Co-Segmentation Methods for Improving Tumor Target Delineation in PET-CT Images

A Thesis Submitted

to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of Master of Science

in the Department of Electrical and Computer Engineering

University of Saskatchewan

by

Zexi Yu

Saskatoon, Saskatchewan, Canada

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, it is agreed that the Libraries of this University may make it freely available for inspection. Permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised this thesis work or, in their absence, by the Head of the Department of Electrical and Computer Engineering or the Dean of the College of Graduate Studies and Research at the University of Saskatchewan. Any copying, publication, or use of this thesis, or parts thereof, for financial gain without the written permission of the author is strictly prohibited. Proper recognition shall be given to the author and to the University of Saskatchewan in any scholarly use which may be made of any material in this thesis.

Request for permission to copy or to make any other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Electrical and Computer Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
S7N 5A9

Abstract

Positron emission tomography (PET)-Computed tomography (CT) plays an important role in cancer management. As a multi-modal imaging technique it provides both functional and anatomical information of tumor spread. Such information improves cancer treatment in many ways. One important usage of PET-CT in cancer treatment is to facilitate radiotherapy planning, for the information it provides helps radiation oncologists to better target the tumor region. However, currently most tumor delineations in radiotherapy planning are performed by manual segmentation, which consumes a lot of time and work. Most computer-aided algorithms need a knowledgeable user to locate roughly the tumor area as a starting point. This is because, in PET-CT imaging, some tissues like heart and kidney may also exhibit a high level of activity similar to that of a tumor region. In order to address this issue, a novel co-segmentation method is proposed in this work to enhance the accuracy of tumor segmentation using PET-CT, and a localization algorithm is developed to differentiate and segment tumor regions from normal regions. On a combined dataset containing 29 patients with lung tumor, the combined method shows good segmentation results as well as good tumor recognition rate.

Acknowledgments

For this thesis, I want to thank my supervisors, Prof. Francis Bui and Dr. Paul Babyn, for their patience, guidance and support during my M.Sc studies. Their guidance and patience are essential for me to complete this work, and I could not thank them enough for this.

I also want to thank Dr. Niranjan Venugopal for his kindness, and help on clinical ideas and manual segmentation of PET-CT images; Mike Kalyn, Suresh Kalle, Hussein Moosavi and my other colleagues in Information Processing & Optimization (IPOP) group, for their support and encouragement.

Last but not least, I want to thank my family members and my landlords for their understanding and care.

Dedication

To my mother and the memory of my father.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	artificial neural network
AP	Affinity Propagation
BGD	background intensity
Bq	becquerel
CCRG	condensed connected region growing
CT	Computed Tomography
DSC	Dice similarity coefficient
DT	Decision Tree
EM	Expectation Maximization
FAR	false acceptance rate
FCM	Fuzzy C - Means
FDG	Fludeoxyglucose
FLAB	fuzzy locally adaptive Bayesian
FRR	false rejection rate
FWHM	full width half maximum
GMM	Gaussian Mixture Models
HD	Hausdorff distance
HU	Hounsfield scale
k-NN	k-nearest neighbor

LS	Level Set
MRF	Markov random field
MRI	Magnetic resonance imaging
MST	minimum spanning tree
NGTDM	neighborhood gray-tone-difference matrix
PDE	partial differential equation
PET	Positron Emission Tomography
PET-CT	Positron Emission Tomography-Computed Tomography
QDA	Quadratic Discriminant Analysis
ROC	receiver operating characteristic
ROIs	region of interests
RW	Random walk
S/B	source-to-background ratio
SBR	source-to-background ratio
SGLDM	spatial gray-level dependence matrices
SPNs	solitary pulmonary nodules
STAPLE	Simultaneous Truth and Performance Level Estimation
SUV	standard uptake value
SVM	support vector machine
TBR	tumor-to-background ratio
TCGA-LUAD	The Cancer Genome Atlas Lung Adenocarcinoma

TCGA-LUSC	The Cancer Genome Atlas-Lung Squamous Cell Carcinoma
TCIA	The Cancer Imaging Archive
TLE	total lesion evaluation

List of Mathematical Symbols

I, I'	CT image, PET image
u, u'	voxel node in CT and PET
v, v'	voxel node in CT and PET
l_u	segmentation label assigned to node
$E_s(l_u)$	self cost term
$E_n(l_u, l_v)$	neighbor cost term
N	neighbor pair
L_m	label set
$C_{source}, C_{sink}, C_{nodes}$	cost functions
t_u, t_v	normalized PET image value
$G_{uv}, G_{u'v'}$	gradient value on CT image and PET image
$t_{source}, t_{sink}, \lambda_1, \lambda_2, \alpha$	parameters
i	image gray tone value
N_i	set of all voxels with the gray tone i
R	region of interest
$\bar{A_u}$	average gray-tone of the neighbor area of a voxel u in the set N_i
M_u	intersection set of the neighbor set and the ROI R in the boundary area
w	number of total voxels in the set M_u

$I(v)$	gray tone of voxel v in the neighbor set M_u of u
G_R	highest gray tone level presented in the ROI R
n	number of voxels in R
N_g	number of gray levels with at least one voxels in R
$P_i = N_i/n$	probability distribution of the gray level i.

1. Introduction

Nowadays, cancer has become one of the major threats to human health. As of 2014, cancer is responsible for around 30% of all deaths, and there are 11 million cases emerging every year worldwide [2]. Radiation therapy, one of the dominant form of therapies, plays an important role in countering cancer [3]. Radiation is effective in terms of killing abnormal cells. However, it can also kill normal cells and damage adjacent body tissues in the process. In order to help clinicians to better define tumor in radiation therapy process, a multi-modal imaging technique known as Positron Emission Tomography-Computed Tomography (PET-CT) has been introduced [4].

As a structural imaging technique, Computed Tomography (CT) is widely used in clinical practice. It helps clinicians to better examine structural abnormalities caused by disease. However, in cases where recognition of abnormal cellular activity is more significant than structural abnormalities, structural imaging becomes less effective [4]. The need to characterize functional information has led to the development of imaging techniques such as Positron Emission Tomography (PET), which provides molecular biological information of many diseases. Although PET has good sensitivity in terms of radiotracer uptake, and by specific design such uptake can reflect the biochemical activity of tumors, structural imaging is still needed to differentiate tumor regions from normal physiological high uptake regions such as brain, liver, heart, kidneys. etc [4].

With the help of PET-CT, clinicians can localize and contour tumor regions for radiation therapy. However, most tumor segmentation is still done manually on a slice by slice basis requiring well-trained expertise; and many computer-aid algorithms also need human interaction [4]. In this work, a two-stage algorithm is proposed to localize and segment lung tumor, freeing clinicians from tedious slice by slice manual segmentation process.

In this chapter, a brief introduction to PET-CT imaging and segmentation is given. Due to the

dominating role of PET imaging in PET-CT, the introduction will be focused on PET imaging. Then the challenges in PET-CT imaging segmentation are addressed and analyzed. Given these analyses, a brief overview of our proposed method is provided, and the contents of subsequent chapters are described.

Background on PET-CT imaging and Segmentation

PET imaging plays an important role in cancer treatment. By revealing chemical activity of specially designed radiotracers, PET imaging is capable of detecting a wide range of disease activity in human body, including tumor activity. It has been widely applied to cancer clinics as a tool for tumor detection and diagnosis. However, PET imaging has low spatial resolution in its nature, making it very difficult for precise localization of tumor. Precise localization is important for radiation therapy, a mainstream tumor treatment method, for treatment efficiency and avoiding side-effect. Therefore, in early clinical application, PET was often combined with CT, an imaging modality with limited functional information but higher anatomic spatial resolution. Shortly after, specially designed scanner combining PET and CT was developed, and PET-CT has become a mainstream imaging modality ever since [3]. However, many challenges remain to be solved in precise tumor segmentation and localization using PET-CT imaging.

In this section, the underlying mechanisms of PET imaging are introduced at first as the background of this work. After that, challenges in segmentation using PET or PET-CT to be addressed by this work are described.

Radiotracers

PET imaging is designed to reveal the bio-distribution of different types of positron emitted radio-pharmaceutical compound with a biologically active ligand. Such compounds are also called radiotracers. Radiotracers must be injected into a subject's body for imaging. As PET scanners use positron emitting radioisotopes, the radiation activity of those radiotracers can be picked up through imaging in vivo [5, 6]. Among the radiotracers that have been developed, the most commonly used is ^{18}FDG [7], which consists of ^{18}F combined with deoxyglucose. Metabolically active tumors often utilize higher than normal glucose or FDG levels, so revealing the distribution

of ^{18}FDG reveals tumor distribution as well. Although other radiotracers have been developed as well, ^{18}FDG remains the most commonly used [8].

Quantitative evaluation of radiotracer uptake in PET

A qualitative assessment of the PET image done by experts is often sufficent for tumor assessment and detection. However, accurate tumor diagnosis and assessment require quantitative evaluation of the PET scanner records, since such records vary with time and dose of radiotracer [9, 10]. Popular semi-quantitative and quantitative methods are standard uptake value (SUV), tumor-to-background ratio (TBR), nonlinear regression techniques, total lesion evaluation (TLE), and the Patlak-derived methods [11]. Among them, SUV is the most popular technique, for it relates to physiological activity of cellular metabolism [6, 9].

SUV represents the tissue concentration of the radiotracer at a given time divided by a few normalization factors. These factors include injected dose, patient size in terms of weight, and a decay factor related to the radiotracer. It is defined as:

$$SUV = \frac{C_{img}(t)}{ID/BW}, \tag{1.1}$$

where $C_{img}(t)$ stands for total radiotracer concentration in time t, ID represents injected dose (Bq) and BW is body weight (in g or kg). SUV is used in localization for normalization purposes. For details of the available different quantitative evaluation methods, the reader can refer to the review work [12].

Challenges in segmentation of PET images

Segmentation of tumor with PET can be subdivided into two parts: initial tumor recognition and its sequential delineation [13].

Recognition determines the tumor location from other similar regions in the image. Traditionally, experienced clinicians identify the high uptake regions in PET images from other normal tissues with high uptake. In many modern algorithms, a rough area of tumor region defined by clinicians is still needed. Subsequently, delineation focuses on drawing a spatial extent of tumor

in the area defined by recognition, in order to achieve a precise separation between tumor and background areas [14].

Several difficulties can be encountered throughout the segmentation process. In recognition, whether a high-uptake area represents a tumor or normal tissue is also determined by a clinician's subjective judgment, i.e., a decision that is dependent on both tumor appearance and clinician experience.

On the other hand, in delineation, several factors affect segmentation. First, PET images typically suffer from low resolution and high smoothing factor. What is more, PET image may be further filtered in order to address factors like motion artifacts. Second, tumors have large variations in their shape and texture, and this makes generalizing PET segmentation rules more difficult. Last, the noise in PET is high because the nature of PET itself. As indicated by work [15], noise is considered as the most significant factor affecting segmentation performance.

Given the challenges above, several algorithms have emerged to help clinicians in the process. A literature review details them in the following chapter.

Research Overview

Problem Overview

Using PET/CT, localizing tumor and drawing its boundary becomes more effective. In this process, over-segmenting puts normal tissues within tumor region, leading radiation to damage those tissues, and further hardening side effects. On the other hand, under-segmenting results in missing abnormal cells, and affects treatment efficiency. Therefore, accurate tumor contour and localization is vitally important [4].

For now, the standard process of tumor segmentation still involves heavy manual work of experienced experts on a slice by slice basis. Even with existing computer-aid algorithms, such a process still takes a lot of time and effort. On the other hand, reports also indicate disadvantages of manual segmentation. It is pointed out that there are large inter and intra variations among manual segmentations made by different clinicians. Naturally, many computer algorithms have been

developed to improve the efficiency of the segmentation process. However, most of them need a pre-defined tumor area in order to work correctly, since those methods rely solely on information extracted from PET intensity information, and therefore wouldn't be able to tell the differences between tumor region and other normal high uptake regions. Another disadvantage of many methods relying solely on PET intensity is that they are not able to use the high spatial resolution of CT to help define the tumor boundary [4].

Given the above considerations, an algorithm is developed in this work to address the issue of tumor localization while enhancing tumor contouring. A brief summary of the method is provided in the section below, and the reader can refer to chapter 3 and chapter 4 for further details.

Method Overview

Basically the algorithm in this work contains two stages. The first stage is segmentation or delineation, meaning that in this stage, the algorithm will segment all the regions with a high uptake value, no matter if it is a tumor region or a normal region. In this stage, an algorithm based on graph-cuts is used, but the cost function of graph-cuts is defined in a way so that the boundary information from CT is contained along with intensity and boundary information from PET into the cost function, resulting in a more detailed tumor boundary consistent with CT boundary.

After the segmentation stage, a machine learning based localization stage, i.e., tumor recognition, is employed on both PET and CT data, differentiating tumor regions from normal regions. The features used in this stage include first-order features (such as mean and median), second-order features (such as variance) and higher-order features (such as neighborhood gray-tone-difference matrix (NGTDM) features). For evaluation, the dataset is partitioned based on leave-one-out technique, and subjected to three different classifiers for result evaluation.

In terms of results, the segmentation stage gives a Dice index close to 92%, and all tumor regions are successfully separated from non-tumor regions in localization stage, more details can be found in chapter 3 and chapter 4.

Thesis Organization

The remainder of the thesis is organized as follows: chapter 2 provides a literature review, chapter 3 and chapter 4 are manuscript-based chapters covering the details of the methodology in two stages. Lastly, chapter 5 serves as a conclusion.

2. Literature Survey

In chapter 1 we subdivided segmentation into two parts: tumor recognition and tumor delineation. In manual segmentation tumor recognition process typically precedes delineation. However, using an automatic system, which first segments high uptake regions and then recognizes tumor location, saves computational time and resource overall. Accordingly, the proposed algorithms put delineation stage before localization stage. In the following sections, previous works related to these two stages will be reviewed.

Delineation Methods

In terms of delineation process, the goal is to divide the image into two regions of interest (ROIs): foreground and background. Foreground contains high uptake regions, including tumors, while background contains tissues with low uptake. In delineation process, algorithms can be roughly classified into four categories: manual segmentation, thresholding methods, stochastic and learning-based methods and region based methods.

Manual Segmentation

Before going to various segmentation methods, it is necessary to know what is the standard way of evaluating and comparing the accuracy of these methods. In order to do such evaluation, the true boundary of the tumor must be achieved. When using a phantom, such boundary can always be found [16]. However, human cases are much more complex than typical phantoms, making evaluation results on phantom less convincing. Unfortunately in clinical cases, such true boundary cannot be established if there are no histopathologic samples. A common way in such evaluation is to use manually segmented results and compare them with results given by algorithm [17].

Ground Truth Construction and Segmentation Evaluation

It is not uncommon to see variations between segmentations provided by different clinicians. Therefore, in order for results evaluation it is important to compare the result from algorithm with as many manual segmentations as possible. However, as such comparisons increase the computational complexity, it is common to combine all the manual segmentations into one single ground truth by weighting each segmentation using statistical methods. One method called Simultaneous Truth and Performance Level Estimation (STAPLE) has been used for such combination. It incorporates a prior spatial distribution model and adjusts weighting by using a performance estimation level [18].

After ground truth is established, the algorithm result is compared to the ground truth in terms of similarity and difference. Naturally better results are closer to the ground truth. Many metrics are utilized for such comparisons. Most of them falls into three categories: Volumetric Difference Quantification, Shape based Similarity Index and Regression Based Statistical Methods.

Volumetric Difference Quantification

The core idea of volumetric difference quantification is computing the absolute total volume difference percentage between two segmentations. It is commonly used as it is both simple and intuitive. However, using this kind of metric alone may not give enough information in terms of similarity. For example, a segmentation method may produce same volume as the one of ground truth, while producing unsatisfactory results such as having some amount of volume into non-object area. Several metrics are derived from the concept of confusion matrix, which is commonly used in machine learning area for error evaluation. The Dice similarity coefficient (DSC) is one of the most widely used metrics among this kind [19, 20]. By measuring spatial overlap between segmented area and ground truth in percentage, a higher DSC value represents better segmentation performance. Given X as our method result and Y as ground truth, the Dice index can be computed using the following equation:

$$d_I(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{2.1}$$

with $|.|$ defined as the area of the graph. A higher DSC represents more similarity, and thus better result.

Shape based Similarity Index

For segmentation evaluation based on shape similarity, the idea is to quantify the shape dissimilarity based on boundary-based measures [21, 22]. Geometric metrics measure how far two boundaries are from each other. Among them, the Hausdorff distance (HD) is most commonly used [23]. Given x as voxel in the image X and y as voxel in the image Y, the Hausdorff distance is defined as:

$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\}. \tag{2.2}$$

A lower HD represents smaller boundary difference distance, thus better result.

In conclusion, a good segmentation should have a high DSC (high overlap) and a low HD (low boundary distance). These two metrics are commonly used together in research for a thorough evaluation.

Regression Based Statistical Methods

In clinical literatures, using regression based indices such as Spearman and/or Pearson correlation coefficients and simple mean volume difference or relative volume ratio for segmentation evaluation is more common than DSC and HD based evaluation [24, 25]. However, the disadvantage of using regression based indices is that comparing statistics on the absolute volume difference does not provide complete information on segmentation accuracy. Therefore, even though regression based indices are more popular in clinical medical imaging literatures, in this work we use the DSC and HD pair for segmentation evaluation.

Problem with Manual segmentation

As previously mentioned, manually drawing the boundary for a tumor is perhaps the easiest and most direct way of obtaining region of interests (ROIs), and is commonly used for ground

truth establishment. However, there have been ongoing debates around manual segmentation due to its disadvantages: It is time consuming, labor intensive, and operator-dependent [4]. What is more, a recent study shows high intra- and inter-operator variability of segmentation results, After investigating cases involving 18 physicians from 4 different departments, researchers found that the volume overlap of 70% was found only in 21.8% of radiation oncologists and 30.4% of haematologic oncologists. Such high intra-observer variability may come from small size tumors (that is, smaller than $4cm^3$) [26]. More studies on various cases show intra and inter variations could be as high as 78% [27–31]. However, even with those problems manual segmentation remains the only choice under many circumstances.

Thresholding Methods

Thresholding technique is very popular in image segmentation area. The main idea is to convert a gray-level image into a bi-level image by defining all the voxels higher than a threshold value into one class and the rest into other class [32].

Among all the computer-aid segmentation methods, thresholding methods are the first ones being developed. This method is suitable for PET segmentation, as PET images have low resolution with high contrast, and tumors usually have much higher intensities than their surrounding areas. However, some uncertainty arise when using these methods. Because of the difficulties we mentioned in the last chapter, that PET image has large variability in terms of pathologies, low resolution, high noise from imaging nature and from imaging precess, there are no general rules in selecting an optimal threshold level (especially in automatic cases). As a result, the thresholding methods are still under development towards optimal thresholding selection. Thresholding based methods are commonly seen in clinical usage for they are easy to implement. The state-of-art thresholding methods can be classified into the following groups: Fixed Thresholding, Adaptive Thresholding, and Iterative Thresholding Methods.

Fixed thresholding

As the name implies, fixed thresholding sets all voxels with intensities above a fixed value to foreground and everything else into background. The thresholding level can be set by a clini-

cian, derived from a given phantom, or learned from a training set of similar cases. The resulting boundary may contain certain amount of uncertainty because of the noise of PET imaging.

In many studies, a value of 2.5 SUV is set as a threshold. However, this value is quite arbitrary and does not allow one to achieve superior performance in most cases. In order to incorporate variance between patients, in other cases, using a specific percentage of SUV_{max} (which is the local maximum value of SUV) is preferred. Depending on various factors such as scanner type, image properties and noise, the percentage may vary significantly. Regardless of these factors, in many studies a number between 40-60% SUV_{max} is often preferred. In other works, this percentage can vary from 15% to 78% [33–35]. Beside the uncertainty of boundary, fixed thresholding methods tend to give overestimated boundaries, especially for small tumors. Because of that, adaptive thresholding has been developed. The following table 2.1 shows a collection of these fixed thresholding methods and their accuracies, where BGD (background intensity) $=$ $70\%\text{SUV}_{max}+$ mean background intensity, $\text{BGD}_\alpha = \text{BGD} + \alpha * (\text{SUV}_{max} - \text{BGD})$. Note that for all the tables in this chapter, for the cases where multiple results occur under the Evaluation Performance column, these results correspond to the respective methods described in the Method column.

Table 2.1: Fixed thresholding methods and their reported accuracies

Reference	Method	Sample descripition	Evaluation Performance
[36]	42% SUV_{max}	Static phantom, 3 spheres	Mean volume deviation (%): 8.4%
[37]	34% SUV_{max}	Moving phantom, 3 spheres 3 motions	Difference from ideal ranged from 3 to 94 cm^3 for motion volumes of 1.2 to 243 cm^3
[38]	34% SUV_{max}	Moving lollipop phantom, 1 sphere, 3 longitudinal movements	Volume deviation from ground truth (%): 1.4 ± 8.1%
[39]	50% SUV_{max}	Intact squamous cell carcinoma, 40 patients	Volume deviation from CT (%): 54.5%

[40]	50% SUV_{max}	NSCLC, 101 Patients	Volume deviation from CT (%): 27 ± 3%
[41]	42%, 24%, 15% SUV_{max}	NSCLC, 20 Patien	Determined threshold values such that the volumes were exactly the ground truth
[42]	40% SUV_{max}, 2.5 SUV	NSCLC, 19 Patients	Median volume deviation from CT (%): −140%, −20%
[33]	Manual, 40% SUV_{max}, 50% SUV_{max}, , SBR SUV_{max}	Oral cavity, oropharynx, hypopharynx, larynx, 78 cases	Mean overlap fraction (CT): 0.61, 0.55, 0.39, 0.43
[35]	Manual, 2.5 SUV, 40% SUV_{max}, , SBR SUV_{max}	NSCLC, 25 cases	Mean GTV (cm3): 157.7, 164.6, 53.6, 94.7 Mean Radius (cm): 3.03, 3.05, 2.18, 2.52
[43]	43% SUV_{max}, 2.5 SUV	Rectal and anal Cancer, 18 patients	55.4 ± 18.3, 36.7 ± 38.4 Volume difference compared to manual delineation
[44]	SUV_{max}, BGD, Iterative Theshold	Nonspherical simulated tumors inserted into real patient PET scans, 41 cases	Mean error volume (%): −50% ± 10%, −40% ± 40%, 4% ± 10%, 24% ± 20%
[45]	SBR SUV_{max}	NSCLC, 23 Tumors	Compared to Histopathology: Sensitivity: 66.7% and Specificity: 95.0% Compared to Manual Segmentation: Sensitivity: 55.6% Specificity: 88.3%

[29]	42%, 50%,SUV_{max}, FCM	Sphere phantoms with diameters 13 - 37 (mm), 6 Spheres from 4 scanners	Classification error (%): 42.6 ± 51.6, 20.3 ± 18.5, 27.8 ± 25.6
[34]	Manual, SBR SUV_{max}, 40% SUV_{max}, 50% SUV_{max}, 2.5 SUV	High-grade gliomas, 18 patients	Mean Overlap Fraction (PET): 0.61, 0.62, 0.57, 0.67, 0.67 Mean Overlap Fraction (MRI): 0.45, 0.44, 0.54, 0.36, 0.14
[46]	Manual, $BGD_{40\%}$, $BGD_{20\%}$, 2.5 SUV, 40% SUV_{max}	Esophageal Carcinoma, 96 cases	Mean length of tumors (cm): 6.30 ± 2.69, 5.55 ± 2.48, 6.80 ± 2.92, 6.65 ± 2.66, 4.88 ± 1.99, 5.90 ± 2.38
[47]	Gradient Based, 40% SUV_{max}	Stage I-II NSCLC, 10 patients	DSC: 66%, 65%

Adaptive thresholding

As already mentioned, the arbitrary value in fixed thresholding varies from case to case and therefore such thresholding comes from either the interpreting clinician, derived from a given phantom, or learned from a training set of similar cases. Adaptive thresholding techniques arise where thresholds are adapted based on various parameters: image quality metrics such as source-to-background ratio (SBR or S/B) and full width half maximum (FWHM), motion artifacts measured using oscillating spheres instead of static sphere.

However, these parameters come from a thorough prior examination, and depend on various factors including the specific scanner, the image reconstruction method, and even the patient's size. It means that the segmentation results are not reproducible on similar cases with different scanners, patients or reconstruction algorithms [48]. Also, since introducing adaptive parameters introduce the construction and calibration of analytic expressions, uncertainties are also introduced in finding

optimal thresholding level, making this kind of methods unacceptable for clinical applications [49].

Some typical methods and results of the adaptive thresholding literature are listed below. One can note that the analysis parameters get more and more complex in recent studies. Volume difference here refers to relative volume difference between ground truth and algorithm reults.

Table 2.2: Adaptive thresholding methods and their accuracies

Reference	Method (T is the threshold value)	Parameter	Evaluation Performance (Volume Difference)
[36]	$T = Ae^{-CV_{ml}}$	Parameters A,C are coefficients computed for each SBR. First work to use the SBR to estimate the thresholding level.	8.4%, NA
[50]	$T = 0.307\text{SUV}_{mean} + 0.588$	Iteratively selected the mean target SUV	21%, 67%
[35]	$T = 0.15\text{SUV}_{mean} + BGD$	$\text{SUV}_{mean}$is approximated in advance	NA, 60.1%

[41]	$T = 59.1 - 18.5\log(\text{SUV}_{mean})$	Fitted a logarithmic regression curve to threshold values in PET that resulted in the same volume from CT	NA
[51]	$T = SBR * (\text{SUV}_{max} - BGD) + BGD$	Considered influence of difference between target and background intensities	47.3%, NA
[52]	$T = 0.078V + 0.617SBR + 0.316\text{SUV}_{mean}$	Initialization for an iterative thresholding method, V stands for volume of object	10%, 16.6%
[53]	$T(\%) = 90.787e^{-0.0025V}$ $T(\%) = 0.00154V + 28.77$	These two functions corresponds to area above and under $448mm^2$cases	4.7%, 7.5%

[54]	$T(\%) = \frac{a*\text{SUV}_{70\%}+b*\text{SUV}_{background}}{\text{SUV}_{max}}$	For diameter ¿ 30 mm, $a = 0.5, b = 0.5$, for diameter ≤ 30 mm, $a = 0.67, b = 0.6$	4.7%, 7.5%
[55]	$T = 309.2 - 171.9 * (1 - \frac{1}{SBR}) - 9.643 * D$, $T = 151.1 - 101.8 * (1 - \frac{1}{SBR}) - 0.976 * D$	D stands for diameter of object, two functions for cases with D under and above 10mm, considered scan duration and initial injected activity in model.	NA
[56]	$T = 5 + e^{3.568+0.197/V-0.1069*\log V}$	An iterative technique based on Monte Carlo simulation studies	5.1%, 11.1%
[57]	$\ln T(\%) = 0.0634x^{1/3} + 0.1202y^{1/3} + 0.7327\ln z + 0.0597x^{1/3}\ln z - 0.1221y^{1/3}\ln z - 0.0248(xy)^{1/3} - 0.9504$	T is normalized to background. x is Volume in cm3, y is the motion in mm, z is the SBR, considered how motion in moving lung tumors affects the thresholding	NA

Iterative thresholding method (ITM)

One problem for adaptive methods is that such methods typically require precise prior analytic assessment extracted from a phantom or CT [4]. However, the iterative thresholding method proposed by Jentzen does not need such complex and precise knowledge [52]. The core idea behind these methods is to determine the optimal threshold for the specific image. Before applying the method, the calibrated threshold-volume curve at varying S/B ratio is acquired by measuring phantom under different conditions. Then the image S/B ratio is estimated, the corresponding curve is then used for iterative coverage: first a fixed threshold T_1 is selected, then the volume of object is then estimated in image. Using this volume, the curve gives another threshold T_2. If $T_1$and T_2 are close to each other, the algorithm stops. Otherwise it continues such process.

Using iterative thresholding methods, one only needs to measure the S/B-threshold-volume curve and S/B ratio of the image, instead of all the complex parameters in adaptive thresholding methods. However, the S/B curve still heavily depends on the spatial resolution of the imaging device and reconstruction method used. Moreover, the way of applying one S/B curve indicates the assumption that the imaged activity distribution is homogeneous: asymmetric activity leads to underestimated volume. Another assumption is the spherical tumor shape, which may lead to uncertain results if the shape of tumor is irregular.

Stochastic and learning-based methods

Using pattern recognition, machine learning and stochastic technique in the segmentation problem leads to another kind of segmentation methods. Stochastic and learning-based methods emphasize on using the differences between uptake regions and surrounding tissues, therefore employing more information than thresholding methods into segmentation.

Mixture models

Using Gaussian Mixture Models (GMM), it is assumed that any distribution of intensities within a PET image can be roughly modeled by a summation of Gaussian densities. The goal of segmentation then becomes trying to identify and separate these densities. The problem can

thus be solved by the Expectation Maximization (EM) algorithm. Hatt et al. [58] and [59] employ this kind of analysis, describing a GMM based algorithm in [59]. The authors consider three classes: the background, the tumor and the uncertain regions. It needs a user-defined ROI for initialization, and the underlying GMM is analyzed. The performance details of this kind of methods can be found in the table 2.3 below.

Fuzzy locally adaptive Bayesian (FLAB) method

The work [58] also provides another way of using GMM known as fuzzy locally adaptive Bayesian (FLAB) segmentation. In this method, there are two classes as "background" and "foreground", and a finite level of "fuzzy levels" in between. The fuzzy level means a voxel in this level has a mixture existence of two classes, and its final class depends on the portion of this mixture. As in the previous methods, a pre-defined ROI is needed for method initialization. FLAB method has been shown to be quite robust and reproducible. However, it needs a significant number of background voxels for accurate modeling. Also the model is limited to two classes.

Recently an improved version of FLAB has been developed. It allows up to three classes with fuzzy statuses, and has higher accuracy and robustness. Its performance has been demonstrated in several papers [29, 60].

Clustering/Classification of PET image intensities

Classifier-based methods seek to find a set of features derived from images, and use them to partition images into classes. Classifier based methods are known as supervised methods, for the set of feature comes from images with known labels, i.e., these images are training data. Due to limited computational capability, the most common feature is the image intensity itself. As long as the feature set remains sufficient to distinguish background from foreground, the method can continue its work on new data. The disadvantage of supervised methods is that very little spatial information is used, and manual labeling for training data is also time consuming.

Clustering methods do not use training data. Therefore they are called unsupervised methods. Because of that, no time consuming manual labeling is needed in this kind of method. These methods emphasize more on spatial information. However, the nature of unsupervised process

makes them very sensitive to noise and outliers in the image.

Examples of supervised and unsupervised methods include k-nearest neighbor (k-NN) [61–63], support vector machine (SVM) [64,65], Fuzzy C-Means (FCM) [66], artificial neural network (ANN) [61], and more recently Affinity Propagation (AP) [67] and spectral clustering [68]. Using classifiers or clustering technique, voxels with similar features (intensity value, spatial location) are gathered into same group, and label them with background, foreground or more classes. They are especially helpful when non-convex boundaries [67] appear in the problem, which is a common circumstance in many cases. Therefore, many algorithms have been developed using this technique.

Among these methods, spectral clustering method shows good performance in cases containing inhomogeneous activities in the presence of a heterogeneous background in [68]. However, the similarity choice in this method is intensity value, which may not be optimal. In another popular clustering method known as FCM [66], a voxel is given a mixture of two classes. The algorithm assigns possibilities that this voxel belongs to each classes. and the final decision is made based on these possibilities.

The problem with many methods belong to this category is that they still need ROI definition, and it is often hard for them to deal with multi-focal uptake patterns. A recent study seeks to address this issue by using a novel affinity estimation function with AP platform. It then uses these functions to draw several thresholds to divide an image into different regions. Table 2.3 lists some common stochastic and learning-based methods for PET segmentation.

Table 2.3: Clustering and classifier methods and their accuracies

Reference	Method	Parameter	Evaluation Performance (Volume Difference)
[69]	k-Means, MRF, Multiscale MRF	4 Lesions	Relative Volume Difference: 9.09%, 6.97%, 5.09%

[62]	k-NN	Monte Carlo simulation using Zubal whole-body phantom as prior	DSC: 0.8-0.85
[70]	k-Means, MRF, Multiscale MRF	6 Phantom spheres	Relative Volume Difference: 42.86%, 32.59%, 15.36%
[63]	k-NN, SUV 2.5, 50% SUV_{max}, SBR	10 Head and neck cancer patients	Sensitivity-Specificity: 0.90-0.90, 0.95-0.93, 0.84-0.48, 0.98-0.68, 0.96-0.96
[58]	FLAB, FHMC, FCM, 42% SUV_{max}	10 Spherical Phantoms	Classification errors (%): 25.2, 31.2, 51.6, 55.8
[66]	FCM	Simulated lesions from the NCAT Phantom 21 NSCLC and 7 LSCC patients	Classification Error (%): 0.9 ± 14.4
[61]	k-NN, ANN, Adaptive Thresholding	6 Phantom spheres	Error Rate: 6.83, 0.28, 7.61
[71]	Standard GMM, SVFMM, CASVFMM, ICM, MEANF, Dirichlet Gaussian mixture model	PET Image of dog lung and spherical phantoms	Misclassification ratio(%): 32.15, 12.43, 11.85, 3.52, 1.19, 0.67
[68]	Spectral Clustering, Adaptive Thresholding	30 Simulated lesions	DSC: 95%, 92%

Geometry Methods: Region and Boundary based methods

Another group of researchers seek to incorporate even more spatial information from PET/CT image into segmentation process. This idea forms the origin of region based methods and boundary based methods. Region based methods emphasize the homogeneity of the image, while boundary based methods focus on identifying and locating boundaries in the image.

Region based Methods

Besides using pixel or voxel intensities of the image, region based methods focus much more on homogeneity of the intensities on the image. Mainstream region based methods can be divided into two groups: region growing methods and graph based methods.

Region Growing Methods

As we mentioned previously, in thresholding and classifier methods, spatial information is barely used while intensity value is the only feature used for segmentation. In clustering methods, spatial information is used but making the methods sensitive to noise and outliers. Region growing seeks to use the spatial information along with intensity value in a way different than clustering methods. A user-defined seed is placed on the ROI to initiate the algorithm. Based on mean and standard deviation, the connected voxels (also known as neighbor voxels) of the seed is either selected as foreground (with seed) or background. Then a homogeneity metric is used to examine if the new voxels are statistically close enough to be included with seed region. As the seed region grows larger, new neighbor voxels are selected and examined in the same way, until all the voxels in the ROI are examined. Upon development, region growing has been used on various occasions. One assumption for this method is that the segmented target has constant or slowly varying intensity values therefore homogeneity metrics work better. If this assumption is satisfied, region growing gives superior outcome with small variance when compared to fixed thresholding methods such as $SUV_{2.5}$ and $SUV_{43\%}$ [43]. Also the results from region growth are highly reproducible. Same as other region based method, region growth has high false positive rates, that means it cannot differentiate normal high uptake region with tumor region especially when they are close. While this problem remains a challenge, region growth also has several dis-

advantages. It does not work well on heterogeneous structures. Initial seeds and parameter settings have significant impact on the outcome, and the algorithm will fail if the parameters are not properly set. Sometimes leakage occurs in the image. These difficulties are either addressed by adding another refined active contour based method after region growing is performed [72], or integrating a threshold method with region growing after background voxels are removed. Such process may introduce some non-target structure into segmentation, as [73] suggests. Recently a method called condensed connected region growing (CCRG) has been developed. Using a novel metric including information for mean and standard deviation of selected region, it only requires ROI selection, and gives good results when compared to thresholding based methods [43].

Graph-based methods

By using background and foreground seeds specified by the user, or automatically selected by an algorithm based on intensity values, graph-based methods can use more information for recognition than other segmentation methods. These seeds are capable of combining global information with local pairwise voxel similarities. Two types of graph-based methods are often used: graph-cuts and random walk

Graph-cuts

By creating cost functions related to voxel intensity and neighbor similarities, graph-cuts model the segmentation problem into a cost minimization problem [74]. By constructing a graph with each node representing a voxel in the image and the connection capacity between nodes representing strength of pairwise pixel similarities, the cost minimization problem can be solved by maximum flow algorithm. It has been shown to achieve optimal segmentation with local pairwise pixel similarities. However the outcome also largely depends on seed placement. Recently a new set of segmentation appears in [75], using graph-cuts to incorporate information from PET and CT to achieve better performance, which is the basis for the delineation method in this work.

Random walk

Random walk (RW) is another region based method used in segmentation. It is robust against noise and weak boundaries. When compared with two threshold based segmentation methods (de-

scribed previously in this review), FLAB and FCM, RW outperforms them [74]. Its disadvantage is that it may not be able to handle multi-focal tumor uptakes and human involvement may still be necessary under certain cases [23].

Boundary-based methods

Different from using homogeneity of the image, boundary based methods focuses on analysis of the boundaries of objects in PET images. Two kinds of methods are used mainly: Gradient Based methods and Level Set/Active Contours methods.

Gradient-based methods

Gradient is a suitable tool to determine how and where the image intensity changes. Since the image intensity usually changes significantly at boundaries, gradient is a good tool for detecting them. In PET image case, because of its low resolution and high noise, boundaries are smoothed and contain significant noise. So simply applying gradient may result in less optimal results. Despite that, some works seek to compensate for these factors. In [76] an edge preserving filter is applied before gradient analysis. Watershed transform is also used for segmentation [76], and shows superior performance when compared with threshold based techniques [47]. Another method, named GRADIENT, requires a user defined initial starting point and an user defined ellipse which is then used for the initial bounding region for gradient detection [16] and was validated in several studies demonstrating higher performance than manual and constant threshold methods in a phantom study. The problem for this kind of method is the pre- and post- processing may cause over-segmented regions, and further fuse those regions together. Besides, noise remains to be a problem.

Active Contours and Level Set (LS)

Another more advanced boundary based method is based on active contours. The core idea of this kind of method is deforming and moving edges (contours). The deformation of the contour is handled by what is termed as the energy function. The energy function consists of two set of terms: internal and external energies. The internal energy guarantees the smoothness of the contour, whereas the external term forces the contour to move to the desired features (gradient,

texture, edge information, etc.) of the image. Classical active contour methods rely on the gradient information and their performance is highly dependent on the location of the initial contour, i.e., the initial contour must be as close to the object of interest as possible so that the external energy is strong enough to move contour towards the target object boundaries. Moreover, the classical model cannot handle the topological changes of the curve. Geometric active contours capable of handling these curves were later introduced by [77]. Their model utilizes the gradient information to define an edge; whereas the energy functional minimization procedure is carried out using the level set formulation. As its PET applications, [78] applied the classic active contour model to liver PET images. The active contour method is first proposed to deform an initial contour around the object, moving towards the desired object's edges. In this approach, the authors estimated the external energy by solving a Poisson partial differential equation (PDE) and the algorithm was initialized by Canny edge detection.

First proposed in [79], the Level Set (LS) method traces different phase interfaces of fluid flows over time and intensity, therefore provides an active contours model. The method employs the intensity gradient information using evolving level sets given by iteratively solving the Euler-Lagrange partial differential equation (2.3):

$$\frac{\partial}{\partial t}\varphi = V(k)|\nabla_{\varphi}|, \tag{2.3}$$

where φ stands for an function that monitors the change of level set such as distance function, V stands for the velocity function that controls the expansion or shrinkage of level set. The method has been further used in many applications, and is used for PET imaging segmentation and registration problem in [80, 81]. In imaging application, the velocity function is a function of curvature and inversely proportional to intensity gradient. Specifically, in [82] a multi-phase level set method is developed to utilize both spatial and temporal information in dynamic PET data. The method weighs data between frames to counter the changes in tracer distribution over time and noise level. When compared to k-means algorithm, the LS method shows higher accuracy. Another method using LS to refine the result from a region growing method in [72] outperforms iterative threshold methods on phantom and real images. The LS method also shows high reproducibility in these cases. The method only suffers from computational complexity from certain function choices and

initial condition set up.

Joint Segmentation Methods

As previously mentioned, combining PET and CT can potentially give advantages of both modalities: high spatial resolution as well as functional information. There are more and more studies considering this idea lately. These methods are known as co-segmentation methods. As a matter of fact, there are multiple benefits of co-segmentation. To start, co-segmentation methods tend to be more robust than single modality methods because of unified information. Also by combing two modalities, they provide more spatial and temporal information options for segmentation process. Last but not least, the uncertainty for the segmentation result is less [4]. In order to incorporate CT information into PET segmentation, researchers attempt to mimic human visual recognition of images, by taking features such as corners, texture, edges, and lines available in the images.

Mainstream methods that employ CT information into PET recognition can be classified into four classes:

Firstly, in work [83], the proposed method combines the result from PET and CT segmentation into one final result using a multi-value LS method.

Secondly, in works [84, 85], textural features from CT are used to delineate cancer tissue from normal in a PET defined interested region. These efforts are questionable, for there is no standard for combining feature sets from different imaging modalities. Also the algorithms give sub-optimal solutions in certain cases, and it takes relatively longer to process.

Thirdly, an semi-automated co-segmentation method is proposed in [21], where high uptake region is given by PET and boundaries are refined by CT. The method is based on random walk and has been applied to PET, PET/CT, PET/MRI(Magnetic resonance imaging) and MRI/PET/CT to examine the performance. The co-segmentation method achieves $91.44 \pm 1.71\%$ DSC on PET/CT using data from patients, and outperforms the state-of- the-art co-segmentation method in [86]. However it should be pointed out that user-defined regions of interest are still needed in order for the method to perform.

Reference	Method	Database Description	Evaluation Performance
[83]	active contours on on CT, MRI, PET, and PET/MRI/CT	static phantom, 3 spheres	DSC 0.74± 0.046, 0.89± 0.018, 0.85± 0.053, 0.90± 0.017
[84]	textural features classification	head and neck cancer, 18 patients	NA
[43]	confidence connected region growing method	Rectal and anal Cancer, 18 patients	absolute percentage mean difference 9.0% ± 70.%
[86]	graph-cuts for PET and sub-graphed graph cuts for PET/CT	hand-and-neck cancer, 16 patients	DSC 0.78±0.045 (PET) 0.86±0.051 (PET/CT)
[74]	random walk	static phantom, 4 cases	92.7±2.99%, 96.7±3.18%, 88.9±2.62%, and 83.6±2.68%
[23]	automatic seed planting random walk on PET/CT	lung cancer, 75 cases	DSC 0.91, HD 4.5mm

Table 2.4: Region and boundary based methods and their accuracies

A fourth co-segmentation method is proposed in [86]. It uses the idea of graph cuts where a Markov random field (MRF) algorithm is defined on a graph. The method uses two sub-graphs for the segmentation of PET and CT images respectively and defines an error term to incorporate the information from both modalities. In a later study [75], the results from both modalities are combined into the same result, by adding a third sub-graph as result outcome. The shortcoming of the method is it requires user interaction in order to perform well, and it has not been tested with small tumors. It also suffers from its assumption of one-to-one correspondence between PET and CT delineations. Apart from that, this kind method shows potential to fully incorporate information from both modalities under same structure, and achieves globally optimized result.

Results Table

Some typical results from the co-segmentation methods and region based methods are summarized in Table 2.4.

Conclusion for Tumor Delineation Methods

Given the reviews above, the choice for delineation method in this work is a variation of graph-cuts methods, for it shows good results when compared to state-of-art thresholding methods and learning methods. Also it facilitates the incorporation of CT information into the PET segmentation process. In terms of disadvantages, a conventional graph-cut approach may be susceptible to performance variations, since it depends on human interaction for completeness. In this thesis, the proposed framework addresses this potential pitfall by further implementing a tumor recognition process after delineation, as explored in the next section.

Tumor Recognition Methods

In the above literature review for delineation, almost all of the methods require a pre-defined ROI for initialization. If the ROI is not available or incorrectly defined, the delineation methods may actually recognize some normal tissues—most notably kidney, heart and brain—as tumors, since these tissues tend to exhibit high uptake. In the recognition process of this work, we seek to partition genuine tumors from normal tissues with high uptake region. There is a rich clinical literature on automated detection of tumor (also known as tumor auto diagnosis methods). However, most of the existing works use CT for recognition, while very few works perform this process on PET/CT. In this section, the literature of tumor automated detection is reviewed. Also, many works focus on differentiating benign nodules from malignant ones. By contrast, in this work, we focus on differentiating tumors from normal tissues.

CT image based Recognition: Learning based on Shape and Appearance

For CT image based recognition methods, many are based on the concept of tumor growth rate: comparing CT images with segmented tumor region from the same patient but at different times to see if segmented area changes over time, and accordingly judge if such a change comes from malignant, benign or normal tissues. These methods rely on morphologic features for shape recognition. Several works reveal the relation between shape features of segmented region and their underlying pathology. In [87], a total number of 193 pulmonary nodules on HRCT are manually examined and subjectively rated in several types, including round, lobulated, densely spiculated,

ragged, and halo. It is discovered that certain shapes have a high possibility of being malignancy: lobulated (82%), spiculated (97%), ragged (93%), and halo (100%). On the other hand, around 66% of round nodules proved to be benign.

As this approach demands a large amount of manual work, researchers have employed pattern recognition to automate the process. Prior works [88, 89] extracted features such as surface curvature and degree of surrounding radiating patterns from benign and malignant nodules, resulting in a good separation between classes. In the work [90], the method is further developed into an automated system. The system employs features such as CT image intensity and curvature shape index extracted from joint histograms. The system was tested on a database with 248 objects, and shows reasonable result. The authors of [91] analyze lung nodule using fractal features in a database containing 117 objects with various conditions. A shortcoming for these methods is that the authors did not directly assess the accuracy of their method in diagnosis prediction. However these works indicates the relationship between nodule shape and nodule's potential pathology.

These works implement a classical pattern recognition structure, and are evaluated using ROC curves. Table 2.5 shows a list of these methods.

Table 2.5: Shape based methods

Reference	Method Aiming	Method	Database	Performance
[92]	To discriminate between B (benign) and M (malignant) nodules using neural networks (NNs)	Statistical-multiple-object detection and location system (S-MODALS) NNs technique developed for automatic target recognition (ATR)	CT images of 28 pulmonary nodules, 14 B and 14 M, each having a diameter less than 3 cm were selected	Correctly identify all but three B nodules, but did not misclassify any M nodule

[93]	To classify nodules into benign or malignant	LDA with stepwise feature selection based on nodule's shape, size, attenuation, distribution of attenuation, and texture	HRCT scans of 17 M and 14 B nodules	Correct classification rate of 90.3%
[94]	To discriminate between B and M nodules	LDA with stepwise feature selection based on nodule's features (density and curvatures) and surrounding structure features	CT images of 248 pulmonary nodules including 179 M and 69 B nodules	Nodule's features ($A_z = 0.88$) were more effective than the surrounding structure features ($A_z = 0.69$) in classification. Combing both features achieves $A_z = 0.94$
[95]	To classify nodules into benign or malignant	ANN with 16 subjective features determined by radiologists and 7 clinical data	155 HRCT scans of 99 M and 56 B nodules	$A_z = 0.951$in a leave-one-out test
[96]	To quantify lung nodules in thoracic CT	A NNs based on geometrical features, intensity, and texture features	CT images of 48 cases of lung nodules (24 B, 24 M)	$A_z = 0.89$

[97]	To classify nodules into benign or malignant	LDA withWilks' lambda stepwise feature selection	Thick-slice (10 mm) screening LDCT scans of 76 M and 413 B nodules	$A_z = 0.849$ in a leave-one-out test
[98]	To classify nodules into benign or malignant	Two NNs: one trained with 8 subjective features recorded by radiologist rating and the other with 12 matched computerized objective features	56 radiographs of 34 M and 22 B nodules	$A_z = 0.854$using subjective features and $A_z = 0.761$ using objective features. The reported radiologist accuracy was $A_z = 0.752$
[99]	To classify the shape of pulmonary nodules using computer analysis of HRCT	LDA with 2 features (circularity and second moment)	HRCT images from 102 patients with 102 nodules classified as round or oval, lobulated, polygonal, tentacular, speculated, ragged, and irregular	For 95 of 102 cases, the shape classification by the two radiologists was the same. For the seven mismatched cases, pulmonary nodules with circularity $\leq$0.75 and second moment $\leq$0.18 were very likely to reveal lung cancer

[100]	To classify nodules into benign or malignant	Logistic regression or QDA with stepwise feature selection from 31 features	Thin-slice ($\leq$3mm) CE-CT scans of 19 M and 16 B nodules	$A_z = 0.69$ and 0.92 with logistic regression and QDA, respectively, in a leave-one-out test
[101]	To classify nodules into benign or malignant	LDA using 3 features: shape index, curvedness values, and attenuation	Thin-slice (2 mm) CE-CT scans of 35 M and 27 B nodules	$A_z = 0.91$ and 1.0 with non-CE CT and CE-CT, respectively, in a leave-one-out test
[102]	To classify nodules into benign or malignant	Multiple MTANNs using pixel values in a 9×9 subregion	Thick-slice (10 mm) screening LDCT scans of 76 M and 413 B nodules	$A_z = 0.88$ in a leave-one-out test
[103]	To classify nodules into benign or malignant	LDA based on nodule's circularity and second moment features	HRCT (0.5–1mm slice) scans of 52 M and 55 B nodules	Sensitivity of 76.9% and a specificity of 80%
[104]	To classify nodules into benign or malignant	LDA or SVM with stepwise feature selection	CT scans of 124 M and 132 B nodules in 152 patients	$A_z = 0.857$ in a leave-one-out test
[105]	To classify nodules into benign or malignant	ANN ensemble	CT scans (slice thickness of 2.5 or 5 mm) of 19 M and 13 B nodules	$A_z = 0.915$ in a leave-one-out test

[106]	To classify nodules into benign or malignant	GA-based feature selection and a random subspace method	Thick-slice (5 mm) CT scans of 62 M and 63 B nodules	$A_z = 0.889$ in a leave-one-out test
[107]	To classify nodules into benign or malignant	Analysis of the spatial distribution of the nodule Hounsfield values	CT scans (2mm slice) of 51 M and 58 B nodules	Sensitivity of 92.3% and a specificity of 96.6%
[108]	To classify nodules into benign or malignant	Analysis of the SHs needed to delineate the lung nodule	CT scans (2mm slice) of 153 Mand 174 B nodules	$A_z = 0.9782$

PET/CT

Compared to many CT based recognition methods, few works focus on combining PET with CT to achieve higher performance. PET reinforces the recognition process, for it reveals functional information of the region despite its low spatial resolution nature. In works [10, 109–117], PET characteristic is used to detect malignant solitary pulmonary nodules (SPNs) which are single, spherical, well-circumscribed, radiographic opacity that measure $\leq 3cm$ in diameter. The studies shows a sensitivity of 88–96% and a specificity of 70–90%.

In order to further evaluate detection performance on PET/CT, the work [118] uses an ANN approach to classify benign and malignant pulmonary nodules on PET/CT. Results indicate 0.95 accuracy. The work [119] compared the diagnosis accuracy among CT, side-by-side PET/CT and software-fused PET/CT. The results indicate software-fused PET/CT gives highest performance when it comes to lung cancer cases. The work [114, 115, 120] indicate similar results on various cases. The table 2.6 gives a summary of these methods.

Recently, more complicated methods are employed on this matter using more sophisticated features. In [63] and [121], high order features, such as neighbourhood grey-tone difference matrix

Reference	Method	Database	Evaluation Performance (Sensitivity-Specificity)
[109]	Shape analysis on CT and PET intensity	30	95%-80%
[110]	Shape analysis on CT and PET intensity	61	93%-88%
[10]	Shape analysis on CT and PET intensity	89	92%-90%
[111]	Shape analysis on CT and PET intensity	71	95%-82%
[112]	Shape analysis on CT and PET intensity	36	93%-77%
[113]	Shape analysis on CT and PET intensity	28	94%-89%
[114]	Shape analysis on CT and PET intensity	42	96%-82%
[115]	Shape analysis on CT and PET intensity	119	96%-88%
[118]	ANN approach for PET and CT features	92	95%-95%
[122]	Shape analysis on CT and PET intensity	53	94%-75%
[121]	14 PET features and 15 CT features based on NGTDM	20	NA
[63]	DT based KNN method based on PET and CT features	20	70%,-100%

Table 2.6: PET/CT methods

(NGTDM), are used for recognition of tumors on PET/CT images. The NGTDM feature shows superior performance in terms of separation, and is worth further investigation.

Conclusion for Tumor Recognition Methods

Mainstream computer aided recognition methods are machine learning methods, differentiated by the features and classifier used. However, few works among them are applied on PET/CT simultaneously, and the features selected in these works are not guaranteed for best performance. Also most works emphasize on differentiating benign nodules from malignant nodules while, in our work, we aim to differentiate tumor nodules from normal tissues. Also, in order to examine

them in depth, 3 classifiers and 60 features are used in the recognition process in this work for a through examination.

3. A 3D Semi-Automated Co-Segmentation Method for Improved Tumor Target Delineation in 3D PET-CT Imaging

As mentioned previously, this chapter is organized and presented as a manuscript. It is thus essentially self-contained, and has a modicum of overlap with the literature survey in chapter 2. In this chapter, the proposed tumor delineation method is described. The method utilizes a graph-cuts technique based on Markov Random Field theory to combine the information from PET and CT together to achieve high performance. In order for this method to function properly, a user must first draw a rough area, i.e., a region of interest (ROI) around the tumor, and the method is then performed in that area. The ROI selection procedure is later automated by the localization method described in Chapter 4.

Abstract

As a multi-modal imaging technique, positron emission tomography (PET)-computed tomography (CT) can provide not only functional but also anatomical tumor information. Therefore, given this important information, radiotherapy planning is becoming increasingly reliant on PET-CT. In this work, a novel co-segmentation method is proposed to enhance the accuracy of tumor segmentation using PET-CT, to further improve the performance of radiation treatment. The method seeks to use features from both imaging modalities, by modeling the segmentation problem into minimization of a Markov random field model. Based on graph-cut theory, the problem can then be solved via a maximum flow algorithm approach. The proposed method is evaluated using both a phantom with a high-radiation area, and patient data with lung cancers. Both qualitative and quantitative results show significant improvement when compared to existing segmentation

methods.

Introduction

Cancer has been a well-known threat to human life since its discovery. According to a recent World Health Organization report [2], approximately 13.1 million new cancer cases occurred in 2014, causing 8.2 million deaths worldwide. Among treatment solutions against cancer, radiation therapy remains a significant tool. It involves the use of ionizing radiation in an attempt to either cure or improve the symptoms of cancer, and is used in about half of all cancer cases [3]. To enhance the segmentation of tumor targets in radiation therapy, multi-modal imaging modalities, such as positron emission tomography (PET)-computed tomography (CT), are used. The PET modality can provide functional tumor information by showing the biochemical uptake of a metabolic radiotracer, while its disadvantage of low spatial resolution is enhanced by CT.

While PET-CT is becoming increasingly popular, tumor delineation is largely still performed manually by oncologists on a slice-by-slice basis. Such a manual procedure is time consuming, labor intensive, and operator-dependent [4]. Many segmentation algorithms have been developed to address this issue. Reference [4] provides a comprehensive review of these algorithms. However, since many existing algorithms work solely on one modality (i.e., either PET or CT), they are unable to combine advantages of both modalities. Accordingly, a novel semi-automated method for tumor delineation is proposed in this work, exploiting information from both PET and CT. It is based on the graph-cut method, reformulating the segmentation process as an energy minimization problem. The energy encodes both the boundary and region costs, as defined by the information from PET and CT. The minimization problem can then be solved by the classical graph-cut method. Compared to related works in the literature, the novelty of this paper is in the construction of the energy function. While adhering to a standard graph structure, the energy function we propose is suitable for incorporating the region and boundary information from both PET and CT. For the purpose of exposition, the proposed method will be evaluated on lung cancer cases — as lung cancer has high incidence, and the highest death rate among all cancer cases [2].

The remainder of the paper is organized as follows. In section 3.3, we briefly survey popular approaches for the tumor segmentation problem. We describe the dataset used for this work in

section 3.4, and our proposed refinement of the classical graph-cut method for PET-CT in section 3.5. Then, we present our proposed tumor delineation framework in section 3.6, with performance evaluation in section 3.7. Conclusion and future work are discussed in section 3.8.

Background and Related Works

Overview of Tumor Segmentation in PET-CT Imaging

From a medical imaging perspective, the segmentation procedure can be divided into two constituent parts: recognition and delineation [13]. In the recognition procedure, the location of the object is determined; and in the delineation procedure, the spatial boundary of the object is drawn [14]. Specifically for the PET-CT context, in the recognition procedure, the high uptake regions in the PET image are located. These regions correspond to either the tumor regions, or other regions with high metabolism, and are considered as ROIs (Regions of Interest). In the delineation procedure, the aim is to find a precise separation boundary between high uptake regions and regions with non-significant uptakes or belonging to the background [13]. Compared to recognition, delineation is relatively challenging for the following two main reasons [4, 123]: (1) large variability of pathologies in terms of shape, texture, and location, which is primarily caused by the differences of organ morphology and texture from person to person [4]; and (2) image noise, which comes from the image scanner (hardware perspective) and the way the images are reconstructed (software perspective). The noise in PET is considered as one of the most significant factors for not having a reproducible standard uptake value (SUV) measurement, which is defined by the ratio of the radioactivity concentration of image and the whole body concentration of injected radio dose. [15, 123].

Motivated by the increasing usage of PET-CT in clinical applications, there have been many attempts to improve PET image segmentation methods in order to overcome the above difficulties. These improvements may provide more accurate and robust performance under a PET-CT multimodal image frame [4].

In the delineation process, high uptake regions are precisely separated from the background and regions with non-significant uptakes. If the background and non-significant uptake regions

are considered as one class, the separation process is said to be a two-class segmentation problem. Otherwise, the problem turns into a multi-class segmentation problem. Since the high update regions are the actual objects of interest, regardless of how other regions are treated, without loss of generality, a two-class approach is often pursued for notational convenience. Specifically, target delineation in PET-CT images involves dividing the image into two classes: tumor region and non-tumor region. From an image processing perspective, such a division is tantamount to labeling each pixel (2D case) or voxel (3D and 4D cases) in the image with 1 (being tumor) or 0 (being non-tumor). Some works [4] suggests multiple region segmentation, therefore

In recent years, using both PET and CT images for accurate target delineation has attracted considerable attention. Various methods and algorithms have been applied and developed to address this issue theoretically or practically. These methods may be classified into four general categories, as described in the sequel.

Manual Segmentation Methods

For obtaining ROIs in clinical applications, the manual approach involves an experienced physician drawing a boundary that separates the tumor and non-tumor regions by hand. Manual segmentation is also the most common method for obtaining the surrogate truth.

However, there are several inherent drawbacks making the manual approach not ideal for clinical applications. The most significant disadvantage is that the method is highly subjective, and therefore is typically not reproducible. For a particular image, even the same physician may produce segmentation results with minor difference at different points in time. Therefore, the intra- and inter-operator agreement rates are some metrics used to assess the reliability of the obtained surrogate truths, and the level of difficulty of the segmentation problem [27,28,124–126]. Another problem for manual segmentation is that it is time consuming and labor intensive, which in turn may cause the delineation to be less precise, because of the concentration the physician must exert during the segmentation procedure. In a recent study involving 18 physicians from 4 different departments [127], the agreement, in segmentation volume overlap of 70%, was found only in 21.8% of radiation oncologists and 30.4% of hematologic oncologists. From these studies, there seems to be no consensus on the variability of manual segmentation, or on the required experience level

of the experts conducting the manual delineation. However, manual segmentation is nevertheless most widely used and recognized by physicians, therefore manual segmentation is still considered as golden standard to determine the quality of segmentation result given by other segmentation methods.

Threshold Methods

Among the non-manual segmentation methods, thresholding is perhaps the most simple and popular one. By labeling all the voxels that are greater than a certain threshold value to be the foreground and others as the background, this technique effectively converts a gray-level image into a binary image [32]. The rationale of using it in PET-CT is straightforward: the ROIs in PET images usually have far higher uptake values than the normal areas. The PET intensity histogram usually provides enough information for separating the ROIs from other areas. However, no general agreement is reached among various works for the threshold settings due a wide range of reasons, including large variability of pathologies, image nosies and the blurred target boundaries [4]. What is worse, the low spatial resolution of the PET image often limits the thresholding method from giving accurate results. The state-of-art thresholding methods usually fall into one of the following categories: fixed thresholding, such as [33–35]; adaptive thresholding, such as [49]; and iterative thresholding as presented in [53].

Stochastic and Learning-Based Methods

In general, stochastic and learning methods seek to exploit the differences between uptake regions and surrounding tissues in a statistical manner. Among this category, mixture-model methods assume the PET image can be viewed as a summation of Gaussian densities, and attempt to separate these densities using an optimization algorithm such as the Expectation Maximization (EM) algorithm [59]. Locally adaptive Bayesian methods such as [58, 60] use the mixture model in another way, by building a fuzzy boundary and adapting it with a Bayesian approach. Many classical clustering methods, such as k-nearest neighbor (K-NN) [61], fuzzy C-Means [66], artificial neural network [61], support vector machine [64] and Affinity Propagation [67], have been applied to PET-CT segmentation by viewing the foreground and background as two different classes. As

a high-light method under this catergories, Affinity propagation methods proposed by [67] gives good result on multi-focal and small tumor cases, while other methods generally show good performance for the delineation of focal uptake, but their performances are more limited with diffused and multi-focal cases, as shown in work [4].

Region-Based Methods

Last but not least, region-based methods emphasize the use of the information from local intensity distribution (homogeneity or the boundary information), while still utilizing the global intensity distribution (the region information). Recently, region-based methods such as region growing, random walk, and graph-cuts have gained significant attention. The graph-cut method generated by Boykov and his colleagues was originally proposed for general imaging processing [128], and is later applied to the PET-CT scenario [75]. Since our proposed method falls into this category, we will explore it further in the next section.

Graph-Based Segmentation Methods

Region-based methods are typically referred to as graph-based methods, because they often utilize graphs for incorporating homogeneity information. By using foreground and background seeds specified by a user, graph-based approaches locate objects in the image by solving an optimization problem, generated by implementing local pairwise voxel similarities (homogeneity) with global information. Among the graph approaches, two methods are most often used for PET segmentation: Random Walk and Graph-Cuts.

Random Walk

Although originally designed for computer vision, the random walk has been applied to image segmentation in many cases. The advantage of Random Walk related approaches is that they are robust against noise and weak boundaries, making them potentially suitable for PET-CT segmentation because of the low resolution and high noise characteristic of PET [31, 74]. However, the use of random walk is limited, since the method are not able to deal with multi-focal uptake regions properly.

Graph-Cuts

In general, graph-cuts have been widely used for image segmentation. The main idea is to formulate the segmentation problem into a discrete energy minimization problem. Image features are incorporated by designing proper cost functions for the so-called energy terms. Different segmentations will lead to different cost values based on the defined cost function. Therefore, the segmentation problem can be solved by minimizing the cost. In order to solve the minimization problem, a graph is constructed based on the image, where nodes in the graph represent the voxels in the image, and the connections between the nodes represent the degrees of similarity between the nodes. Two special nodes, the source node and the sink node, are added in order to utilize the region information. The constructed graph can be used to solve the energy minimization problem defined to deliver optimal segmentation. By using local pairwise pixel (or voxel) similarities, the graph-cut approach has shown immense promise in optimally segmenting images [4].

Graph-cut segmentation methods are differentiated by their cost function definitions, and the corresponding ways of solving the minimization problem. For example, the work by Felzenswalb and colleagues [129] used pairwise region comparison for image segmentation, based on a minimum spanning tree (MST) algorithm. Grady et al. [130] showed how to extend the shortest-path method for 3-D surface segmentation. Xu et al. [131] proposed an approach using shortest-path algorithm for multiple surfaces segmentation in 3-D.

Inspiration for Our Proposed Method

Taking a similar approach to [75], our work utilizes a graph-cut method. The basic idea is to reformulate the segmentation problem as an energy minimization problem. The rationale for selecting this approach is that it facilitates incorporating the boundary and region information from both PET and CT into one graph-cut framework, and the corresponding maximum flow algorithm is a mature algorithm that guarantees attaining a global optimal of the minimization. Our approach is largely inspired by [75]. Both works use the graph-cut model, and design the cost function from the boundary and region information from PET and CT. However, there are several drawbacks of Song's work. First, in their work, it is claimed that the tumor volume in PET may not be identical to CT, and therefore two potentially different segmentation results are produced. However, this

PET(^{18}F – FDG)					CT		
Frame Length	Width/length	depth	pixel size	slice depth	Width/length	depth	
7 mins	192pixels	83slices	1.8229mm by 1.8229mm	3.27mm	512pixels	135slices	

Table 3.1: Phantom Descrption

point has been criticized by several works [4, 21], since the PET and CT are aligned and registered in Song's method. More importantly, regardless of the method applied, physicians ultimately still need one segmentation result for the upcoming application such as radiation therapy. Also Song's method uses a dual-graph for implementing graph-cuts. As a result, the computational cost for the method is rather significant, and even higher if expanded to the 3D version.

Given the above drawbacks, one novelty of our work is that we use a single graph model, combining both PET and CT information into one graph, instead of two graphs as in Song's work, thus avoiding the unnecessary penalty term for the label difference between PET and CT. This reduces the computational cost, in turn making a 3D graph-cut approach more accessible. Secondly, unlike traditional graph-cuts we do not use initial seeds for the foreground in the tumor selection. And perhaps the most notable difference between our approach and Song's work is that they take the region cost from CT into consideration while we do not. As will be seen in section 3.7, experimental results show that even under the circumstances where the involvement from CT into PET in the regional cost is 10%, the results are worse in adverage.

Dataset Description

Both phantom and patient data are used for evaluation of our proposed method. In this section, the details of the dataset utilized are provided.

Phantom Description

We first validated our method on a specifically designed phantom. The structure of the phantom object and the specifications of the PET-CT image are shown in figure 3.1 and table 3.1.

Here, there are 6 small balls in the phantom, the diameters of which are 4mm, 6mm, 8mm, 10mm, 12mm, 14mm. The phantom helps to evaluate the method performance on the relatively

Figure 3.1: Phantom

small-sized simulated tumors.

The main body of the phantom is filled with radiation, and the balls are filled with normal water. The ground truth is established using the phantom measurements.

Patient Data Description

A selection of 29 sets of patient data was extracted from The Cancer Genome Atlas-Lung Squamous Cell Carcinoma (TCGA-LUSC) data collection on The Cancer Imaging Archive (TCIA) website. Although the original datasets contain more patients, only 29 of them have PET-CT images that were taken simultaneously. The patients suffered from lung cancer, and the tumor sizes vary from 8mm to 33mm.

The ground truth is drawn on a slice-by-slice basis for each patient, and the 3D results are divided into slices for comparison. The patient IDs as well as some details are listed in table 3.2. For more complete details on the dataset, the interested reader can refer to the TCIA website [1].

Proposed Graph-Cut Method

The main differences between various graph-based segmentation methods are found in the cost function definition, and in the approach of consequently solving the minimization problem. For instance, consider the graph-cut method: it models the cost function as a combination of "self cost" and "neighbor cost".

To demonstrate the graph-cut model, suppose that we have a 3D PET-CT image pair (I, I'), with I representing the CT image and I' representing the PET image. Let us further assume that the images are perfectly aligned. Thus, for each voxel $u \in I$, we have a corresponding voxel $u' \in I'$. Let l_u represent the binary labels assigned to each voxel, since we consider the segmentation problem as a two-class problem. Specifically, $l_u = 0$ means that the voxel u belongs to the background (i.e, non-tumor in our case), while $l_u = 1$ indicates that the voxel u is put into the foreground (i.e., tumor tissue). Furthermore, t_u represents the image intensity value at the voxel u in the corresponding image.

With the above notations, the self-cost term $E_s(l_u)$ is constructed as follows. For each voxel $u \in I \quad \text{or} \quad I'$, the self cost function is defined as

$$E_s(l_u) = \begin{cases} C_{source}(t_u), & \text{if } l_u = 1 \\ C_{sink}(t_u), & \text{if } l_u = 0 \end{cases}. \tag{3.1}$$

Where C_{source} and C_{sink} are self-cost functions. There definition can be determined by user and decides performance of graph-cuts algorithm. Also, for a neighbor voxel pair (u, v), the neighbor cost function $E_n(l_u, l_v)$ is defined as

$$E_n(l_u, l_v) = \begin{cases} C_{nodes}(t_u, t_v), & \text{if } l_u \neq l_v \\ 0, & \text{otherwise} \end{cases}. \tag{3.2}$$

We further define a neighbor set N in the image pair (I, I'), where $N = \{(u, v) | u \text{ and } v \text{ are neigbors}\}$. Then, the main idea of graph-cuts can be formed as finding a labeling set L_m among all the possibe sets $L = \{l_u\}$, that the set L_m minimizes the sum of self and neighbor costs on all the voxels in the

image pair (I, I'):

$$L_m = \arg\min_{L} \left(\sum_{l_u \in L} E_s(l_u) + \sum_{(l_u, l_v) \in L, (u,v) \in N} E_n(l_u, l_v) \right). \tag{3.3}$$

Traditionally, the cost functions C_{source} and C_{sink} are defined based on the intensity value of PET or CT. Some works use the cost terms of the PET image alone to decide the segmentation [74]. Others combine the cost terms later to decide the optimal solution [75].

In our work, we seek to encapsulate the information from PET and CT together into the self cost term and neighbor cost term by redefining the cost function C_{source}, C_{sink} and C_{nodes} based on both PET and CT information. To this end, we first describe the graph model of the graph cuts.

Graph Construction

In graph cuts, the function (3.3) is solved by a maximum flow algorithm after modeling the function into the graph form. The model is shown here in figure 3.2. As shown in the figure, S is the source of flow and the sink T is the end of flow, each voxel in the image is modeled as a voxel node, with tube connecting them to the source, sink and themselves. Self costs (i.e., the C_{source} and the C_{sink} values) are represented by the tube capacity from the source to the voxel node (the corresponding value given by C_{source}, and the orange lines in the 2nd subfigure), and from the voxel node to the sink (the corresponding value given by C_{sink}, the blue lines in the 2nd subfigure). And neighbor cost (i.e., the corresponding given by C_{nodes}, the black lines) is modeled as the tube capacity between the nodes themselves. By doing this, equation (3.3) can be effectively solved by solving the corresponding maximum flow problem (that is, using which set of tubes gives maximum flow capacity from source to sink) under this graph [128]. The solution ultimately gives the segmentation result (The selected tubes are selected costs), as described in further detail in [128].

Naturally, we have $l_u = l_{u'}$ if u and u' are corresponding voxels in PET and CT, so we only need to segment one image for an image pair. Therefore, each node of the graph used in this work represents a voxel pair. This model is much like the one used when performing graph-cuts using

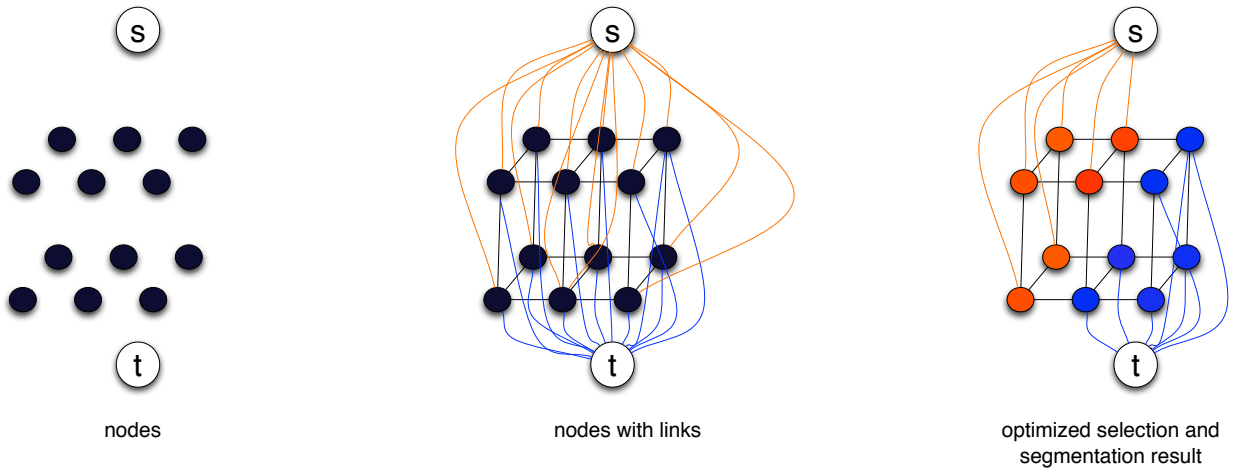


Figure 3.2: Graph-cuts modal process

This graph shows the modal process of graph-cuts. The nodes are constructed based on image pixels/voxels, two nodes source and sink are added. The single direction links/tubes are then added between pixel/voxel nodes and source/sink nodes, the link/tube capacity is decided by source self cost function C_{source} (orange lines) and sink self cost function C_{sink} (blue lines). These link/tubes are single directed, only from source node to pixel/voxel nodes (orange) and from pixel/voxel nodes to sink node (blue)

Also double direction links/tubes are added between neighbor pixel/voxel nodes (black lines). The capacity of these links/tubes are decided by neighbor cost function C_{nodes}. Then maximum flow algorithm is performed and the link/tubes that actually has flow in them will decide the node segmentation.

a single modality, and the difference is the choice of cost functions. Here we choose to define self cost functions from PET and neighbor cost functions from PET and CT, as it gives better performance than other combination

The most important factor for good performance in this method is the appropriate selection of the cost functions. Under this model, there are certain rules for the cost functions definition to ensure the success of the segmentation. Self cost represents the intensity similarity between the voxel and region (foreground or background), can be computed using intensity of the voxel itself (two self costs are given to one voxel, one relates to foreground and one relates to background). For instance of a voxel with higher intensity, the foreground self cost is low and background self cost is high, making algorithm tend to label voxel as foreground, since selecting it into background result in higher cost in energy minimization function 3.3. On the other hand, neighbor cost represents that pixels with same region tend to be close to each other, and only occurs when the neighbor pixels belong to different regions.

Energy function

The major differences between our proposed method and existing ones are in the energy function selections. Traditionally seeds are placed for the foreground and background in order for the graph-cuts to work robustly. In this context, the seeds refer to certain voxels that can invariably only be segmented into foreground or background, in all segmentation solutions. Unlike the traditional graph-cut approach, we employ a relatively simple method to place seeds. In our approach, we do not place any seed for the foreground, i.e., for the tumor. Instead, we place seeds for the background, by roughly marking the known high uptake value regions without tumor. Such regions, for example the heart and and urinary bladder, are given the cost $C_{source} = \infty$ and $C_{sink} = 0$.

Subsequently, we combine both CT and PET information into the cost function. Further details are described in the sequel.

Self cost

The self cost in this case still comes from the PET image itself (and is the flow capacity from the node to the source node and to the sink node).

We use PET and CT images, and after normalizing them, we get the image pair I_n, I'_n. Then, for each node corresponding to the voxel pair (u, u'), the self cost term for our approach is defined as:

$$C_{source}(t_u) = \begin{cases} |t_u - \tau_{source}|, \\ \qquad \text{if u is not marked as a background seed, ,} \\ \qquad \text{and if } t_u < \tau_{source}, \\ 0, \quad \text{if } t_u > \tau_{source} \\ \infty, \quad \text{otherwise} \end{cases} \tag{3.4}$$

$$C_{sink}(t_u) = \begin{cases} |t_u - \tau_{sink}|, \\ \qquad \text{if u is not marked as a background seed,} \\ \qquad \text{and if } t_u > \tau_{sink}, \\ 0, \quad \text{otherwise} \end{cases}$$

Here C_{source} is the flow capacity from the node to the source, and C_{sink} is the flow capacity from the node to the sink. t_u is the normalized PET image value on the voxel u (i.e., the radiation intensity on the voxel). τ_{source} and τ_{sink} are given parameters defined by users.

Neighbor cost

We use a parametrized neighbor cost in our method. The neighbor cost in our method is:

$$C_{nodes}(u, v) = \begin{cases} \left[1 + \lambda_1 (G_{uv} + \lambda_2 \dfrac{t_u + t_v}{2} G_{u'v'})\right]^{-\alpha}, \\ \qquad \text{if u and v are neighbor pixels} \\ 0, \quad \text{otherwise} \end{cases}$$

Here, G_{uv} and $G_{u'v'}$ are the gradient values of normalized PET and CT between the voxel pairs (u, v) and (u', v') correspondingly. α is a given parameter defined by user. The main idea is to only include the gradient information from PET and the gradient information from CT for high uptake value regions in PET.

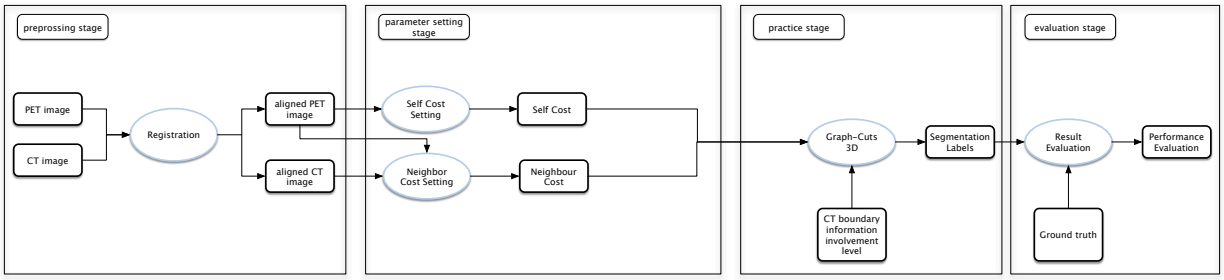


Figure 3.3: High-Level Block Diagram of the Proposed Segmentation Framework

Proposed Tumor Delineation System Framework

Figure 3.3 shows a block diagram of the proposed framework involving four stages: preprocessing, parameter setting for the graph-cuts, practice (performing 3D graph-cuts) and evaluation.

Preprocessing

As mentioned in the previous section, one assumption of the proposed method is that the PET image and CT image are aligned. This assumption is ensured by the preprocessing stage, where the CT image is aligned with the PET image. The stage is also called registration, resulting in an aligned image pair.

For the registration phase, we have the following options to consider: rigid transformation, meaning the image is only allowed to rotate and translate; similarity, meaning the image is allowed to rotate, scale and translate; affine transformation, meaning the image is allowed to rotate, scale, translate and cut; b-spline transformation, meaning all the features from affine transformation plus a certain amount of non-linear transformation defined by b-spline function.

It is obvious that b-spline transformation is the transformation with most freedom; however, the results from phantom indicates this transformation tends to lead to overfitting of the image, and in turn further poor segmentation. Accordingly, the affine transformation is implemented instead in this step. The registration is performed using the elastix toolbox [132].

Parameter Setting

In this section we decide the value of τ_{source}, τ_{sink}, λ_1 , λ_2 and α. In order to determine τ_{source} and τ_{sink} , an Otsu's thresholding algorithm is first applied to PET volume, then τ_{source} is set to be $\frac{7\tau}{4}$, and τ_{sink} is set to be $\frac{\tau}{4}$, while τ being the threshold value decided by the thresholding algorithm. In this application we set the λ_1 to be 10 and λ_2 to be 10, $\alpha$is set to be 1.5.

Practice Stage

A 3D graph-cuts algorithm is performed based on maximum flow algorithm as we discussed in the graph-cuts section. The neighbor set is defined as the 8 neighbor set, the most common configuration in 3D image processing scenario. The algorithm is performed on Matlab 2014.

Evaluation Stage

Here we evaluate the performance of the method, more details will be shown in the following section.

Experimental Results

This section describes how the experiments are set up and how the results are evaluated. For evaluation, the method is performed on both phantom and patient data with established ground truth. The results from the proposed method and other comparison methods are then further evaluated using Dice index and Hausdorff distance, two of the most widely used evaluation metrics in medical imaging.

Experimental Setup

Ground truth establishment

The true boundary of the object of interest should be identified in order to evaluate an image segmentation algorithm, such process is often referred to as ground truth establishment.

For the phantom case, ground truth establishment is an easy task. An advantage for using

phantom image is that the exact dimensions of the phantom can be measured before and after imaging procedure, so that in most cases one know exactly what a perfect segmentation result should be. When it comes to patient data, it is often hard or impossible to acquire ground truth eve with histopathologic samples. Therefore, in many cases surrogate truths (or reference standards) are used for measuring the quality of a segmentation algorithm. Even the same physician shows variability in segmentation results when segmenting the same image many times over. Nevertheless this is the best option we have for now, representing the closest starting point upon which we can compare our result. After all, most of today's radiation therapy treatment plans are made according to the physicians' segmentation.

Evaluation Matrix

With the ground truth established, the comparison of how close the segmentation result is to the ground truth can be made. There are two major classes for such a comparison: one is similarity indices, which show how similar the ground truth and the result are, e.g., the Dice index; the other is error indices, which show the maximum offset of the result from the ground truth, e.g., the Hausdorff distance.

As mentioned, Dice index is one of the most widely used similarity indices. Given X as our method result and Y as ground truth, the Dice index can be computed using the following equation:

$$d_I(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{3.5}$$

with $|.|$ defined as the area of the graph.

As can be seen in the definition, Dice index represents the overlay area of the ground-truth and the result over the total area. As a similarity index, it reflects the overall performance of the segmentation accuracy.

Another evaluation metrics used in this paper is Hausdorff distance, one of the most popular error indices. Given x as voxel in the image X and y as voxel in the image Y, the Hausdorff distance is defined as:

$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\}, \tag{3.6}$$

with $d(.,.)$ defined as the distance (usually Euclidean distance)

The Hausdorff distance, by definition, is the largest offset from the result to the ground truth. It represents the detailed accuracy of the result. In this work, the Hausdroff distance is shown using the real distance instead of pixel distance. That is, after calculating the Hausdroff distance by definition, we multiply the result by the voxel size.

By using Dice index and Hausdorff distance together, both overall and detailed performances of the method can be evaluated. This explains why one can see this combination often used in the existing literature as the evaluation metric. In general, better result gives higher Dice index and lower Hausdorff distance.

Comparing Methods

As mentioned, we perform our proposed method on both the phantom data and the patient data, then evaluate the performance using the established ground truth. With the evaluation complete we run several classical and state-of-art methods and evaluate their performance.

Before we come to the results, here is a list of comparison methods and their description.

- Thresholding methods (T1,T2) As described above the thresholding method is the most popular method and thus included in our comparison. The classic methods used in this paper come from the work [25] (marked as T1) and the work [48] (marked as T2).
- Machine Learning methods (AP) we use a novel machine learning method based on Affinity Propagation principle, that is the method from the work [67], marked as AP.
- Graph-cuts based Approaches (GA, CO, PM) The graph-cut approach is the basis of our method. As such, it is vital for our method to be compared with other similar approaches based on the same principle. The approaches chosen for comparison are: the basic graph-cut method that performs graph cuts on PET only, based on the work [128], marked as GA; a

recent co-segmentation method based on the graph-cut principle [75], marked as CO; and our proposed method (PM).

Results and Discussions

We run our methods and the comparison methods using Matlab 2014b on a PC in our lab. The configuration of the PC is Intel Core i7-4790 CPU @ 3.60 GHz with 16GB RAM, also with a NVIDIA Quadro K2000 for GPU computing.

The Overall Performance

First, in Table 3.3, a summary of the overall performance results can be found. Since we only have one phantom, no variation value can be shown on phantom results.

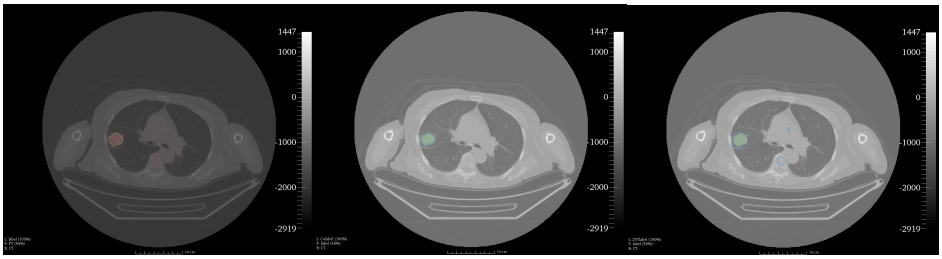
From the table we can see that our method attains the best performance over all the methods used, with a Dice index close to 0.93 and HD lower than 3mm. Considering the voxel size of PET-CT image, 2.5mm HD means an error with less than a voxel deviation in PET image. The co-segmentation method achieves second best performance, makes sense because of the similairity of method ideas. Compared to CO method, the proposed method has higher variation of HD at patient dataset. but it does not matter that much since proposed method has much lower average HD value. However it may implies that proposed method has a little less robust than CO method.

In order to illustrate how the method works, we provide an image of the PM result on the phantom in Figure 3.4. We can see from this image that the result created by PM is very accurate. The boundary between the radiation and the normal tissue is very well separated. Here the high radiation dose in injected into the liquid that surrounds the spheres, while pure water is in spheres themselves. The PM method successfully find the right boundary of the high radiation dose area and water in spheres with plastic shell of spheres, despite the fact that plastic shell has higher CT read, while high dose area and water has similar CT read.

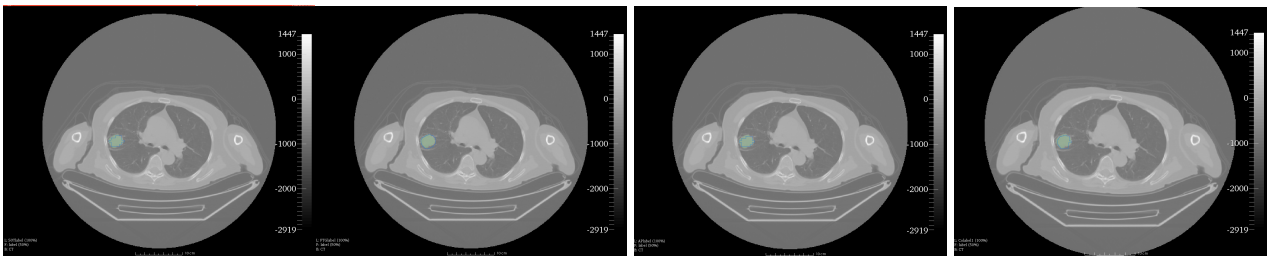
In order to give an impression of the overall performance of the methods in real life application, the result images of one slice from all the approaches are shown in Fig. 3.5.

From the image we can see that our method gives clear separation while keeping the tumor

Figure 3.4: Resulting Image of the Proposed Method on the Phantom

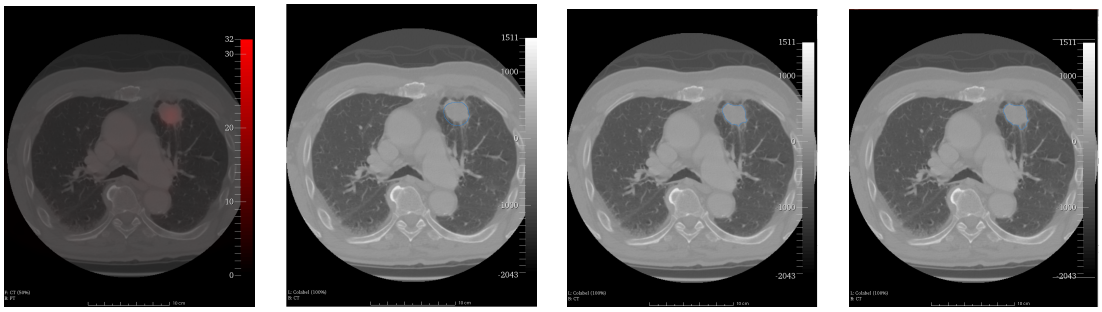


(a) Ground Truth **(b)** Proposed Method (PM) **(c)** Threshold Method (T1)



(d) Threshold Method (T2) **(e)** Graph-Cuts PET (GA) **(f)** Affinity Propagation(AP) **(g)** Co-Segmentation (CO)

Figure 3.5: Result on the Patient data.
On the top left we present our original slice, with CT represented by the black-and-white image and PET represented by red region. The green line in the original data represents the ground truth. The blue line on other images are the boundaries given by the methods, and the green regions are the ground truth



(a) original image **(b)** PT-CT information level $\lambda_2 = 0.1$ **(c)** PT-CT information level $\lambda_2 = 10$ **(d)** PT-CT information level $\lambda_2 = 100$

Figure 3.6: Phantom Description

details. The results from two thresholding methods and AP method suffer greatly from low resolution of PET, therefore give relatively low separation details. Also in the first thresholding method the in-relative aspen area is wrongly segmented. Graph related methods give higher details because of the CT involvement, because of a pre-performed implementing value on PET image, GA method achieves smoother boundary regardless of performing only on PET. CO method gives conservative result compared to PM method, which explains its better robust and lower HD variation. When comparing PM method to GA and CO, we can see that different level of CT information involvement leads to over-segmentation and under-segmentation. Part of this effect is demonstrated below.

The level of CT information

From above we find different level of CT involvement can lead to over or under segmentation. While the results are good, we should carefully control the level of CT information involvement since the over-involvement. For this section, we seek to analyze this effect under our proposed method framework. From figure 3.6. we can see how the level of involvement impact the result. The ratio indicates the value of the parameter λ_2 in the equation neighbor cost function.

Firstly, when compared results from first figure which has very little CT involvement, to the last figure which has highest level of CT involvement, we can see more CT information gives richer boundary detail. However, highest level of CT involvement put too much weight on CT boundary in algorithm, resulting a boundary closer to a boundary of overall tissue than a boundary of tumor

tissue, since in many cases tumor tissues have similar CT reading to its surrounding tissues, too high CT involvement level in this case will surely lead to an over-segmentation result.

Given above, we can conclude that over-committing such an involvement can lead the segmentation boundary close to the tissue boundary in CT and thus results in over-segmentation, while under-commitment will lead to a result that is closer to thresholding, thus losing the details of the tumor. Therefore selecting the appropriate level of CT involvement is important in this work. In this paper we use decide the involvement level using experiment approach. A more automated approach for this decision could be a work for future.

The involvement of CT Region Cost

In the last section CT involvement impact is shown for neighbor cost. In our framework there is no CT involvement for self cost functions, this is also one of the differences between our work and the work [75], which is the CO method. Given the above analysis, it would be interesting to see how CT involvement affect segmentation if put in self cost function. Since self cost represents region similarity, it is also noted as region cost in this paper. Table 3.4 shows the results of our method if CT Region Cost is included.

In this table One can conclude that involvement of CT information in Regional Cost, no matter how small it is, will lower the method performance. It is decided by tumor segmentation nature since tumor has PET reading and may has similar CT reading to its surrounding tissues. Since self cost represents the reading similarity between pixel/voxel and the region it belong to, put CT information in self cost may result in mis segmentation since tumor may has similar CT reading to its surrounding tissues. As a result, it is not wise to use CT Regional Cost in our method.

Conclusion and Future Work

Among the potential ways to further improve our proposed method in the future, two specific directions that we believe to be the most promising are described below. One direction is to develop a pre-training approach to decide the parameters of the cost function, therefore changing the method into an adaptive approach. For another direction, note that we currently rely on human to differentiate the tumor region from the other high uptake regions such as kidneys using seed place-

ments. If the seeds are too close to the tumor region, the method becomes potentially inaccurate. Accordingly, we plan to develop a localization approach that will automatically divide the tumor from other high uptake regions. This will remove the requirement for marking the background in the graph-cut approach, therefore delivering a fully automated and hopefully more robust method.

Patient ID	Tumor description	note
2609	Lung Squamous Cell Carcinoma	mixed regions
2695		tumor region only
2696		mixed regions
2711		mixed regions
2719		tumor region only
2721		tumor region only
2722		mixed regions
2723		mixed regions
2724		mixed regions
4079		mixed regions
4080		tumor region only
5236		mixed regions
5239		tumor region only
5240		mixed regions
6867		mixed regions
7340		tumor region only
8064		tumor region only
A5IX		mixed regions
5045	Lung Adenocarcinoma	mixed regions
5066		tumor region only
5072		mixed regions
5941		tumor region only
6591		mixed regions
6594		tumor region only
6595		mixed regions with 3 non-tumor regions
Z050		mixed regions
Z053		tumor region only
Z060		2 correct tumor regions
Z062		mixed regions

Table 3.2: Patient data details, consisting of a selection of 29 patients (with simultaneous PET-CT images) from the TCGA-LUSC data collection [1], the image voxel sizes are 4.6875 by 4.6875 by 3.2701 (mm) for PET, and 0.9765 by 0.9765 by 3.2701 (mm) for CT

Table 3.3: Result Evaluation

Methods	Phantom		Patient	
	DSC	HD	DSC	HD
T1	0.96	13.6720	0.7660 ± 0.0841	44.2301 ± 28.1491
T2	0.92	12.3048	0.8121 ± 0.1261	7.1835 ± 3.3331
AP	0.97	7.5196	0.8385 ± 0.0870	6.2840 ± 1.0519
GA	0.95	8.2032	0.8623 ± 0.0780	5.1273 ± 1.1993
CO	0.99	5.4688	0.8760 ± 0.0534	4.2830 ± 0.6182
PM	0.99	4.1016	0.9238 ± 0.0347	2.5154 ± 0.9249

Table 3.4: CT Region Result Evaluation

Ratio of CT Region involvement (CT:PET)	Phantom		Patient	
	DSC	HD	DSC	HD
none	0.99	4.1016	0.9398 ± 0.0247	2.5154 ± 0.9249
1:100	0.96	6.1524	0.9126 ± 0.0447	2.9185 ± 1.2279
1:10	0.90	7.1778	0.9015 ± 0.0678	3.1154 ± 2.0360
1:1	0.90	9.2286	0.8579 ± 0.0716	3.5126 ± 2.5260
10:1	0.40	31.7874	0.3816 ± 0.1247	20.6420 ± 20.8574

4. A Lung Tumor Localization Framework Based On Image Feature Analysis From PET-CT Chest Imaging

This chapter is also organized and presented as a manuscript, which can be considered as a follow-up work of the previous chapter. Specifically, a tumor localization method is proposed. This method seeks to eliminate the necessity of a user-defined Region of Interest. It proceeds as follows: first, the algorithm in chapter 3 is performed *without* a pre-defined Region of Interest; then, the method in this chapter is applied to locate the tumors, while ruling out normal regions with high uptakes.

Abstract

Positron emission tomography-computed tomography (PET-CT) has become increasingly popular in clinical applications as a tool for tumor diagnosis and delineation for radiation therapy. Many computer-aiding algorithms are developed for the tumor delineation task. However, it is often hard for algorithms to differentiate tumor regions from normal regions with high uptake value. In order to address this issue, in this paper a tumor detection and localization method is proposed. After the segmentation is done, the pattern recognition approach is performed on the segmented tumor regions as well as the segmented non-tumor regions. The method extracts 60 features from PET and CT image, and the sequential forward feature selection is performed using the resulting accuracy from three different classifiers. The method shows good performance by achieving 100% accuracy on all the classifiers.

Introduction

Lung cancer remains to be one of most lethal diseases in modern society. In 2014, the disease causes 14.6% of all the human deaths [2] worldwide. Currently, early diagnosis plays a critical role for cancer treatment. Positron emission tomography-computed tomography (PET-CT) is an important tool for cancer diagnosis and assessment of treatment, thus becoming more and more popular. By showing the uptake of specially designed radiation marker (most commonly $^{18}F-FDG$), PET-CT is capable of detecting tumors at an early stage. On the other hand, radiation therapy remains a main method of cancer treatment. As accurate tumor delineation plays an important role in radiation therapy, the role of PET-CT to limit potential radiation damage to normal tissues around the tumor region is growing [3, 133].

In our previous work, a delineation method was proposed for tumor delineation for radial therapy planing based upon PET-CT imaging using the graph-cuts method [134]. The method combines regional information from both PET and CT and achieves good performance in terms of delineation accuracy. However, the problem of this method as well as many other computer-aided tumor segmentation methods is that it is often hard for many algorithms to differentiate tumor region from other high uptake regions such as heart and liver [84, 121]. As a result, all these regions are marked as region of interests (ROIs) in the method's output. In order to address this issue, foreground and background seeds are placed by user in advance to mark the tumor regions as well as non-tumor regions for the algorithm, and a good performance can thus be guaranteed [121].

However, placing seeds means that manual input is still needed in the tumor delineation process. Meanwhile, since the algorithm outcome varies with the seed placement, the methods are not robust if the seed are mistakenly placed [29]. Hoping to address this issue, a more automated tumor detection and localization method is proposed in this paper. In order to simplify the problem, we limit the question scope to the lung cancer localization in the chest PET-CT modality, The method is shown to be able to recognize the tumor region from non-tumor high uptake regions such as heart and liver.

There are quite a few studies [63, 84, 121, 135–137] that have researched PET-CT features for tumor recognition and localization. In the work [84], the high-order features based on the

neighborhood gray-tone-difference matrix (NGTDM) are used to recognize head and neck tumors and achieves good result using decision tree (DT)-based KNN classifier. In the work [121], a single feature named NGTDM contrast is used for the lung tumor recognition and achieves good classification rate. In this work, we take more features and more classifiers into consideration for the lung cancer, and see if higher classification accuracy can be achieved.

This proposed method follows the classic procedure of pattern recognition, selecting the features from the first, second, and higher order features extracted from both PET and CT, and uses different clustering methods to do classification to see how the features fit with those classifiers.

The method is evaluated based on the two datasets from the TCIA database [138]. Due to the relatively small number of this dataset the leave-one-out strategy is used. The method achieves 100 % accuracy on all the classifiers used, with different combination of features. The results are shown and discussed in detail.

The rest of the paper is organized in the following sections: In section 4.3 the information of the dataset and the preprocessing of the data is shown. Then in section 4.4 the method is proposed and discussed in detail. Results are shown in section 4.5, and discussed further along with some issues in the method in section 4.6. Finally the paper is concluded at section 4.7 and feature work is discussed.

Image and Problem Description

Basic Information of the Database

The dataset used in this work is downloaded from the TCIA website, a public database. Two datasets are selected: The Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) and The Cancer Genome Atlas Lung Squamous Cell Carcinoma (TCGA-LUSC) [1]. There are data of 72 studies from 36 patients in the TCGA-LUSC dataset, and data of 147 studies from 64 patients in the TCGA-LUAD dataset. However, only 18 patients from TCGA-LUSC and 11 patients from TCGA-LUAD had simultaneous PET and CT, these were selected to build the dataset used in this paper. After manual review of the segmentation result, 28 tumor regions and 19 non-tumor regions are extracted from these data. We will talk about segmentation method in detail in the section 4.4.

A list of the data and their details are listed in the table 4.2. The size of the transaxial PET image is 128×128 with the size of transaxial CT image being 512×512.

Preprocessing of the Database

Two major preprocessing processes are performed before the method:

1. To simplify the problem we specify the problem to be localizing lung tumor in the chest PET-CT image. As some of the database consists of whole-body PET-CT, this is limited to the one PET-CT of the chest.

2. Prior to analysis, the PET image is registered and rescaled to the same size of CT image using the elastix toolbox [132], that is, we consider the PET image has the same spatial resolution and size with CT image in the following part of paper if not specifically defined.

Problem Description

As mentioned above, segmentation methods tend to segment non-tumor region with high uptake value as output. For example, the method introduced in [134], if no seeds are placed for background, the segmentation output for chest PET-CT image in certain cases will look like the blue regions in Fig. 4.1.

The blue region in the figure is the output of segmentation result. We can clear see that part of liver, stomach and part of spine are segmented as tumor region.

Method Description

As mentioned in the previous section, the goal of this work is to differentiate the tumor region from other normal high uptake regions located in the chest or around lung area in a tumor delineation method. These healthy regions include the heart and the liver.

Despite the fact that these healthy regions have different pathology features, the problem is still modeled as a classification problem with two classes: tumor region and non tumor region for the sake of simple implementation.

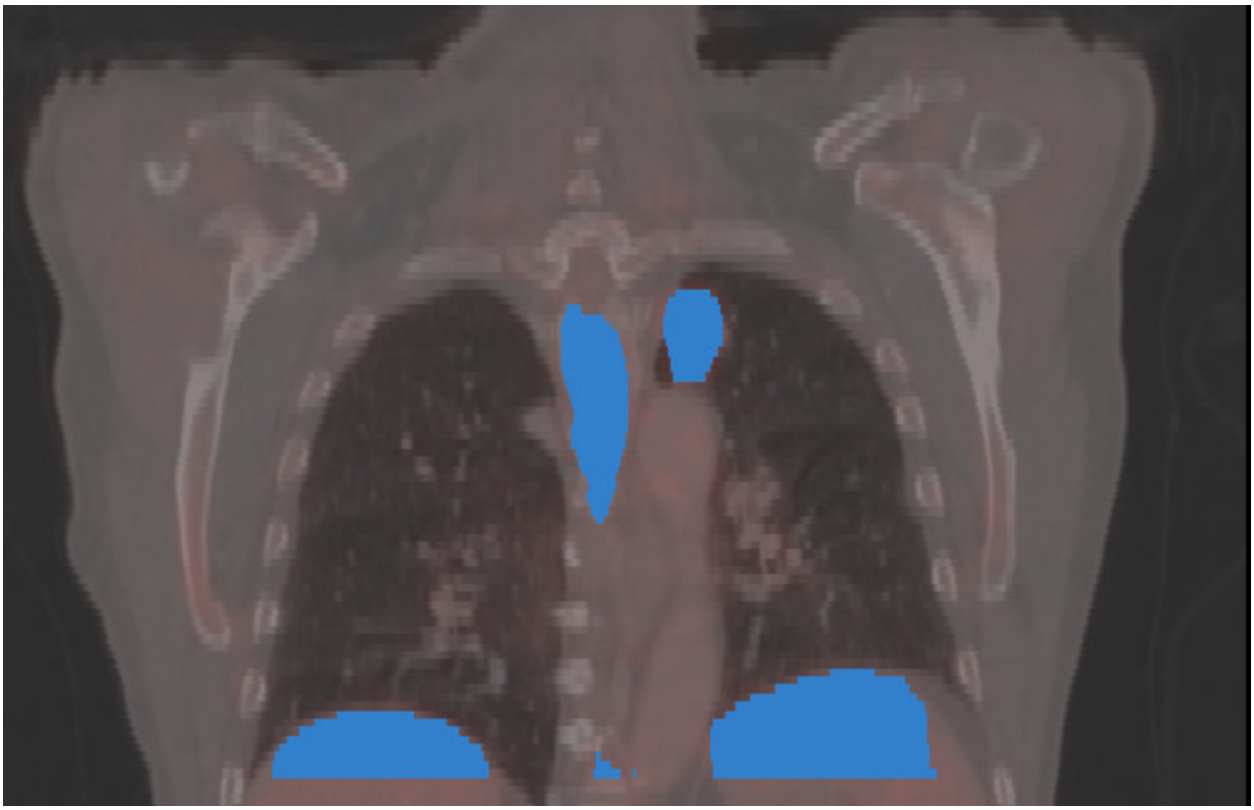


Figure 4.1: Mis-segmentation

Figure 4.2 shows the block diagram for the method. The dataset is divided into the training set and the testing set. For the training set, the processing goes as the left process: The images are segmented using the segmentation method, and all the features are extracted from the training set. Feature selection is performed to select the features that have most discriminant performance, and classifier is trained based on the result. For the testing set, the images are segmented using the same approach, and only the features that are selected in the feature selection are extracted, the trained classifier then uses the features to decide which class the testing objects fall into, and the results are given. In a supervised learning problem, we can perform an evaluation approach based the result and the ground truth classes that have already been established.

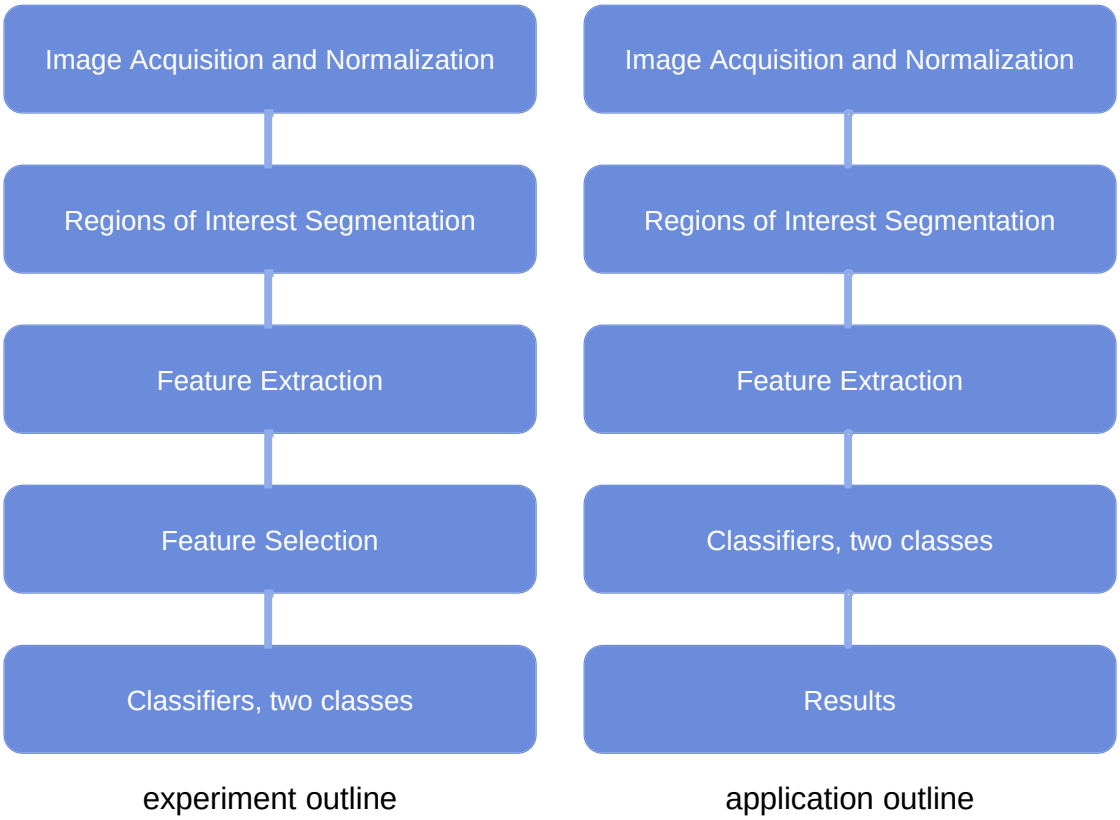


Figure 4.2: method outline

We will talk about these processes in detail in the rest of this section.

Segmentation Method Overview

In our previous work, a segmentation method was proposed based on the graph-cuts method [134, 139]. The method is utilizing the information extracted from both the PET and CT images. In short, the method seeks to solve the tumor delineation problem by modeling it as an energy minimization problem. Suppose we have a PET-CT image pair (I, I'), u is a voxel of the image I, N is the neighbor voxel sets of I, and the tumor delineation means to assign a label l_u to the voxel u, with $l_u = 1$ means the voxel is a part of tumor (foreground), and $l_u = 0$ means the voxel belongs to non-tumor (background) region. The method seeks to find the delineation label map $L_m = \{l_u | u \in I\}$ by solving

$$L_m = \arg\min_{L} \Sigma_{u \in I} E_s(l_u) + \Sigma_{(u,v) \in N} E_n(l_u, l_v), \tag{4.1}$$

where in this equation $E_s(l_u)$ is the self cost defined as

$$E_s(l_u) = \begin{cases} C_{source}(t_u), & \text{if} \quad l_u = 1 \\ C_{sink}(t_u), & \text{if} \quad l_u = 0 \end{cases} \tag{4.2}$$

and the neighbor cost $E_n(l_u, l_v)$ is defined as

$$E_n(l_u, l_v) = \begin{cases} C_{nodes}(t_u, t_v), & \text{if} \quad l_u \neq l_v \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

In this definition it can be shown that the equation 4.1 can be solved using a maximum flow algorithm by modeling each voxel in the image as the flow node, adding a source and sink nodes, modeling the self cost function C_{source} and C_{sink} as the flow capacity of the tunnel connecting the source and sink nodes to the corresponding voxel node, and modeling the neighbor cost C_{nodes} as the flow capacity of the tunnel connecting the voxel nodes.

Under this basic model of graph-cuts, our method then further encodes the information from both PET and CT images into the graph-cuts model by defining the self and neighbor cost functions

as:

$$C_{source}(t_u) = \begin{cases} |t_u - t_{source}|, & \text{if u not background seeded} \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

$$C_{sink}(u) = \begin{cases} |t_u - t_{sink}|, & \text{if u not background seeded} \\ \infty, & \text{otherwise} \end{cases} \tag{4.5}$$

$$C_{nodes}(u,v) = \begin{cases} \frac{1}{1+\lambda_1(G_{uv}+\lambda_2 \frac{t_u+t_v}{2} G'_{uv})}, \\ \text{if u and v are neighbor pixels} \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{4.6}$$

here u and v are the nodes and G_{uv} and G'_{uv} are the gradient value of normalized PET and CT between the nodes u and v correspondingly. t_u is the normalized PET image value on that node (i.e, the radiation intensity on the pixel). t_{source} and t_{sink} is a user-specified value.

For further details reader can read the paper [134].

The segmentation approach works well with seeds [29], however the goal in this paper is to find a way to automatically decide which part of segmentation outcome is tumor and which part is not, so after manually reviewing the segmentation outcome with no seeds placed, 28 tumor regions and 19 non-tumor regions are extracted using the segmentation method as labeled in the table 4.2. The number of non-tumor regions is decided by the segmentation method, on some patient the method works perfectly and gives only tumor regions, on some patient the method gives multiple non-tumor regions. On average in half cases only the tumor is segmented.

Feature Extraction

One drawback of many existing tumor delineation algorithms, including our graph-cuts based co-segmentation algorithm, is the fact that in some cases these algorithms are not able to differ-

entiate the tumor regions from other normal regions with high uptake value, therefore mistakenly label them as tumor regions in the output [121]. In order to address these issues, user interaction is introduced in many algorithms. In these algorithms, background seeds are placed in and around normal high uptake areas to prevent the algorithm from labeling these areas as tumor regions. In some applications, foreground seeds are placed in tumor region, leading the algorithm to correctly segment the tumor region, therefore increasing the robustness of the algorithm [29]. The use of user placed interactive seeds have proven to be effective, but makes the algorithm less automatic. If the computer can be trained to differentiate the tumor region from non-tumor region, it would further save the physician's effort.

In application, human experts differentiate tumor from normal high uptake regions by its textures, location and shape. Theoretically, tumor grows very different from the normal tissues and therefore has very different biological features. It is therefore reasonable to assume that the tumors may show different textures and patterns in PET and CT image from normal regions [84]. Given this assumption, tumor regions may able to be distinguished from other normal regions using a pattern recognition approach based on image texture features. Besides texture features, other features that shows good performance to distinguish tumor and non-tumor regions on similar works are also included.

A total of 60 features are extracted from PET and CT images, with 30 features extracted from each modality. The features can be classified into 3 groups listed as follows:

1. First-Order Feature

   The first-order features reflect the characters of density distribution of the given area. They are the simplest and most classic features used in image processing. The first-order features used in this work are mean, median, standard deviation, kurtosis and skewness.

2. Second-Order Features

   Second-order features reflect the characteristics of the second-order joint conditional density distribution of two voxels with the parameter of distance and direction angel. Originally proposed by Haralick, these features are calculated using spatial gray-level dependence matrices (SGLDM) and are used in many image processing applications [84, 140]. In these

applications various features based on SGLDM are used, and we include 22 of them in this work. They are: autocorrelation, contrast, correlation (2), cluster prominence, cluster shade, dissimilarity, energy (angular second moment), entropy, homogeneity, maximum probability(2), sum of squares: variance, sum average, sum variance, sum entropy, difference variance, difference entropy, inforamtion measure of correlation(2), inverse difference homo, inverse difference normalized, inverse difference moment normalized. For details of these features, readers can refer to the papers [141–143].

In all cases, SGLDM are calculated using a 7 by 7 2D window in each trans-axial slice, and at angles of 0, 45, 90, 135. The mean of the SGLDM matrices are used later to calculate the features.

3. High-Order Features

The second-order features are suggested to reflect the texture features of the image in many applications. However, in PET-CT image specifically, texture features from another source have proven to be more effective in some work [63, 84, 121]. These include features calculated using the neighborhood gray-tone-difference matrix (NGTDM) proposed in Amadasm and King's work [144]. The features are more effective, perhaps due to they are claimed to correlate more with human impressions. The NGTDM are calculated using a 7 by 7 by 3 neighborhood, since the thickness of a CT slice is about twice than the spatial resolution of a trans-axial CT image. The features used in this works are the coarseness, contrast and busyness.

Th NGTDM is defined as:

$$S(i) = \begin{cases} \sum |i - \bar{A_u}|, & \text{for } u \in N_i \text{ if } N_i \neq \varnothing \\ 0, & \text{if } N_i = \varnothing \end{cases} \tag{4.7}$$

where N_i corresponds to the set of all voxels with the gray tone i in the specified region of interests (ROI) R. $\bar{A_u}$ is the average gray-tone of the neighbor area of a voxel u in the set N_i:

$$\bar{A_u} = \frac{1}{w-1} \sum_{v \in M_u} I(v), \tag{4.8}$$

where $I(v)$ is the gray tone of voxel v in the neighbor set M_u of u, and w is the number of total voxels in the set M_u. Note that M_u is defined as the intersection set of the neighbor set and the ROI R in the boundary area, so the this number is 146 in the center area of R in our work for the 7 by 7 by 3 neighbor set, and a bit smaller at the boundary area.

It is also worth noting that the gray levels of PET and CT image are linearly scaled to roughly 256 gray levels prior to the NGTDM calculation. Details of this will be discussed in detail in the section 4.6.

We select three high-order features calculated based on the NGTDM defined above, they are coarseness, contrast and busyness.

Coarseness is defined as:

$$\text{Coarseness} = [\varepsilon + \sum_{i=0}^{G_R} S(i)/n]^{-1}, \tag{4.9}$$

where ε is a small number, G_R is the highest gray tone level presented in the ROI R, n is the number of voxels in R. Coarseness represents to the density of edge elements, the finer texture gives higher value of coarseness.

Contrast is defined as:

$$\begin{aligned}\text{Contrast} =& [\frac{1}{N_g(N_g-1)} \sum_{i=0}^{G_R} \sum_{i=0}^{G_R} P_i P_j (i-j)^2] \\ & \times [\sum_{i=0}^{G_R} S(i)/n],\end{aligned} \tag{4.10}$$

where N_g is the number of gray levels with at least one voxels in R, $P_i = N_i/n$ is the probability distribution of the gray level i. Contrast relates to the dynamic range of gray levels in the area. Abnormal tissues may have higher contrast than normal ones on PET.

Busyness is defined as

$$\text{Busyness} = \sum_{i=0}^{G_R} P_i S(i) \sum_{i=0}^{G_R} \sum_{i=0}^{G_R} |iP_i - jP_j|,$$
$$P_i \neq 0, P_j \neq 0. \tag{4.11}$$

Busyness reflects the spatial frequency of the gray level, abnormal tissues are assumed to have more rapid change than normal tissue, therefore are assumed to have higher busyness.

In order to better process features in the upcoming feature selection process, we perform zero mean normalization for all the features, i.e,

$$\text{new feature} = \frac{\text{feature} - \text{feature mean}}{\text{feature standard deviation}} \tag{4.12}$$

Classifiers

Even for the same set of features, using different classifier may result in different outcome. A classifier that fits the data leads to better performance of the classification. In order to evaluate the features used in this work in multiple aspects, three classifiers are applied and evaluated in this work. They are Quadratic Discriminant Analysis, Support Vector Machine, and K-Nearest Neighbors (K = 5).

The classifier:

1. Quadratic Discriminant Analysis (QDA):

   Quadratic Discriminant Analysis (QDA) assumes that the data of different classes has different Gaussian distributions with different means and different covariances. It then employs the likelihood ratio to determine a best separation surface between the classes. Assume that we have a set of features $\{x\}$ and known class corresponding labels y with two classes. The QDA method assumes the conditional probability density functions of classes $p(x|y=0)$ and $p(x|y=1)$ are both normally distributed with the means and covariances $(\bar{\mu_0}, \Sigma_0)$ and $(\bar{\mu_1}, \Sigma_1)$, where the means and covariances are determined by the training data. The method

then employs the Bayes optimal solution to predict the classes based on the log likelihood ratio. i.e, for a preset threshold T the observation x has a prediction label $y = 1$ if

$$(\bar{x} - \bar{\mu_0}^T)\Sigma_0^- 1(\bar{x} - \bar{\mu_0}) + \ln|\Sigma_0| -$$
$$(\bar{x} - \bar{\mu_1}^T)\Sigma_1^- 1(\bar{x} - \bar{\mu_1}) - \ln|\Sigma_1| < T,$$

The resulting separation surface is a quadratic. Quadratic Discriminant Analysis is widely used in pattern recognition. It is effective in most cases; However it is affected if there are outliers in the dataset.

2. Support Vector Machine (SVM):

   Support Vector Machine is widely used in pattern recognition problems. Generally speaking, the method seeks to find the Maximum-margin separation surface between the classes. The linear SVM models the maximum-margin separation as finding the separation hyperplane $wx - b = 0$, so that

$$(w, b) = \min||w||, \text{subject to}$$
$$y_i(wx_i - b) \geq 1 \text{ for all } i$$

   where (x_i, y_i) marks the ith element in the training data, with x_i is the feature vector, and y_i is the class label that is neither 1 or -1. The dual form of this equation is shown to be a function of support vectors, the subset of the training data that lie on the margin.

   Because of the hyperplane correlates only to the support vectors, SVM is robust against outliers, giving it advantage against QDA.

3. K-Nearest Neighbors (KNN):

   The K-Nearest Neighbors (KNN) method is a non-parametric method. The method determine the label of the object based on the most common among its k nearest neighbors. If k

= 1, the method simply labels the object to the same label of the nearest data in the training set. Nearest means the minimum distance defined by the distance metrics, most commonly Euclidean distance.

The KNN classifier is easy to implement and takes very little resource to train. But it is not robust against outliner in some cases and gives poor performance when the feature number is high.

Feature Selection

In feature selection, we seek to select features that are most correlated to the classes. Features with most discriminatory power are selected and the redundant features are removed. The reason for feature selection is to save future computational resources by using the least number of features. The robustness of certain classifiers can also be improved when the feature number is less. Many feature selection algorithms are proposed [145]. However, some methods, like the genetic algorithm approach, assume the features have Gaussian distributions of data within the classes. Using these methods can have pitfalls with small dataset size [146]. Since the data size is relatively small in our work (17+11 cases), we decides to use the sequential forward selection combined with three classifiers. Since sequential forward selection is a sub-optimal selection algorithm, it is normal to see performance up and down during the selection process. For each classifier we perform the training and testing processing using leave-one-out technique, and select the features combination with highest accuracy. The accuracy is defined as the ratio between the data of correct classification and all the training data, i.e:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{Correct Classifications}}{\text{ALL Classifications}} \\ &= \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False postive} + \text{False Negative}} \end{aligned} \tag{4.13}$$

Needless to say, Accuracy = 1 means the sensitivity and specificity of the classification system both equal to one. After the features are selected, results are evaluated using the evaluation metrics described below.

Evaluation Metrics

Evaluation metrics measures the performance of the system. Five metrics are used in this work. They are accuracy, the false acceptance rate (FAR) and false rejection rate (FRR), the receiver operating characteristic (ROC) curve, the score histogram and the feature scatter plots.

Accuracy

Accuracy is already defined in equation 4.13. Strange as it may appear, using the same metrics for feature selection and evaluation is not uncommon [132]. The accuracy in this work is calculated using different classifiers, and can be seen as a trade-off between FAR and FRR, as will be shown in the following part.

FAR and FRR

Before talking about FAR and FRR, it is best to introduce the concept of false acceptance and false rejection. False acceptance refers to the cases that a object is accepted by recognition system when it should not be. False rejection on the other hand, refers to the cases that a object is rejected when it should be accepted.

With the definitions established, the FAR is then defined as:

$$\text{FAR} = \frac{\text{total number of false acceptances}}{\text{total number of acceptances}}, \tag{4.14}$$

while the FRR is defined as:

$$\text{FRR} = \frac{\text{total number of false rejections}}{\text{total number of rejections}}. \tag{4.15}$$

In a recognition system, different recognition strategies lead to different FAR and FRR pairs, while classifiers seek for a balance between FAR and FRR using their unique techniques, it is often beneficial to plot all the pairs into one graph, which is the ROC curve.

ROC Curve

By definition, ROC curve is a plot of FAR versus FRR. It gives more detailed evaluation of system in terms of robust and reliability, and allows better examination of system. It is very complicated to draw ROC curve for all possible feature combinations, so in this work we only draw ROC curve for the top single feature and for the best feature combinations selected by the three classifiers.

Score Histogram

The score histogram is made by plotting the frequency of scores calculated from a predefined score function for non-matches and matches over the match ranges. It shows the distributions of classes under a certain type of classifier strategy. In this work we will show the score histogram for the top single feature as well as the best feature combinations selected by classifiers. The score function in the later case is defined as the distance between the object's features and the thresholding hyperplane defined by classifiers.

Feature scatter plots

Feature scatter plot is much like score histogram, but unlike calculating score using predefined functions, it simply plot the feature values of objects. The dimension of the plot depends on the number of features. Same as histogram it shows the general distribution of objects on selected feature sets. However it is more direct. In this work we only shows the plot for best feature combination with least feature numbers.

Results

In this section we will show the results from the feature selection, and evaluate them using the metrics mentioned above.

A total number of 60 features are extracted from the segmentation result. We perform a sequential forward selection based on the accuracy given by the classifiers. Since we use the classification accuracy as the evaluation metrics for the feature selection, the feature selection result of one clas-

sifier includes the method performance on the classifier. Because of that, in this section we show the result of feature selection of each classifier and draw conclusion of the method performance based on that.

As mentioned in the data description section, 28 tumor regions and 19 non-tumor regions are extracted by the segmentation method from 28 patient cases. By using leave-one-out technique, it means a training set with $46 \times 47 = 2162$ samples and a testing set with 47 samples. The results are as follows:

Direct impact

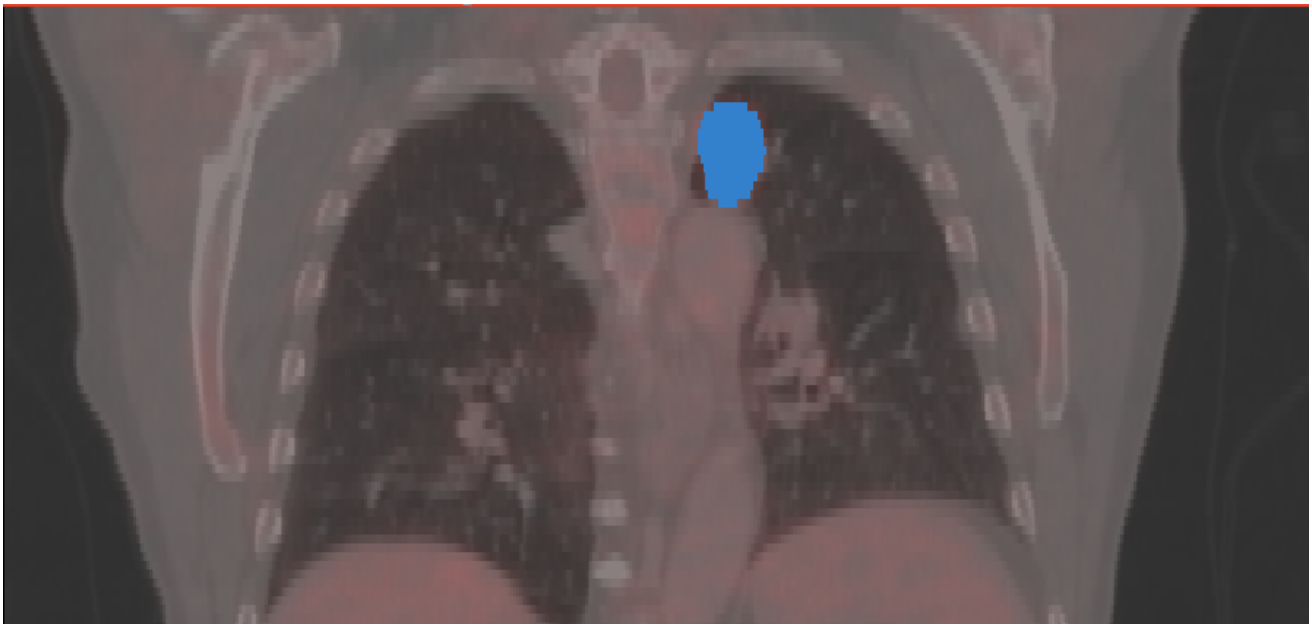
For first impression, Here is the algorithm performance for the previous mis-segmented case (same case at another coronal slice), the delineation method finds regions at sub-figure a, they are part of liver, stomach, spine and tumor. After the localization method, tumor region is recognized and the normal tissues from liver, stomach and spine is ruled out.

Feature Selection Results Using Classifiers

Results Based on QDA Classifier

The figure 4.4 shows the classification accuracy of QDA classifier using the sequential forward selection approach under leave-one-out technique.

(a) Before



(b) After

Figure 4.3: Algorithm Performance First Impact

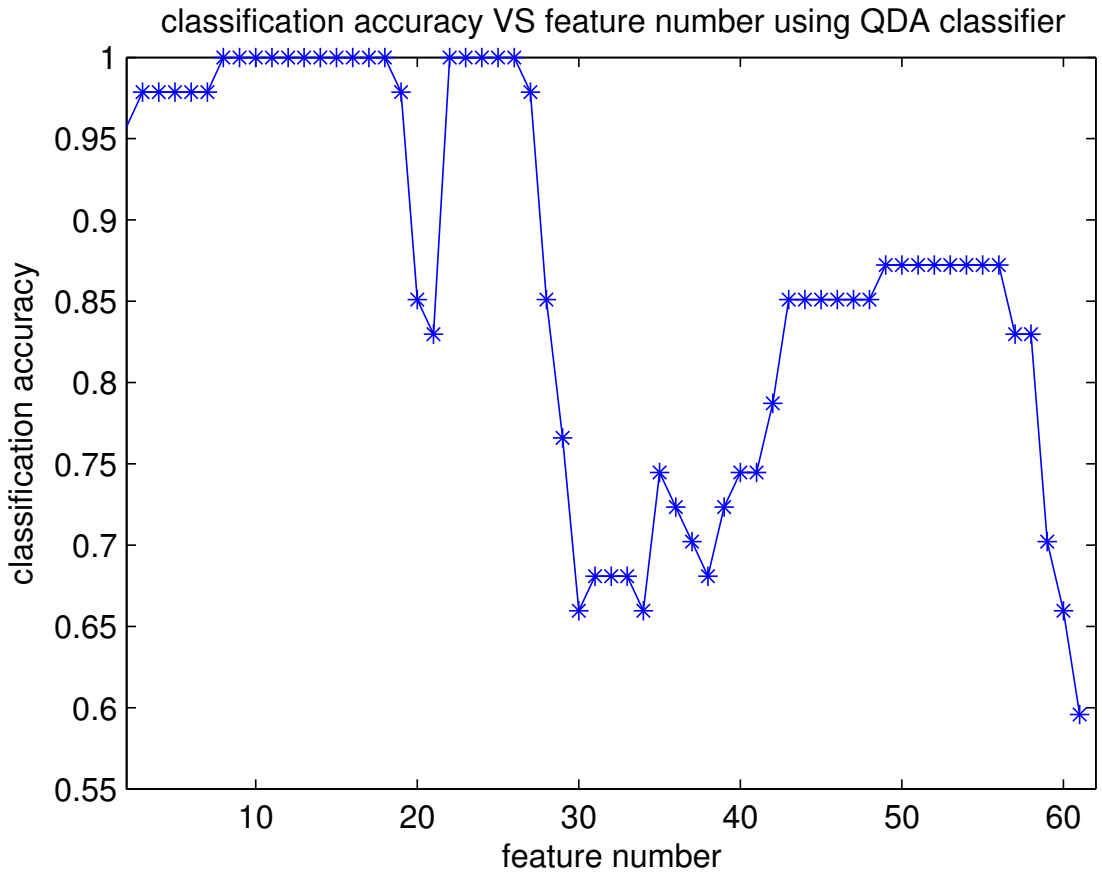


Figure 4.4: the classification accuracy of QDA classifier

In Fig. 4.4, we can see that the classifier achieve 100% accuracy (meaning that the algorithm perfectly separate tumor from non-tumor) at feature number = 7, corresponding to a feature set of PET NGTDM busyness, CT median, PET NGTDM contrast, PET mean, PET standard deviation, PET SGLDM entropy and PET SGLDM autocorrelation.

Results Based on SVM Classifier

The figure 4.5 shows the classification accuracy of SVM classifier using the sequential forward selection approach under leave-one-out technique.

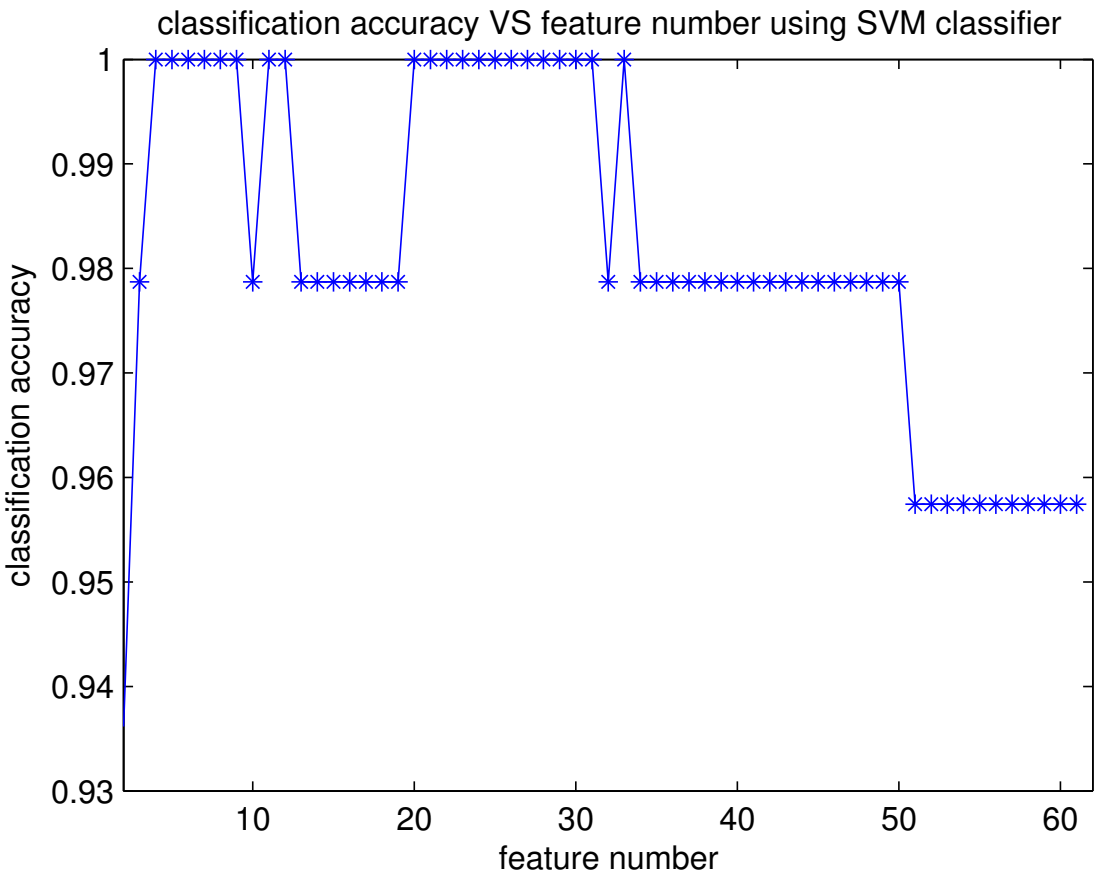


Figure 4.5: the classification accuracy of SVM classifier

In Fig. 4.5, we can see that the classifier achieve 100% accuracy at feature number = 3, corresponding to a feature set of PET NGTDM busyness, CT median and CT standard deviation. Note that SVM shows significant improvement from QDA, probably because the dynamic range of the features is still relatively large after normalization.

Results Based on KNN Classifier

The figure 4.6 shows the classification accuracy of KNN classifier using the sequential forward selection approach under leave-one-out technique.

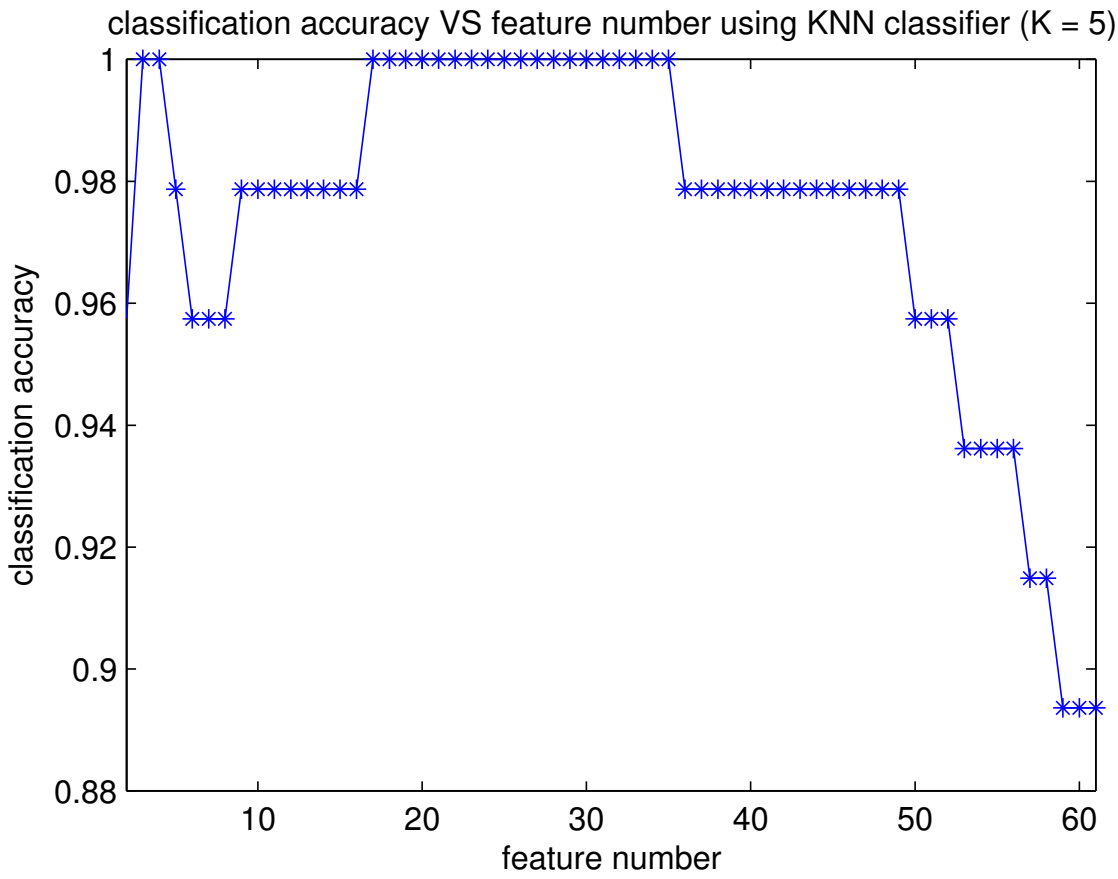


Figure 4.6: the classification accuracy of KNN classifier

In Fig. 4.6, we can see that the classifier achieve 100% accuracy at feature number = 2, corresponding to a feature set of PET NGTDM busyness and CT median. Note the performance goes unstable as the feature number increases.

Singe Feature Analysis

Although the classifiers gives the optimal feature sets leading to better results, using a single feature leads to a faster system and can be a potential way of fast application in the future. So it is beneficial for future assessment.

The table below lists the top features selected by classifiers. All three classifiers gives same top features. They are listed below with their accuracy numbers:

From the table only one feature gives an accuracy over 0.9, meaning less than 4 mismatch

Table 4.1: top features with 0.8 accuracy or higher and their accuracy

Feature name	Accuracy
PET NGTDM busyness	0.9362
PET NGTDM contrast	0.8511
CT SGLDM variance Sum	0.8298
PET NGTDM coarseness	0.8085
PET standard derivation	0.8085
CT median	0.8085

cases. In order to further analyze the system performance on signal feature, we employs the ROC curve as well as score histogram on the PET busyness.

The ROC curve and score histogram is as follows:

We can see from the figures that, although only using one feature, there is till good segmentation between tumor regions and non tumor regions on PET busyness.

Multiple Features Analysis

The results of the classifiers suggest different feature combinations. Since KNN classifier is known to be not robust as more features are included and the number of optimal features set given by LDA is a bit large, we use the results from SVM for analysis.

Below is the ROC curve 4.9 for optimal sets of SVM, those are PET NGTDM busyness, CT median and CT standard deviation, and score histogram 4.10 based on score function of SVM. And since it is more convinent to show 2 dimension scatter plot, the feature scatter plot in figure 4.11 only shows PET NGTDM busyness and CT median the discrimination to demonstrate performance of the combination.

From Fig. 4.11 it is clear that using the feature combination, the dataset is already a separable dataset. Such result suggests that it is realistic to separate tumor region from non-tumor region at least in the case of chest PET-CT image in the lung tumor cases.

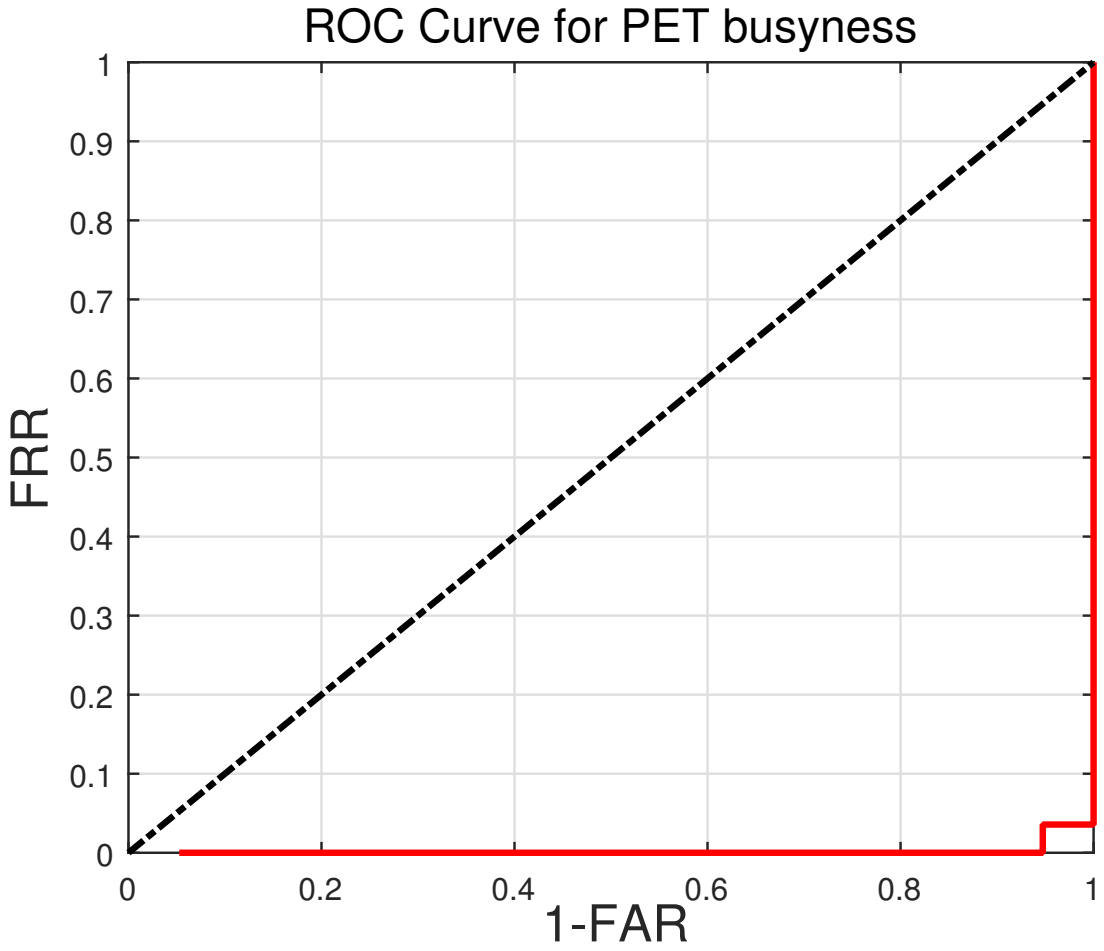


Figure 4.7: ROC curve for PET busyness, it is calculated that the area under curve (AUC) is -0.49812

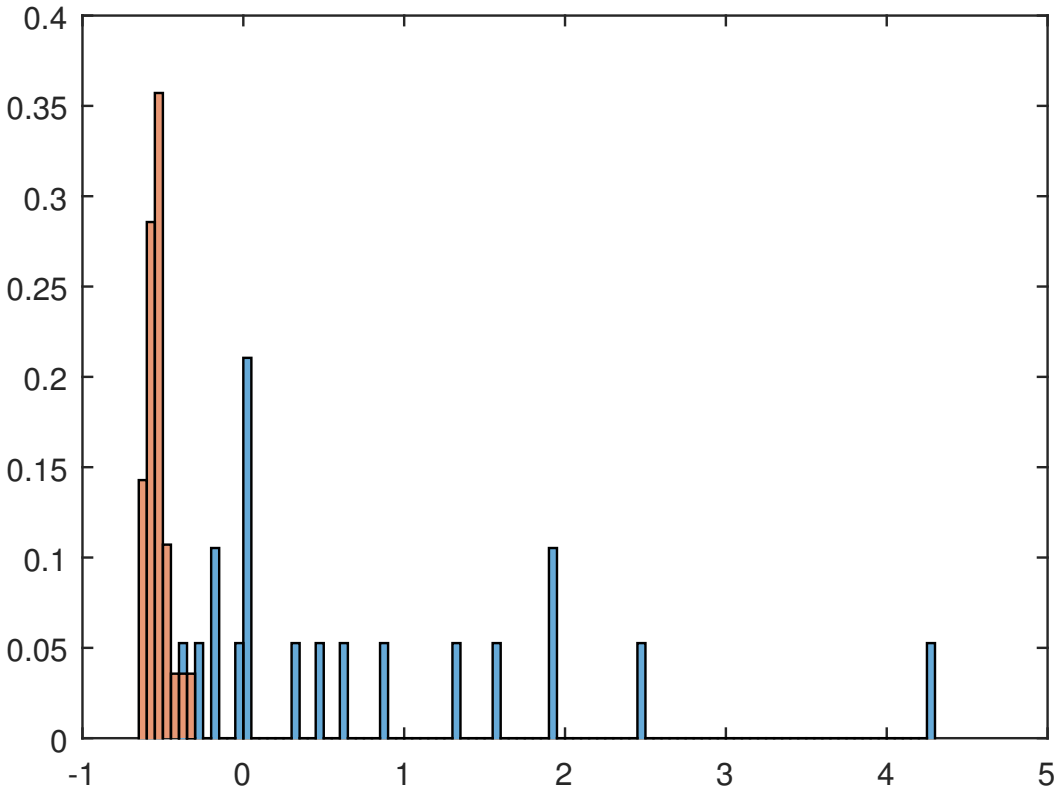


Figure 4.8: Score histogram of PET busyness, x axis is the feature value of the feature and y axis is the distribution percentage, red bars represents non-tumor regions and blue bars represents tumor regions

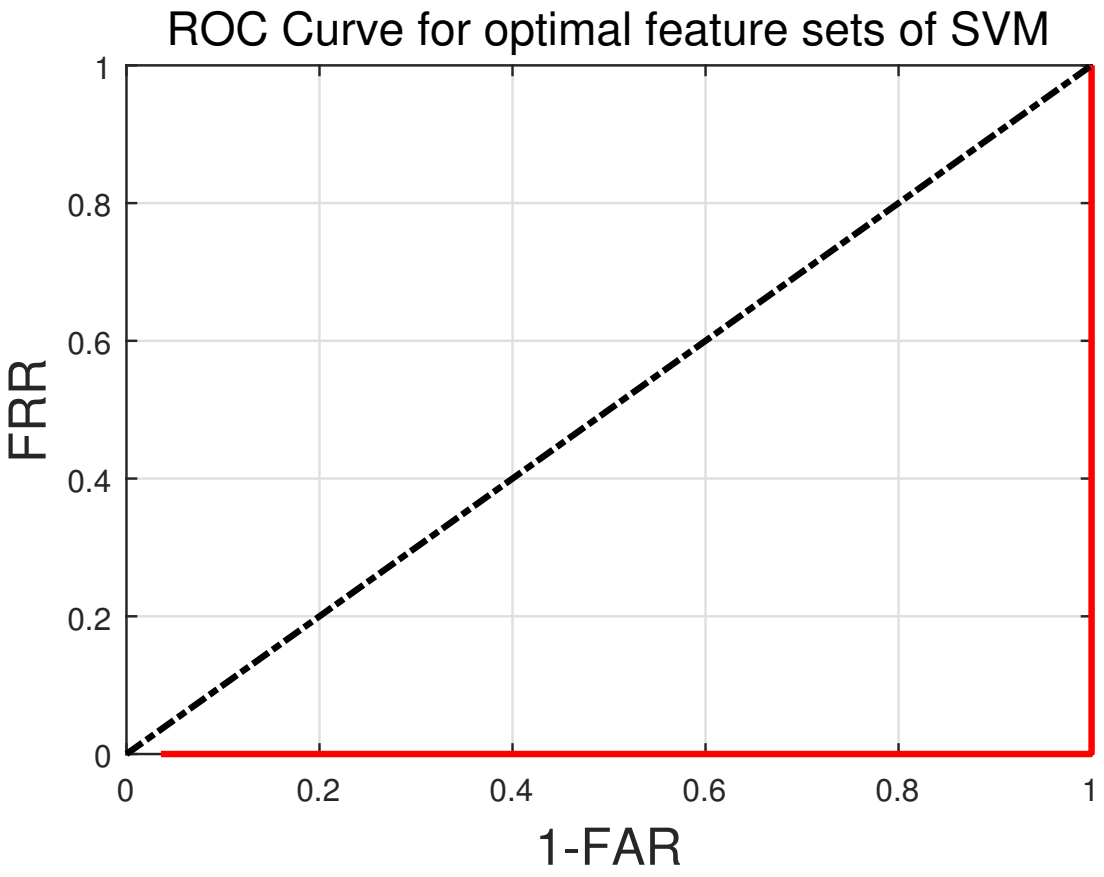


Figure 4.9: ROC curve for SVM optimal feature sets, the area under curve (AUC) is -0.5

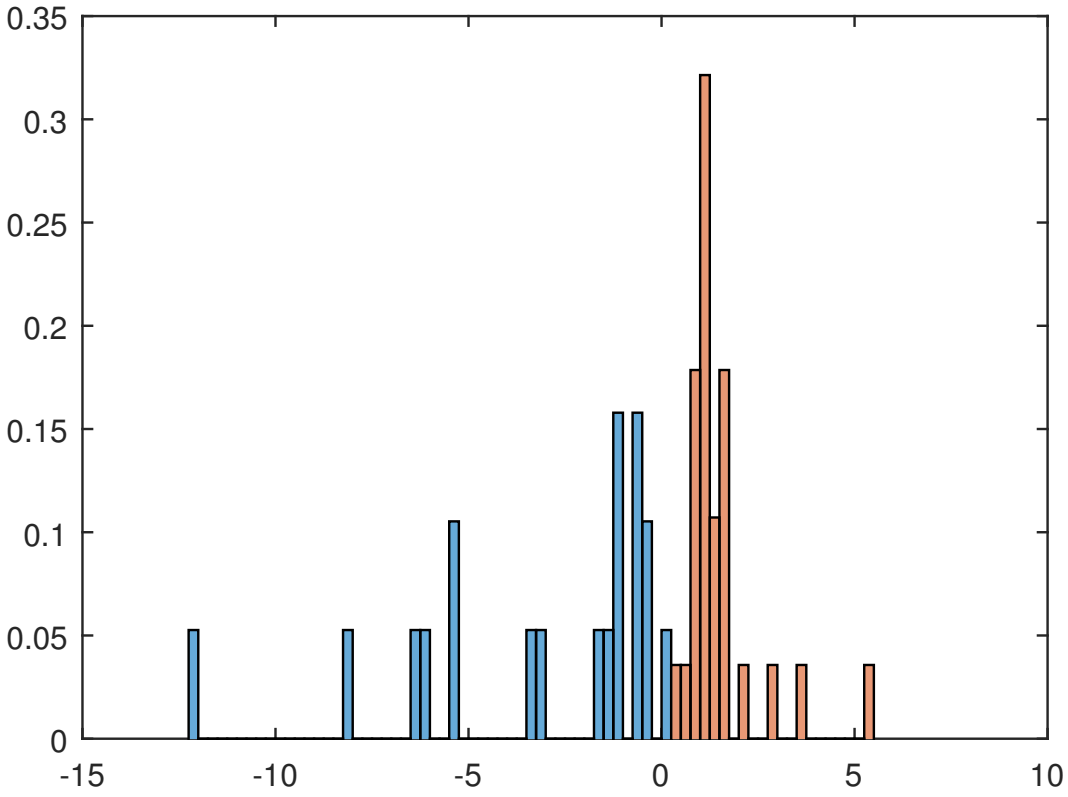


Figure 4.10: Score histogram, x axis is the score given by SVM, y axis is distribution precentage

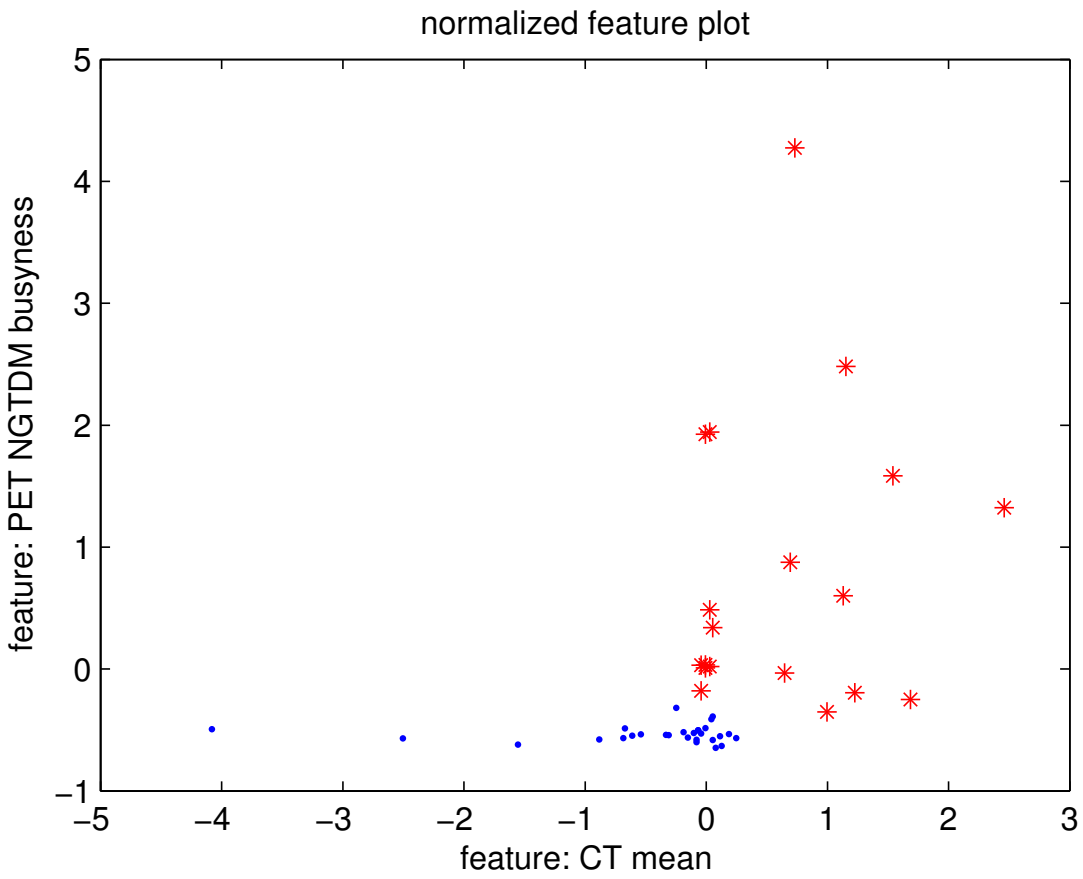


Figure 4.11: the feature scatter plot using features PET NGTDM busyness and CT median, the blue dot represents the non-tumor regions and the red stars represents the tumor regions

Discussion

In this section we generalize the results as well as address some problems occurs when performing the result from the following prospectives: dataset, feature extraction and feature selection.

Dataset Issues

We start the discussion with the size of the data set. In this work we evaluate our work based on 28 tumor regions and 19 non-tumor regions extracted from 28 patients that has both PET and CT record from two public datasets TCGA-LUSC and TCGA-LUAD from TCIA website. Regardless of the clinical perspective, This is indeed a small size for a dataset from pattern recognition point of view, so instead of classically dividing the dataset into the training set and the testing set, we use leave-one-out technique. Also it is worth noting that both datasets are lung cancer, so the

conclusions in this work may not apply to tumors belongs to other region.

Feature Extraction Issues

Another issue we would like to address is the issue of NGTDM scaling. Originally NGTDM is performed on the gray image with 256 density level, From the NGTDM's definition we know that the value of NGTDM features highly depends on the number of gray levels of the image. In our case, the dynamic range of PET image is around 1e5 (Bq/ml, in original PET image) or around 30 (in standard uptake value PET image), the dynamic range of CT image is around 2000 (HU, in original CT image). In this work, the PET and CT image are roughly scaled into the same dynamic range of the gray image by using a constant scaler (100 for original PET image and 10 for original CT image) instead of restrictively scale the image to 256 gray level based on their local maximum and minimum. For a constant value may better preserve the similarity of features of data from the same classes.

Feature Selection and Classifier Issues

The first thing should be mentioned is, although up to 23 second-order features are extracted in this work, they are hardly selected by any classifiers in early order. Most likely this suggests that the second-order features show less discrimination performance than the first-order and high-order ones. However, it must be pointed out that this absent may also be a result from the sequential forward selection, since we put the first and high-order features in the front of our selection sequence.

Secondly, from the results, it is clear that QDA is not as suitable to the features as the other two classifiers. However, it is hard to determine the better one between SVM and KNN. Surely KNN achieves 100% accuracy with only 2 features, which is less than the SVM's 3, and will surly save computational resources in application phase, speed up the processing time for the system. However, less feature number often means less stableness when the two-classes system expand into multi-classes system in the future, and that may become a problem when we expand this method into the other type of tumor or expand the PET-CT image type from chest PET-CT to whole-body PET-CT. On the other hand, choosing SVM will not increase the system cost dramatically, and the SVM is known to be stable against outliers in training phase regardless of the feature

number [145]. Also as a plus, SVM is known to give highest inter-class discriminant distance, so SVM is preferred. However, KNN may still be used under the need real-time determination.

Lastly, readers may already notice the bouncing performance of accuracy for the classifiers at the figure 4.4, figure 4.5 and figure 4.6. In theory such bouncing should not occur if we perform an exhaustive search. But since we perform the sequential forward selection, it is normal to have these bouncing performance [146]. In the future we hope to refine this performance by using more sophisticated feature selection technique and metrics.

Feature Analysis

From the result, we learn that the system already achieves 0.93 accuracy using single feature, and would further achieves 100% accuracy when combined with other features. Depending on the classifier used, the system achieves such accuracy at 2, 3 or 7 features and will maintain such performance. The analysis from classifier aspect is already covered in last subsection. From feature's perspective, there are still a few interesting fact.

Firstly, from ROC curve perspective, the figure 4.7 shows good trade-off between FAR and FRR, and also indicates a closer distribution between two classes, as can be seen from score histogram and scatter plot. On the other hand, 100% accuracy means both FAR and FRR can achieve 0 at the same time, so it is not surprising to see an ROC curve in figure 4.9.

Secondly, From the kind of feature itself, busyness weights much more than other features, which may indicates that tumor tissues have more complicated and busy texture character than normal tissues, one can see that PET mean or median is not among the most relevant features. It is decided by the structure of this work. We apply a segmentation method and select tumor and non-tumor regions within the high responding regions. Since segmentation mainly relies on PET intensity to identify tumor, it leads to high PET intensity (and therefore, close PET mean and PET median) in both tumor and non tumor regions. Therefore it is not strange that features that characterize to PET intensity does not give good separation.

Thirdly, the score histogram and scatter plot shows if the classes are separated on the given feature or features and how well such separation can be. We can see that although accuracy of

segmentation is high in this work, the scores of both classes have large variations, and the gap between them is narrow. It implies the variety nature of tissues involved in this work. However, despite the overlapping trend on the margin of classes, the main part of classes remains quite significant from each other.

Future Work and Conclusion

For the future, we hope to expand the method from localizing lung tumor in chest PET-CT image to localizing lung tumor in whole body image, and further to localizing any type of tumor in whole body image. This is the blueprint for this work and needs time and effort in order to get it done. For now, the method itself can be refined. For instance, the sequential forward selection can be replaced by more sophisticated technique, and the selection metrics can be further improved as well. Instead of using just accuracy, we can add more metrics to decide the feature to pick when they gives the same accuracy. For instance, such metrics can be the maximum margin value when using the SVM classifier.

In conclusion, in this work, a tumor localization framework is proposed based on a former developed tumor delineation system. The proposed framework is further evaluated. From them one can conclude that the lung tumor can be differentiate from other non tumor regions in the chest area by using the features such as PET NGTDM busyness and CT median with appropriate classifiers such as SVM, KNN and QDA. Some issues of dataset, feature extraction and feature selection are then addressed and discussed.

Patient ID	Tumor description	PET voxel size (mm)	CT voxel size (mm)	note
2695	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
2696	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
2711	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
2721	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
2722	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
2723	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
2724	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
4079	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
4080	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
5236	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
5239	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
5240	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
6867	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
7340	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
8064	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
A5IX	Lung Squamous Cell Carcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
5045	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
5066	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
5072	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
5941	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
6591	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
6594	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
6595	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions with 3 non-tumor regions
Z050	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions
Z053	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	tumor region only
Z060	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	2 correct tumor regions
Z062	Lung Adenocarcinoma	4.6875 by 4.6875 by 3.2701	0.9765 by .09765 by 3.2701	mixed regions

Table 4.2: Patients data details

5. Conclusion

Contributions

In this thesis, the problem of avoiding pre and post manual work in tumor segmentation using PET-CT image is examined. Specifically, the focus is on how to automatically differentiate high uptake regions due to a tumor from other non-tumor high uptake regions like kidney, liver and heart [40, 43].

To this end, this thesis focuses on building an fully automated segmentation method, by adding a tumor recognition procedure after a co-segmentation procedure. In the co-segmentation procedure, by incorporating CT boundary information into PET segmentation using a graph-cuts based method, the segmentation result is more accurate when compared with result based on PET solely. In the recognition procedure, features are extracted from the segmented regions resulting from the previous procedure with no user defined ROI. Three classifiers are used for feature set selection and evaluation.

With respect to the obtained results, when the methods are tested on non-small cell lung cancer database, 100% accuracy is reported for recognition procedure with different classifiers and feature combinations. Therefore the combined segmentation result is the same as that reported in chapter 3, that is, the segmentation evaluation result with pre-defined ROI: roughly 92% DSC and 2.5mm HD are obtained.

In particular, the main findings of this work are:

- In chapter 3, CT boundary information is used under a classical graph-cuts structure, with a novel neighbor cost function. Compared to other similar works, it simplifies the graph,

therefore requiring less computational time and resources. Specifically, when compared to thresholding methods, state-of-art learning method (AP), and other co-segmentation methods, the proposed method shows better performance in terms of DSC and HD.

- In chapter 4, a tumor detection method is proposed using a pattern recognition framework. First order, second order, and higher order features are examined under three classifiers. The result shows that, when using features from NGTDM, a properly designed feature set with very few features can clearly separate between normal tissues and tumor nodes.

Future Work

The proposed framework can be extended from the following perspectives.

- Examined on other databases: the present work is only examined on two combined non-small cell lung cancer databases. However, because of tumor metastasis, even when restricted to chest images of lung tumor diagnosis, it is highly possible that different types of tumor can occur (although there won't be too many cases). Therefore, it is beneficial to further examine the capability of this method with multiple types of tumor.

- Examined for Benign and Malignant lung nodules: as a tumor metastasizes, very few benign nodules appear in the datasets we select. However this may be possible in other occasions. Recognition methods for this separation have been developed based on PET-CT feature subjective score as reviewed in literature review, and the idea of automatic feature extraction on this issue seems quite promising. For this endeavor, a decision tree approach could be used as a replacement for classifiers.

- Merge Localization and Segmentation Procedure: theortically, segmentation and localization can both be viewed as classifying objects into different classes. In our case, the class is identical: tumor and non tumor . Therefore, if in segmentation process a voxel can be directly recognized as tumor region instead of just high uptake region, the procedure can be greatly reduced.

References

[1] T. C. G. Atlas, "**TCGA** research network." http://cancergenome.nih.gov/.

[2] B. W. Stewart and C. Wild, *World cancer report 2014*. World Health Organization, 2014.

[3] C. K. Bomford and I. H. Kunkler, *Walter and Miller's Textbook of Radiotherapy: Radiation Physics, Therapy, and Oncology*. Churchill Livingstone London, 1993.

[4] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A Review on Segmentation of Positron Emission Tomography Images," *Computers in biology and medicine*, vol. 50, pp. 76–96, 2014.

[5] M. F. McNitt-Gray, "AAPM/RSNA Physics Tutorial for Residents: Topics in CT: Radiation Dose in CT 1," *Radiographics*, vol. 22, no. 6, pp. 1541–1553, 2002.

[6] S. Basu, H. Zaidi, S. Holm, and A. Alavi, "Quantitative techniques in PET-CT imaging," *Current Medical Imaging Reviews*, vol. 7, no. 3, pp. 216–233, 2011.

[7] E. Miele, G. P. Spinelli, F. Tomao, A. Zullo, F. De Marinis, G. Pasciuti, L. Rossi, F. Zoratto, and S. Tomao, "Positron Emission Tomography (PET) radiotracers in oncology–utility of 18F-Fluoro-deoxy-glucose (FDG)-PET in the management of patients with non-small-cell lung cancer (NSCLC)," *Journal of Experimental & Clinical Cancer Research*, vol. 27, no. 1, p. 1, 2008.

[8] G. Kramer-Marek, M. Bernardo, D. O. Kiesewetter, U. Bagci, M. Kuban, A. Omer, R. Zielinski, J. Seidel, P. Choyke, and J. Capala, "PET of HER2-positive pulmonary metastases with 18F-ZHER2: 342 affibody in a murine model of breast cancer: comparison with 18F-FDG," *Journal of Nuclear Medicine*, vol. 53, no. 6, pp. 939–946, 2012.

[9] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors," *Journal of nuclear medicine*, vol. 50, no. Suppl 1, pp. 122S–150S, 2009.

[10] V. J. Lowe, J. M. Hoffman, D. M. DeLong, E. F. Patz, R. E. Coleman, *et al.*, "Semiquantitative and visual analysis of FDG-PET images in pulmonary abnormalities," *Journal of nuclear medicine*, vol. 35, no. 11, pp. 1771–1776, 1994.

[11] S. Basu, T. C. Kwee, S. Surti, E. A. Akin, D. Yoo, and A. Alavi, "Fundamentals of PET and PET/CT imaging," *Annals of the New York Academy of Sciences*, vol. 1228, no. 1, pp. 1–18, 2011.

[12] M. Kelly, "SUV: advancing comparability and accuracy," *White Paper Siemens*, 2009.

[13] U. Bagci, X. Chen, and J. K. Udupa, "Hierarchical scale-based multiobject recognition of 3-D anatomical structures," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 3, pp. 777–789, 2012.

[14] P. K. Saha and J. K. Udupa, "Scale-based diffusive image filtering preserving boundary sharpness and fine structures," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 11, pp. 1140–1155, 2001.

[15] R. Boellaard, N. C. Krak, O. S. Hoekstra, and A. A. Lammertsma, "Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study," *Journal of Nuclear Medicine*, vol. 45, no. 9, pp. 1519–1527, 2004.

[16] M. Werner-Wasik, A. D. Nelson, W. Choi, Y. Arai, P. F. Faulhaber, P. Kang, F. D. Almeida, Y. Xiao, N. Ohri, K. D. Brockway, *et al.*, "What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom," *International Journal of Radiation Oncology* Biology* Physics*, vol. 82, no. 3, pp. 1164–1171, 2012.

[17] U. Bagci, B. Foster, K. Miller-Jaster, B. Luna, B. Dey, W. R. Bishai, C. B. Jonsson, S. Jain, and D. J. Mollura, "A computational pipeline for quantification of pulmonary infections in small animal models using serial PET-CT imaging," *EJNMMI Res*, vol. 3, no. 1, p. 55, 2013.

[18] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 7, pp. 903–921, 2004.

[19] D. P. Dogra, A. K. Majumdar, and S. Sural, "Evaluation of segmentation techniques using region area and boundary matching information," *Journal of Visual Communication and Image Representation*, vol. 23, no. 1, pp. 150–160, 2012.

[20] J. A. Lee, "Segmentation of positron emission tomography images: some recommendations for target delineation in radiation oncology," *Radiotherapy and oncology*, vol. 96, no. 3, pp. 302–307, 2010.

[21] U. Bagci, J. K. Udupa, J. Yao, and D. J. Mollura, "Co-segmentation of functional and anatomical images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 459–467, Springer, 2012.

[22] U. Bagci, J. Yao, K. Miller-Jaster, X. Chen, and D. J. Mollura, "Predicting future morphological changes of lesions from radiotracer uptake in 18F-FDG-PET images," *PLoS One*, vol. 8, no. 2, p. e57105, 2013.

[23] U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D. 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*, vol. 17, no. 8, pp. 929–945, 2013.

[24] C. Lin, E. Itti, C. Haioun, Y. Petegnief, A. Luciani, J. Dupuis, G. Paone, J.-N. Talbot, A. Rahmouni, and M. Meignan, "Early 18F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis," *Journal of Nuclear Medicine*, vol. 48, no. 10, pp. 1626–1632, 2007.

[25] D. Hellwig, T. P. Graeter, D. Ukena, A. Groeschel, G. W. Sybrecht, H.-J. Schaefers, and C.-M. Kirsch, "18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense?," *Journal of Nuclear Medicine*, vol. 48, no. 11, pp. 1761–1766, 2007.

[26] B. Shah, N. Srivastava, A. Hirsch, G. Mercier, and R. Subramaniam, "Intra-reader reliability of FDG PET volumetric tumor parameters: effects of primary tumor size and segmentation methods," *Annals of nuclear medicine*, vol. 26, no. 9, pp. 707–714, 2012.

[27] J. L. Fox, R. Rengan, W. O¡¯Meara, E. Yorke, Y. Erdi, S. Nehmeh, S. A. Leibel, and K. E. Rosenzweig, "Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non–small-cell lung cancer?," *International Journal of Radiation Oncology* Biology* Physics*, vol. 62, no. 1, pp. 70–75, 2005.

[28] C. Fiorino, M. Reni, A. Bolognesi, G. M. Cattaneo, and R. Calandrino, "Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning," *Radiotherapy and oncology*, vol. 47, no. 3, pp. 285–292, 1998.

[29] M. Hatt, C. C. Le Rest, N. Albarghach, O. Pradier, and D. Visvikis, "PET functional volume delineation: a robustness and repeatability study," *European journal of nuclear medicine and molecular imaging*, vol. 38, no. 4, pp. 663–672, 2011.

[30] J. J. Erasmus, G. W. Gladish, L. Broemeling, B. S. Sabloff, M. T. Truong, R. S. Herbst, and R. F. Munden, "Interobserver and intraobserver variability in measurement of non–small-cell carcinoma lung lesions: Implications for assessment of tumor response," *Journal of Clinical Oncology*, vol. 21, no. 13, pp. 2574–2582, 2003.

[31] U. Bagc, J. Yao, J. Caban, E. Turkbey, O. Aras, and D. J. Mollura, "A graph-theoretic approach for segmentation of PET images," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 8479–8482, IEEE, 2011.

[32] P. K. Saha and J. K. Udupa, "Optimum image thresholding via class uncertainty and region homogeneity," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 7, pp. 689–706, 2001.

[33] D. A. Schinagl, W. V. Vogel, A. L. Hoffmann, J. A. Van Dalen, W. J. Oyen, and J. H. Kaanders, "Comparison of Five Segmentation Tools for 18 F-Fluoro-Deoxy-Glucose–Positron Emission Tomography–Based Target Volume Definition in Head and Neck Cancer," *International Journal of Radiation Oncology* Biology* Physics*, vol. 69, no. 4, pp. 1282–1289, 2007.

[34] H. Vees, S. Senthamizhchelvan, R. Miralbell, D. C. Weber, O. Ratib, and H. Zaidi, "Assessment of various strategies for 18F-FET PET-guided delineation of target volumes in high-grade glioma patients," *European journal of nuclear medicine and molecular imaging*, vol. 36, no. 2, pp. 182–193, 2009.

[35] U. Nestle, S. Kremp, A. Schaefer-Schuler, C. Sebastian-Welsch, D. Hellwig, C. Rübe, and C.-M. 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*, vol. 46, no. 8, pp. 1342–1348, 2005.

[36] Y. E. Erdi, O. Mawlawi, S. M. Larson, M. Imbriaco, H. Yeung, R. Finn, and J. L. Humm, "Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding," *Cancer*, vol. 80, no. S12, pp. 2505–2509, 1997.

[37] C. B. Caldwell, K. Mah, M. Skinner, and C. 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*, vol. 55, no. 5, pp. 1381–1393, 2003.

[38] C. Nagel, G. Bosmans, A. Dekker, M. Öllers, D. De Ruysscher, P. Lambin, A. Minken, N. Lang, and K. Schäfers, "Phased attenuation correction in respiration correlated computed tomography/positron emitted tomography," *Medical physics*, vol. 33, no. 6, pp. 1840–1847, 2006.

[39] A. C. Paulino, M. Koshy, R. Howell, D. Schuster, and L. W. Davis, "Comparison of CT-and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 61, no. 5, pp. 1385–1392, 2005.

[40] E. Deniaud-Alexandre, E. Touboul, D. Lerouge, D. Grahek, J.-N. Foulquier, Y. Petegnief, B. Grès, H. El Balaa, K. Keraudy, K. Kerrou, *et al.*, "Impact of computed tomography and 18 F-deoxyglucose coincidence detection emission tomography image fusion for optimization of conformal radiotherapy in non–small-cell lung cancer," *International Journal of Radiation Oncology* Biology* Physics*, vol. 63, no. 5, pp. 1432–1441, 2005.

[41] K. J. Biehl, F.-M. Kong, F. Dehdashti, J.-Y. Jin, S. Mutic, I. El Naqa, B. A. Siegel, and J. 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*, vol. 47, no. 11, pp. 1808–1812, 2006.

[42] R. Hong, J. Halama, D. Bova, A. Sethi, and B. Emami, "Correlation of PET standard uptake value and CT window-level thresholds for target delineation in CT-based radiation treatment planning," *International Journal of Radiation Oncology* Biology* Physics*, vol. 67, no. 3, pp. 720–726, 2007.

[43] E. Day, J. Betler, D. Parda, B. Reitz, A. Kirichenko, S. Mohammadi, and M. Miften, "A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients," *Medical physics*, vol. 36, no. 10, pp. 4349–4358, 2009.

[44] P. Tylski, S. Stute, N. Grotus, K. Doyeux, S. Hapdey, I. Gardin, B. Vanderlinden, and I. Buvat, "Comparative assessment of methods for estimating tumor volume and standardized uptake value in 18F-FDG PET," *Journal of Nuclear Medicine*, vol. 51, no. 2, pp. 268–276, 2010.

[45] A. Van Baardwijk, G. Bosmans, L. Boersma, J. Buijsen, S. Wanders, M. Hochstenbag, R.-J. Van Suylen, A. Dekker, C. Dehing-Oberije, R. Houben, *et al.*, "Pet-ct–based auto-contouring in non–small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes," *International Journal of Radiation Oncology* Biology* Physics*, vol. 68, no. 3, pp. 771–778, 2007.

[46] W. Yu, X.-L. Fu, Y.-J. Zhang, J.-Q. Xiang, L. Shen, G.-L. Jiang, and J. Y. Chang, "GTV spatial conformity between different delineation methods by 18 FDG PET/CT and pathology in esophageal cancer," *Radiotherapy and Oncology*, vol. 93, no. 3, pp. 441–446, 2009.

[47] M. Wanet, J. A. Lee, B. Weynand, M. De Bast, A. Poncelet, V. Lacroix, E. Coche, V. Grégoire, and X. 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*, vol. 98, no. 1, pp. 117–125, 2011.

[48] J.-F. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonneux, and V. Grégoire, "Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms," *Radiotherapy and Oncology*, vol. 69, no. 3, pp. 247–250, 2003.

[49] B. Yaremko, T. Riauka, D. Robinson, B. Murray, A. Alexander, A. McEwan, and W. Roa, "Thresholding in PET images of static and moving targets," *Physics in medicine and biology*, vol. 50, no. 24, p. 5969, 2005.

[50] Q. C. Black, I. S. Grills, L. L. Kestin, C.-Y. O. Wong, J. W. Wong, A. A. Martinez, and D. Yan, "Defining a radiotherapy target with positron emission tomography," *International Journal of Radiation Oncology* Biology* Physics*, vol. 60, no. 4, pp. 1272–1282, 2004.

[51] L. Drever, D. M. Robinson, A. McEwan, and W. Roa, "A local contrast based approach to threshold segmentation for PET target volume delineation," *Medical physics*, vol. 33, no. 6, pp. 1583–1594, 2006.

[52] W. Jentzen, L. Freudenberg, E. G. Eising, M. Heinze, W. Brandau, and A. Bockisch, "Segmentation of PET volumes by iterative image thresholding," *Journal of Nuclear Medicine*, vol. 48, no. 1, pp. 108–114, 2007.

[53] L. Drever, W. Roa, A. McEwan, and D. Robinson, "Iterative threshold segmentation for PET target volume delineation," *Medical physics*, vol. 34, no. 4, pp. 1253–1265, 2007.

[54] A. Schaefer, S. Kremp, D. Hellwig, C. Rübe, C.-M. Kirsch, and U. Nestle, "A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: derivation from phantom measurements and validation in patient data," *European journal of nuclear medicine and molecular imaging*, vol. 35, no. 11, pp. 1989–1999, 2008.

[55] M. Brambilla, R. Matheoud, C. Secco, G. Loi, M. Krengli, and E. Inglese, "Threshold segmentation for PET target volume delineation in radiation treatment planning: the role of target-to-background ratio and target size," *Medical physics*, vol. 35, no. 4, pp. 1207–1213, 2008.

[56] S. Nehmeh, H. El-Zeftawy, C. Greco, J. Schwartz, Y. Erdi, A. Kirov, C. Schmidtlein, A. Gyau, S. Larson, and J. Humm, "An iterative technique to segment PET lesions using a Monte Carlo based mathematical model," *Medical physics*, vol. 36, no. 10, pp. 4803–4809, 2009.

[57] A. C. Riegel, M. K. Bucci, O. R. Mawlawi, V. Johnson, M. Ahmad, X. Sun, D. Luo, A. G. Chandler, and T. Pan, "Target definition of moving lung tumors in positron emission tomography: correlation of optimal activity concentration thresholds with object size, motion extent, and source-to-background ratio," *Medical physics*, vol. 37, no. 4, pp. 1742–1752, 2010.

[58] M. Hatt, L. Rest, C. Cheze, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 6, pp. 881–893, 2009.

[59] M. Aristophanous, B. C. Penney, M. K. Martel, and C. A. Pelizzari, "A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography," *Medical physics*, vol. 34, no. 11, pp. 4223–4235, 2007.

[60] M. Hatt, C. C. Le Rest, P. Descourt, A. Dekker, D. De Ruysscher, M. Oellers, P. Lambin, O. Pradier, and D. Visvikis, "Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications," *International Journal of Radiation Oncology* Biology* Physics*, vol. 77, no. 1, pp. 301–308, 2010.

[61] M. S. Sharif, M. Abbod, A. Amira, and H. Zaidi, "Artificial neural network-based system for PET volume segmentation," *Journal of Biomedical Imaging*, vol. 2010, p. 4, 2010.

[62] J. Kim, L. Wen, S. Eberl, R. Fulton, and D. D. Feng, "Use of anatomical priors in the segmentation of PET lung tumor images," in *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*, vol. 6, pp. 4242–4245, IEEE, 2007.

[63] H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. 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*, vol. 75, no. 2, pp. 618–625, 2009.

[64] A. Kerhet, C. Small, H. Quon, T. Riauka, R. Greiner, A. McEwan, and W. Roa, "Segmentation of lung tumours in positron emission tomography scans: a machine learning approach," in *Artificial Intelligence in Medicine*, pp. 146–155, Springer, 2009.

[65] E. Yoshida, K. Kitamura, Y. Kimura, F. Nishikido, K. Shibuya, T. Yamaya, and H. 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*, vol. 571, no. 1, pp. 243–246, 2007.

[66] S. Belhassen and H. Zaidi, "A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET," *Medical physics*, vol. 37, no. 3, pp. 1309–1324, 2010.

[67] B. Foster, U. Bagci, B. Luna, B. Dey, W. Bishai, S. Jain, Z. Xu, and D. J. Mollura, "Robust segmentation and accurate target definition for positron emission tomography images using affinity propagation," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pp. 1461–1464, IEEE, 2013.

[68] F. Yang and P. W. Grigsby, "Delineation of FDG-PET tumors from heterogeneous background using spectral clustering," *European journal of radiology*, vol. 81, no. 11, pp. 3535–3541, 2012.

[69] D. W. Montgomery, A. Amira, and H. Zaidi, "Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model," *Medical Physics*, vol. 34, no. 2, pp. 722–736, 2007.

[70] A. Amira, S. Chandrasekaran, D. W. Montgomery, and I. S. Uzun, "A Segmentation concept for Positron Emission Tomography imaging using Multiresolution analysis," *Neurocomputing*, vol. 71, no. 10, pp. 1954–1965, 2008.

[71] T. M. Nguyen and Q. J. Wu, "Dirichlet gaussian mixture model: Application to image segmentation," *Image and Vision Computing*, vol. 29, no. 12, pp. 818–828, 2011.

[72] H. Li, W. L. Thorstad, K. J. Biehl, R. Laforest, Y. Su, K. I. Shoghi, E. D. Donnelly, D. A. Low, and W. Lu, "A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours," *Medical physics*, vol. 35, no. 8, pp. 3711–3721, 2008.

[73] J. B. Davis, B. Reiner, M. Huser, C. Burger, G. Székely, and I. F. Ciernik, "Assessment of 18 F PET signals for automatic target volume definition in radiotherapy treatment planning," *Radiotherapy and Oncology*, vol. 80, no. 1, pp. 43–50, 2006.

[74] U. Bagci, J. Yao, J. Caban, E. Turkbey, O. Aras, and D. J. Mollura, "A graph-theoretic approach for segmentation of PET images," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 8479–8482, IEEE, 2011.

[75] Q. Song, J. Bai, D. Han, S. Bhatia, W. Sun, W. Rockey, J. E. Bayouth, J. M. Buatti, and X. Wu, "Optimal co-segmentation of tumor in PET-CT images with context information," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 9, pp. 1685–1697, 2013.

[76] X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," *European journal of nuclear medicine and molecular imaging*, vol. 34, no. 9, pp. 1427–1438, 2007.

[77] V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische mathematik*, vol. 66, no. 1, pp. 1–31, 1993.

[78] C.-Y. Hsu, C.-Y. Liu, and C.-M. Chen, "Automatic segmentation of liver PET images," *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 601–610, 2008.

[79] J. A. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, vol. 3. Cambridge university press, 1999.

[80] B. Vemuri, J. Ye, Y. Chen, and C. Leonard, "A level-set based approach to image registration," in *Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on*, pp. 86–93, IEEE, 2000.

[81] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard, "Image registration via level-set motion: Applications to atlas-based segmentation," *Medical image analysis*, vol. 7, no. 1, pp. 1–20, 2003.

[82] J. Qi, *PET Image Segmentation and Reconstruction Using Level Set Method*. University of Toronto, 2011.

[83] I. El Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J. D. Bradley, P. Grigsby, and J. O. Deasy, "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planninga)," *Medical physics*, vol. 34, no. 12, pp. 4738–4749, 2007.

[84] H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 3, pp. 374–383, 2009.

[85] D. Markel, C. Caldwell, H. Alasti, H. Soliman, Y. Ung, J. Lee, and A. Sun, "Automatic segmentation of lung carcinoma using 3D texture features in 18-FDG PET/CT," *International journal of molecular imaging*, vol. 2013, 2013.

[86] D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu, "Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method," in *Information Processing in Medical Imaging*, pp. 245–256, Springer, 2011.

[87] K. Furuya, S. Murayama, H. Soeda, J. Murakami, Y. Ichinose, H. Yauuchi, Y. Katsuda, M. Koga, and K. Masuda, "New classification of small pulmonary nodules by margin characteristics on highresolution CT," *Acta Radiologica*, vol. 40, no. 5, pp. 496–504, 1999.

[88] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Classification of pulmonary nodules in thin-section CT images based on shape characterization," in *Image Processing, 1997. Proceedings., International Conference on*, vol. 3, pp. 528–530, IEEE, 1997.

[89] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Quantitative surface characterization of pulmonary nodules based on thin-section CT images," *Nuclear Science, IEEE Transactions on*, vol. 45, no. 4, pp. 2132–2138, 1998.

[90] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Mori, K. Eguchi, M. Kaneko, and N. Moriyama, "Curvature based analysis of pulmonary nodules using thin-section CT images," in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, vol. 1, pp. 361–363, IEEE, 1998.

[91] S. Kido, K. Kuriyama, M. Higashiyama, T. Kasugai, and C. Kuroda, "Fractal analysis of small peripheral pulmonary nodules in thin-section CT: evaluation of the lung-nodule interfaces," *Journal of computer assisted tomography*, vol. 26, no. 4, pp. 573–578, 2002.

[92] C. I. Henschke, D. F. Yankelevitz, I. Mateescu, D. W. Brettle, T. G. Rainey, and F. S. Weingard, "Neural networks for the analysis of small pulmonary nodules," *Clinical imaging*, vol. 21, no. 6, pp. 390–399, 1997.

[93] M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results," *Medical Physics*, vol. 26, no. 6, pp. 880–888, 1999.

[94] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, H. Nishiyama, K. Eguchi, M. Kaneko, and N. Moriyama, "Computerized analysis of 3-D pulmonary nodule images in surrounding and internal structure feature spaces," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, pp. 889–892, IEEE, 2001.

[95] Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, and K. Doi, "Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: evaluation with receiver operating characteristic analysis," *American Journal of Roentgenology*, vol. 178, no. 3, pp. 657–663, 2002.

[96] S.-C. B. Lo, L.-Y. Hsu, M. T. Freedman, Y. M. F. Lure, and H. Zhao, "Classification of lung nodules in diagnostic CT: an approach based on 3D vascular features, nodule density distribution, and shape features," in *Medical Imaging 2003*, pp. 183–189, International Society for Optics and Photonics, 2003.

[97] M. Aoyama, Q. Li, S. Katsuragawa, F. Li, S. Sone, and K. Doi, "Computerized scheme for

determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images," *Medical Physics*, vol. 30, no. 3, pp. 387–394, 2003.

[98] K. Nakamura, H. Yoshida, R. Engelmann, H. MacMahon, S. Katsuragawa, T. Ishida, K. Ashizawa, and K. Doi, "Computerized Analysis of the Likelihood of Malignancy in Solitary Pulmonary Nodules with Use of Artificial Neural Networks 1," *Radiology*, vol. 214, no. 3, pp. 823–830, 2000.

[99] S. Iwano, T. Nakamura, Y. Kamioka, and T. Ishigaki, "Computer-aided diagnosis: a shape classification of pulmonary nodules imaged by high-resolution CT," *Computerized Medical Imaging and Graphics*, vol. 29, no. 7, pp. 565–570, 2005.

[100] S. K. Shah, M. F. McNitt-Gray, S. R. Rogers, J. G. Goldin, R. D. Suh, J. W. Sayre, I. Petkovska, H. J. Kim, and D. R. Aberle, "Computer Aided Characterization of the Solitary Pulmonary Nodule Using Volumetric and Contrast Enhancement Features 1," *Academic radiology*, vol. 12, no. 10, pp. 1310–1319, 2005.

[101] K. Mori, N. Niki, T. Kondo, Y. Kamiyama, T. Kodama, Y. Kawada, and N. Moriyama, "Development of a novel computer-aided diagnosis system for automatic discrimination of malignant from benign solitary pulmonary nodules on thin-section dynamic computed tomography," *Journal of computer assisted tomography*, vol. 29, no. 2, pp. 215–222, 2005.

[102] K. Suzuki, F. Li, S. Sone, *et al.*, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 9, pp. 1138–1150, 2005.

[103] S. Iwano, T. Nakamura, Y. Kamioka, M. Ikeda, and T. Ishigaki, "Computer-aided differentiation of malignant from benign solitary pulmonary nodules imaged by high-resolution CT," *Computerized Medical Imaging and Graphics*, vol. 32, no. 5, pp. 416–422, 2008.

[104] T. W. Way, B. Sahiner, H.-P. Chan, L. Hadjiiski, P. N. Cascade, A. Chughtai, N. Bogot, and E. Kazerooni, "Computer-aided diagnosis of pulmonary nodules on CT scans: improvement

of classification performance with nodule surface features," *Medical physics*, vol. 36, no. 7, pp. 3086–3098, 2009.

[105] H. Chen, Y. Xu, Y. Ma, and B. Ma, "Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on CT images: clinical evaluation," *Academic radiology*, vol. 17, no. 5, pp. 595–602, 2010.

[106] M. C. Lee, L. Boroczky, K. Sungur-Stasik, A. D. Cann, A. C. Borczuk, S. M. Kawut, and C. A. Powell, "Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction," *Artificial intelligence in medicine*, vol. 50, no. 1, pp. 43–53, 2010.

[107] A. El-Baz, G. Gimel¡¯farb, R. Falk, M. A. El-Ghar, and J. Suri, "Appearance analysis for the early assessment of detected lung nodules," *Lung imaging and computer aided diagnosis*, vol. 17, pp. 395–404, 2011.

[108] A. El-Baz, M. Nitzken, F. Khalifa, A. Elnakib, G. Gimel'farb, R. Falk, and M. A. El-Ghar, "3D shape analysis for early diagnosis of malignant lung nodules," in *Information Processing in Medical Imaging*, pp. 772–783, Springer, 2011.

[109] N. A. Dewan, N. C. Gupta, L. S. Redepenning, J. J. Phalen, and M. P. Frick, "Diagnostic efficacy of PET-FDG imaging in solitary pulmonary nodules. Potential role in evaluation and management.," *CHEST Journal*, vol. 104, no. 4, pp. 997–1002, 1993.

[110] N. C. Gupta, J. Maloof, and E. Gunel, "Probability of malignancy in solitary pulmonary nodules using fluorine-18-FDG and PET.," *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, vol. 37, no. 6, pp. 943–948, 1996.

[111] J. Lee, J. M. Aronchick, and A. Alavi, "Accuracy of F-18 fluorodeoxyglucose positron emission tomography for the evaluation of malignancy in patients presenting with new lung abnormalities: a retrospective review," *CHEST Journal*, vol. 120, no. 6, pp. 1791–1797, 2001.

[112] G. J. Herder, R. P. Golding, O. S. Hoekstra, E. F. Comans, G. J. Teule, P. E. Postmus, and E. F. Smit, "The performance of 18F-fluorodeoxyglucose positron emission tomography in

small solitary pulmonary nodules," *European journal of nuclear medicine and molecular imaging*, vol. 31, no. 9, pp. 1231–1236, 2004.

[113] A. Halley, A. Hugentobler, P. Icard, E. Porret, F. Sobrio, J.-P. Lerochais, G. Bouvard, G. Zalcman, and D. Agostini, "Efficiency of 18F-FDG and 99mTc-depreotide SPECT in the diagnosis of malignancy of solitary pulmonary nodules," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 32, no. 9, pp. 1026–1032, 2005.

[114] Z. Keidar, N. Haim, L. Guralnik, M. Wollner, R. Bar-Shalom, A. Ben-Nun, and O. Israel, "PET/CT using 18F-FDG in suspected lung cancer recurrence: diagnostic value and impact on patient management," *Journal of Nuclear Medicine*, vol. 45, no. 10, pp. 1640–1646, 2004.

[115] A. Y. Chin, K. S. Lee, B.-T. Kim, J. Y. Choi, O. J. Kwon, H. Kim, Y. M. Shim, and M. J. Chung, "Tissue characterization of solitary pulmonary nodule: comparative study between helical dynamic CT and integrated PET/CT," *Journal of Nuclear Medicine*, vol. 47, no. 3, pp. 443–450, 2006.

[116] J. J. Erasmus, H. P. McAdams, and J. E. Connolly, "Solitary pulmonary nodules: Part II. Evaluation of the indeterminate nodule.," *Radiographics: a review publication of the Radiological Society of North America, Inc*, vol. 20, no. 1, pp. 59–66, 1999.

[117] J. M. Goo, J.-G. Im, K.-H. Do, J. S. Yeo, J. B. Seo, H. Y. Kim, and J.-K. Chung, "Pulmonary Tuberculoma Evaluated by Means of FDG PET: Findings in 10 Cases 1," *Radiology*, vol. 216, no. 1, pp. 117–121, 2000.

[118] Y. Nie, Q. Li, F. Li, Y. Pu, D. Appelbaum, and K. Doi, "Integrating PET and CT information to improve diagnostic accuracy for lung nodules: a semiautomatic computer-aided method," *Journal of Nuclear Medicine*, vol. 47, no. 7, pp. 1075–1080, 2006.

[119] P. Tao, F. Griess, Y. Lvov, M. Mineyev, B. Zhao, D. Levin, and L. Kaufman, "Characterization of small nodules by automatic segmentation of X-ray computed tomography images," *Journal of computer assisted tomography*, vol. 28, no. 3, pp. 372–377, 2004.

[120] D. Lardinois, W. Weder, T. F. Hany, E. M. Kamel, S. Korom, B. Seifert, G. K. von Schulthess, and H. C. Steinert, "Staging of non–small-cell lung cancer with integrated positron-emission tomography and computed tomography," *New England Journal of Medicine*, vol. 348, no. 25, pp. 2500–2507, 2003.

[121] H. Cui, X. Wang, and D. Feng, "Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 5384–5387, IEEE, 2012.

[122] Y. Nakamoto, M. Senda, T. Okada, S. Sakamoto, T. Saga, T. Higashi, and K. Togashi, "Software-based fusion of PET and CT images for suspected recurrent lung cancer," *Molecular Imaging and Biology*, vol. 10, no. 3, pp. 147–153, 2008.

[123] P. Kinahan, D. Townsend, T. Beyer, and D. Sashin, "Attenuation correction for a combined 3D PET/CT scanner," *Medical physics*, vol. 25, no. 10, pp. 2046–2053, 1998.

[124] C. B. Caldwell, K. Mah, Y. C. Ung, C. E. Danjoux, J. M. Balogh, S. N. Ganguli, and L. E. Ehrlich, "Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of 18 FDG-hybrid PET fusion," *International Journal of Radiation Oncology* Biology* Physics*, vol. 51, no. 4, pp. 923–931, 2001.

[125] P. Giraud, S. Elles, S. Helfre, Y. De Rycke, V. Servois, M.-F. Carette, C. Alzieu, P.-Y. Bondiau, B. Dubray, E. Touboul, *et al.*, "Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists," *Radiotherapy and oncology*, vol. 62, no. 1, pp. 27–36, 2002.

[126] R. J. Steenbakkers, J. C. Duppen, I. Fitton, K. E. Deurloo, L. J. Zijp, E. F. Comans, A. L. Uitterhoeve, P. T. Rodrigus, G. W. Kramer, J. Bussink, *et al.*, "Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis," *International Journal of Radiation Oncology* Biology* Physics*, vol. 64, no. 2, pp. 435–448, 2006.

[127] H. Vorwerk, G. Beckmann, M. Bremer, M. Degen, B. Dietl, R. Fietkau, T. Gsänger, R. M. Hermann, M. K. A. Herrmann, U. Höller, *et al.*, "The delineation of target volumes for radiotherapy of lung cancer patients," *Radiotherapy and Oncology*, vol. 91, no. 3, pp. 455–460, 2009.

[128] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004.

[129] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[130] L. Grady, "Minimal surfaces extend shortest path segmentation methods to 3D," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 2, pp. 321–334, 2010.

[131] L. Xu, B. Stojkovic, Y. Zhu, Q. Song, X. Wu, M. Sonka, and J. Xu, "Efficient algorithms for segmenting globally optimal and smooth multi-surfaces," in *Information Processing in Medical Imaging*, pp. 208–220, Springer, 2011.

[132] S. Klein, M. Staring, K. Murphy, M. Viergever, J. P. Pluim, *et al.*, "Elastix: a toolbox for intensity-based medical image registration," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 1, pp. 196–205, 2010.

[133] A. C. Riegel, A. M. Berson, S. Destian, T. Ng, L. B. Tena, R. J. Mitnick, and P. S. Wong, "Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion," *International Journal of Radiation Oncology* Biology* Physics*, vol. 65, no. 3, pp. 726–732, 2006.

[134] Z. Yu, F. M. Bui, and P. Babyn, "A 3D Semi-Automated Co-Segmentation Method for Improved Tumor Target Delineation in 3D PET-CT Imaging," in *Computing and Communication, 2015 6th International Conference and Workshop on*, IEEE, 2015.

[135] B. Ganeshan, S. Abaleke, R. C. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a

relationship with tumour glucose metabolism and stage," *Cancer Imaging*, vol. 10, no. 1, p. 137, 2010.

[136] H. Ng, S. Huang, S. Ong, K. Foong, P. Goh, and W. Nowinski, "Medical image segmentation using watershed segmentation with texture-based region merging," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 4039–4042, IEEE, 2008.

[137] J. Zhang, W. Zhang, C. Chen, Y. Guan, and C. Wang, "Computed diagnosis system for lung tumor detection based on PET/CT images," in *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, vol. 1, pp. 166–170, IEEE, 2010.

[138] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[139] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[140] C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Computer Vision, Graphics, and Image Processing*, vol. 23, no. 3, pp. 341–352, 1983.

[141] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 2, pp. 780–795, 1999.

[142] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 6, pp. 610–621, 1973.

[143] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of remote sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[144] M. Amadasun and R. King, "Textural features corresponding to textural properties," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 19, no. 5, pp. 1264–1274, 1989.

[145] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 1999.

[146] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 2, pp. 153–158, 1997.

Appendix

A 3D Semi-Automated Co-Segmentation Method for Improved Tumor Delineation in 3D PET/CT Imaging

A conference version of chapter 3 has been published at IEMCON 2015, and is provided in this appendix for reference. Major differences between this paper and chapter 3 include: The conference paper is based on a dataset of 5 patients, while in chapter 3 the dataset is expanded into almost 30 patients. Also there are improvements on the cost function used in chapter 3, compared to the cost function originally proposed in the conference publication.

A 3D Semi-Automated Co-Segmentation Method for Improved Tumor Target Delineation in 3D PET/CT Imaging

Zexi Yu and Francis M. Bui
Department of Electrical and Computer Engineering
University of Saskatchewan,
Saskatoon, SK S7N 5A9, Canada
Email: {zexi.yu,francis.bui}@usask.ca

Paul Babyn
Department of Medical Imaging
Royal University Hospital
Saskatoon, SK S7N 0W8, Canada
Email: paul.babyn@saskatoonhealthregion.ca

***Abstract*—The planning of radiotherapy is increasingly based on multi-modal imaging techniques such as positron emission tomography (PET)-computed tomography (CT), since PET/CT provides not only anatomical but also functional assessment of the tumor. In this work, we propose a novel co-segmentation method, utilizing both the PET and CT images, to localize the tumor. The method constructs the segmentation problem as minimization of a Markov random field model, which encapsulates features from both imaging modalities. The minimization problem can then be solved by the maximum flow algorithm, based on graph cuts theory. The proposed tumor delineation algorithm was validated in both a phantom, with a high-radiation area, and in patient data. The obtained results show significant improvement compared to existing segmentation methods, with respect to various qualitative and quantitative metrics.**

***Index Terms*—Context information, global optimization, graph cuts, image segmentation, tumor delineation, positron emission tomography-computed tomography (PET/CT)**

I. Introduction

Although positron emission tomography-computed tomography (PET/CT) has become increasingly popular in clinical oncology, tumor delineation is still manually performed on a slice-by-slice basis with little support from automated segmentation tools. The challenge with manual segmentation is that it is operator-dependent, time consuming and labor intensive [1]. The fact that manual segmentation is still widely used is not because of a lack of segmentation algorithms, but rather the functional is not appropriate. Many algorithms have been developed to address this issue, e.g., reference [1] provides a good review of these algorithms. However, the problem is that many existing algorithms work based on single modality (i.e., either PET or CT). As a result, these algorithms do not take advantage of both imaging modalities. In this work, we propose a novel semi-automated method for tumor delineation based on information derived from both PET and CT. The method is based on the graph cuts method. The idea is to model the information from both PET and CT into the segmentation process, and reformulate the segmentation process as an energy minimization problem. The energy encapsulates both boundary and region cost, defined by the information from PET and CT. The minimization problem can be solved by the classic graph cuts method. The main novelty in this paper is the energy definition. While the general structure of graph cuts is maintained, the definition we use is capable of forming the region and boundary information from both PET and CT. For simplification of the performance evaluation, the method will be evaluated on lung cancer, one of the commonest cancers.

The paper is organized as follows. In section II, we briefly survey approaches for tumor segmentation, followed by our proposed refinement of the classic graph cuts method for PET/CT in section III. Then, we present our proposed tumor delineation framework in section IV, with performance evaluation in section V. Conclusion and future work are discussed in section VI.

This work was supported in part by funding from NSERC and the Royal University Hospital Foundation.

II. Background and Related Works

A. Overview of Tumor Segmentation in PET/CT Imaging

From an image processing perspective, image segmentation can be subdivided into two related parts: object recognition and delineation [2]. Generally speaking, recognition is the process of determining the location of the object, while delineation is the process of finding the spatial boundary of the object in the image [3]. In a PET/CT image specifically, the recognition process typically means finding the high uptake regions in the PET image. These regions are considered as ROIs (Regions of Interest). This process is relatively easy to do by hand or software. When it comes back to delineation part, our goal is to achieve precise separation of high uptake regions from the background and non-significant uptake [2]. If the background and non-significant uptake regions are considered as one region, the task can be viewed as a two-region segmentation problem. If the background and non-significant regions are considered as different regions, the segmentation problem turns into a multi-region problem. For simplification, In this paper we consider the segmentation as a two-region problem, and we will come back to this in the future. Specifically, target delineation in PET/CT images involves dividing the image into two parts: tumor region and non-tumor region. From an image processing point of view, such a division is tantamount to labeling each pixel on the image with 1 (being tumor) or 0 (being non-tumor).

The delineation process is relatively tricky as there are various factors that increase the difficulty of this process, including: low image resolution, due to the way the image

978-1-4799-6908-1/15/$31.00 ©2015 IEEE

is produced and filtered; motion artifacts during image acquisition [1], [4]; and individual factors related to the large variability of pathologies in terms of shape, texture, location, and differences in organ morphology and texture between individuals [1]; image noise, which comes from the image scanner and the way the images are reconstructed. The noise in PET is considered as the most significant factor for not having a reproducible SUV measurement [4], [5].

Even with all these difficulties mentioned above, because of increasing PET/CT usage in clinical applications, there have been considerable improvements in PET image segmentation methods, making these methods more accurate and robust under PET/CT multi-modal image frame. For example, standard uptake values are introduced in PET imaging for better thresholding [6]; Statistical and machine learning tools are introduced to take further advantage from the differences between uptake regions and their surrounding tissues [1]. These methods generally achieve around a 90 % Dice index similarity compared to established ground truth on patient data [1], [7]. In this paper we seek to achieve even higher segmentation accuracy by implementing the spatial information of CT into the PET/CT segmentation problem, and graph cuts, a graph based segmentation method, allows a relative easy way for implementation.

B. Graph-based Segmentation

Region based methods are typically referred to as graph-based methods, because they often utilize graph based theory for implementing information from homogeneity. By using foreground and background seeds specified by user, graph-based approaches locate objects in the image by solving an optimization problem generated by implementing local pair-wise pixel similarities (homogeneity) with global information. Among the graph approaches, two methods are often used for PET segmentation: Random Walk and Graph cuts.

Although originally designed for computer vision, the random walk has been applied to image segmentation in many cases. The advantage of Random Walk related approaches is that they are robust against noise and weak boundaries, making them suitable for PET/CT segmentation with the low resolution and high noise characteristic of PET [8], [9]. However, the use of random walk is limited, since the method may not be able to deal with multi-focal uptake regions properly.

Alternatively, taking a similar approach to Song et al.'s work [10], our work utilizes a graph cuts method. The basic idea is to reformulate the segmentation problem as an energy minimization problem. Similar to Song, our work uses a graph cuts model and constructs the cost function from the boundary and region information from PET and CT. Unlike Song, we use a single graph model to combine both PET and CT information under one graph instead of two sub-graphs. This simplifies the algorithm, reduces its computational cost and gives only one segmentation output instead of two as Song's work. Secondly, we do not use foreground seeds to mark the tumor and only use background seeds, this reduce the level the human interaction. Thirdly, after experiment we decide to remove the region cost from CT from the algorithm, because the involvement of this cost gives worse results, as shown in the results section.

III. Proposed Graph Cuts Method

Graph-based segmentation methods are differentiated by their ways of defining cost functions and the ways of solving the minimization problem. As the basis of our method, the graph cuts method models the cost function as a combination of "self cost" and "neighbor cost".

Consider that we have a 3D PET/CT image pair (I, I'), where I represents the CT image and I' represents the PET image. We assume the images are perfectly aligned, thus for each voxel $u \in I$ we have a corresponding voxel $u \in I'$. Let l_u represent the binary labels assigned to each voxel, and as we consider the segmentation problem as a two-region segmentation, $l_u = 0$ suggests the voxel u is segmented into the background (i.e, non-tumor in our case), while $l_u = 1$ denotes the voxel u is put into the foreground (i,e, tumor tissue in our case). t_u represents the image intensity value at the voxel u.

With the notations above we can further define the self cost term $E_s(l_u)$. For each voxel $u \in I$ or I' the self cost term is defined as

$$E_s(l_u) = \begin{cases} C_{source}(t_u), & \text{if } l_u = 1 \\ C_{sink}(t_u), & \text{if } l_u = 0 \end{cases}$$

We can also define neighbor cost term in graph cuts. For a neighbor voxel pair (u, v), the neighbor cost $E_n(l_u, l_v)$ is defined as

$$E_n(l_u, l_v) = \begin{cases} C_{nodes}(t_u, t_v), & \text{if } l_u = l_v \\ C_{nodes}(t_u, t_v), & \text{if } l_u = l_v \end{cases}$$

We further define a neighbor set N in the image pair (I, I'), where $N = \{(u, v) / u \text{ and } v \text{ is neighbor}\}$. Then the main idea of graph cuts can be formed as finding a labeling set $L_m = \{l_u\}$ that minimizes the sum of self and neighbor costs on all the voxels in the image pair (I, I'):

$$L_m = \arg\min_L \Sigma_{u \in I} E_s(l_u) + \Sigma_{(u,v) \in N} E_n(l_u, l_v); \quad (1)$$

Traditionally the cost functions C_{source} and C_{sink} are defined based on the intensity value of PET or CT. Some works use the cost terms of PET alone to decide the segmentation [9]. Others combine the cost terms from PET and CT together to decide the optimal solution [10].

In our work, we seek to encapsulate the information from PET and CT together into the self cost term and neighbor cost term by redefining the cost function C_{source}, C_{sink} and C_{nodes} based on both PET and CT information. To this end, we first describe the graph model of the graph cuts.

A. Graph Construction

In graph cuts, the function (1) is solved by maximum flow algorithm after modeling the function into the graph form. The model is shown here in figure 1. As shown in the figure, S is the source of flow and sink is the end of flow, each voxel in the image is modeled as a voxel node, with tube connecting them to the source , sink and themselves. self cost (i.e, the C_{source} and the C_{sink}) are represented by the tube capacity from the source to the pixel node (the C_{source}, and the orange lines in the 2nd subfigure) and from the voxel node to the sink (the C_{sink}, the blue lines in the 2nd subfigure). And neighbor cost (i.e the C_{nodes} when the nodes have different label, the black lines) is modeled as the tube capacity between the nodes themselves. By doing this, equation 1 can be effectively solved by solving maximum flow problem under this graph [11], and ultimately gives the segmentation result, as described in further detail in [11].

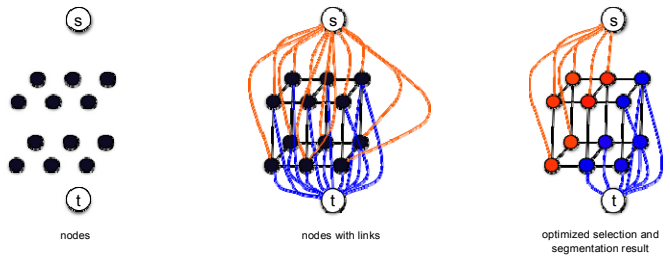


Fig. 1. Graph cuts

B. Energy function

The major differences between our proposed method and existing ones are in the energy function selection. Traditionally seeds are placed for the foreground and background in order for the graph cuts to work robustly. The seeds here refer to certain voxels, which can only be segmented into foreground or background in all segmentation solutions. Unlike traditional graph cuts we use a relatively simple way to place seeds. In our approach we do not place any seed for foreground, i.e. the tumor. Instead we just place seeds for the background by roughly marking the known high uptake value regions without tumor. Such regions, for example as heart and urinary bladder, are given the cost $C_{source} = 0$ and $C_{sink} = \infty$.

Then we combine both CT and PET information into the cost function. The details are listed below.

1) Self cost: The self cost in this case still comes from the PET image itself (and is the flow capability from the node to the source node and to the sink node).

We use PET and CT image, and after normalizing them, we get the image pair I_n, I'_n, and then we define the self cost our approach as:

$$C_{source}(t_u) = \begin{cases} |t_u - \quad \text{if u is not marked as background seeds} \\ 0, \quad \text{otherwise} \end{cases}$$

$$C_{sink}(t_u) = \begin{cases} |t_u - \quad \text{if u is not marked as background seeds} \\ \infty, \quad \text{otherwise} \end{cases}$$

for a single node representing the pixel on image. Here C_{source} is the flow capability from the node to the source, and C_{sink} is the flow capability from the node to the sink. u represents the node and t_u is the normalized PET image value on that node (i.e, the radiation intensity on the pixel). t_{source} and t_{sink} is pre-set based on a per-segmentation method or by the user.

2) Neighbor cost: We use a parametrized neighbor cost in our method. The neighbor cost in our method is:

$$C_{nodes}(u, v) = \begin{cases} \dfrac{1}{1 + \lambda_1 (G_{uv} + \lambda_2 \frac{t_u + t_v}{2} G_{uv})}, \\ \quad \text{if u and v are neighbor pixels} \\ 0, \quad otherwise \end{cases}$$

Here, u and v are the nodes and G_{uv} and G_{uv} are the gradient value of normalized PET and CT between the nodes u and v correspondingly. The main idea is to only include the gradient information from PET and the gradient information from CT at high uptake value region.

IV. Proposed Tumor Delineation System Framework

Figure 2 shows a block diagram of the proposed framework, with five stages involved: preprocessing stage, parameter setting stage for the graph cuts, and the practice stage (performing 3D graph cuts) and evaluation.

A. Preprocessing

In a preprocessing stage what we want to do is to align or register the CT image with the PET image.

For the registration we have the following options: rigid transformation, meaning the image is only allowed to rotate and translate; similarity, meaning the image is allowed to rotate, scale and translate; and affine transformation, meaning the image is allowed to rotate, scale, translate and cut. The b-spline transformation is the most complex one, giving the images the most freedom to align to each other. However, this transformation performs poorly with the phantom and is thus not considered further. The affine transformation is implemented instead in this step using elastix toolbox [12].

B. Parameter Setting

In this section we decide the value of t_{source}, t_{sink}, λ_1 and λ_2. We run an iterative thresholding method for these two values. If t is the iterative thresholding result, we set t_{source} to be $\frac{3t}{2}$, and the t_{sink} to be $\frac{t}{2}$

In application such as lung cancer, we find the value of t_{source} around 0.4 and t_{sink} around 0.1. In this application we set the λ_1 to be 10 and λ_2 to be 10.

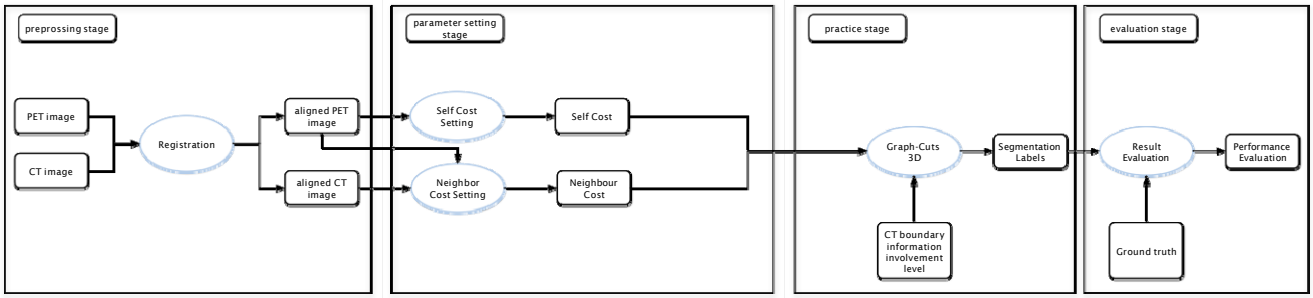


Fig. 2. Proposed Segmentation Method Blocks

C. Practice Stage

We perform a 3D graph cuts algorithm based on maximum flow algorithm as we discussed in the graph cuts section. We set the neighbor set as the 8 neighbor set, the most common configuration in 3D image processing scenario.

D. Evaluation Stage

Here we evaluate the performance of the method, we will show more details in the following section.

V. Experimental Results

In this section, we describe how the experiments are set up and how the results are evaluated. The method is performed on both phantom data and patient data, and the ground truth is established for both cases. With this ground truth, the result from the proposed method is evaluated using Dice index and Hausdorff distance, two of the most widely used evaluation metrics in medical imaging. Then the performance will be discussed, with comparisons to several classical and state of the art methods.

A. Experimental Setup

1) Ground truth establishment: In order to evaluate an image segmentation algorithm, the true boundary of the object of interest should be identified, the process is often referred to as ground truth establishment.

For the phantom image case this is an easy task, for we know the exact dimensions of the object in the image and thus we know exactly what the segmentation result should be. This is perhaps the most important reason of using an phantom image to evaluate the method performance in many works, because when it comes to the real patient image, there is no ground truth available unless there are histopathologic samples. In many cases, surrogate truths (or reference standards) are used for measuring the quality of a segmentation algorithm. One may argue that even the same physician can hardly be expected to give the same segmentation result for segmenting the same image many times over. Nevertheless this is the best option we have for now, representing the closest starting point upon which we can compare our result. After all, most of today's radiation therapy treatment plans are made according to the manual segmentation by physicians.

2) Evaluation Matrix: With the ground truth established, the comparison of how close the segmentation result is to the ground truth can be made. There are two major classes for such a comparison: one is similarity indices, which show how similar the ground truth and the result are, e.g., the Dice index; the other is error indices, which show the maximum offset of the result from the ground truth, e.g., the Hausdorff distance.

Dice index is one of the most widely used similarity indices [12]. Given X as our method result and Y as ground truth, the Dice index can be computed using the following equation:

$$d_I(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}, \quad (2)$$

with $|.|$ defined as the area of the graph.

As can be seen in the definition, Dice index represents the overlay area of the ground-truth and the result over the total area. As a similarity index, it reflects the overall performance of the segmentation accuracy.

Another evaluation metric used in this paper is the Hausdorff distance, one of the most popular error indices. Given x as voxel in the image X and y as voxel in the image Y, the Hausdorff distance is defined as:

$$d_H(X,Y) = \max\{\sup_{x \in X} \inf d(x,y), \sup_{y \in Y} \inf d(x,y)\}, \quad (3)$$

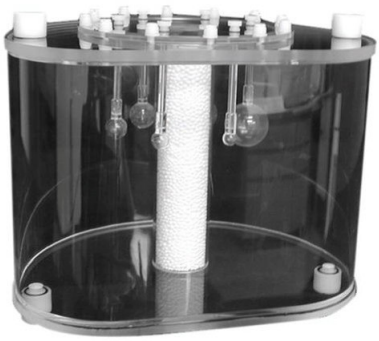
with $d(.,.)$ defined as the distance (usually Euclidean distance)

The Hausdorff distance, by definition, is the largest offset from the result to the ground truth. It represents the detailed accuracy of the result. In this work, the Hausdroff distance is shown using the real distance instead of pixel distance. That is, after calculating the Hausdroff distance by definition, we multiply the result by the voxel size.

By using Dice index and Hausdorff distance together, both overall and detailed performances of the method can be evaluated. This explains why one can see this combination often used in the existing literature as the evaluation metric.

3) Phantom Description: We test our method on a specifically designed phantom. The structure of the phantom and the details of the PET/CT image are shown in figure 3.

Here, there are 6 small balls in the phantom, the diameters of which are 4mm, 6mm, 8mm, 10mm, 12mm, 14mm. The



PET(18F FDG)								
Frame length	Width/ length	depth	Pixel size	Slice depth	Width/ length	depth	Pixel size	Slice depth
7mins	192 pixels	83 slices	1.8229mm by 1.8229 mm	3.27 mm	512 pixels	135 slices	0.6836mm by 0.6836mm	2.50 mm

Fig. 3. Phantom Description

phantom helps to evaluate the method performance on the relatively small size of tumor construction.

The main body of phantom is filled with radiations and the balls are filled with normal water. The ground truth is established using the phantom measurements. It should be noted this is controversial to the normal phantom image. By using this image, We hope to evaluate the method's capability of differentiating small structures and gaps near the tumor region.

4) Patient Data Description: 5 sets of patient data are obtained from The Cancer Genome Atlas-Lung Squamous Cell Carcinoma (TCGA-LUSC) data collection on The Cancer Imaging Archive (TCIA) website [13], [14]. The patients suffered from lung cancer, and the size of tumors varies from 8mm to 23mm.

Due to limited source, the ground truth is only drawn on 2 slices for each patient. Therefore even though our method is applied on the entire 3D patient data, only 2 slices with the established ground truth is used for each patient for comparison.

5) Comparing Methods: As mentioned, we perform our proposed method on both the phantom data and the patient data, then evaluate the performance using the established ground truth. With the evaluation complete we run several classical and state-of-art methods and evaluate their performance.

Before we come to the results, here is a list of the comparison methods and their description.

- Thresholding methods (T1,T2) As described above the thresholding method is the most popular method and thus included in our comparison. The classic methods used in this paper come from the work [6] (marked as T1) and the work [15] (marked as T2). Also we use a novel adaptive thresholding method based on Affinity Propagation principle, that is the method from the work [7], marked as AP.
- Graph cuts based Approaches (GA, CO, PM) The graph cuts approach is the basis of our method. As such, it is vital for our method to be compared with other similar approaches based on the same principle. The approaches chosen for comparison are: the basic graph cuts method that performs graph cuts on PET only, based on the work [11], marked as GA; a recent co-segmentation method based on the graph cuts principle [10], marked as CO; and our proposed method (PM).

B. Results and Discussions

We ran our methods and the comparison methods using a Matlab 2014b on a PC in our lab. The configuration of the PC is Intel Core i7-4790 CPU @ 3.60 GHz with 16GB RAM, also with a NVIDIA Quadro K2000 for GPU computing.

1) The Overall Performance: First, in Table I, a summary of the overall performance results can be found. It should be noted that the method of [10] gives two segmentation outcomes, one for PET and one for CT, without delivering a final outcome which is actually needed for clinical application. Accordingly, this approach is not implemented for the phantom in this paper. For comparison, the results from the paper [1] are listed directly in the table.

TABLE I
RESULT EVALUATION

Methods	Phantom		Patient	
	DSC	HD	DSC	HD
T1	0.96	13.6720	0.7515 ± 0.0739	34.1617 ± 38.0619
T2	0.92	12.3048	0.8090 ± 0.1190	6.7854 ± 3.2686
AP	0.97	7.5196	0.8541 ± 0.0523	5.2988 ± 1.3489
GA	0.95	8.2032	0.8695 ± 0.0455	4.8610 ± 1.1993
CO	–	–	0.8934 ± 0.0195	5.600 ± 0.600
PM	0.99	4.1016	0.9398 ± 0.0247	2.5154 ± 0.9249

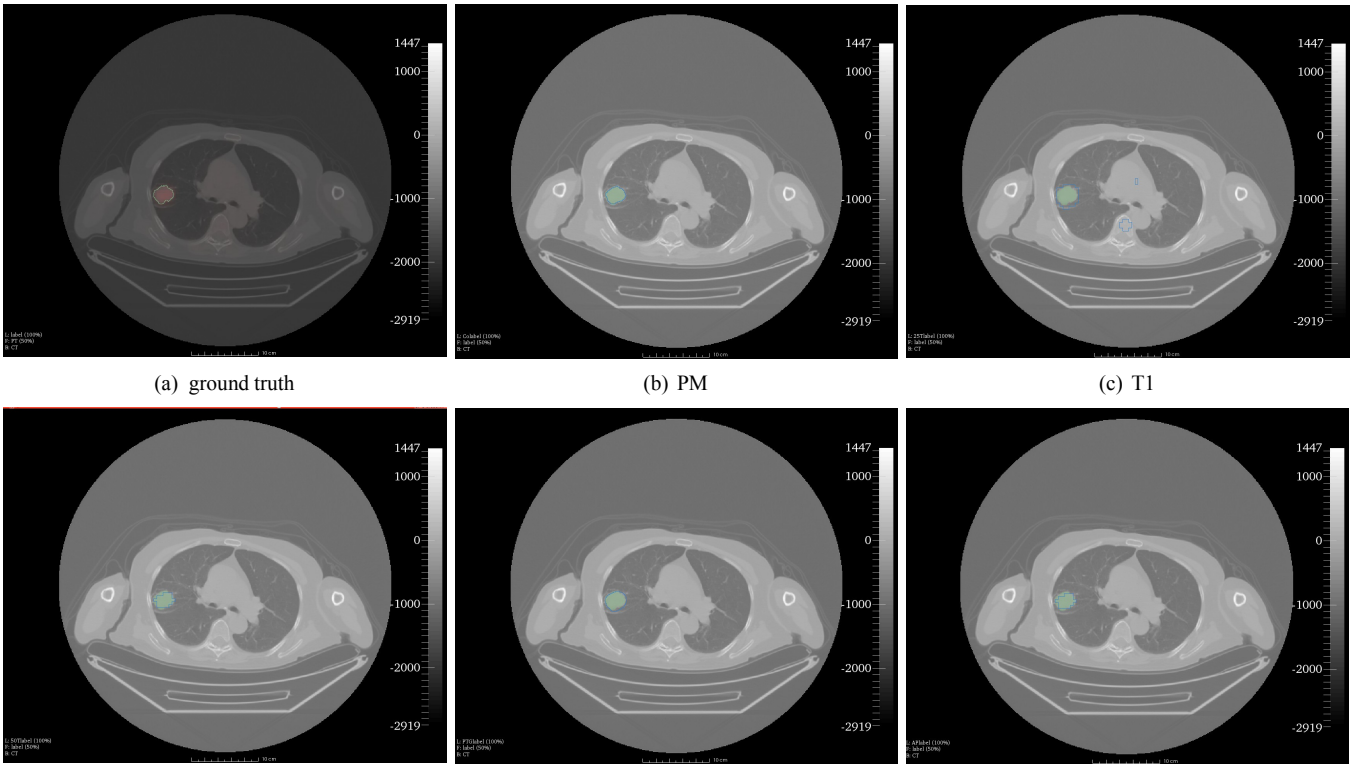
Also note that for the patient results we ran our approach on a 3D frame but only use 2 slices for the evaluation since we only have 2D ground truth. We only have one set of phantom so the result on phantom do not come with variation.

From the table we can see that our method attains the best performance over all the methods used, with a Dice index close to 0.94 and HD lower than 3mm.

In order to illustrate how the method works, we provide an image of the PM result on the phantom in Figure 4. We can see from this image that the result created by PM is very accurate. The boundary between the radiation and the normal tissue is well separated.

In order to give an impression of the overall performance of the methods in real life application, the resultant images of one slice from all the approaches are shown in Figure 5. Note that we did not reproduce the CO method so there are only five images.

From the image we can see that our method gives clear separation while keeping tumor detail. The other results either suffer from inaccurate detail due to the low resolution of PET, or give an over/under segmented result. This conclusion can also be drawn from the table. We can clearly see that the CT



(a) ground truth (b) PM (c) T1

(d) T2 (e) GA (f) AP

Fig. 5. Result on the Patient data, on the top left we present our original slice, with CT represented by the black-and-white image and PET represented by red region. The green line in the original data represents the ground truth. The blue line on other images are the boundaries given by the methods, and the green regions are the ground truth

Fig. 4. Resulting Image of the Proposed Method on the Phantom

information gives our method the advantage of more precise detail.

2) The level of CT information: While the results are good, we should carefully control the level of CT information involvement since over-involvement of CT will result in poor segmentation. We can see this effect from Figure 6. Here we present the impact of the level of involvement on the result. The ratio indicates the value of the parameter λ_2 in the equation neighbor cost function.

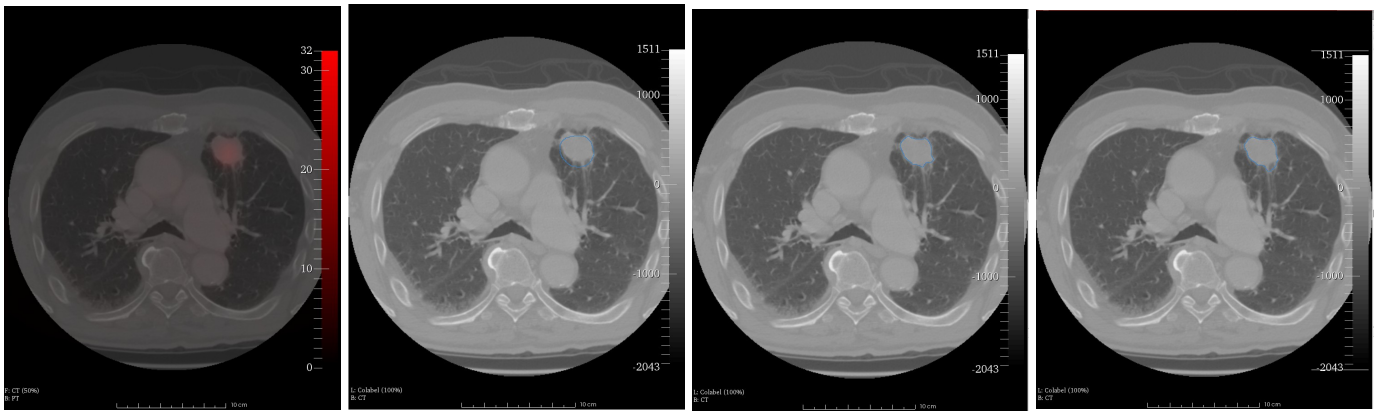
It is clear from these figures that higher level of CT information gives better boundary accuracy, but over-committing such an involvement can lead the segmentation boundary close to the tissue boundary in CT and thus results in over-segmentation as indicated in the ratio 1 to 100. Under-commitment will lead to a result that is closer to thresholding, thus losing the details of the tumor. Therefore selecting a good level of CT involvement is important in this work. In this paper we use a pre-training approach to address this issue and set the ratio to be 1 to 10. In the future, a more automated approach will be explored.

3) The involvement of CT Region Cost: As mentioned previously, the difference between our work and the work [10] is that we exclude the CT Region Cost (Self Cost) in our frame. Table II shows the results of our method if CT Region Cost is included.

One can clearly see the impact of CT Regional Cost. Even a slight level of involvement will cause a lower performance. Accordingly, we do not use CT Regional Cost in our method.

VI. Conclusion and Future Work

There are several potential ways to further improve our proposed method, Two specific directions that we believe to be most promising are described. One is to develop a per-training approach to decide the parameters in the cost function, therefore changing the method into an adaptive approach. For



(a) original image (b) PT-CT information level 10 to 1, $\lambda_2 = 0.1$ (c) PT-CT information level 1 to 10, $\lambda_2 = 10$ (d) PT-CT information level 1 to 100, $\lambda_2 = 100$

Fig. 6. Effect of CT Information

TABLE II
CT REGION RESULT EVALUATION

ratio of CT Region involvement (CT:PET)	Phantom		Patient	
	DSC	HD	DSC	HD
none	0.99	4.1016	0.9398 ± 0.0247	2.5154 ± 0.9249
1:100	0.96	6.1524	0.9126 ± 0.0447	2.9185 ± 1.2279
1:10	0.90	7.1778	0.9015 ± 0.0678	3.1154 ± 2.0360
1:1	0.90	9.2286	0.8579 ± 0.0716	3.5126 ± 2.5260
10:1	0.40	31.7874	0.3816 ± 0.1247	20.6420 ± 20.8574

another direction, note that we currently rely on manual seed planting to differentiate the tumor region from the other high uptake regions such as kidneys. If the seeds are too close to the tumor region, the method becomes potentially inaccurate. Accordingly, we plan to develop a localization approach that will automatically divide the tumor from other high uptake regions. This will remove the requirement for marking the background in the graph cuts approach, therefore delivering a fully automated and hopefully more robust method.

REFERENCES

[1] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in biology and medicine*, vol. 50, pp. 76–96, 2014.

[2] U. Bagci, X. Chen, and J. K. Udupa, "Hierarchical scale-based multiobject recognition of 3-d anatomical structures," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 3, pp. 777–789, 2012.

[3] P. K. Saha and J. K. Udupa, "Scale-based diffusive image filtering preserving boundary sharpness and fine structures," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 11, pp. 1140–1155, 2001.

[4] P. Kinahan, D. Townsend, T. Beyer, and D. Sashin, "Attenuation correction for a combined 3d pet/ct scanner," *Medical physics*, vol. 25, no. 10, pp. 2046–2053, 1998.

[5] R. Boellaard, N. C. Krak, O. S. Hoekstra, and A. A. Lammertsma, "Effects of noise, image resolution, and roi definition on the accuracy of standard uptake values: a simulation study," *Journal of Nuclear Medicine*, vol. 45, no. 9, pp. 1519–1527, 2004.

[6] D. Hellwig, T. P. Graeter, D. Ukena, A. Groeschel, G. W. Sybrecht, H.-J. Schaefers, and C.-M. Kirsch, "18f-fdg pet for mediastinal staging of lung cancer: which suv threshold makes sense?" *Journal of Nuclear Medicine*, vol. 48, no. 11, pp. 1761–1766, 2007.

[7] B. Foster, U. Bagci, B. Luna, B. Dey, W. Bishai, S. Jain, Z. Xu, and D. J. Mollura, "Robust segmentation and accurate target definition for positron emission tomography images using affinity propagation," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. IEEE, 2013, pp. 1461–1464.

[8] U. Bag˘c, J. Yao, J. Caban, E. Turkbey, O. Aras, and D. J. Mollura, "A graph-theoretic approach for segmentation of pet images," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 8479–8482.

[9] U. Bagci, J. Yao, J. Caban, E. Turkbey, O. Aras, and D. J. Mollura, "A graph-theoretic approach for segmentation of pet images," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 8479–8482.

[10] Q. Song, J. Bai, D. Han, S. Bhatia, W. Sun, W. Rockey, J. E. Bayouth, J. M. Buatti, and X. Wu, "Optimal co-segmentation of tumor in pet-ct images with context information," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 9, pp. 1685–1697, 2013.

[11] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004.

[12] S. Klein, M. Staring, K. Murphy, M. Viergever, J. P. Pluim *et al.*, "Elastix: a toolbox for intensity-based medical image registration," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 1, pp. 196–205, 2010.

[13] T. C. G. Atlas, "**TCGA** research network," http://cancergenome.nih.gov/.

[14] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[15] J.-F. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonneux, and V. Gre´goire, "Tri-dimensional automatic segmentation of pet volumes based on measured source-to-background ratios: influence of reconstruction algorithms," *Radiotherapy and Oncology*, vol. 69, no. 3, pp. 247–250, 2003.