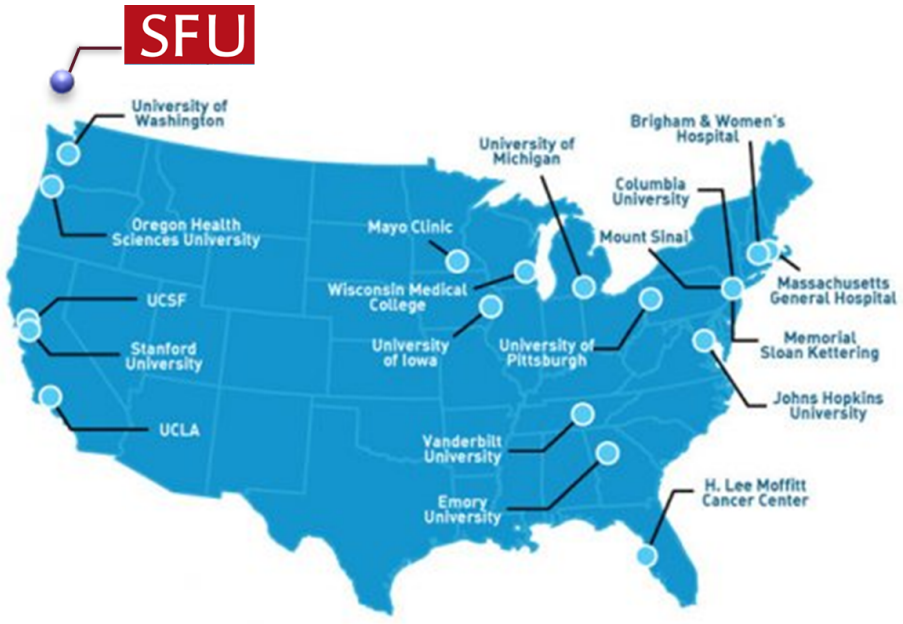


Figure 3.3: The Quantitative Imaging Network for Evaluation of Responses to Cancer Therapies [25].

frames which do not contain any voxels of the segmentation. Therefore, those empty slices are omitted in order to decrease the size of the encoded objects. SEG objects include references to the SOP Instance UIDs of the images (slices) that were segmented [22]. The reconstruction parameters were 2 iterations/ 8 subsets (iterative reconstruction algorithm: OSEM) with a 5 mm smoothing Gaussian filter and voxel size $0.354 \times 0.354 \times 3.37$ mm. For testing, data from the original scanner as well as additional data from Siemens Biograph 40 scanner was used (also with OSEM reconstruction algorithm, parameters 4 iterations, 8 subsets) with a 7 mm smoothing Gaussian filter and voxel size $0.339 \times 0.339 \times 5$ mm.

3.2.2 Subset of QIN-HEADNECK

QIN-HEADNECK manual segmentations are prepared for a semi-automatic method that required a lot of user interaction [8]. Therefore, we defined the following five criteria for exclusion to select a subset that is suitable for a fully automatic method:

1. Lesions that are closely connected but are segmented separately in manual segmentations, while our method connects them as one lesion Fig. 3.4.

2. There are obviously high uptake (abnormal) areas, but they have not been segmented manually as lesions. They are tow cases:
   (a) Separate isolated lesions (or what appear to be lesions) which are not segmented Fig. 3.5.
   (b) Portions/parts of lesions which are not segmented Fig. 3.6.
3. Misalignment: Good agreement between experts but the segmented region does not appear to be a high-uptake region (low intensity region) Fig. 3.7.
4. Expert agreement:
   (a) Low agreement between experts (we didn't quantify this, but based on qualitative assessment) Fig. 3.8.
   (b) Some lesions have been segmented by one expert, but not by others Fig. 4.13.
5. Patient scanned in unconventional bed-positions e.g. arms overhead.

Note: There are two cases (QIN-HEADNECK-01-0184, QIN-HEADNECK-01-0046) where there are lesions in the chest. The current training data does not have examples like this, therefore the method is not expected to perform well on these images. Table 3.3 represents a list of QIN-HEADNECK collection studies that have manual segmentation. The exclusion criteria and usage of each image are presented in this table.

3.3 QIN challenge data

The QIN challenge organized by University of Iowa was based on the two groups of data:

3.3.1 Phantom data

The University of Iowa (UI) and University of Washington (UW) provided two series of phantom based studies with different PET/CT scanner models. These Phantom are the modified version of the NEMA IEC Body Phantom SetTM(PET/IECBODY/P) in which instead of the typical arrangement of all spherical inserts, they utilized a mixture of elliptical and spherical inserts to create a clinically more relevant segmentation challenge which is different from the standard NEMA spheres which can be seen in Fig. 3.10. 18F-FDG object to background contrast ratios of 9.8:1 and 4.9:1 was used in a single 30 min acquisition. University of Iowa reconstructed the images using a reconstruction parameter with highly smoothed images with a 7mm Gaussian filter 145 kernel (UW used a narrow 3 mm Gaussian filter) to create a difficult challenge for segmentation methods. These phantom data can be found in the Cancer Imaging Archive (TCIA) under the collection called QIN PET Phantom. Fig. 3.11 shows the phantom images in different statistics and contrasts.

3.3.2 Real data

In the second part of the challenge, clinically acquired PET scans from patients with head and neck (HN) cancer was utilized. The real data was a subset of QIN-HEADNECK collection. All lesions that need to be segmented were shown by means of indicator images to attendees. Fig. 3.12 shows a sample of ground truth in which lesion with its corresponding labels are defined. 47 lesions were asked to be segmented with the complexity of segmentation tasks ranges from low to complex, and the number of lesions in each image varied.

Table 3.1: Volumes and activity levels of the spheres S1 – S6 for the NEMA-IEC phantom

Sphere	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
Volume (mL)	26	11.5	5.8	2.6	1.1	0.5
Activity (kBq/mL)(SBR-A)	999.7	441.6	223.0	100.0	42.3	19.2
Activity (kBq/mL)(SBR-W)	1190.4	525.9	265.5	119.0	50.4	22.9
Activity (kBq/mL)(SBR-8)	747.7	330.3	166.8	74.8	31.6	14.4

Table 3.2: Volumes and activity levels of of spheres Sp1 - Sp6 and B1 - B2 for the Elliptical Lung-Spine Body Phantom

Sphere	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**B1**	**B2**
Volume (mL)	0.5	1.0	2.0	4.0	8.3	16	196	199
Activity (kBq/mL)(SBR-A) A	36.8	70.2	143.8	289.2	597.1	1162.2	5931.0	5811.3
Activity (kBq/mL)(SBR-W)	29.5	46.2	115.2	231.7	478.3	931.02	7403.9	7254.5
Activity (kBq/mL)(SBR-8)	22.5	42.9	87.8	176.7	364.8	710	4523.2	4432.0

Table 3.3: List of QIN-HEADNECK collection studies that have manual segmentation. The exclusion criteria column are divided into 5 criteria defined in the text (Section 3.2.2). C2 and C4 have two conditions that are shown with two letters a and b.

	Exclusion Criteria					**Usage**	
Images	C1	C2	C3	C4	C5	Training	Test
QIN-HEADNECK-01-0003						*	
QIN-HEADNECK-01-0017		a					
QIN-HEADNECK-01-0024	*	a b		b		*	
QIN-HEADNECK-01-0026	*	a b		b			
QIN-HEADNECK-01-0028	*			b		*	
QIN-HEADNECK-01-0034		a					*
QIN-HEADNECK-01-0041		a		b			
QIN-HEADNECK-01-0046							
QIN-HEADNECK-01-0047	*	a b		a b		*	
QIN-HEADNECK-01-0055							*
QIN-HEADNECK-01-0057	*	a b	*	a b		*	
QIN-HEADNECK-01-0060		b	*			*	
QIN-HEADNECK-01-0062		a					
QIN-HEADNECK-01-0064					*		
QIN-HEADNECK-01-0065							*
QIN-HEADNECK-01-0067		a b	*			*	
QIN-HEADNECK-01-0071	*	b		b		*	
QIN-HEADNECK-01-0074		b		b		*	
QIN-HEADNECK-01-0079							*
QIN-HEADNECK-01-0081							*
QIN-HEADNECK-01-0086		b	*	b			
QIN-HEADNECK-01-0089	*	a		a		*	
QIN-HEADNECK-01-0091		b					*
QIN-HEADNECK-01-0098		b	*				
QIN-HEADNECK-01-0100			*	b			
QIN-HEADNECK-01-0112	*		*				
QIN-HEADNECK-01-0115				a b			
QIN-HEADNECK-01-0116		a					
QIN-HEADNECK-01-0118		a b		a b			
QIN-HEADNECK-01-0120							*
QIN-HEADNECK-01-0127							*
QIN-HEADNECK-01-0129		a	*				
QIN-HEADNECK-01-0135	*	a	*	b			
QIN-HEADNECK-01-0137							*
QIN-HEADNECK-01-0139		a b	*	b			
QIN-HEADNECK-01-0140		a		b			
QIN-HEADNECK-01-0157		a					
QIN-HEADNECK-01-0164	*	b		a b			
QIN-HEADNECK-01-0165		a					*
QIN-HEADNECK-01-0172		b		b			
QIN-HEADNECK-01-0180	*	b	*	a b			
QIN-HEADNECK-01-0184							
QIN-HEADNECK-01-0186	*	a					
QIN-HEADNECK-01-0190		b		b			
QIN-HEADNECK-01-0199							*
QIN-HEADNECK-01-0200							*
QIN-HEADNECK-01-0201		b	*	b			
QIN-HEADNECK-01-0203		a b		ab			
QIN-HEADNECK-01-0212							*
QIN-HEADNECK-01-0214							*
QIN-HEADNECK-01-0220			*	b			
QIN-HEADNECK-01-0224							*
QIN-HEADNECK-01-0228		b					
QIN-HEADNECK-01-0239	*	b		b			
QIN-HEADNECK-01-0244		b	*				
QIN-HEADNECK-01-0275		a					*
QIN-HEADNECK-01-0366		ab	*	b			

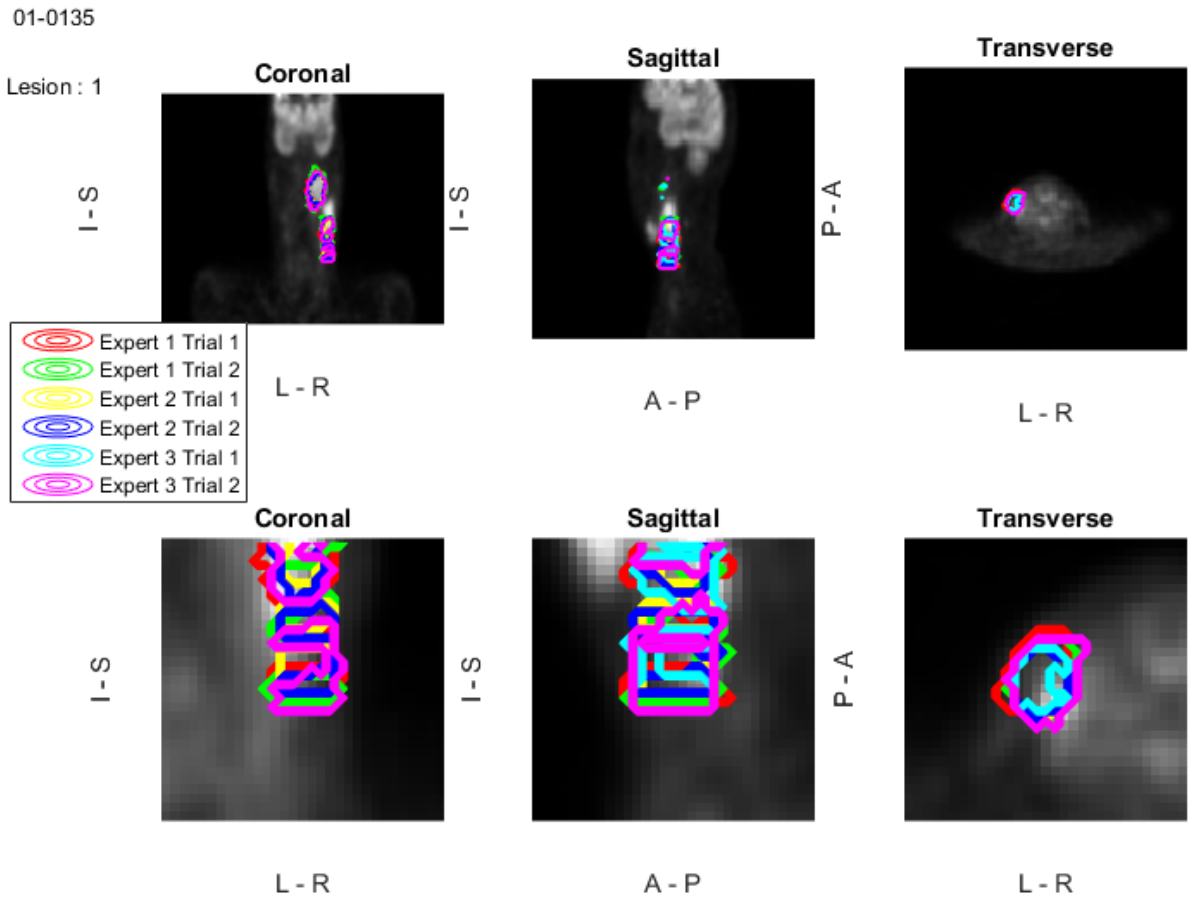


Figure 3.4: Closely connected lesions segmented separately in manual segmentations.

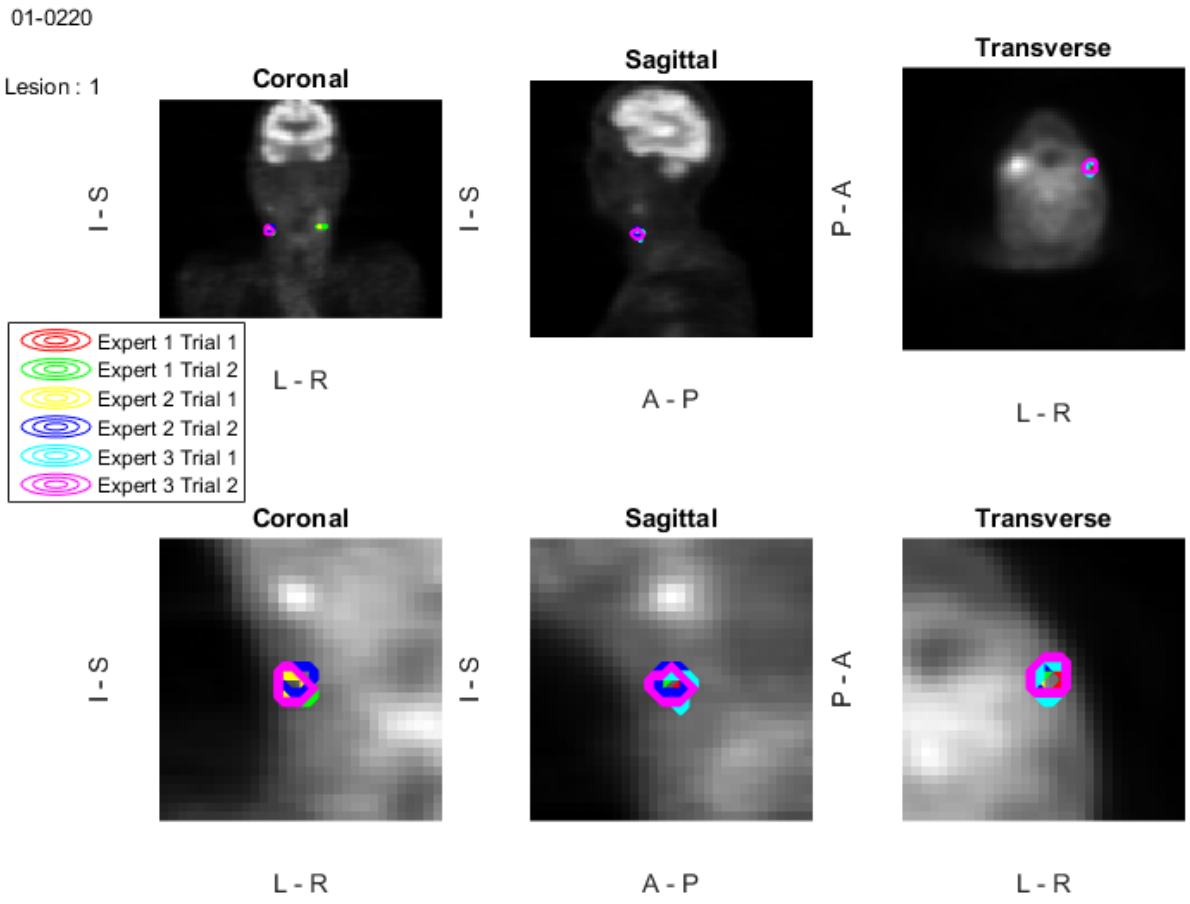


Figure 3.5: Separate isolated lesions which are not segmented.

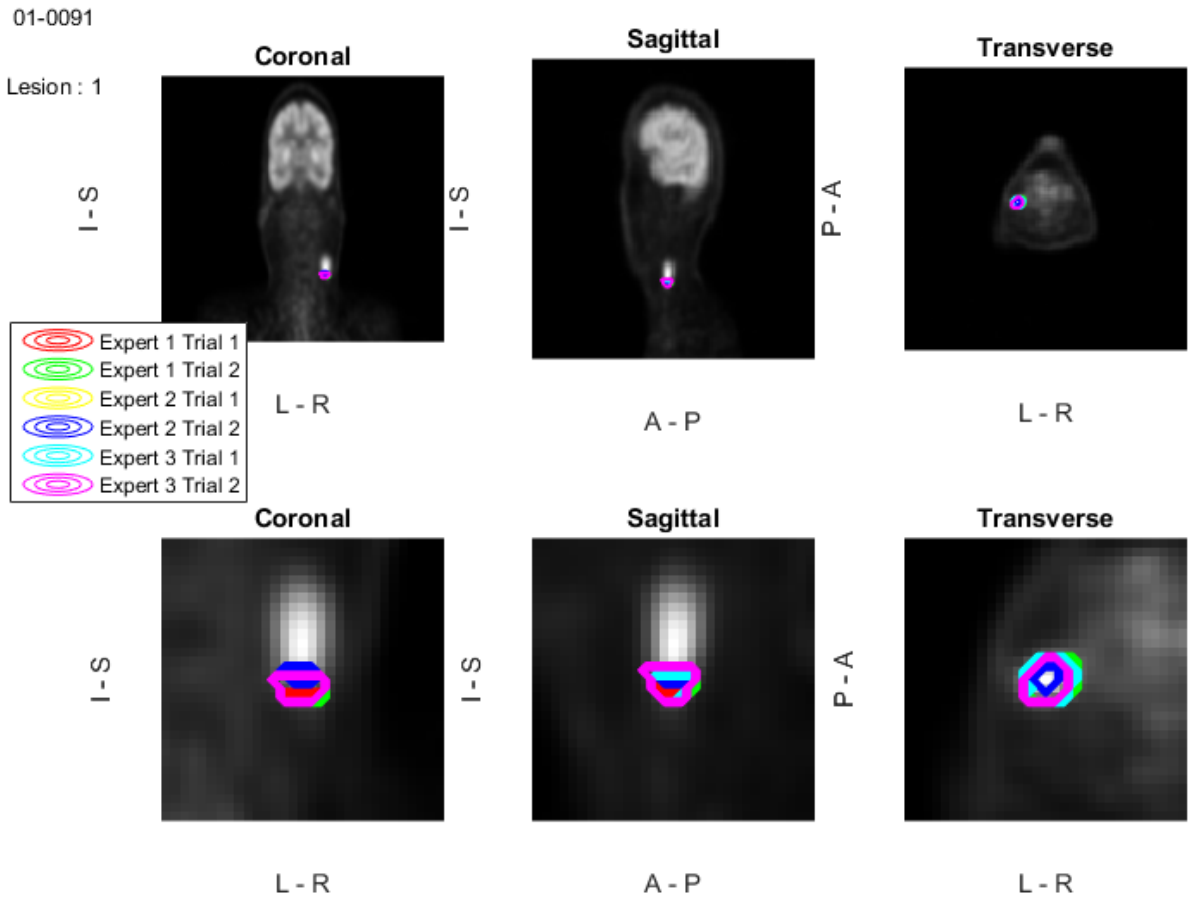


Figure 3.6: Portions/parts of lesions which are not segmented.

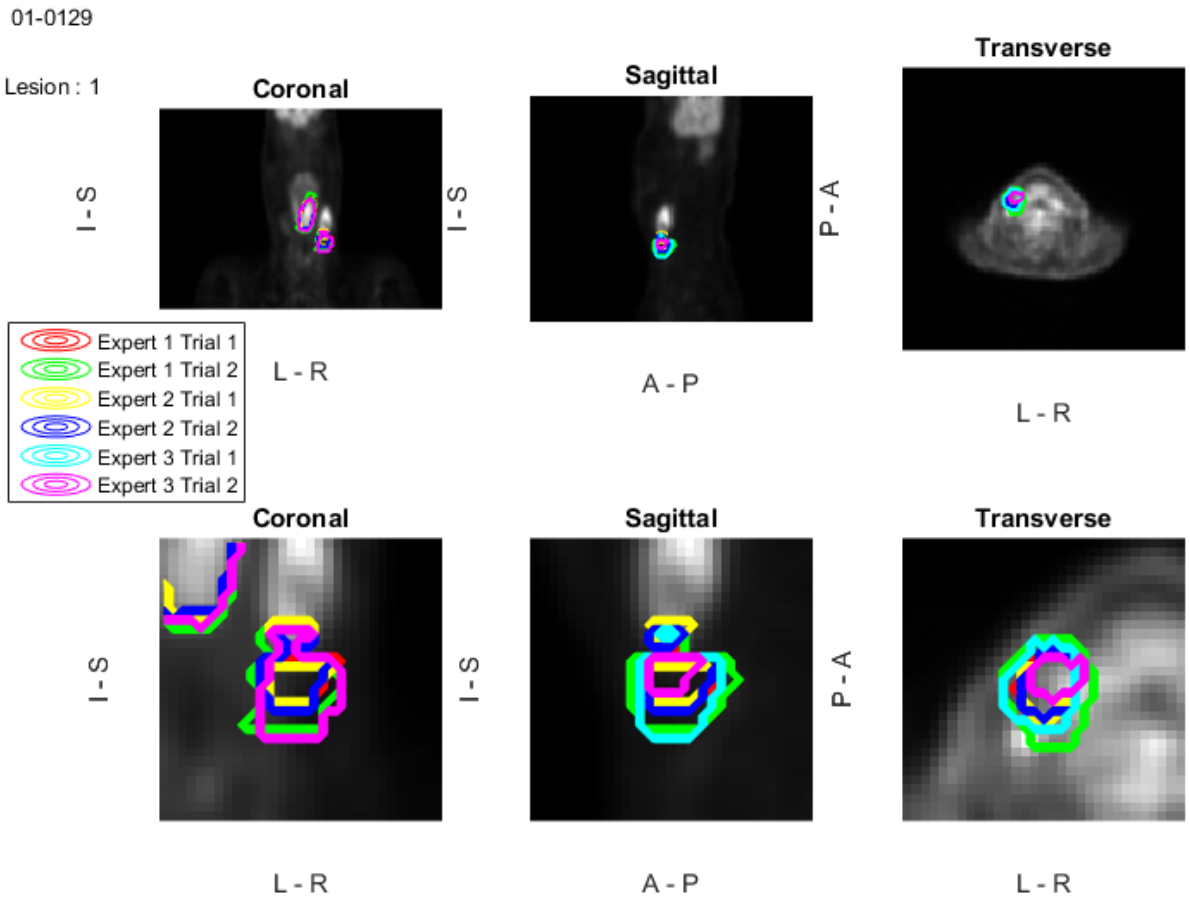


Figure 3.7: Misalignment: region segmented does not appear to be a high-uptake region.

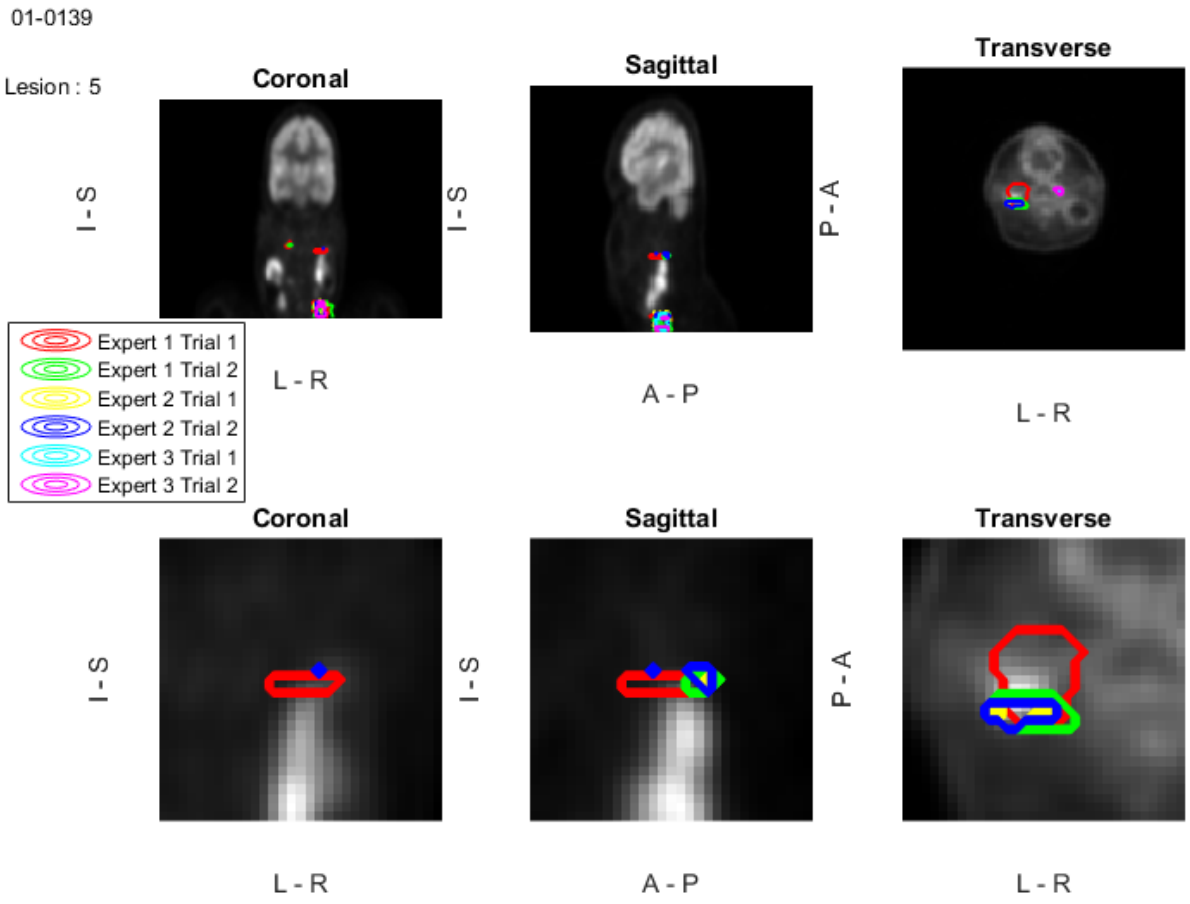


Figure 3.8: Low agreement among experts.

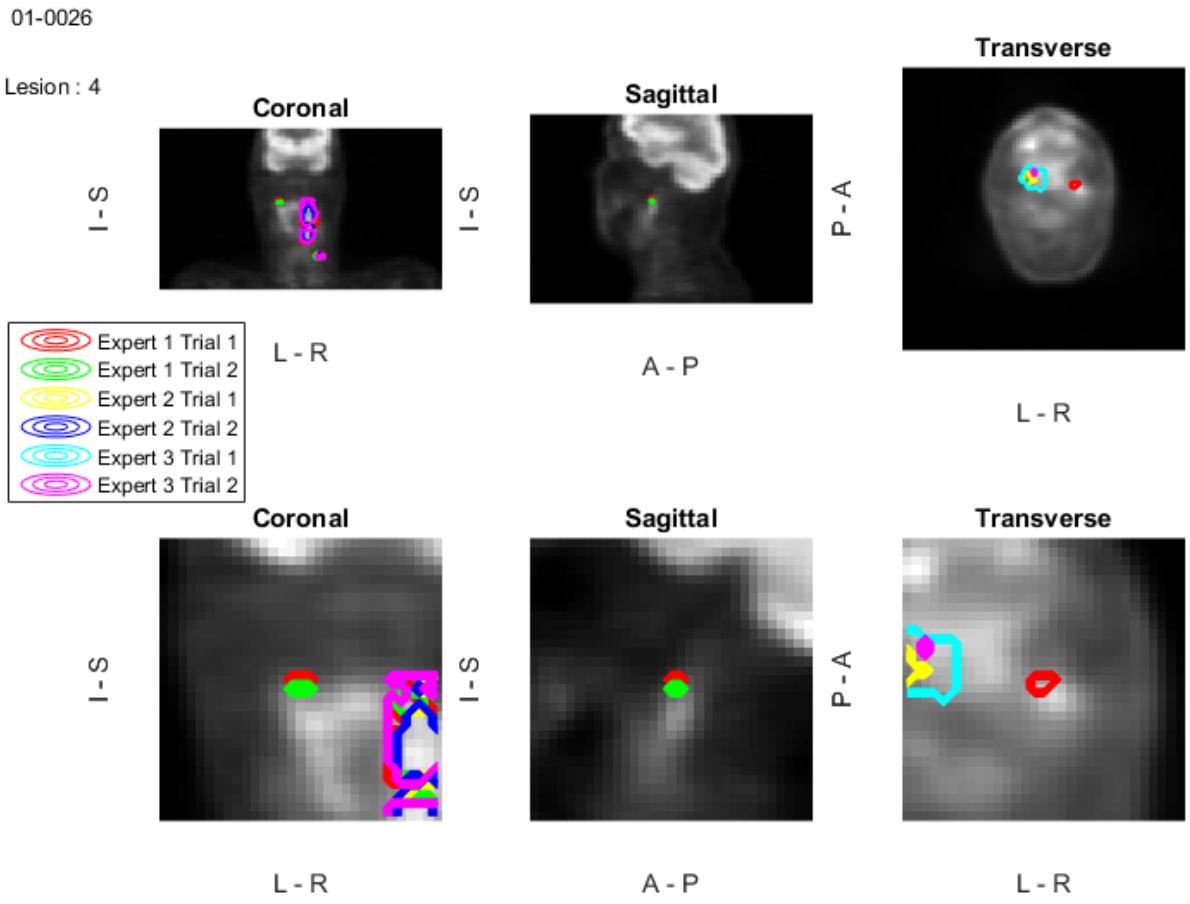


Figure 3.9: Some lesions have been segmented by one expert, but not by others.

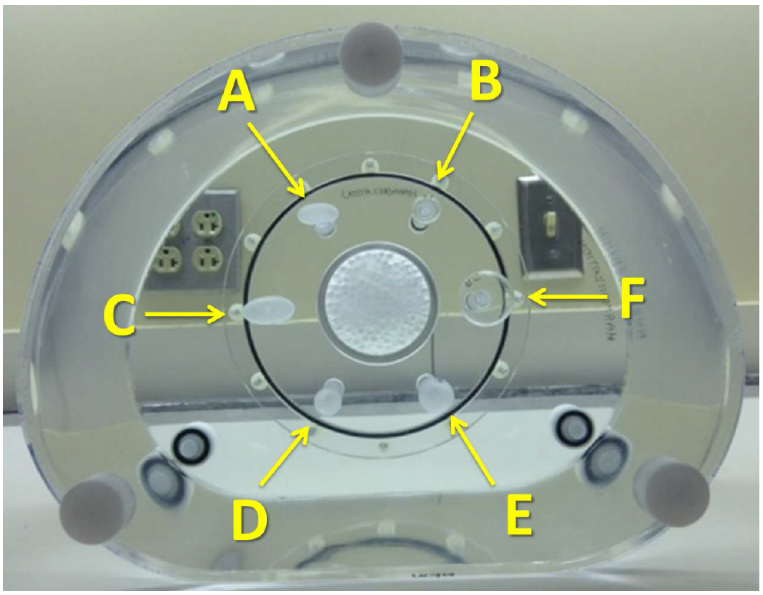


Figure 3.10: Phantom used in the QIN PET challenge.

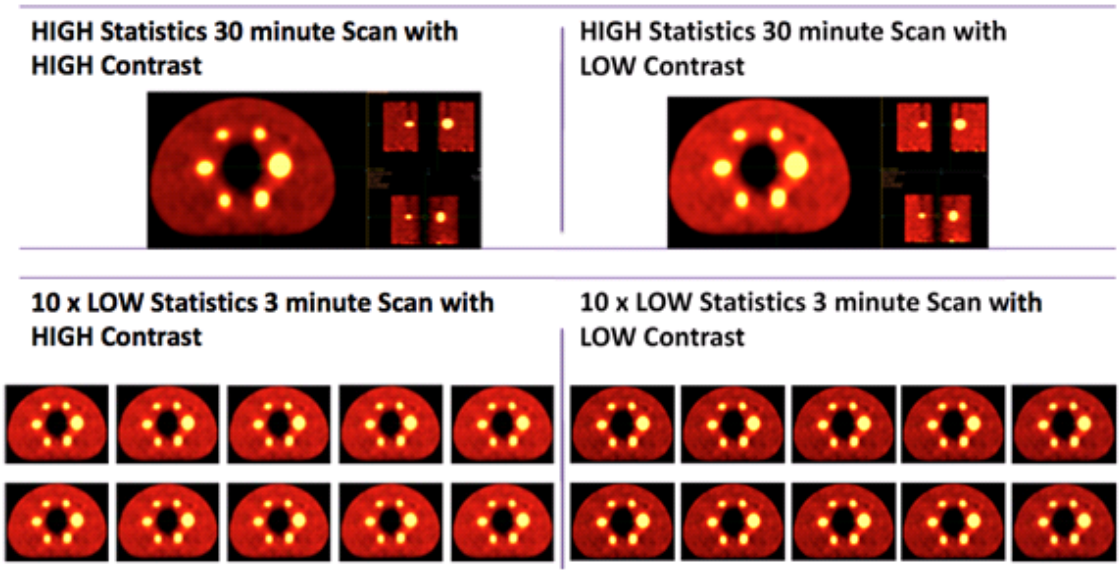


Figure 3.11: QIN challenge phantom data [7].

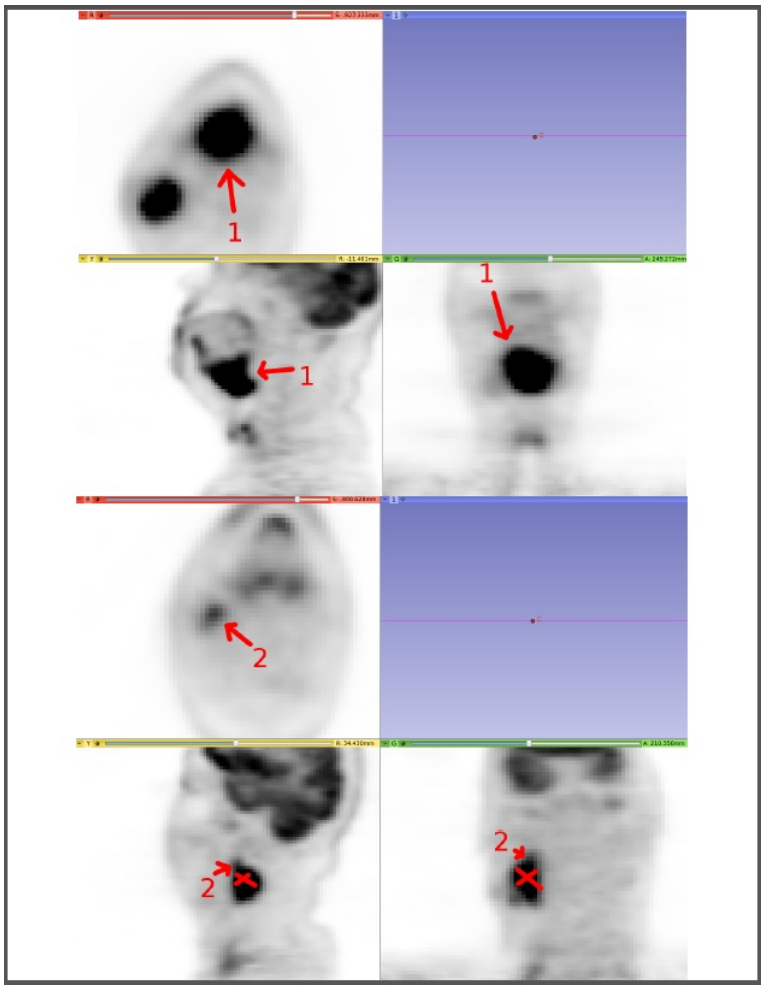


Figure 3.12: Ground truth provided in QIN challenge. The location of lesions required to be segment are shown.

Chapter 4

Results

4.1 Overview

In this chapter, the results of two proposed methods on three sets of data are discussed. In the first section, we evaluate the level set approach on UBC phantom data. The second section is concerned with the results of the QIN challenge in which we employ our method on phantom and real data. Finally, the results of convex segmentation with learnt likelihoods on the QIN head and neck data are presented. Our results are compared with state-of-the-art methods.

4.2 Level set method evaluated on UBC phantom

Level set method described in section 2.1 performed on UBC phantom data discussed in section 3.1. Since the measured activity in a dose calibrator is provided in the phantom data, we are able to quantify the accuracy by the percentage difference of segmentation to the ground truth. The quantification result is presented in Table 4.1, and the final segmentation can be seen in Fig 4.1. As can be inferred from Table 4.1 (LevelSet column), the difference between the activity estimated from the obtained segmentation and ground truth is high. To improve our results, we learn a constant to be added to the converged level set. By adding a positive or negative constant to the converged level set the zero level set that represents the final segmentation would expand or shrink. The results are reflected in the fixed-thresholding column in Table 4.1. Applying a fixed-thresholding step on level set improved the final segmentation to reach a reasonable estimation of the measured activity in a dose calibrator (Fig 4.2 shows the improvement of the segmentation using the fixed-thresholding). We also tried to segment the bottle and the sphere inside of the bottle that simulated a lesion inside of an organ. In the bottle case, the region of interest was segmented using the same experimental setup shown in Fig.4.3. However, targeting the sphere inside the bottle was a bit challenging, since

three different levels of activity (air, hot water inside the phantom, and bottle) were considered as the background. As a solution, two background terms (one for air and hot water together, and one for the) were defined in the objective functional. By using two separate background terms, our method was able to segment the small sphere inside the bottle (Fig 4.3).

Table 4.1: Quantification accuracy represented by the percent difference to the ground truth for level set and fixed-thresholding methods

Interst			LevelSet		Fixed-thresholding	
Region	Volume (mL)	Ground Truth	Calculated	Error	Calculated	Error
Sphere 1	0.5	22.49	14.26	36.6	25	11.1
Sphere 2	2	87.84	67.4	23.3	90	2.5
Sphere 3	4	176.71	135.83	23.1	177.89	0.7
Sphere 4	8	364.75	288.17	36.6	348.96	4.3
Sphere 5	16	709.99	582.47	18	683.71	3.7
Bottle 1	205	4523.24	3254.1	28.1	4323.8	4.4
Bottle 2	199	4431.96	3252.9	26.6	4265	3.8

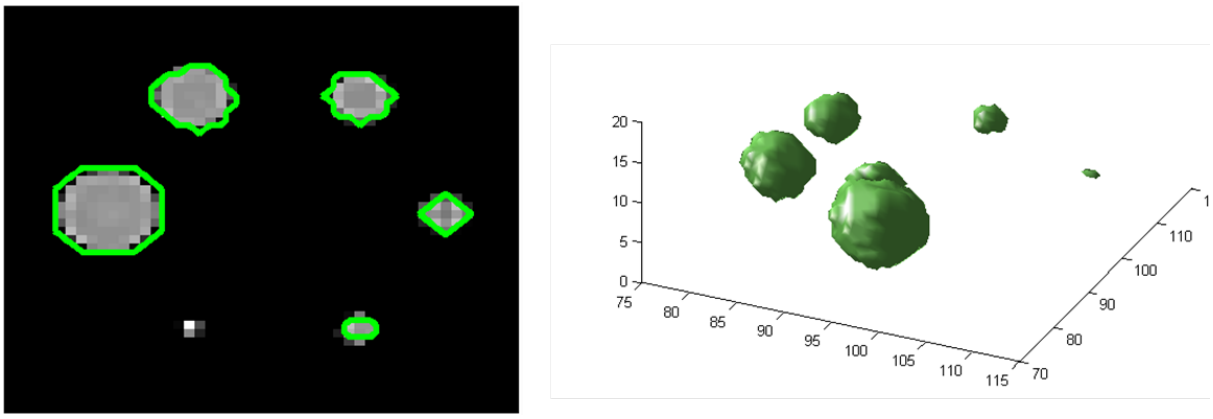


Figure 4.1: The segmentation of a phantom image. Six inserts with different level of radio activity were segmented.

4.3 Level set on QIN challenge

This section presents a reprint of: Beichel, R.R., Smith, B.J., Bauer, C., Ulrich, E.J., Ahmadvand, P., Budzevich, M.M., Gillies, R.J., Goldgof, D., Grkovski, M., Hamarneh et al. ” Multisite Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data”, Medical Physics journal, 2016 [7]. In this work, seven QIN sites (the method and site names

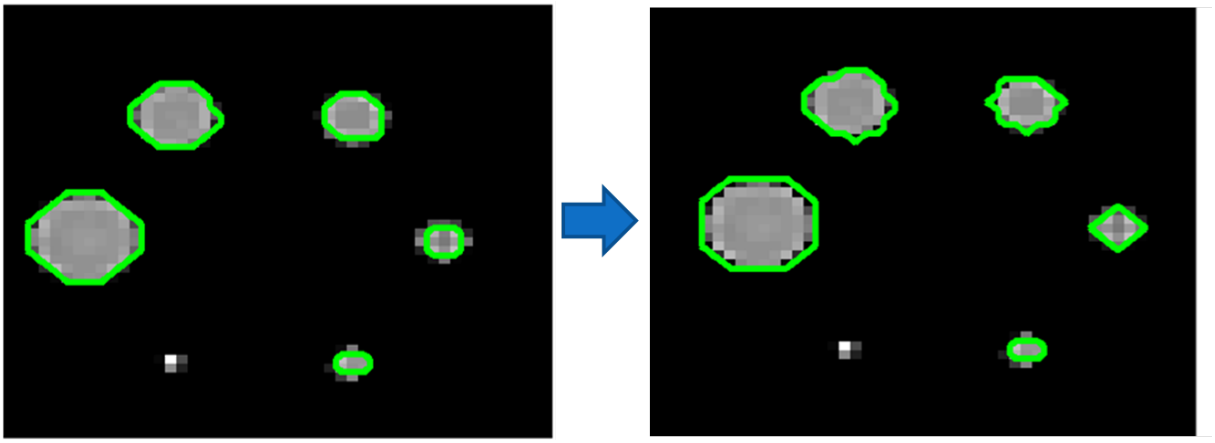


Figure 4.2: The segmentation of a phantom image using fixed-thresholding. A learnt constant was added to the converged level set (left image) to make more accurate segmentation (right image).

are listed in Table 4.3) attended the QIN challenge in which attendees were asked to segment the phantom and real images described in 3.3. We applied level set method described in 2.1 (with the implementation in section 2.1.3) on the QIN challenge data. The same hyperparameters settings were used for all images except for a few real images that required parameter-tuning.

Accuracy of volume measurement

Phantom scan: As reflected in Fig. 4.7 and Table. 4.2 (adopted from [7]), our proposed level set has the second lowest relative error in phantom data. The proposed method could segment all inserts in phantom images even in images with the high level of noise. Fig. 4.5 compares the result of our segmentation with the best method (proposed by Beichel et al. [8]) on three phantom images that were corrupted with different levels of noise.

Real scan: According to Fig. 4.7 and Table. 4.2 (adapted from [7]), segmentation of real images were more challenging. Particularly, our level set method performed worse in comparison with other methods. Dice coefficient is reported in Table IX in [7] ranks level set as the fourth method between seven QIN site's methods. Fig. 4.6 shows examples in which competing methods gave similar segmentation results in some cases. In other cases, there are varying degrees of variability in performance.

Reproducibility

All QIN challenge attendees were asked to repeat their experiments one week after the challenge event in order to evaluate the reproducibility of their segmentation methods. The errors of all QIN

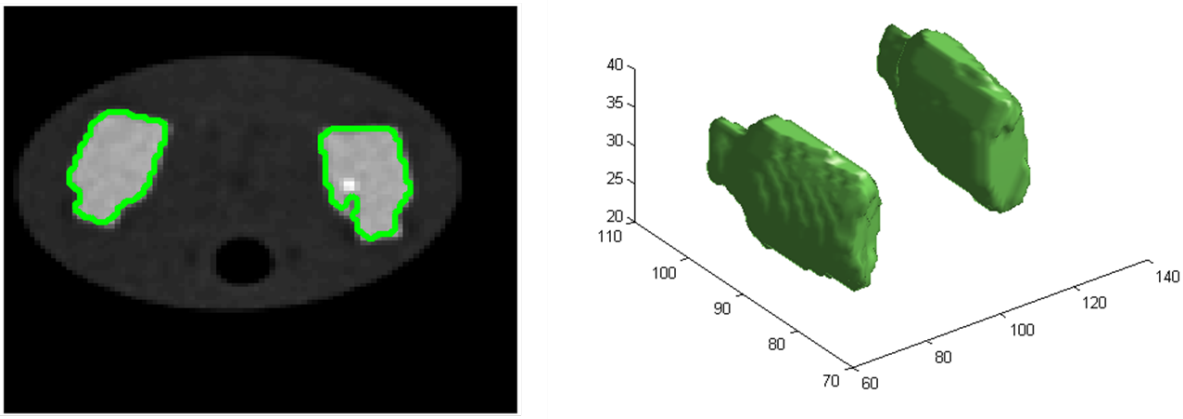


Figure 4.3: The segmentation of a phantom image. Two bottles with different levels of radio activity were segmented.

sites' segmentations for this second round can be seen in Fig. 4.7 and Table. 4.2, as reported in [7]. In addition, the differences in the Dice coefficients are reported in Table IX in [7]. While some methods have lower repeat error, none of the proposed methods are completely reproducible.

4.4 Level set with learnt likelihood on TCIA

4.4.1 Experimental Setup

We evaluated our approach on the H&N cancer collection provided by The Cancer Imaging Archive (TCIA) described in 3.2.2. Ten images from this collection were selected by the Quantitative Imaging Network (QIN) team for the QIN H&N challenge, in which all lesions and lymph nodes with high activity were delineated by experts (details can be found in section 3.2.2). We chose these 10 images for training and tested on 15 new images with one or two lesions in each image from the same collection. The training data came from Siemens Biograph Duo scanner. Six manual segmentations (three users, two trials each) are provided by TCIA for all 25 images. Training seed voxels were collected from the lesions, background, and regions of normal activity, for a total of 1108 seeds. The seed voxels and the manual segmentation of the training data (only) were used to tune hyper-parameters and select features; these hyper-parameters were fixed across all test images.

Table 4.2: Relative errors resulted by our approach for (a) phantom scans and (b) HNC scans.

Phantom						
	Relative Error (%)			Repeat Difference (%)		
Approach	Mean	SD	RMS	Mean	SD	IRMS
1	70.5	87.5	112	5.9	24.8	25.2
2	58.1	55.3	80	0.1	5.3	5.3
3	33.8	58.4	67.2	-36.4	54.6	65.2
4	87.8	179.2	198.7	-0.1	7.2	7.1
5	41.3	48.9	63.8	4.4	19.1	19.4
6	31	83.8	89.1	39.9	57.7	69.7
7(Our method)	29.9	58	65	-6.3	26	26.5
HNC scans						
	Relative Error (%)			Repeat Difference (%)		
Approach	Mean	SD	RMS	Mean	SD	RMS
1	82.1	169.6	187.6	-37.7	196.5	198
2	-5.6	35.7	36	-1.5	30.4	30.1
3	-21	44.4	48.9	31.5	56.2	63.8
4	701.5	1402.6	1560.8	-0.6	332.4	328.5
5	-25.4	48.6	54.6	21.8	56.3	59.8
6	-50.5	26	56.8	2.1	26.8	26.6
7(Our method)	48.6	119.4	128.3	6.5	37.7	37.8

Table 4.3: Methods proposed for PET segmentation by QIN sites

Method	**Technique**	**Handles * Normal Activity**	**Reproducible$^{\mp}$**	**Automatic**
Columbia University Medical Center	In-house developed software based on an active contour segmentation approach	✗	✗	✗
University of Iowa	In-house developed software utilizing a graph-based optimization approach	✗	✗	✗
University of South Florid	Commercial software package Mirada Medical RTx	✗	✗	✗
Memorial Sloan Kettering Cancer Center	Combination of commercial software packages VCAR and PMOD	✗	✗	✗
University of Pittsburgh	Commercial software package MIM	✗	✗	✗
University of Washington Medical Center	Commercial software package PMOD	✗	✗	✗
Simon Fraser University	Active surface	✗	✗	✗
Our Method (level set with learnt likelihood)	Machine Learning & Convex seg.	✓	✓	✓

*Handles normal activity automatically *i.e.* without user seed points $^{\mp}$Accuracy dependent on parameter choice, which includes size of ROI.

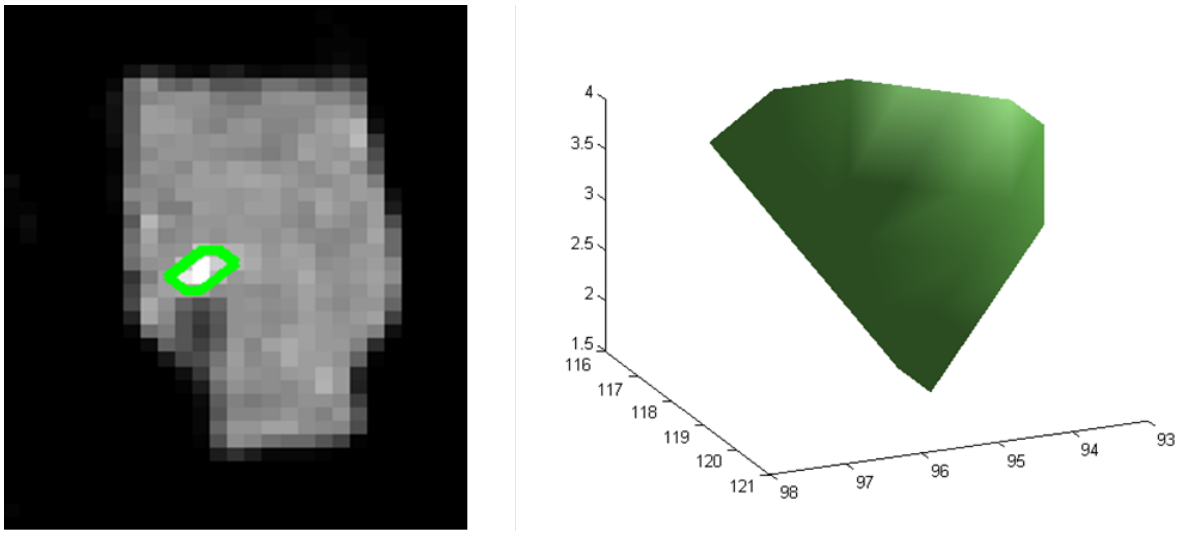


Figure 4.4: The segmentation of a phantom image. The sphere inside of the bottle were segmented correctly.

4.4.2 Learning stage evaluation

During the learning stage, the accuracy of the classifier is evaluated. The average of classifier accuracy over 20 times resampling reached to $86\% \pm 3\%$, as evaluated using leave-one-image-out cross-validation. The accuracy of the segmentation is calculated using three training images to fix the parameter of convex active surface. Fig.4.8 shows a sample of the training result.

4.4.3 Quantitative segmentation results

Our segmentation method was validated against manual segmentations using the Dice similarity coefficient (DSC), Jaccard index, false positive (FP), and false negative (FN) rates. These results are presented in Table 4.4. Different combinations of classes for training the classifier prior to segmentation were evaluated (Rows 1-5). Note that Row 6 of Table 4.4 reports the result of the performance of our method without the final segmentation step (i.e. a voxel is assigned to the class with highest random forest probability). Note also that DSC, Jaccard, FP, and FN are the average values obtained from comparing the automatic segmentation with the 6 manual delineations (3 users, 2 delineations each). We compared our method with the state-of-the-art work of Foster et al. [24] and Beichel et al. [8] using their publicly available segmentation software (Row 7 and 8). In Table 4.4, Row 8 is the average agreement computed over all scans between manual segmentations produced by the three expert users where each combination of users was included i.e.

(user1,user2), (user1,user3) etc.

Comparing with Foster's method

We compared our method with the state-of-the-art work of Foster et al. (2014) [24] using their publicly available PET-only segmentation software. Foster et al. method requires selecting a 3D region of interest around the lesions and tweaking parameters until the best performing results are obtained. This is an extra burden that took an expert with knowledgeable in PET imaging who understands the meaning of parameters on average 5 minutes per image. As can be seen from Table 4.4 row 7 and Table 1.1 row 3, our method performs statistically better with higher DICE and Jaccard along with lower TP and FN in comparison to Foster's method, while the proposed method does not need user interaction and is reproducible as well.

Comparing with semi-automatic method

We also compared the proposed method with the state-of-the-art work of Beichel et al. (2016) [8] in which graph-based optimization is used to solve the segmentation problem. This method needs an approximate lesion center point provided by a user, then a graph is structured around the center point and local image statistic information is used to derive the cost function. The semi-automated algorithm is implemented as an extension for 3D Slicer [6], which is an open source software for medical-image computing and visualization. An experienced user spent on average 3 minutes per image to click on the center of each lesion. The semi-automated algorithm has some parameters, but all parameters are fine tuned on the QIN-HEADNECK data which is exactly the same data used for the proposed method. Semi-automated method statistically is doing better with higher DICE and Jaccard. However, firstly, this method is not automatic and need a user interaction per lesion, secondly, it is not reproducible. Although the Semi-automated method reduces considerably the segmentation variation in compare with manual segmentation, it is not still a reproducible method [8].

Segmentation agreement

We evaluated the inter- and intra-user agreement in lesion delineation and compared it to results from our automated method and those obtained from expert users. We found that user2 has the highest intra-user variability with an average DSC of 0.844, and that user3 agreed the least with the two other users with an average DSC of 0.787. Our fully automated method had an average DSC agreement of 0.741 with other users, i.e. our method falls short by only 4% from performing as well as expert user3. Note Zijdenbos et al [67] suggested that a DSC higher 0.7 indicates good agreement.

4.4.4 Quantitative segmentation results

Fig. 2.13 shows examples of segmentation results from four cases. In the third case, our method segmented a lesion that was missed in the manual segmentation. In the fourth case, a small segmentation leaks into the inferior part of the brain is observed.

4.4.5 Comparing with QIN challenge methods

We were not able to provide a quantitative comparison with QIN site methods, because the images provided by QIN challenge are used for training propose. However, in Table 4.3 we compare our level set with learnt likelihood with seven other methods (including our level set method) proposed in the QIN challenge. Our fully automated method is the only method that can handle normal activity, is reproducible, and does not need any user interaction.

Table 4.4: Segmentation results. (Row 1-6) Variants of our proposed method evaluated on different combinations of classes, with each class surrounded by parentheses (e.g. 2 classes in row 1); (Row 7 and 8) Competing method; (Row 8) Average segmentation agreement for 3 users. Items with the same class label are shown with the same color.

	Method	DSC	Jaccard	FP	FN
1	[Air + Body + Kidney + Heart + Brain] [Lesion]	0.71	0.57	0.66	0.03
2	[Air + Body] [Kidney + Heart + Brain] [Lesion]	0.71	0.56	0.60	0.04
3	[Air] [Body] [Kidney + Heart + Brain] [Lesion]	0.73	0.59	0.41	0.05
4	[Air + Body] [Kidney + Heart] [Brain] [Lesion]	0.72	0.57	0.57	0.04
5	[Air] [Body] [Kidney + Heart] [Brain] [Lesion]	0.74	0.60	0.46	0.04
6	[Air] [Body] [Kidney + Heart] [Brain] [Lesion] w/o seg	0.71	0.56	0.70	0.03
7	Foster et al. (2014) [24]	0.71	0.58	9.8	0.05
8	Beichel et al. (2016) [8]	0.79	0.67	0.30	0.17
9	Average pairwise agreement between 3 expert segmentations	0.80	0.67	0.27	0.17

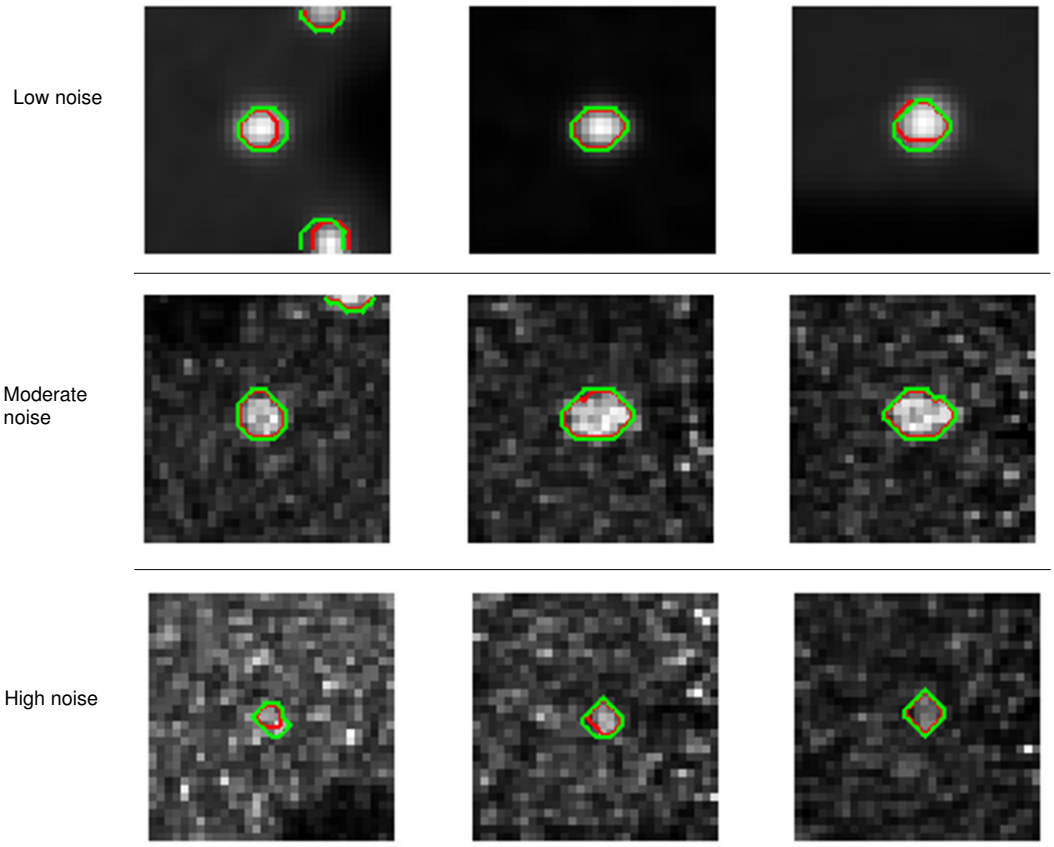


Figure 4.5: Comparing the performance of our method (the green contour) with Beichel et al. [8] (red contour). The three rows are three different phantom images with different levels of noise.

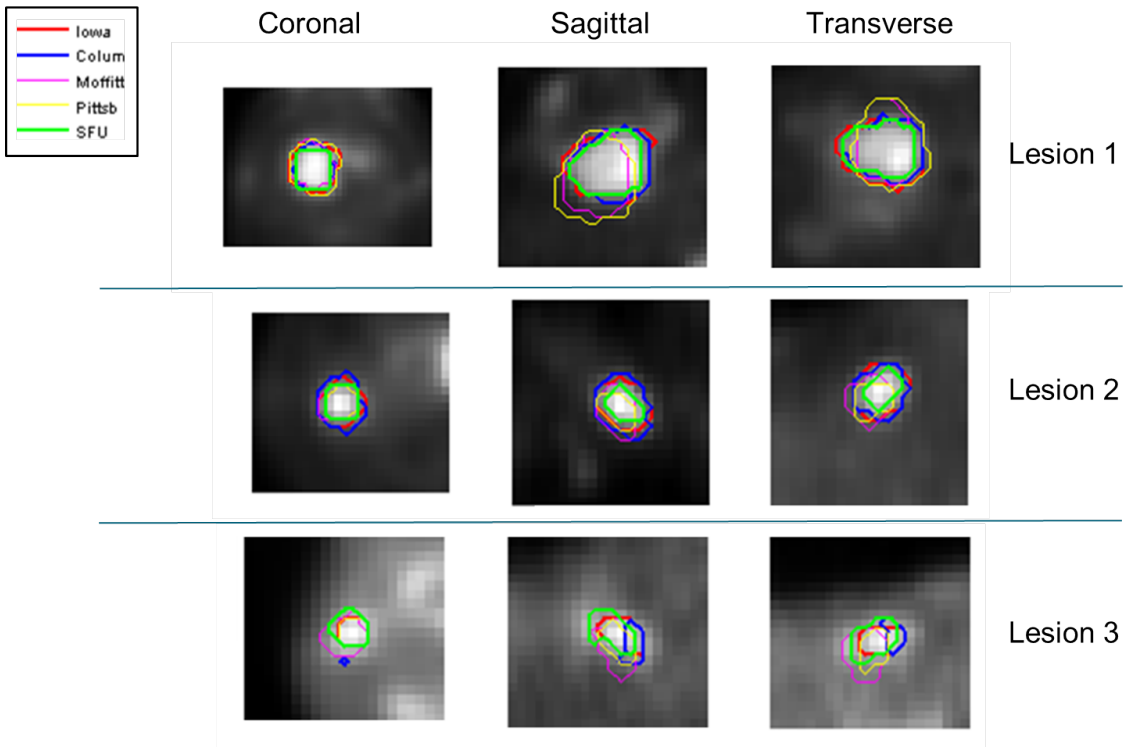


Figure 4.6: Comparing the performance of our method (the green contour) with other QIN site's methods.

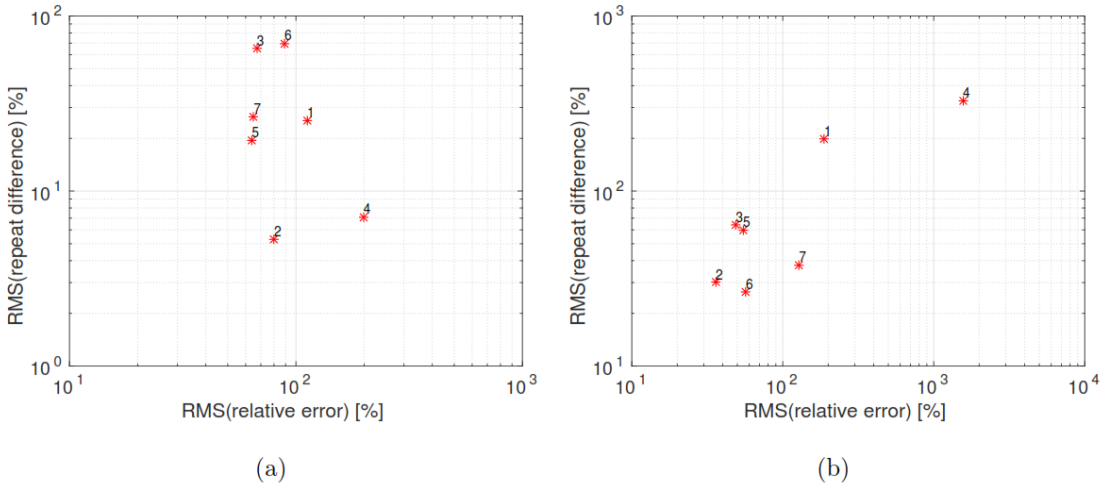


Figure 4.7: Comparing the performance of seven QIN methods segmentation on (a) phantom and (b) Head and Neck [7]. Number 7 is our method.

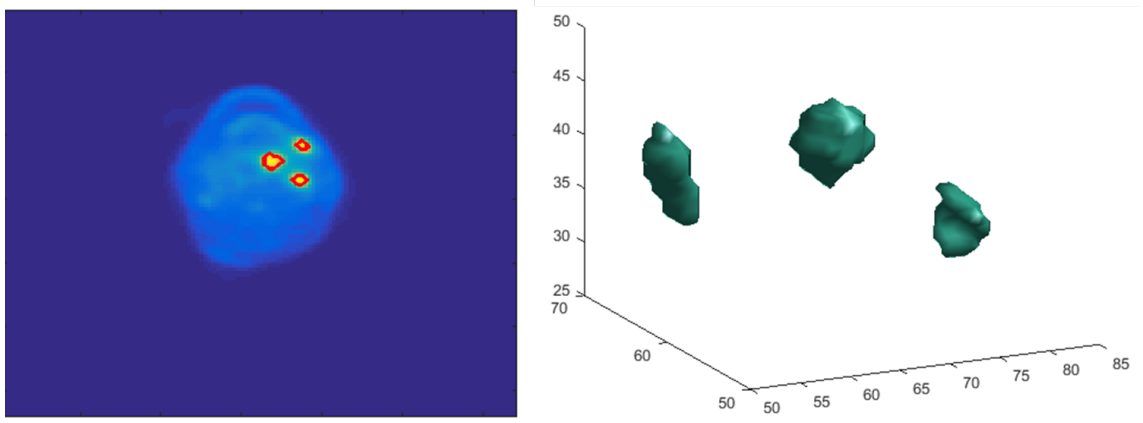
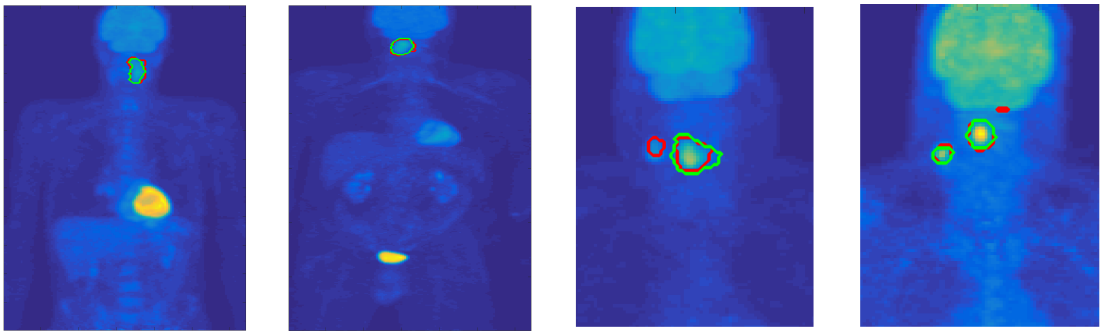


Figure 4.8: The segmentation result of a training image. Three lesions are segmented; the left image shows the segmentation from axial plane, and right images is a 3D view of the segmentation.



(a) (b) (c) (d)

Figure 4.9: Qualitative segmentation results. The PET image is rendered using maximum intensity projection. Our proposed segmentation is shown as red contours, while an example manual segmentation is shown in green. Note in (c) that our method captures a valid lesion missed by the user. In (d), we see an example of segmentation leakage into the inferior part of the brain.

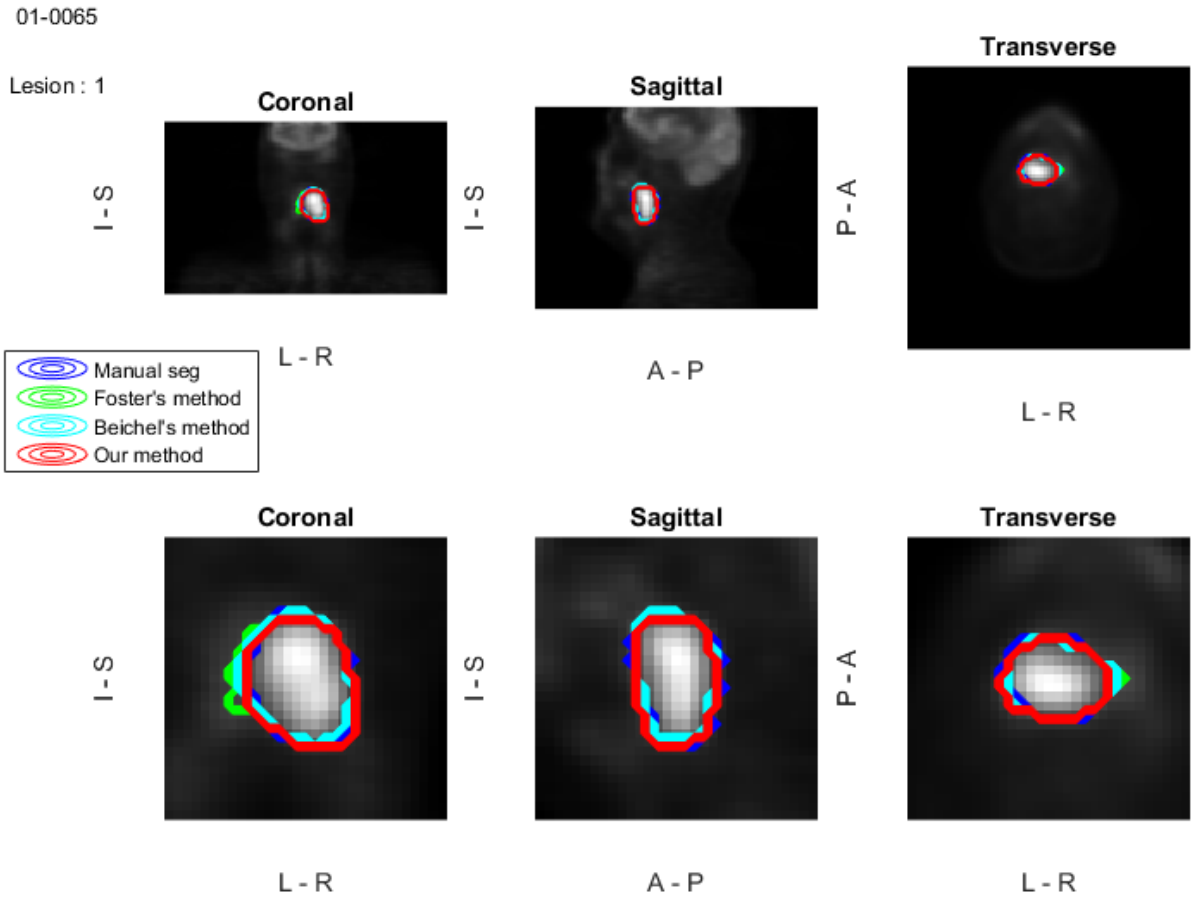


Figure 4.10: Comparison of our method with the state-of-the-art work of Foster et al. [24], Beichel et al. [8], and manual segmentation.

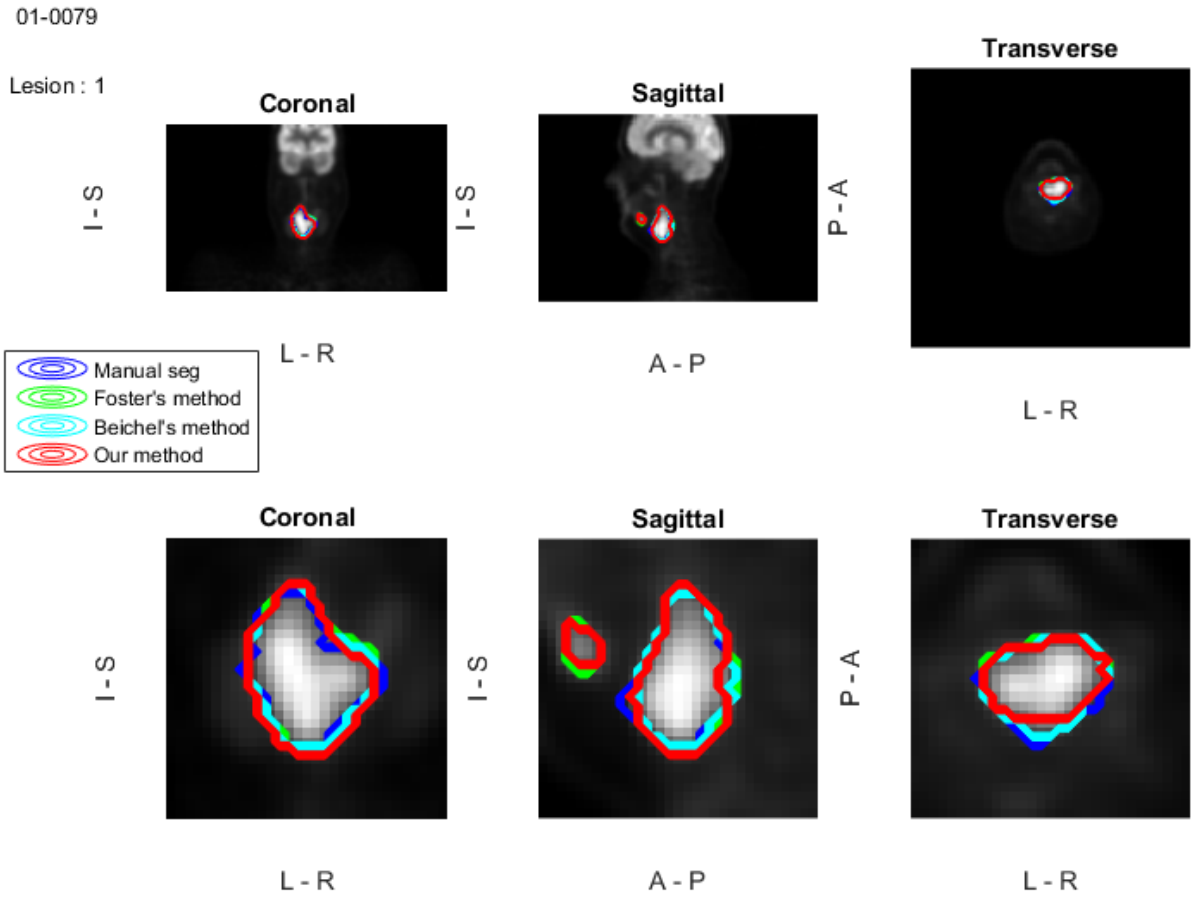


Figure 4.11: Comparison of our method with the state-of-the-art work of Foster et al. [24], Beichel et al. [8], and manual segmentation.

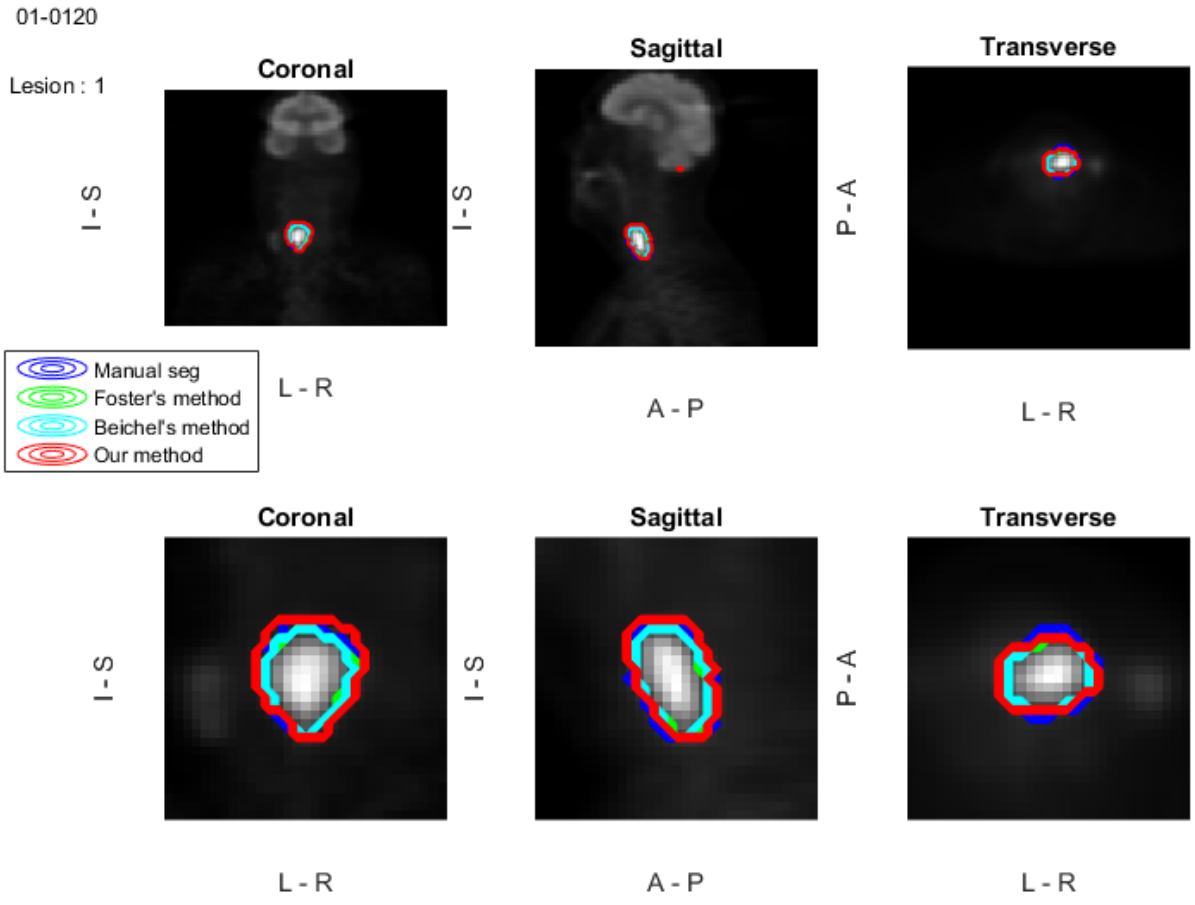


Figure 4.12: Comparison of our method with the state-of-the-art work of Foster et al. [24], Beichel et al. [8], and manual segmentation.

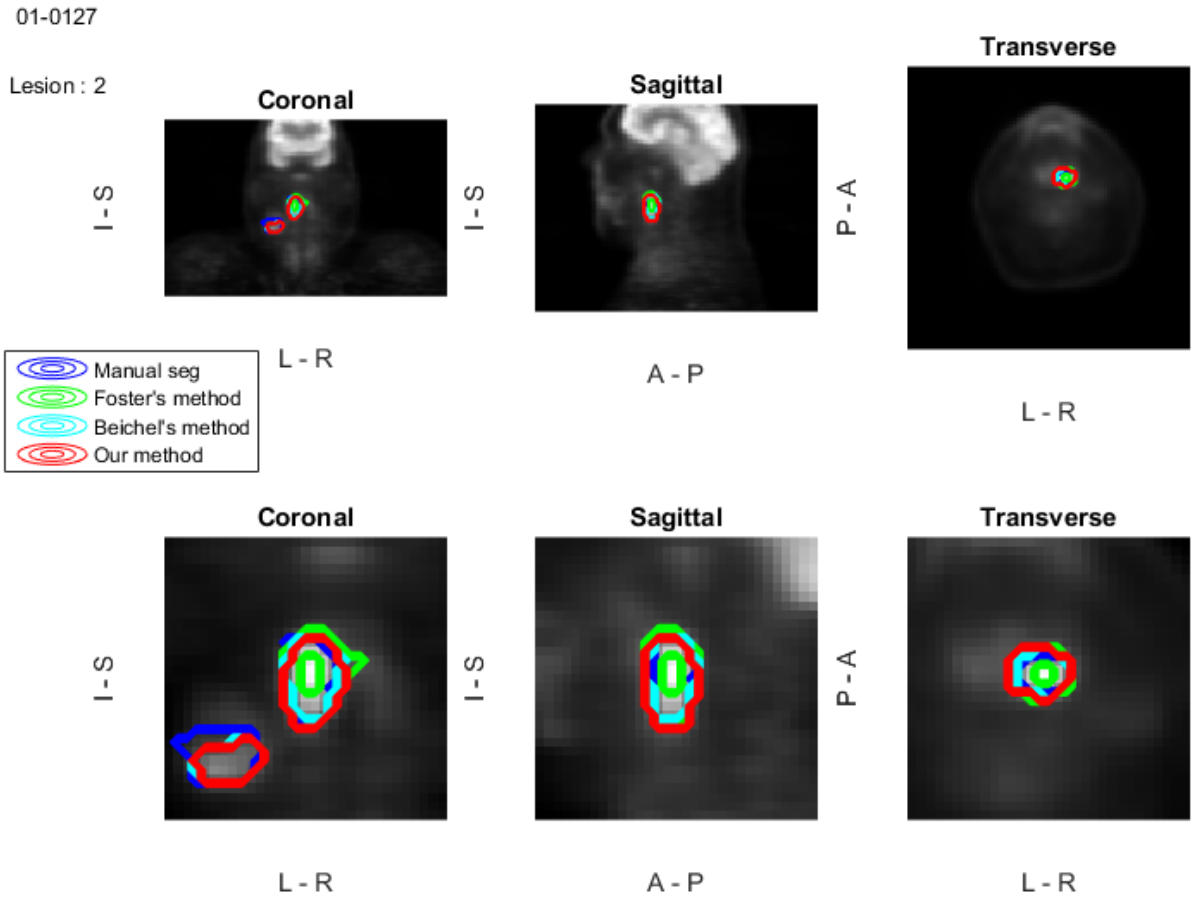


Figure 4.13: Comparison of our method with the state-of-the-art work of Foster et al. [24], Beichel et al. [8], and manual segmentation.

Chapter 5

Conclusions and Future Work

The primary problem that this thesis focuses on is the reproducible and automatic delineation of lesions for radiation treatment of head and neck cancer. Accuracy in target segmentation is essential for the success of radiotherapy. Also, reproducible result improves radiation therapy. We present the first work that fully automates the segmentation of lesions relying solely on PET.

The proposed method uses a random forest classifier to construct a prior model of the data. The proposed method has been shown to successfully distinguish functional abnormalities from normal uptake of PET radiotracers while also providing a segmentation of these regions. Our method is able to isolate abnormal lesion activity from the background and from other tissue regions of normal radiotracer uptake. Our method is a convex segmentation technique that is guided by learnt likelihood terms. The learning is based on the classification model trained on anatomically and physiologically meaningful cues. The use of convex formulation together with a trained classifier to learn all parameters precludes the need for human interaction (i.e. initialization and parameter tuning) and results in fully reproducible results on new test images (obtained from a different scanner with different reconstruction parameters).

Our approach was compared with six other QIN sites who participated in the QIN challenge. Contrary to other methods, our approach is fully automatic and reproducible. Our approach outperforms the recently published state-of-the-art method for this application and differed in average Dice similarity coefficient (DSC) by just 4% compared with an expert user.

In future work, the approach can be extended to other radiotracers of interest in oncology. It is also interesting to incorporate anatomical information from CT, when available. The results may be improved by employing a deep neural network to avoid designing features manually. However, this needs more training data.

Bibliography

[1] Mehrsima Abdoli, RAJO Dierckx, and Habib Zaidi. Contourlet-based active contour model for PET image segmentation. *Medical Physics*, 40(8):082507:1–12, 2013. 7, 9, 10

[2] Payam Ahmadvand, Nóirín Duggan, François Bénard, and Ghassan Hamarneh. Tumor lesion segmentation from 3D PET using a machine learning driven active surface. In *International Workshop on Machine Learning in Medical Imaging*, pages 271–278. Springer, 2016. 16

[3] Payam Ahmadvand, Hillgan Ma, Jesse Tanguay, Anna Celler, Benard Francois, and Ghassan Hamarneh. Level-set based segmentation of tumour lesions in PET images. In *Quantitative Imaging Network (QIN) Annual Meeting*, 2016. 16

[4] Sangtae Ahn and Jeffrey A Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Transactions on Medical Imaging*, 22(5):613–626, 2003. 36, 37

[5] Ulas Bagci, Jayaram K. Udupa, Neil Mendhiratta, Brent Foster, Ziyue Xu, Jianhua Yao, Xinjian Chen, and Daniel J. Mollura. Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. *Medical Image Analysis*, 17(8):929–945, 2013. 8, 10

[6] Christian Bauer and Markus Van Tol. *PET Tumor Segmentation [Online], Available: https://www.slicer.org/wiki/Documentation/4.5/Extensions/PETTumorSegmentation*. 58

[7] Reinhard R Beichel, Brian J Smith, Christian Bauer, Ethan J Ulrich, Payam Ahmadvand, Mikalai M Budzevich, Robert J Gillies, Dmitry Goldgof, Milan Grkovski, Ghassan Hamarneh, et al. Multi-site quality and variability analysis of 3D FDG PET segmentations based on phantom and clinical image data. *Medical Physics*, 44(2):479–496, 2016. xiii, xiv, 16, 50, 53, 54, 55, 62

[8] Reinhard R Beichel, Markus Van Tol, Ethan J Ulrich, Christian Bauer, Tangel Chang, Kristin A Plichta, Brian J Smith, John J Sunderland, Michael M Graham, Milan Sonka, et al. Semi-automated segmentation of head and neck cancers in 18F-FDG PET scans: A just-enough-interaction approach. *Medical Physics*, 43(6):2948–2964, 2016. xiv, 8, 9, 10, 39, 54, 57, 58, 59, 60, 64, 65, 66, 67

[9] Lei Bi, Jinman Kim, Dagan Feng, and Michael Fulham. Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies. In *MICCAI*, pages 569–576, 2014. 7, 10

[10] Kenneth J Biehl, Feng-Ming Kong, Farrokh Dehdashti, Jian-Yue Jin, Sasa Mutic, Issam El Naqa, Barry A Siegel, and Jeffrey D Bradley. 18f-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *Journal of Nuclear Medicine*, 47(11):2006, 1808-1812 2006. 4

[11] Michael A Blake and Mannudeep K Kalra. *Imaging in Oncology*, volume 143. Springer Science & Business Media, 2008. xii, 9, 11, 12

[12] Thomas Bortfeld, Rupert Schmidt-Ullrich, Wilfried De Neve, and David E Wazer. *Image-guided IMRT*, volume 18. Cambridge Univ Press, 2006. 13

[13] Zdravko I Botev, Joseph F Grotowski, and Dirk P Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010. 5

[14] Xavier Bresson, Selim Esedoglu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007. 25

[15] Curtis B Caldwell, Katherine Mah, Matthew Skinner, and Cyril E Danjoux. Can PET provide the 3d extent of tumor motion for individualized internal target volumes? a phantom study of the limitations of CT and the promise of PET. *International Journal of Radiation Oncology* Biology* Physics*, 55(5):1381–1393, 2003. 4

[16] 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. 37, 38

[17] David A Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002. 24

[18] Hui Cui, Xiuying Wang, Weiran Lin, Jianlong Zhou, Stefan Eberl, Dagan Feng, and Michael Fulham. Primary lung tumor segmentation from PET–CT volumes with spatial–topological constraint. *International Journal of Computer Assisted Radiology and Surgery*, 11(1):19–29, 2015. 8, 9, 10

[19] Michel Defrise, Paul E Kinahan, and Christian J Michel. Image reconstruction algorithms in PET. In *Positron Emission Tomography*, pages 63–91. Springer, 2005. 2

[20] A. Dewalle-Vignion, N. Betrouni, R. Lopes, D. Huglo, S. Stute, and M. Vermandel. A new method for volume segmentation of PET images, based on possibility theory. *IEEE Transactions on Medical Imaging*, 30(2):409–423, Feb 2011. 6, 9, 10

[21] Yusuf E Erdi, Osama Mawlawi, Steven M Larson, Massimo Imbriaco, Henry Yeung, R Finn, and John L Humm. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer*, 80(12):2505–2509, 1997. 4

[22] Andriy Fedorov, David Clunie, Ethan Ulrich, Christian Bauer, Andreas Wahle, Bartley Brown, Michael Onken, Jörg Riesmeier, Steve Pieper, Ron Kikinis, et al. DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. *PeerJ*, 4:e2057, 2016. 38, 39

[23] Brent Foster, Ulas Bagci, Awais Mansoor, Ziyue Xu, and Daniel J. Mollura. A review on segmentation of positron emission tomography images. *Computers in Biology and Medicine*, 50:76–96, 2014. 4, 8, 9

[24] Brent Foster, Ulas Bagci, Ziyue Xu, Bappaditya Dey, Brian Luna, William Bishai, Sanjay Jain, and Daniel J Mollura. Segmentation of PET images for computer-aided functional quantification of tuberculosis in small animal models. *IEEE Transactions on Biomedical Engineering*, 61(3):711–724, March 2014. xiv, 5, 10, 57, 58, 59, 64, 65, 66, 67

[25] Evan Godt. QIN network [online], available: http://trimed-cms-prod.s3.amazonaws.com/styles/large/s3/assets/articles/qin.jpg, [accessed: 06-02- 2017]. xiii, 39

[26] Leo Grady. Random walks for image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(11):1768–1783, Nov 2006. 8

[27] Pierre Graff, Michel Lapeyre, Emmanuel Desandes, Cécile Ortholan, René-Jean Bensadoun, Marc Alfonsi, Philippe Maingon, Philippe Giraud, Jean Bourhis, Vincent Marchesi, et al. Impact of intensity-modulated radiotherapy on health-related quality of life for head and neck cancer patients: matched-pair comparison with conventional radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 67(5):1309–1317, 2007. 13

[28] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6):610–621, 1973. 24

[29] Mathieu Hatt, Catherine Cheze le Rest, Patrice Descourt, André Dekker, Dirk De Ruysscher, Michel Oellers, Philippe Lambin, Olivier Pradier, and Dimitris Visvikis. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *International Journal of Radiation Oncology Biology Physics*, 77(1):301–308, 2010. 6, 9, 10

[30] Mathieu Hatt, Catherine Cheze Le Rest, Alexandre Turzo, Christian Roux, and Dimitris Visvikis. A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET. *IEEE Transactions on Medical Imaging*, 28(6):881–893, June 2009. 6

[31] Thi H Ho, Lisa Barbera, Refik Saskin, Hong Lu, Bridget A Neville, and Craig C Earle. Trends in the aggressiveness of end-of-life cancer care in the universal health care system of Ontario, Canada. *Journal of Clinical Oncology*, 29(12):1587–1591, 2011. 13

[32] H Jadvar and JA Parker. Pet physics and instrumentation. *Clinical PET and PET/CT*, pages 1–43, 2005. 2

[33] Ahmedin Jemal, Rebecca Siegel, Elizabeth Ward, Yongping Hao, Jiaquan Xu, Taylor Murray, and Michael J Thun. Cancer statistics, 2008. *CA: a Cancer Journal for Clinicians*, 58(2):71–96, 2008. 13

[34] Walter Jentzen, Lutz Freudenberg, Ernst G Eising, Melanie Heinze, Wolfgang Brandau, and Andreas Bockisch. Segmentation of PET volumes by iterative image thresholding. *Journal of Nuclear Medicine*, 48(1):108–114, 2007.

[35] Wei Ju, Deihui Xiang, Bin Zhang, Lirong Wang, I. Kopriva, and Xinjian Chen. Random walk and graph cut for co-segmentation of lung tumor on PET-CT images. *IEEE Transactions on Image Processing*, 24(12):5854–5867, Dec 2015. 8, 9, 10

[36] Aliaksei Kerhet, Cormac Small, Harvey Quon, Terence Riauka, Russell Greiner, Alexander McEwan, and Wilson Roa. *Artificial Intelligence in Medicine: 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings*, chapter Segmentation of Lung Tumours in Positron Emission Tomography Scans: A Machine Learning Approach, pages 146–155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. 8

[37] Faiz M Khan and Bruce John Gerbi. *Treatment planning in radiation oncology*. Wolters Kluwer Health/Lippincott Williams & Wilkins, 2012. 12, 13

[38] PE Kinahan and JW Fletcher. PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy. *Seminars in Ultrasound*, 31(6):496–505, 2010. 23

[39] Ashnil Kumar, Jinman Kim, Lingfeng Wen, Michael Fulham, and Dagan Feng. A graph-based approach for the retrieval of multi-modality medical images. *Medical Image Analysis*, 18(2):330–342, 2014. 24

[40] Jens Langner. PET schema [online], available: http://upload.wikimedia.org/wikipedia/commons/c/c1/pet-schema.png, [accessed: 05-05- 2017]. xii, 3

[41] Jérôme Lapuyade-Lahorgue, Dimitris Visvikis, Olivier Pradier, Catherine Cheze Le Rest, and Mathieu Hatt. Speqtacle: An automated generalized fuzzy c-means algorithm for tumor delineation in PET. *Medical Physics*, 42(10):5720–5734, 2015. 6, 9, 10

[42] Thomas Layer, Matthias Blaickner, Barbara Knäusl, Dietmar Georg, Johannes Neuwirth, Richard P. Baum, Christiane Schuchardt, Stefan Wiessalla, and Gerald Matz. PET image segmentation using a gaussian mixture model and markov random fields. *EJNMMI Physics*, 2(1):1–15, 2015. 6, 9, 10

[43] Benoît Lelandais, Isabelle Gardin, Laurent Mouchard, Pierre Vera, and Su Ruan. Segmentation of biological target volumes on multi-tracer PET images based on information fusion for achieving dose painting in radiotherapy. In *MICCAI*, pages 545–552, 2012. 6, 9, 10

[44] Hua Li, Wade L Thorstad, Kenneth J Biehl, Richard Laforest, Yi Su, Kooresh I Shoghi, Eric D Donnelly, Daniel A Low, and Wei Lu. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Medical Physics*, 35(8):3711–3721, 2008. xii, 5

[45] Andy Liaw and Matthew Wiener. Classification and regression by RandomForest. *R news*, 2(3):18–22, 2002. 27

[46] TB Lynch. *PET/CT in clinical practice*. Springer Science & Business Media, 2007. 2

[47] Jacques-Antoine Maisonobe, Camilo A Garcia, Hatem Necib, Bruno Vanderlinden, Alain Hendlisz, Patrick Flamen, and Irène Buvat. Comparison of PET metabolic indices for the early assessment of tumour response in metastatic colorectal cancer patients treated by polychemotherapy. *European Journal of Nuclear Medicine and Molecular Imaging*, 40(2):166–174, 2013. 23

[48] Roberta Matheoud, Patrizia Della Monica, Chiara Secco, Gianfranco Loi, Marco Krengli, Eugenio Inglese, and Marco Brambilla. 5

[49] Ursula Nestle, Stephanie Kremp, Andrea Schaefer-Schuler, Christiane Sebastian-Welsch, Dirk Hellwig, Christian Rübe, and Carl-Martin Kirsch. Comparison of different methods for delineation of 18F-FDG PET–positive tissue for target volume definition in radiotherapy of patients with non–small cell lung cancer. *Journal of Nuclear Medicine*, 46(8):1342–1348, 2005. 4, 6, 9

[50] Arnold C Paulino. *PET-CT in radiotherapy treatment planning*. Elsevier Health Sciences, 2008. 12, 13

[51] Myria Petrou and Suresh K Mukherji. Extracranial head and neck neoplasms: role of imaging. In *Imaging in Oncology*, pages 93–117. Springer, 2008. 13

[52] Charnchai Pluempitiwiriyawej, José MF Moura, Yi-Jen Lin Wu, and Chien Ho. Stacs: New active contour scheme for cardiac MR image segmentation. *IEEE Transactions on Medical Imaging*, 24(5):593–603, 2005. 17, 19

[53] Jinyi Qi and Richard M Leahy. Iterative reconstruction techniques in emission computed tomography. *Physics in Medicine and Biology*, 51(15):R541, 2006. 36

[54] Beichel R R, Ulrich E J, Bauer C, Wahle A, Brown B, Chang T, Plichta K A, Smith B J, Sunderland J J, Braun T, Fedorov A, Clunie D, Onken M, Riesmeier J, Pieper S, Kikinis R, Graham M M, Casavant T L, Sonka M, and Buatti J M. Data From QIN-HEADNECK. The Cancer Imaging Archive. 2015. 38

[55] Mhd Saeed Sharif, Maysam Abbod, Abbes Amira, and Habib Zaidi. Artificial neural network-based system for PET volume segmentation. *Journal of Biomedical Imaging*, 2010:4:1–4:11, January 2010. 8

[56] Leen-Kiat Soh and Costas Tsatsoulis. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):780–795, 1999. 24

[57] Qi Song, Junjie Bai, Dongfeng Han, S. Bhatia, Wenqing Sun, W. Rockey, J.E. Bayouth, J.M. Buatti, and Xiaodong Wu. Optimal co-segmentation of tumor in PET-CT images with context information. *IEEE Transactions on Medical Imaging*, 32(9):1685–1697, Sept 2013. 8, 9, 10

[58] RL Wahl, H Jacene, Y Kasamon, and MA Lodge. From recist to percist: Evolving considerations for PET response criteria in solid tumors. *Journal of Nuclear Medicine*, 50(Suppl 1):122S–150S, 2009. 7

[59] Marie Wanet, John Aldo Lee, Birgit Weynand, Marc De Bast, Alain Poncelet, Valérie Lacroix, Emmanuel Coche, Vincent Grégoire, and Xavier Geets. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. *Radiotherapy and Oncology*, 98(1):117–125, 2011. 4

[60] Ziyue Xu, Ulas Bagci, Jurgen Seidel, David Thomasson, Jeff Solomon, and Daniel J. Mollura. Segmentation based denoising of PET images: An iterative approach via regional means and affinity propagation. In *MICCAI*, pages 698–705, 2014. 6

[61] Fei Yang and Perry W. Grigsby. Delineation of FDG-PET tumors from heterogeneous background using spectral clustering. *European Journal of Radiology*, 81(11):3535–3541, 2012. 8

[62] Eiji Yoshida, Keishi Kitamura, Yuichi Kimura, Fumihiko Nishikido, Kengo Shibuya, Taiga Yamaya, and Hideo Murayama. Inter-crystal scatter identification for a depth-sensitive detector using support vector machine for small animal positron emission tomography. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 571(1–2):243–246, 2007. Proceedings of the 1st International Conference on Molecular Imaging TechnologyEuroMedIm 2006. 8

[63] Huan Yu. *Automated Segmentation of Head and Neck Cancer Using Texture Analysis with Co-registered PET/CT images*. PhD thesis, University of Toronto, 2010. 12

[64] Huan Yu, Curtis Caldwell, Katherine Mah, and Daniel Mozeg. Coregistered fdg PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Transactions on Medical Imaging*, 28(3):374–383, March 2009. 7

[65] Huan Yu, Curtis Caldwell, Katherine Mah, Ian Poon, Judith Balogh, Robert MacKenzie, Nader Khaouam, and Romeo Tirona. Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. *International Journal of Radiation Oncology Biology Physics*, 75(2):618–625, 2009. 7, 10

[66] Ziming Zeng, Jue Wang, Bernie Tiddeman, and Reyer Zwiggelaar. Unsupervised tumour segmentation in PET using local and global intensity-fitting active surface and alpha matting. *Computers in Biology and Medicine*, 43(10):1530–1544, Oct 2013. 10

[67] Alex P Zijdenbos, Benoit M Dawant, Richard A Margolin, and Andrew C Palmer. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724, 1994. 58