Topology polymorphism graph for lung tumor segmentation in PET-CT images

Hui Cui[1], Xiuying Wang[1], Jianlong Zhou[2], Stefan Eberl[1,3], Yong Yin[4], Dagan Feng[1,5] and Michael Fulham[3,6]

[1] Biomedical and Multimedia Information Technology research group, School of Information Technologies, University of Sydney, Sydney, Australia
[2] National ICT Australia
[3] Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia
[4] Department of Radiation Oncology, Shandong Tumor Hospital, Jinan, China
[5] Med-X Research Institute, Shanghai Jiao Tong University, China
[6] Sydney Medical School, University of Sydney, Australia

Abstract. Accurate lung tumor segmentation is problematic when the tumor boundary or edge, which reflects the advancing edge of the tumor, is difficult to discern on chest CT or PET. We propose a "topo-poly" graph model to improve identification of the tumor extent. Our model incorporates an intensity graph and a topology graph. The intensity graph provides the joint PET-CT foreground similarity to differentiate the tumor from surrounding tissues. The topology graph is defined on the basis of contour tree to reflect the inclusion and exclusion relationship of regions. By taking into account different topology relations, the edges in our model exhibit topological polymorphism. These polymorphic edges in turn affect the energy cost when crossing different topology regions under a random walk (RW) framework, and hence contribute to appropriate tumor delineation. We validated our method on 40 patients with non-small cell lung cancer (NSCLC) where the tumors were manually delineated by a clinical expert. The studies were separated into an 'isolated' group (n=20) where the lung tumor was located in the lung parenchyma and away from associated structures / tissues in the thorax and a 'complex' group (n=20) where the tumor abutted / involved a variety of adjacent structures and had heterogeneous FDG uptake. The methods were validated using Dice's similarity coefficient (DSC) to measure the spatial volume overlap and Hausdorff Distance (HD) to compare shape similarity calculated as the maximum surface distance between the segmentation results and the manual delineations. Our method achieved an average DSC of 0.881±0.046 and HD of 5.311±3.022 mm for the isolated cases and DSC of 0.870±0.038 and HD of 9.370±3.169 mm for the complex cases. Student's t-test showed that our model outperformed the other methods (p-values <0.05).

1. Introduction

Lung cancer is the leading cause of cancer death in men and women (Ahmedin Jemal 2011). The most common type of lung cancer is non-small cell lung cancer (NSCLC). The five-year survival rate remains poor (15%) and there has been little change in the past two decades (Ahmedin Jemal 2011). Accurate lung cancer staging is essential for best patient management. The best option for cure is limited stage disease that is treated by surgical removal of the tumor if the patient is fit for surgery. Radiotherapy can offer similar survival benefits in patients who cannot tolerate surgery. Involvement of adjacent

structures such as the mediastinum, major vessels, pericardium, diaphragm, vertebral bodies and chest wall upstages the disease and affects the treatment approach. Further, as radiotherapy techniques can now deliver higher radiation doses with improved accuracy, accurate tumor delineation is needed for effective radiation therapy and to decrease the exposure and harm to surrounding healthy tissues (Kirov and Fanchon 2014). However, manual delineation of the tumor boundaries is time consuming and there are inter- and intra-observer variability between the manual segmentations (Kirov and Fanchon 2014), (MacManus, Nestle et al. 2009), (Peng, Zhang et al. 2013). Hence an automated tumor segmentation algorithm would be advantageous.

Positron emission tomography (PET) has an essential role in NSCLC staging and because the primary lung tumors typically take up more ^{18}F-Fluorodeoxyglucose (FDG), the PET tracer, than surrounding normal lung tissue. So sites of active tumor appear as regions of increased FDG uptake or "hot spots" on PET scans (Kunz, Thon et al. 2011), (ACS 2014). The standard uptake value (SUV) is a semi-quantitative expression of tumoral FDG uptake and the SUV has been used to separate tumor from surrounding normal structures (Simona and Peter 2009). A number of methods have been used to help identify the tumor including: thresholding based methods, where a fixed threshold of 40% to 50% of maximum tumor SUV (SUV_{max}) (Vanderhoek, Perlman et al. 2012); adaptive thresholds; fuzzy c-means (FCM) and gradient based methods such as watershed (Tylski, Bonniaud et al. 2006). These methods, however, may result in over segmentation of tumors with where there is heterogeneous (non-uniform) FDG uptake or fail to separate the tumor from neighbouring structures where there is similar FDG uptake such as the heart, liver and brown fat (Ballangan, Wang et al. 2011). We previously investigated the relationship between SUV and gradient magnitude, and reported a tumor customized downhill region growing (TCD) method to separate a tumor from adjacent structures with similar densities (Ballangan, Wang et al. 2011) (Wang, Cui et al. 2012). The TCD method also provided better delineation of tumors with non-uniform FDG uptake compared to thresholding methods. PET, because of limited spatial resolution, still has difficulty in accurately delineating the edges of the tumor. Although CT provides better anatomic detail and higher resolution, the CT abnormality can include the tumor, adjacent lung, which may have changes due to consolidation, and adjacent soft tissues that have similar densities.

The 'hybrid' or dual-modality PET-CT scanner provides functional (PET) and anatomical (CT) data in a single patient scan. PET-CT is now the accepted clinical standard for staging NSCLC. It has also been the focus of intensive research in the segmentation field. Segmentation methods in PET-CT include the co-likelihood modeling of PET-CT intensity distributions (Wang, Ballangan et al. 2014) and the construction of dual-graphs with PET and CT sub-graphs (Han, Bayouth et al. 2011, Bagci, Udupa et al. 2013, Qi, Junjie et al. 2013). The dual-graph models can be solved by graph-based algorithms such as Markov random field (MRF) (Han, Bayouth et al. 2011), graph cut (GC) (Qi, Junjie et al. 2013) and random walks (RW) (Bagci, Udupa et al. 2013). Among the graph-based algorithms, RW provides better segmentation than other graph algorithms with weak boundaries and noise problems (Grady, Schiwietz et al. 2005, Grady 2006, Chen, Helm et al. 2011). The RW is, however, sensitive to the locations and the number of the input seeds. Many investigators have identified that RW lacks the global informative features and depends on the local changes of intensities (Grady 2005) and local pixel-based intensities are insufficient for robust segmentation (Kim, Lee et al. 2008) (Yang, Cai et al. 2010). Since graph-based segmentation methods are affected by the edge weights (Grady and Jolly 2008), one solution was to design weighting functions incorporating prior estimated models. Yang *et al* described a generative image segmentation where the prior foreground and background distribution were estimated and improved the texture image segmentations (Yang, Cai et al. 2010). For PET-CT some background non-tumor tissues, especially those with high FDG uptake, may have a similar intensity distribution to the foreground tumor. Thus they will have similar prior likelihood and when they are adjacent, their separation remains problematic. Other global feature solutions include the random walk with restart (RWR) models (Kim, Lee et al. 2008, Cui, Wang et al. 2013), which capture the global information of the graph instead of only local neighboring pairwise relations in RW. Kim *et al*. incorporated regional information obtained by super pixels as a fully connected region layer to improve segmentation for

textural objects when compared to RW for natural images in (Kim, Lee et al. 2010). We reported, however, that the full regional connections produced misleading groupings for the sub-regions in tumors with heterogeneous FDG uptake (Cui, Wang et al. 2014).

When tumor boundaries are difficult to discern on CT and PET, segmentation based on PET SUV or CT intensity are unlikely to provide accurate tumor margins. Hence, we propose a novel topo-poly graph model, which incorporates the estimation of PET-CT foreground similarity and structural/topological relations, where the edges of the topo-poly graph present topology polymorphism that contributes to the appropriate tumor separation and boundary delineation. Our paper is then organized as follows: a) Section 2 - we describe the related RW segmentation model in single modality images; b) Section 3 - we introduce our topo-poly graph model; c) Section 4 - we evaluate the performance of our model in 40 NSCLC cases; d) Sections 5 and 6 – contain the discussion and conclusions.

2. Related Work: Random Walk Algorithm

The RW represents an image as an undirected graph $G=(V,E)$. In the graph, each image pixel x is represented as a node $v_x \in V$ and the edge $e_{xy} \in E \subseteq V \times V$ connects two nodes v_x and v_y. To indicate affinity of the nodes v_x and v_y, each edge e_{xy} is assigned a weight w_{xy} that is defined as the Gaussian weighting function in (Grady 2006) as (1):

$$w_{xy} = \exp(-\beta \left\| g_x - g_y \right\|^2) \tag{1}$$

where β is the weighting parameter and g_x is the intensity value at pixel x. The higher the weight, the stronger the nodes connection, and greater ease with which the walker travels through the edge. Let $\Pi_t = [\pi_t(x)]_{1\times N} = \Pi_{t-1}\Gamma$ be the status of a walker at each node v_x at a time t, where $\Gamma = [\tau_{xy}]_{N\times N}$ is the transition matrix defined as (2)

$$\tau_{xy} = \frac{w_{xy}}{\sum w_{xy}} \tag{2}$$

Given a set of labeled nodes $V_L \subseteq V$, which are initially specified with labels L by users, the segmentation is to assign the unlabeled nodes $V_U = V \setminus V_L$ a label $l_k \in L = \{l_1, l_2\}$, where l_1 is the label for the object to be segmented (foreground) and l_2 for the background. RW algorithm solves this problem by calculating the probabilities of reaching labeled nodes V_L from V_U. When the stationary status is reached, $\Pi\Gamma \to \Pi$. By decomposing Γ, we can obtain the probability $\pi_x^{l_k}$ of each unlabelled pixel x belonging to a label $l_k \in L$ by solving (4)

$$\Pi_U = (1-\Gamma_{UU})^{-1}\Gamma_{UL}\Pi_L \tag{3}$$

and each pixel is given a label

$$l(x) = \arg\max_{l_k} \pi_x^{l_k} \tag{4}$$

In the RW algorithm, the weight calculation is essential in generating the probability map of each pixel belonging to foreground or background. Since the weights are calculated on the basis of the intensity changes, RW provides better solutions for weak boundary problems. Since the propagation and location of the initially labeled pixels, usually referred to as seeds, are major factors in the calculation of probability maps, the RW segmentation accuracy may vary with the number and location of the seeds.

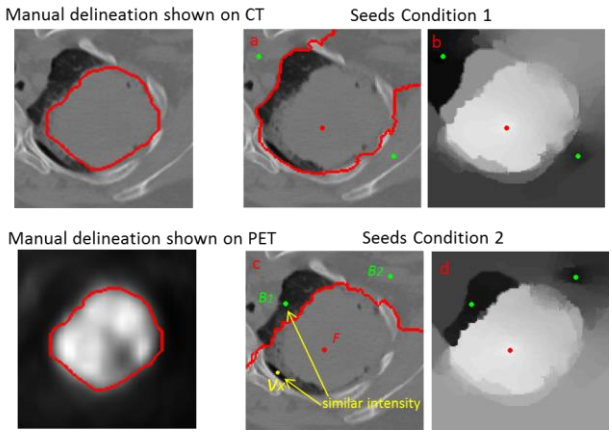

Figure 1: Two different seed inputs and corresponding foreground probability maps of the RW algorithm.

In Figure 1, the foreground probability maps in 1(b) and 1(d) capture the intensity changes; the different seed inputs (red points as foreground seeds; green points as background seeds) lead to variation on foreground (tumor) definitions marked as red curves in 1(a) and 1(c). When compared to the background seeds in 1(a), the background seed B_1 in 1(c) implies that the region with low intensity should be the background. The segmentation, however, incorrectly included the node v_x as the foreground in 1(c) due to the propagation and location of the seeds, while ignoring the absolute density similarity between B_1 and a node v_x to be segmented.

3. Methodology

3.1 Hypothesis and overview of the topo-poly graph model

As discussed above, the conventional RW algorithm neglects the absolute density similarity between the seeds and the unlabeled pixels. In addition, current lung tumor segmentation methods do not take into account the 'high-level' topological information. The topological information is the compact representation of the image. In this paper, we used a contour tree as the topological representation of different regions of the image to provide an inference of the regional relations, such as inclusion/nesting and exclusions that cannot be learnt directly from geometrical information.

Our hypothesis is that a segmentation model, which incorporates the estimated tumor similarity together with topological information, improves tumor delineation from its proximity with similar densities. Hence, we formulate the lung tumor segmentation for PET-CT as a topo-poly graph model that correlates and fuses the structural information from a topology graph with the foreground similarity information from an intensity graph (see Figure 2). With the regional topology changes, the edges of the graph exhibit polymorphic connections to reflect the energy cost when passing through the boundary of different topology regions.

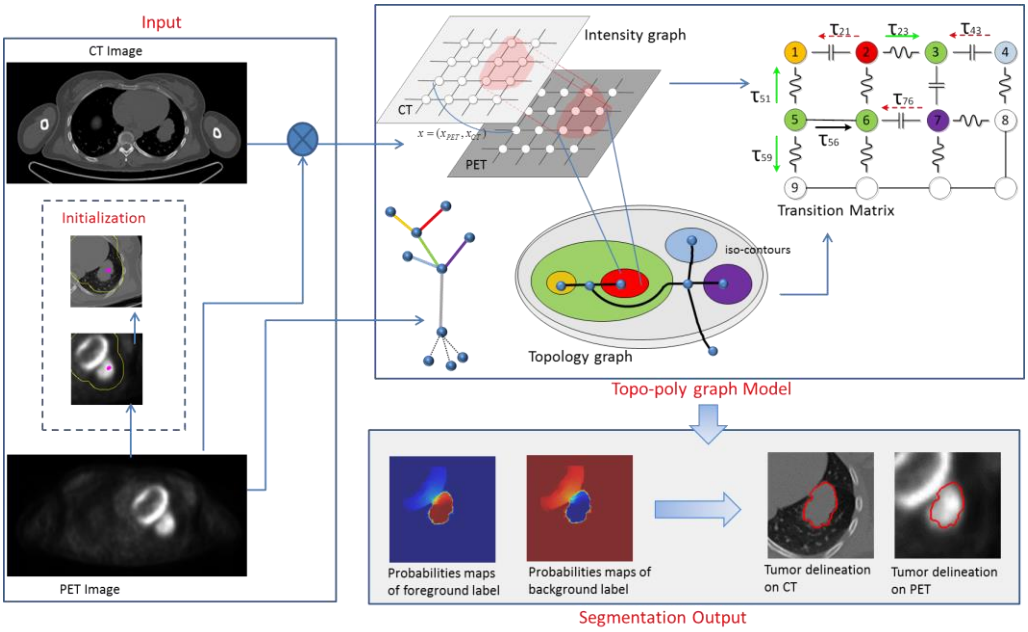

Figure 2: Framework for the topo-poly segmentation model.

3.2 Intensity-based graph

For a pair of input PET-CT images ((I_P, I_C)), the PET image I_P provides functional identification of tumor, and its registered CT counterpart I_C, contains better anatomical depiction. To utilize this complementary information, we estimate the joint foreground similarity p and define an intensity-based graph $G^I = (V^I, E^I)$, where $v_x^I \in V^I$ is a node and x corresponds to a PET-CT pixel pair (x_P, x_C). The weights of the intensity graph are defined to incorporate intensity changes with the similarity information.

Weights definition: To retain the RW feature of capturing the sharp intensity changes at the clear boundary and to take into account the joint foreground similarity p, we define a weight w_{xy}^I for the edge $e_{xy}^I \in E^I$ as:

$$w_{xy}^I = \exp(-\beta \left\| p_x \cdot g_{x_C} - p_y \cdot g_{y_C} \right\|^2) \tag{5}$$

Joint foreground similarity: Based on the corresponding foreground seeds set $F_C = \{x_C \mid l(x_C) = l_1\}$ and $F_P = \{x_P \mid l(x_P) = l_1\}$, we estimate the joint tumor membership p_x for each unlabeled node v_x^I in PET and CT images and p_x is defined as the posteriori likelihood:

$$p_x = \phi(g_{x_C}, F_C) \cdot \phi^{\lambda}(g_{x_P}, F_P) \tag{6}$$

where g_{x_C} and g_{x_P} are the normalized intensity at pixel x_c in CT and x_p in PET respectively; $\phi(g, F) = (2\pi\alpha)^{1/2} \exp(-(g-\mu)^2 / (2\alpha))$ is the Gaussian probability density function where μ and α are the mean and standard deviation estimated by F; and $\lambda \in [0, 1]$ is a free parameter to moderate the PET contribution. When $\lambda = 1$, PET and CT contribute equally to the membership estimation and when $\lambda = 0$, the PET intensity is not considered.

As shown in Figure 3(a) and 3(b), the absolute density differences of v_x and foreground seed F are salient in both PET and CT images and therefore the joint foreground similarity of v_x is low (Figure 3(e)). In contrast, the high PET density difference of v_y and F complements similar CT densities,

which leads to relatively lower joint foreground similarity at v_y. These small joint similarities contribute to tuning down the corresponding weights; and hence in turn correctly classify v_x and v_y as background.

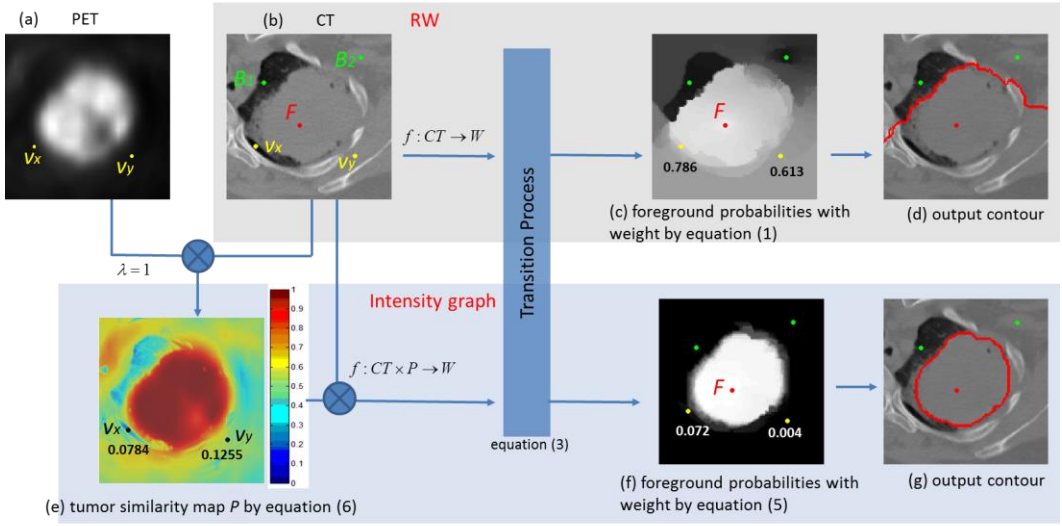

Figure 3: Graph weight calculation in RW (top row) and intensity graph (bottom row).

3.3 Topology graph

The joint foreground similarity, which is based on the joint density differences of PET and CT, allows a better foreground classification when compared to RW. When an adjacent or attached non-tumor region, however, shares similar densities with foreground seeds on both PET and CT (see Figure 4a where the tumor and heart are in close proximity), the joint foreground similarity will be high and will not aid correct separation. Hence, prior knowledge indicating the structural or topological (for example neighboring) relationship of the regions would provide guidance for correct separation of these regions.

We have constructed a weighted topology graph $G^T = (V^T, E^T)$ from PET to reflect the topology relations of inclusion and exclusion on the basis of a contour tree (Carr, Snoeyink et al. 2003) (Zhou 2012). Since PET images are noisy, the direct adoption of a conventional contour tree may include redundant information, which makes data analysis impractical (Zhou, Xiao et al. 2013). Thus we focus only on abstracting topology relations of the region of interest (ROI) that is extracted from PET on the basis of given foreground seeds and SUV thresholding. We detect the local extremes so that we can partition the topology regions. The topology regions are represented as the sets of iso-contours, within the PET ROI. We define the joint point of two iso-contours as a saddle point. A leaf node corresponds to a local extreme; an interior node corresponds to a saddle point; and an edge corresponds to a topology region. On the basis of the topology graph, the topology relations of image regions are derived by the rules as below and shown in Figure 4:

Topology relations

IF (any two edges e_i^T, e_j^T split from a same interior node with stem e_n^T) **THEN**

exclusion relationship - the two corresponding image regions r_i and r_j are neighboring/exclusive regions, i.e., $r_i \cap r_j = \Phi$;

inclusion relationship - r_n includes r_i and r_j , i.e., $r_i \subset r_n$ and $r_j \subset r_n$.

END IF

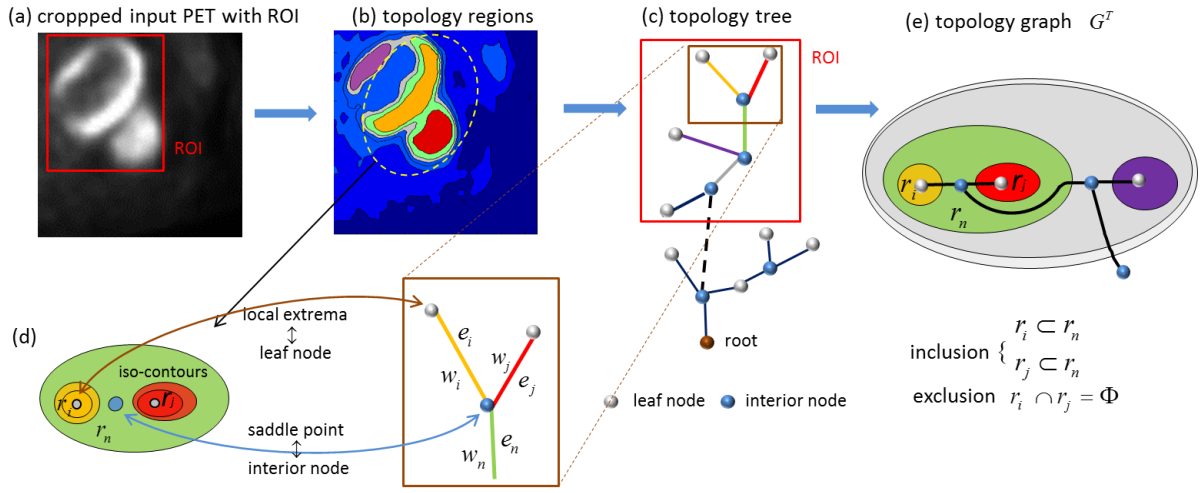

Figure 4: A topology graph for PET foreground ROI. (a) cropped transaxial FDG-PET image with ROI in red bounding box; (b) the corresponding topology regions; (c) corresponding topology tree with edges representing the regions in same colors; (d) cropped topology branches and corresponding topology regions with iso-contours, and interior nodes correspond to saddle points and leaf nodes correspond to local extreme; (e) derived topology structural relations and topology graph.

Weights definition: We calculate the intensity average in a region as the weight for an edge to represent the regional intensity information, and the weight w_i^T for edge e_i^T is

$$w_i^T = \frac{\sum_{r_i} g_p}{N_{r_i}} \tag{7}$$

where r_i is the corresponding region, and N_{r_i} is the number of pixels in r_i.

3.4 Graph model with topology polymorphic connections (Topo-poly graph)

We propagate the topology relations of the regions in $G^T = (V^T, E^T)$ onto $G^I = (V^I, E^I)$ to define a weighted graph $G^* = (V^I, E^*)$. With E^I and E^T as inputs, we define E^* as: $F : (E^I, E^T) \rightarrow E^* = \{e_{xy}^* \mid e_{xy} \in E^I\}$.

Edge definition: With the incorporation of the topology relations as reflected by E^T, e_{xy}^* exhibits topological polymorphism. These polymorphic edges indicate the different types of node connections. The types of connection are determined by the location of the nodes and the topological relations of the regions that the nodes belong to. Polymorphic connections reflect the physical energy costs when crossing the different edges, and hence they help to differentiate the regions that are in close proximity and share similar intensities. We classified the polymorphic connections into: (a) direct connections (see Figure 5(a)); (b) resistance connections (Figure 5(b)); and (c) capacitance connections (Figure 5(c)). The edges are mathematically defined in (8):

$$e_{xy}^* = \begin{cases} e_{xy}^I & \text{if } v_x, v_y \in r_i \quad \text{direct connection} \\ \delta_{xy} \cdot e_{xy}^I & \text{if } v_x \in r_i, v_y \in r_n, r_i \subset r_n \quad \text{resistance connection} \\ e_{xz}^* + e_{zy}^* & \text{if } (v_x \in r_i, v_y \in r_j, r_i \cap r_j = \Phi) \,\& \\ & \quad (v_z^I \in r_n, r_i \subset r_n, r_j \subset r_n) \quad \text{capacitance connection} \end{cases} \tag{8}$$

where $\delta_{xy} = \max(w_i^T, w_j^T)(\min(w_i^T, w_j^T))^{-1}$ is a factor of the energy cost when crossing the two topology edges e_i^T, e_j^T; v_z is an imaginary node to indicate the region r_n. The location of the imaginary node v_z

will not affect the energy cost, which is calculated based on the average intensities of the regions r_i, r_j and r_n. The physical differences between different connections reflect that more energy, for instance, is required when crossing different regions, or passing through an edge of a capacitance connection than direct connection.

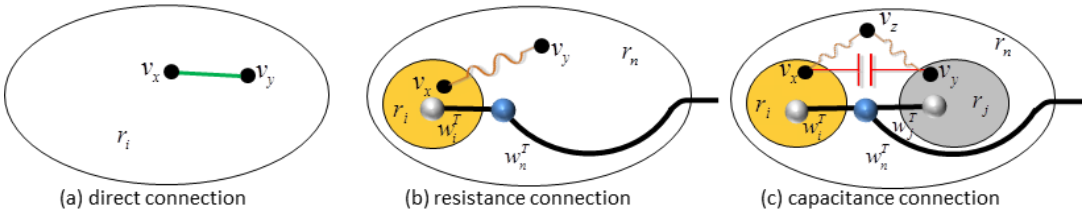

Figure 5: The three node connections.

Weights definition: The weight w^*_{xy} of edge e^*_{xy} is defined as (9) to reflect these polymorphic connections:

$$w^*_{xy} = \begin{cases} w^I_{xy} & \text{direct connection} \\ \delta^{-1}_{xy} \cdot w^I_{xy} & \text{resistance connection} \\ \delta^{-1}_{xz} \cdot \delta^{-1}_{zy} \cdot w^I_{xy} & \text{capacitance connection} \end{cases} \tag{9}$$

The transition process is illustrated in Figure 6, and the topo-poly graph G^* is solved using RW framework.

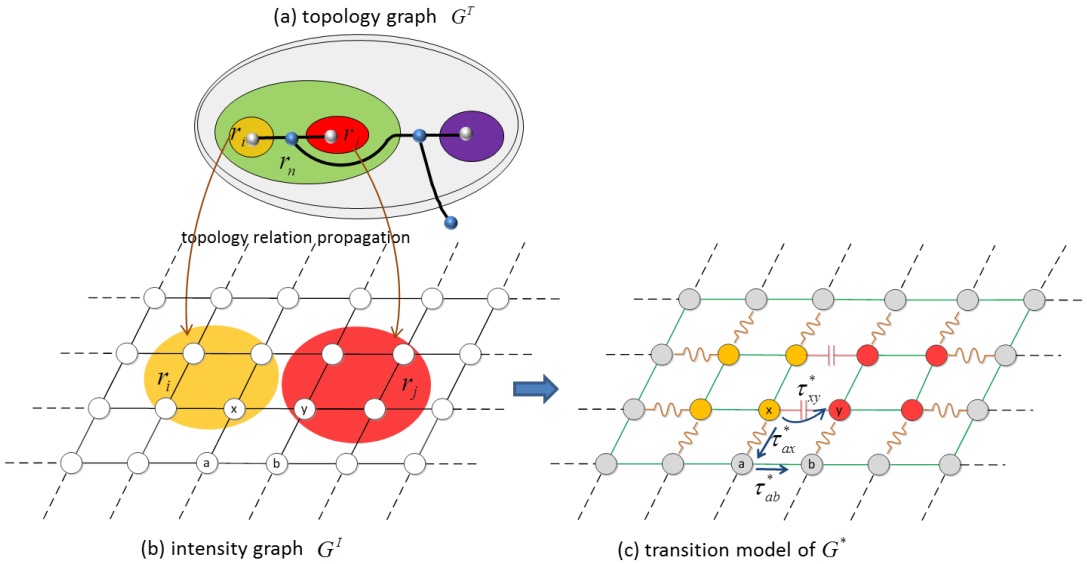

Figure 6: Semantic diagram for topo-poly transition model

3.5 Patient Studies

We analyzed 40 PET-CT studies from patients with NSCLC. The scans were carried out on a Biograph TrueV 64 slice PET-CT scanner (Siemens Medical Solutions, Hoffman Estates, IL, USA). PET data were reconstructed into 168 × 168 matrices with pixel size of 4.07 mm × 4.07 mm. The CT data were reconstructed using a matrix of 512 × 512 pixels with pixel size of 0.98 mm × 0.98 mm. The slice thickness in PET-CT was 2 mm (20 studies) and 3 mm (20 studies). The registration of PET and CT

volumes was based on Insight Segmentation and Registration Toolkit[1] (ITK) affine registration. Mutual information (MI) was used as the similarity metric and the average MI for the 40 patient studies was -0.595 ± 0.0517. The 40 PET-CT studies were separated into two groups: an 'isolated' (n=20) and a 'complex' group (n=20). We defined the 'isolated' group as cases where the lung tumor was located in the lung parenchyma and away from associated structures / tissues in the thorax and the 'complex' group as cases where the tumor abutted / involved a variety of adjacent structures, where the tumor margins were indistinct, and/or had heterogeneous regions of FDG uptake. Among the 20 'complex' cases, 16 cases had tumors abutting the pleura/chest wall and 4 had tumor adjacent to or extending into the mediastinum / pericardial-myocardial region.

Manual delineation was performed by one senior clinical expert (observer 1) and a radiologist (observer 2) using both PET and CT information, and the tumor boundaries were drawn on CT images. The delineation results were used as the "ground truth" (GT).

3.6 Phantom PET-CT studies

Twenty lung PET-CT phantom datasets were collected from the public RIDER[2] collections from the Cancer Imaging Archive. The data were based on a NEMA NU-2 IQ phantom (GE Medical Systems using Ge-68) with the central 'lung' cylinder of the IQ phantom removed. The datasets were part of a research plan for measuring the response to drug or radiation therapy (ITK) and so the decay of the Ge-68 was different between the 20 scans. The target/background ratio was 4:1 with the initial background activity level set to be equivalent to 555 MBq (15 mCi) in a 70 kg patient. The diameters of the 6 spheres were 10 mm, 13 mm, 17 mm, 22 mm, 28 mm and 37 mm. The PET data were reconstructed using a matrix of 128 × 128 with voxel size 2.73 × 2.73 × 3.27mm. The CT data were constructed using a matrix of 512 × 512 with voxels size 0.68 × 0.68 × 2.5 mm.

3.7 Validation methods

To assess the accuracy of the proposed method, we calculated the spatial overlap and shape dissimilarity between the segmentation results and GT by Dice's Similarity Coefficient (DSC) and Hausdorff Distance (HD). DSC was defined as (10)

$$DSC(U_1, U_2) = \frac{2|U_1 \cap U_2|}{|U_1| + |U_2|} \tag{10}$$

where U_1 is the segmented volume, and U_2 is the GT volume. The DSC value is 1 for a perfect segmentation.

HD was defined as

$$HD(U_1, U_2) = \max\{\sup_{i \in S_1} \inf_{j \in S_2} d(i, j), \sup_{j \in S_2} \inf_{i \in S_1} d(i, j)\} \tag{11}$$

where S_1 and S_2 denote the boundary of the segmented volume and the GT volume, and sup represents the least subset element and inf the greatest subset element; *d* is the Euclidean distance between point *i* and *j*. A low HD value indicates high segmentation accuracy.

To compare the volumes of auto-segmentation and manual delineation, we calculated the relative absolute volume differences (RAVD) [3] for each method. The RAVD is calculated as:

$$RAVD = \left| \frac{Vol_1}{Vol_2} - 1 \right| \tag{12}$$

where vol_1 is the volume of the segmentation and vol_2 is the volume of the ground truth. The smaller RAVD, the better segmentation, and RAVD is 0 for a perfect segmentation. However, as discussed[3], the

[1] National Library of Medicine Insight Segmentation and Registration Toolkit (ITK): http://www.itk.org/
[2] https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections
[3] MICCAI grand challenge: http://mbi.dkfz-heidelberg.de/grand-challenge2007/sites/eval.htm

perfect value of 0 can also be obtained from a non-perfect segmentation, as long as the segmented volume is equal to the volume of the ground truth.

3.8 Comparison methods

We evaluated the performance of our method by comparing the segmentation results with the following PET methods: 1) a threshold of 40% SUV_{max}, referred to as RG40; 2) a threshold of 50% SUV_{max} referred to as RG50; 3) an adaptive threshold method (Black, Grills et al. 2004), referred as RGa; and 4) Fuzzy c means (FCM), 5) tumor-customized downhill method (TCD) (Ballangan, Wang et al. 2011). The thresholding methods and FCM only consider PET intensities. The TCD also takes into account the gradient information to improve the tumor delineation and separation. For our graph model we compared it to : 6) RW (Grady 2006) and 7) a high order interactive learning segmentation method (NHLIS) (Kim, Lee et al. 2010). The NHLIS incorporates the regional information from super pixels into graphs. NHLIS and RW were implemented on PET and CT images, and are referred as NHLIS-PT, NHLIS-CT, RW-PT and RW-CT, so we could evaluate the contributions of topology graph and tumor similarity model in our method. All the preceding methods focus on tumor segmentation from monomodality images, i.e., from PET or from CT. As our model utilized complementary PET-CT information, we also compared our method to 8) PET-CT tumor-background likelihood model (referred as TBLM) (Wang, Ballangan et al. 2014). TBLM incorporates the intensity downhill feature in PET as a distance cost into the background likelihood function of CT under a Bayesian framework.

3.9 Initialization and implementation

The algorithms were implemented in MATLAB and a user-input seed was manually provided and the local maximum SUV (SUV_{nax}^{local}) was obtained automatically. The RG40. RG50, RGa, FCM and TCD methods were implemented starting from the SUV_{nax}^{local} .

For our method as well as the other graph-based methods, both the foreground and background seeds were automatically searched as follows:

Algorithm. Initialization

Input: a pair of (I_P, I_C), $SUV_{\max}^{local}$,

foreground and background seeds criterion $(\varepsilon_1, \varepsilon_2) = (0.95, 0.3)$

Output: foreground $F = F_C \cup F_P$ and background seeds $B = B_C \cup B_P$.

Step 1: starting from $SUV_{\max}^{local}$ and perform region growing on I_P

$F_P \leftarrow \{x_p \mid g_{x_p} \geq \varepsilon_1 \cdot SUV_{\max}^{local}\}$

$F_C \leftarrow F_P$ // mapping to I_C

Step 2: starting from contour of F_P and perform region growing on I_P

$temp \leftarrow \{x_p \mid g_{x_p} \geq \varepsilon_2 \cdot SUV_{\max}^{local}\}$, perform image dilation with a disk structuring element

$B_P \leftarrow$ contour of *temp*

$B_C \leftarrow B_P$ // mapping to I_C

4. Experimental Results

4.1 Parameter Setting and Initialization Criterion

To evaluate the robustness of our method with respect to the free parameter $\lambda \in [0, 1]$, we calculated the segmentation accuracy in terms of DSC and HD for the 40 clinical patient studies. As it is plotted in Figure 7, the best results were achieved when $\lambda = 0.2$ for the isolated cases and $\lambda = 0.4$ for the complex cases. When $\lambda = 1$, the isolated cases obtained the lowest accuracy with DSC of 0.826 ± 0.054 and HD

(mm) of 6.228 ± 3.718; and when $\lambda = 0$, the complex cases obtained the lowest accuracy with DSC of 0.811± 0.063 and HD (mm) of 10.726 ± 4.011. Student's t-test ($p > 0.05$) in Table 1 showed that there were no significant differences between the results when $\lambda \in [0.2, 0.4]$. In the experiments λ was set to be 0.2 for the isolated cases and set to 0.4 for the complex cases.

Tabel 1. Student t-test of different λ settings over the patient studies, P(t<=0.05) two-tail

λ comparison	Isolated cases	Complex cases
0.2 VS 0.3	0.730	0.256
0.2 VS 0.4	0.367	0.570
0.3 VS 0.4	0.526	0.100

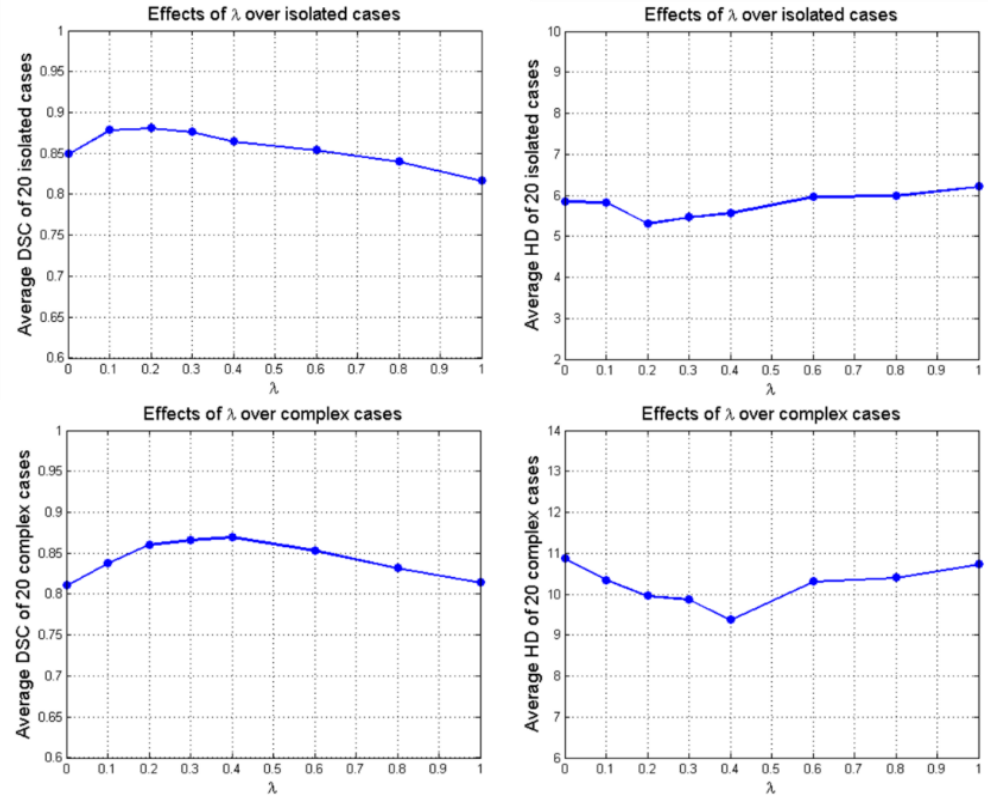

Figure 7. The effects of λ on the segmentation results over isolated and complex cases in DSC and HD

We carried out an evaluation of the seed initialization. We validated and compared the seed criteria of (0.9, 0.4), (0.9, 0.3), (0.95, 0.3), and (0.95, 0.4) on our method and RW-CT. Since the size of the dilation disc for choosing the background seeds was the equivalent radius of the initial foreground region, when we change the criteria (e.g., 90% and 95% SUVmax) for defining the initial foreground region, the size of the disc is changed accordingly. As shown in the Figure 8,in contrast to the consistent segmentation accuracy from our method, RW-CT had varying accuracy over different criteria. This evaluation indicated that our method was not sensitive to the seed initialization criteria.

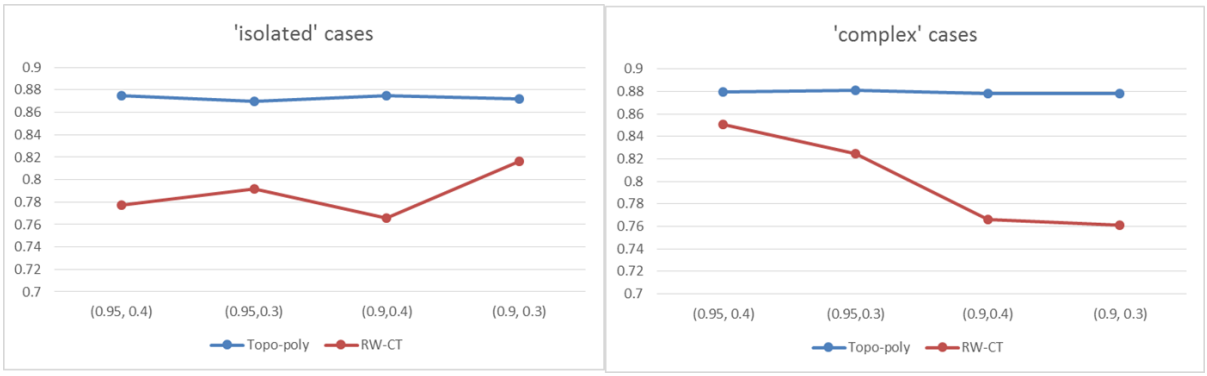

Figure 8. DSC evaluation of seeds criteria

4.2 Evaluation on Isolated Cases

Segmentation results for two isolated cases are shown in Figures 9 and 10. In Figure 9, the contrast between the tumor and surrounding tissue was relatively low in PET and the tumor volume (measured according to GT) was 7.98 ml. FCM based on PET failed to segment the tumor and resulted in leakage.

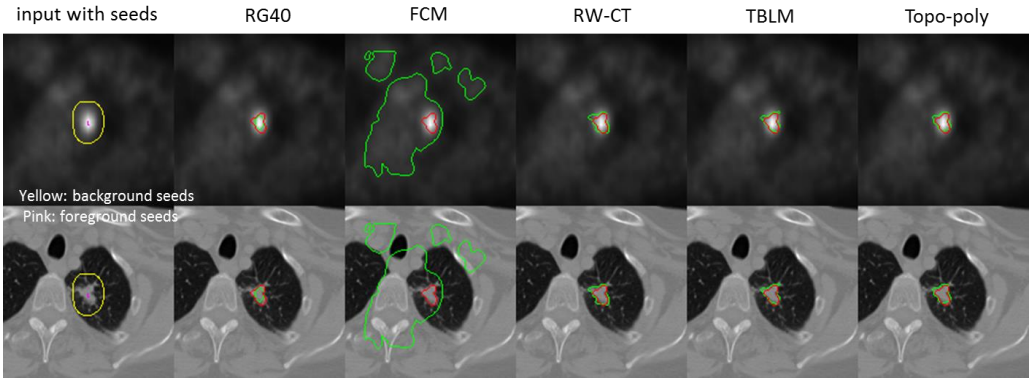

Figure 9: Cropped transaxial FDG PET images on top row show tumor delineation results for an 'isolated' case; corresponding CT is on second row; segmentation results are shown in green and GT in red.

The tumor in Figure 10 was larger with a tumor volume of 53.38 ml and the PET contrast was better. RG40 and FCM, based on PET, resulted in smaller, round tumor delineations. Although RW-CT captured the boundary details in both cases, it included the surrounding "spikes" from the CT which can include normal tissue. TBLM and the topo-poly model achieved better results with the PET-CT information. The DSCs of the topo-poly model were 0.877 for the case in Figure 9 and 0.905 for the case in Figure 10 and the HDs were 1.03 mm for the case in Figure 9 and 1.65 mm for the case in Figure 10.

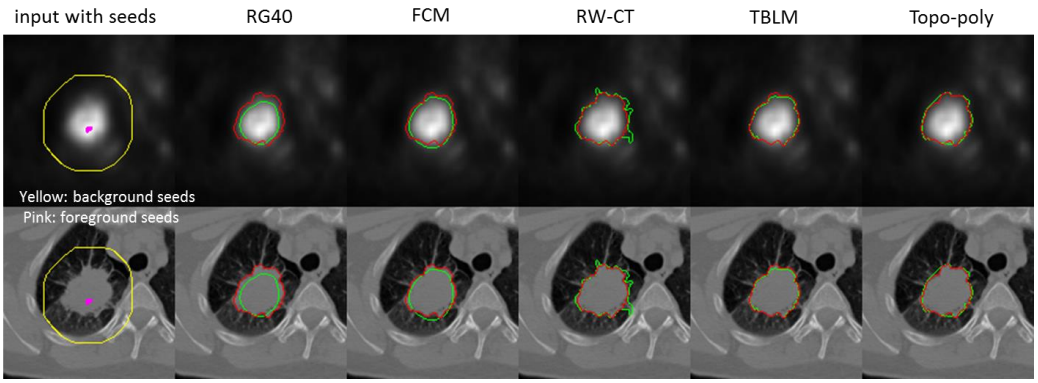

Figure 10: Cropped transxial images show tumor delineation results of an 'isolated' case on PET (first row) and CT (second row); segmentation results are shown in green and GT in red.

Our model had the best results on the basis of DSC and HD (see Table 2). RW on CT was ranked second and then TBLM. The methods using PET generated lower accuracy than methods using CT or PET-CT. The NHLIS resulted in leakage when segmenting both PET and CT images. Figure 11 shows the statistical analysis on the accuracy of the segmentation methods in box plots. The measurements based on median DSC and HD were consistent with the findings in Table 2. For the outlier case for our method (Figure 11), with the lowest DSC and highest HD, the tumor shape was irregular and abutting/involving vessels, and the contrast between the tumor and the background was low on PET. So our model and all the other methods all failed to identify the tumor margins accurately.

Table 2: Overlap measures of eight different methods for the 'isolated' cases

Methods		DSC (mean ± sd)	HD(mm) (mean ± sd)
PET only	RG40	0.622 ± 0.106	15.987 ± 15.809
	RG50	0.578 ± 0.062	11.241 ± 7.964
	RGa	0.551 ± 0.129	20.907 ± 19.423
	FCM	0.555 ± 0.143	33.246 ± 25.989
	TCD	0.669 ± 0.069	9.458 ± 6.894
	NHLIS-PT[a]	0.488 ± 0.095	25.632 ± 23.878
	RW-PT	0.461 ± 0.079	9.698 ± 3.131
CT only	RW on CT	0.825 ± 0.047	5.721 ± 3.274
	NHLIS-CT[b]	0.740 ± 0.060	24.855 ± 39.793
PET-CT	TBLM	0.813± 0.072	7.001 ± 6.442
	Topo-poly	**0.881 ± 0.046**	**5.311 ± 3.022**

[a] 4 cases with leakage (DSC<0.1) have been removed
[b] 9 leakage cases have been removed

Table 3: Tumor volumes measured for 'isolated' cases

Methods	Tumor volume (ml) (mean ± SD)	median	range	ARVD (%) (mean ± SD)
GT	**6.623 ± 12.311**	**3.200**	**0.790 – 53.375**	-
RG40	14.910 ± 35.409	4.126	1.017 – 150.147	230.389 ± 597.700
RG50	4.020 ± 4.089	2.190	0.672 - 16.225	51.298 ±27.390
RGa	28.289 ± 44.943	5.662	2.781 – 149.374	399.443±686.270
FCM	32.565 ± 31.870	15.488	0.744 – 94.551	1040.761±1363.534
TCD	5.896 ± 7.518	3.359	1.016 – 33.194	40.593±34.057
NHLIS-PT[a]	8.980 ±14.102	2.214	0.366 – 49.922	42.436±60.436
RW-PT	2.467 ± 3.621	1.308	0.275 – 15.796	53.724±20.290
RW on CT	8.082 ± 16.024	3.659	0.881 – 69.225	17.013±8.831
NHLIS-CT[b]	38.315 ± 98.948	3.2501	0.595 – 318.108	15.873±14.725
TBLM	5.891 ± 10.162	3.430	0.475 - 44.337	12.830±12.353
Topo-poly	7.151 ± 13.655	3.429	0.807 – 59.135	12.565±8.811

[a] 4 cases with leakage (DSC<0.1) have been removed
[b] 9 leakage cases (DSC<0.1) have been removed

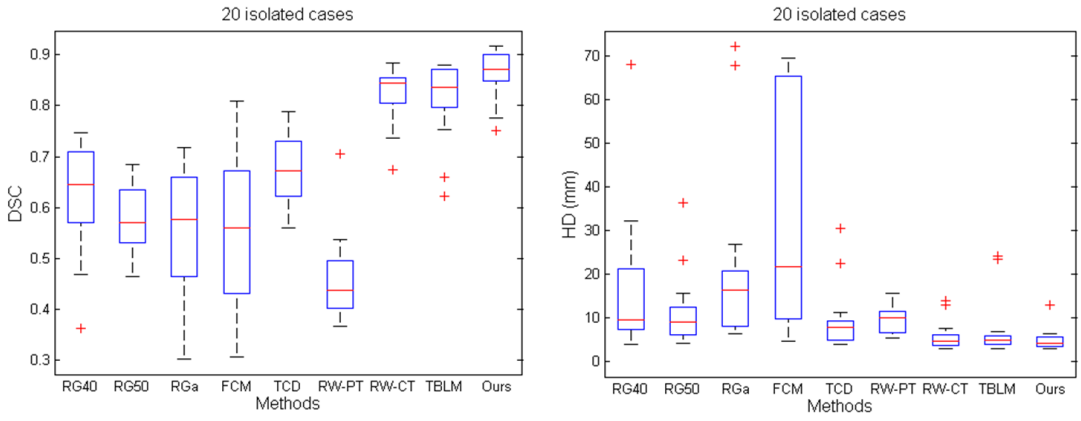

Figure 11: Comparison of DSC and HD for 'isolated' cases using box plots. Blue edges of boxes are the 25th and 75th percentiles; the middle red line is the median; edges of the dashed lines are the extreme data points; red points are outliers.

4.3 Evaluation on Complex Cases

In the example in Figure 12, the tumor with irregular margins lies adjacent to the myocardium, which has similar FDG uptake, and may involve the pericardium. The local tumor extent is difficult to discern on CT and PET. We show by viewing the tumor from different angles that our method was able to accurately separate the tumor from the surrounding structures.

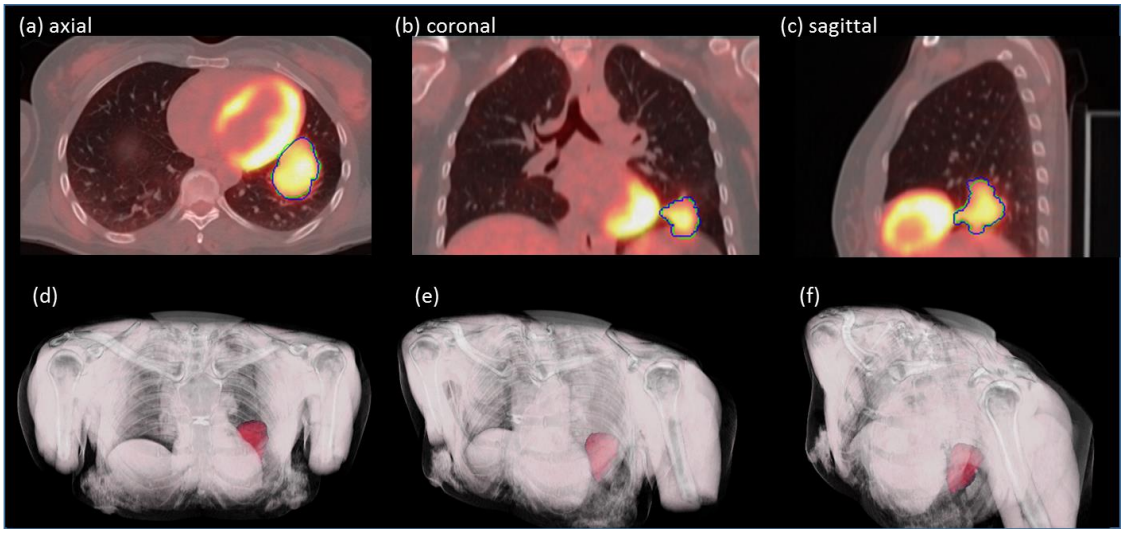

Figure 12: The local tumor extent is difficult to identify on CT and the tumor has similar FDG uptake to the adjacent heart. (a)-(c) axial, coronal, and sagittal planes on fused PET-CT and (d)-(f) the 3D visualization of tumor segmentation by topo-poly model (in red) with three different rotation degrees. Green contours are the segmentation results and blue contours are GT.

The results of graph-based methods and their associated foreground probability maps are seen in Figures 13 and 14. For the case in Figure 13, tumor is attached to the diaphragm and lies immediately adjacent to the underlying liver; it is not clear if the tumor extends through the pericardium into the abdominal cavity. RW-CT incorrectly included the non-tumor region, which had low intensities, as tumor, and in comparison, RW-PT, NHLIS-CT and NHLIS-PT resulted in smaller tumor delineations. Our method provided better tumor delineation with DSC of 0.834 and HD of 2.318 mm.

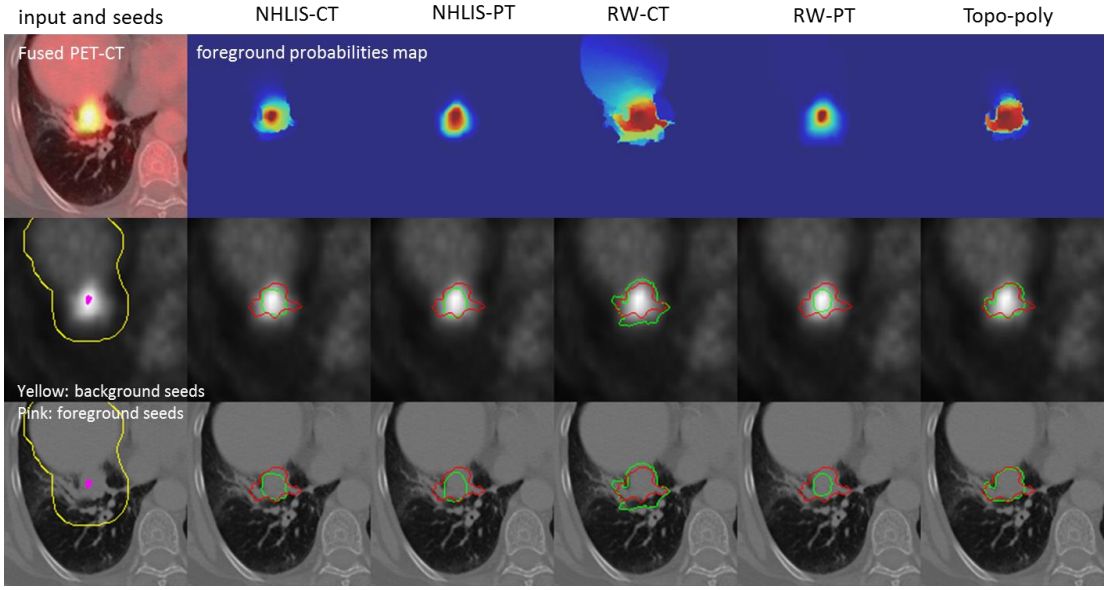

Figure 13: Cropped transaxial image slices; tumor delineation results for a complex case showing graph based methods; the tumor is adjacent to liver and the boundaries on CT are difficult to discern; segmentation results are shown in green and GT in red.

For the case in Figure 14, the large left upper lobe tumor abuts and may involve the adjacent parietal and visceral pleura and the chest wall and it has heterogeneous FDG uptake reflecting regions of necrosis. NHLIS-CT failed in the label propagation, NHLIS-PT and RW-PET excluded the tumor

regions with low SUVs and so were not able to delineate the entire tumor. Although the foreground and background intensities were sufficiently separable, RW-CT failed to separate the background region with this seed initialization. Our method provided better tumor delineation because of its more concentrated probability map with a clear cutoff between foreground and background. The DSC was 0.872 and the HD was 3.122 mm for our method.

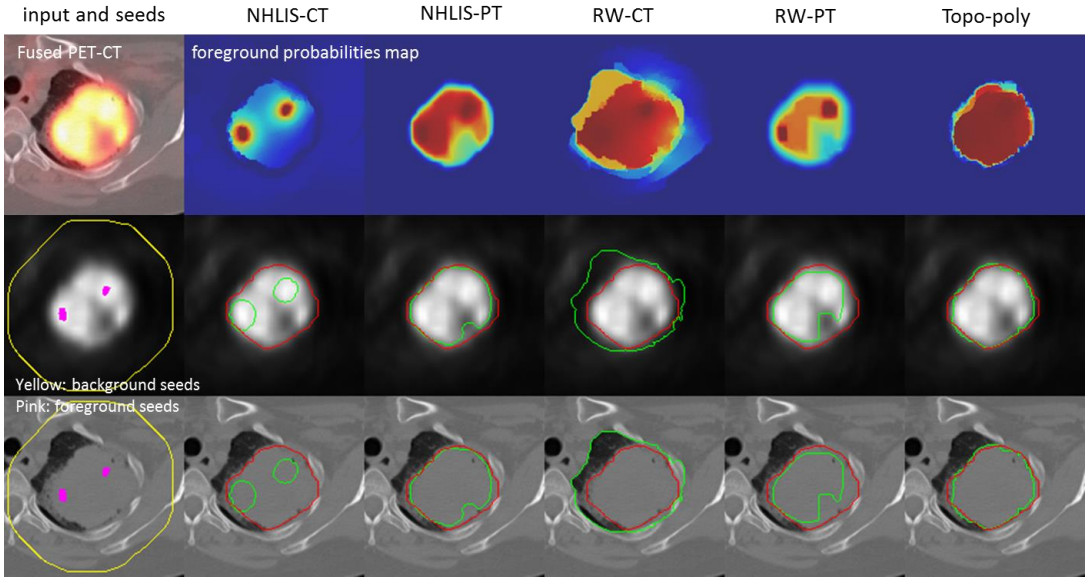

Figure 14: Cropped transaxial image slices; tumor delineation results for a complex case with graph-based methods; tumor boundaries are difficult to discern on CT and the tumor has heterogeneous FDG uptake; segmentation results are shown in green and GT in red.

In Table 4, our method had the best results with the highest mean DSC and lowest HD; TBLM ranked second and then RW-CT. The results in Figure 15 confirm these findings and show that, with the exception of our model, the other methods all had outliers for the cases where the tumor had heterogeneous FDG uptake and/or was adjacent to or involved other tissues.

Table 4. Overlap measures of seven different methods over complex cases

Methods		DSC (mean ± sd)	HD(mm) (mean ± sd)
PET only	RG40	0.651 ± 0.110	17.046 ± 14.842
	RG50	0.562 ± 0.115	18.047 ± 14.333
	RGa	0.567 ± 0.180	23.176 ± 23.944
	FCM	0.614 ± 0.200	23.735 ± 23.447
	TCD	0.707 ± 0.069	14.349 ± 8.667
	NHLIS-PT	0.534 ± 0.163	31.369 ± 43.799
	RW-PT	0.478 ± 0.155	15.045 ± 5.179
CT only	RW-CT	0.792 ± 0.063	12.902 ± 6.991
	NHLIS-CT	0.632 ± 0.159	15.990 ± 6.797
PET-CT	TBLM	0.807 ± 0.093	12.597 ± 7.613
	Topo-poly	**0.870 ± 0.038**	**9.370 ± 3.169**

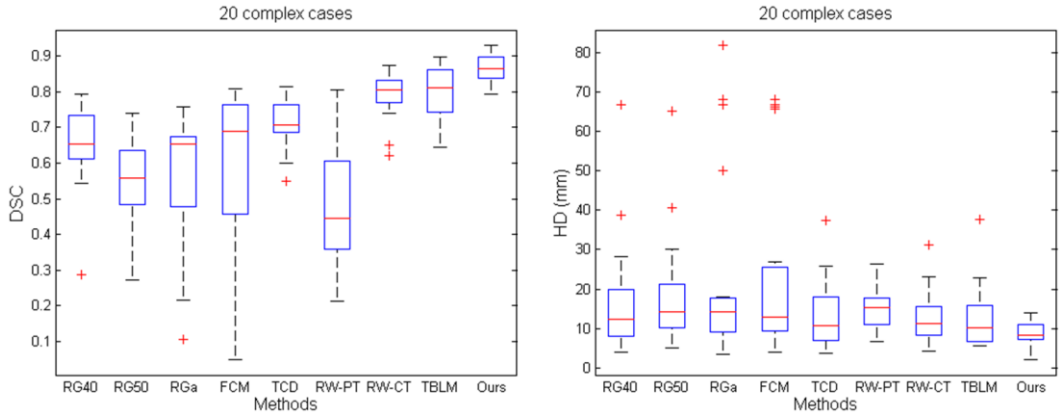

Figure 15: Comparison of DSC and HD for complex cases with box plots; blue edges of boxes are 25th and 75th percentiles, middle red line is median; edges of the dashed lines are extreme data points; red points are outliers.

In Tables 3 and 5, the tumor volumes and the volume difference measured by ARVD show that our method had the best results for the 'complex' cases and was slightly better than TBLM for the 'isolated' cases. We found that the segmented volumes by RGa were consistently higher than GT for the 'isolated' and 'complex' cases. The segmented volumes of RW-PET were consistently smaller than GT for both cases. Student's t-test results ($p<0.05$) in Table 6 show that our method outperformed the other methods for the 'isolated' and the 'complex' groups.

Table 5 Tumor volumes measured for "complex" cases

Methods	Tumor volume (ml) (mean ± SD)	median	range	ARVD (%) (mean ± SD)
GT	38.822 ± 57.036	27.351	1.399 – 264.276	-
RG40	37.764 ± 67.357	14.603	1.082 – 242.941	39.086 ± 47.834
RG50	28.724 ± 52.482	10.752	0.641 – 181.237	53.672± 24.610
RGa	87.877 ± 121.724	37.219	1.531 – 435.191	251.591±616.171
FCM	57.834 ± 75.475	31.609	2.049 – 281.473	274.981 ± 664.620
TCD	33.396 ± 54.624	21.784	0.896 – 255.993	18.172±16.285
NHLIS-PT	25.713 ± 34.383	13.923	0.386 – 138.296	51.888±45.694
RW-PT	16.906 ± 29.592	8.964	0.343 – 136.371	59.088±23.582
RW on CT	38.822 ± 57.036	27.351	1.399 – 264.296	21.033±12.221
NHLIS-CT	15.454 ± 16.822	10.348	0.987 – 67.890	40.560±28.381
TBLM	37.108 ± 57.345	21.938	0.724 – 267.246	18.575±19.296
Topo-poly	32.943 ± 50.266	20.960	1.044- 230.622	13.853±8.194

Table 6 Student t-test our model and the other methods for the patient studies with respect to DSC, P(t<=0.05) two-tail

Methods	Simple cases	Complex cases
RG40	6.85E-08	1.16E-05
RG50	7.78E-14	9.75E-09
RGa	8.74E-07	3.54E-05
FCM	3.49E-07	2.01E-04
TCD	1.07E-09	1.01E-05
NHLIS-PT	3.72E-09	2.4E-07
RW-PT	1.44E-16	2.99E-09
RW on CT	0.00449	1.03E-05
NHLIS-CT	6.1E-05	8.7E--05
TBLM	0.00382	0.00277

4.4 Inter-observer agreement and validation

Since there is inter-observer variability in manual tumor contouring, we validated the auto-segmentation results with respect to the manual segmentation by a radiologist. The inter-observer agreement was measured by DSC overlap, and the average DSC was 0.85 ± 0.059 over 40 datasets.

On the basis of manual segmentation by observer2, the experimental comparison demonstrated that our method consistently achieved the best results. Our method obtained the average DSC of 0.824 ± 0.068 with respect to observer 2 in comparison with achieving average DSC of 0.875 ± 0.059 when using observer1's contouring as ground truth. The DSCs of the comparison methods are summarized in Figure 16.

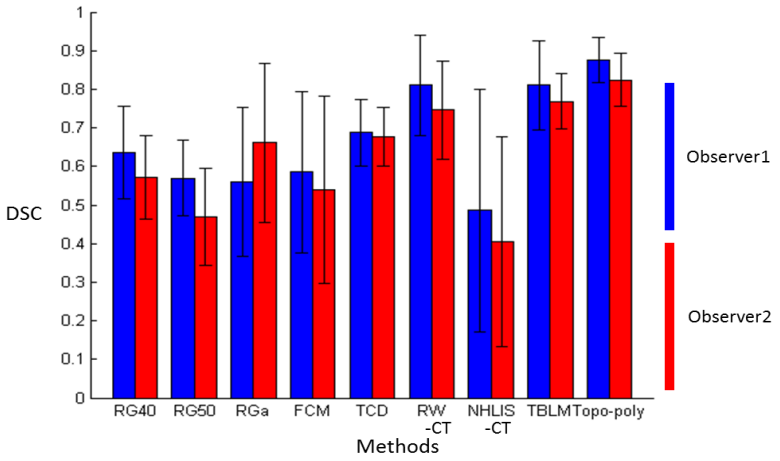

Figure 16. DSCs of the methods with respect to observer1 and observer2 over the 40 patient studies.

4.5 Evaluation on phantom datasets

On 20 phantom datasets, each with 6 simulated tumors of different sizes, our method achieved DSC (mean ± SD) of 0.864 ± 0.085 when λ =0.2 and outperformed the second best method, RW-CT that achieved 0.839 ± 0.108. As shown in Figure 17, RW-CT was not able to delineate the complete biggest tumor.

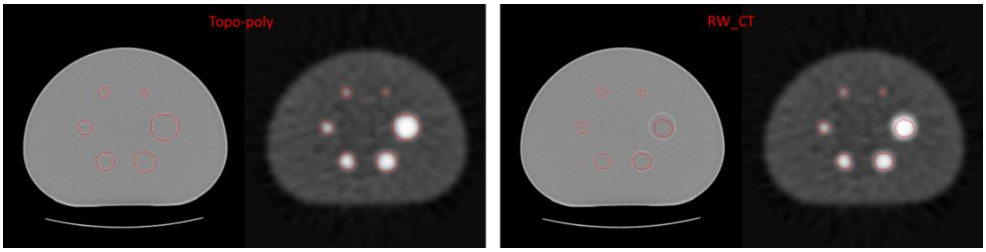

Figure 17. Results of one phantom PET-CT study of our model and RW-CT

5. Discussion

Our main findings are that our model segmented the tumor when the tumor boundaries were not easily discernible on PET and / or CT and it had better segmentation results when compared to other conventional approaches.

The topo-poly graph, with the polymorphic edge connections, was able to outline the entire tumor when the tumor abutted the chest wall and other structures. We explain this finding as follows: a) The joint foreground intensity similarity was embedded in the intensity graph and our method obtained more concentrated final foreground and background probability maps when compared to the graph-based methods. In comparison, RW-CT included the background region as tumor although the intensities were sufficiently separable and the low intensities were already identified as background. b) The topology relations of regions allow our method the ability to separate adjacent different structures with similar intensities. For the case where the tumor was located in the left upper lobe of the lung adjacent to the pericardium and the left ventricle (see Figures 6 and 12), the tumor and the heart have an exclusion relationship and with a capacitance connection in our method, in comparison with the sole direct connection in the conventional graph. Thus in our model, it takes more energy for the foreground label information to reach the heart when crossing these capacitance edges for a more appropriate separation. This also applies for case with heterogeneous FDG uptake and with our method the tumor region with low SUV has a resistance connection with the surrounding high SUV regions, while it is in capacitance connection with the background. So, together with the information provided by the intensity graph, the low SUV region is separated from the background and grouped as the foreground. On the contrary, although NHLIS also incorporated the regional information into the graph model, without the topology structural information, the full connection of the regions misled the grouping of the low SUV region as background and the end result was under-segmentation.

In regard to segmentation, the measurements of spatial overlap (DSC), shape similarity (HD) and volume size (RAVD) showed that our method outperformed the other conventional approaches. When compared to PET-only approaches, our method provided improved results and this related to the poor boundary delineation of the relatively lower resolution PET images as was exemplified in the case in Figure 9. Our method provided better anatomical and functional boundary delineations for small and large tumors and for ‘complex’ cases.

PET images have low signal-to-noise ratios and when the tumor is small, the contrast between the tumor and its neighbouring regions is low. In these instances, as discussed by Soret et al (Soret M et al. 2007) and Kinahan et al (Kinahan et al. 2009), objects smaller than 20 to 30 mm would result in size-dependent loss in accuracy. In such instances it may be difficult for our method to extract appropriate topological information, potentially leading to a reduction in segmentation accuracy. Advances in PET instrumentation including high-resolution reconstruction and time-of-flight capability in current generation PET-CT devices offer the opportunity for us to evaluate our approach in better instrumentation than was available when this study was undertaken. We plan to undertake such work in the future.

6. Conclusion

We present a new graph model for PET-CT lung tumor segmentation. Our model combines joint

foreground density similarity and topology information. We evaluated it on 40 NSCLC patient studies and our results show that we were better able to delineate the tumor margins, in particular, for case where the tumors abutted / involved adjacent structures.

REFERENCES

ACS. (2014). "American Cancer Society How is non-small cell lung cancer diagnosed?", from http://www.cancer.org/cancer/lungcancer-non-smallcell/overviewguide/lung-cancer-non-small-cell-overview-diagnosed.

Ahmedin Jemal, F. B., Melissa M. Center, Jacques Ferlay, Elizabeth Ward and David Forman (2011). "Global cancer satatistics." CA: A Cancer Journal for Clinicians **61**(2): 69-90.

Bagci, U., J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen and D. J. Mollura (2013). "Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images." Medical Image Analysis **17**(8): 929-945.

Ballangan, C., X. Wang, M. Fulham, S. Eberl, Y. Yin and D. Feng (2011). "Automated Delineation of Lung Tumors in PET Images Based on Monotonicity and a Tumor-Customized Criterion." IEEE Transactions on Information Technology in Biomedicine **15**(5): 691-702.

Black, Q. C., I. S. Grills, L. L. Kestin, C.-Y. O. Wong, J. W. Wong, A. A. Martinez and D. Yan (2004). "Defining a radiotherapy target with positron emission tomography." International Journal of Radiation Oncology Biology Physics **60**(4): 1272-1282.

Carr, H., J. Snoeyink and U. Axen (2003). "Computing contour trees in all dimensions." Comput. Geom. Theory Appl. **24**(2): 75-94.

Chen, M., E. Helm, N. Joshi and S. M. Brady (2011). Random Walk-based Automated Segmentation For The Prognosis of Malignant Pleural Mesothelioma. IEEE International Symposium on Biomedical Imaging: From Nano to Macro**:** 1978-1981.

Cui, H., X. Wang, M. Fulham and D. D. Feng (2013). Prior knowledge enhanced random walk for lung tumor segmentation from low-contrast CT images. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society**:** 6071-6074.

Cui, H., X. Wang, J. Zhou, M. Fulham, S. Eberl and D. Feng (2014). Topology constraint graph-based model for non-small-cell lung tumor segmentation from PET volumes. IEEE International Symposium on Biomedical Imaging (ISBI)**:** 1243-1246.

Grady, L. (2005). Multilabel random walker image segmentation using prior models. IEEE Computer Society Conference on Computer Vision and Pattern Recognition**:** 763-770.

Grady, L. (2006). "Random walks for image segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(11): 1768-1783.

Grady, L. and M.-P. Jolly (2008). Weights and topology: A study of the effects of graph construction on 3D image segmentation. Medical Image Computing and Computer-Assisted Intervention**:** 153-161.

Grady, L., T. Schiwietz, S. Aharon and M. Westermann (2005). Random walks for interactive organ segmentation in two and three dimensions: Implementation and validation. Medical Image Computing and Computer-Assisted Intervention. **3750:** 773-780.

Han, D., J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti and X. Wu (2011). Globally Optimal Tumor Segmentation in PET-CT Images: A Graph-Based Co-Segmentation Method. Inf Process Med Imaging. **6801:** 245-256.

ITK. (Jul 21 2014). "Image Similarity Metrics." from http://www.itk.org/Doxygen46/html/ImageSimilarityMetricsPage.html.

Kim, T. H., K. M. Lee and S. U. Lee (2008). Generative Image Segmentation Using Random Walks with Restart. European Conference on Computer Vision**:** 264-275.

Kim, T. H., K. M. Lee and S. U. Lee (2010). Nonparametric higher-order learning for interactive segmentation. IEEE Conference on Computer Vision and Pattern Recognition**:** 3201-3208.

Kirov, A. and L. Fanchon (2014). "Pathology-validated PET image data sets and their role in PET segmentation." Clinical and Translational Imaging **2**(3): 253-267.

Kunz, M., N. Thon, S. Eigenbrod, C. Hartmann, R. Egensperger, J. Herms, J. Geisler, C. la Fougere, J. Lutz, J. Linn, S. Kreth, A. von Deimling, J. C. Tonn, H. A. Kretzschmar, G. Pöpperl and F. W. Kreth (2011). "Hot spots in dynamic(18)FET-PET delineate malignant tumor parts within suspected WHO grade II gliomas." Neuro-Oncology **13**(3): 307-316.

MacManus, M., U. Nestle, K. E. Rosenzweig, I. Carrio, C. Messa, O. Belohlavek, M. Danna, T. Inoue, E. Deniaud-Alexandre, S. Schipani, N. Watanabe, M. Dondi and B. Jeremic (2009). "Use of PET and PET/CT for Radiation Therapy Planning: IAEA expert report 2006–2007." Radiotherapy and Oncology **91**(1): 85-94.

Peng, B., L. Zhang and D. Zhang (2013). "A survey of graph theoretical approaches to image segmentation." Pattern Recognition **46**(3): 1020-1038.

Qi, S., B. Junjie, H. Dongfeng, S. Bhatia, S. Wenqing, W. Rockey, J. E. Bayouth, J. M. Buatti and W. Xiaodong (2013). "Optimal Co-Segmentation of Tumor in PET-CT Images With Context Information." Medical Imaging, IEEE Transactions on **32**(9): 1685-1697.

Simona, B.-H. and E. Peter (2009). "18F-FDG PET and PET/CT in the Evaluation of Cancer Treatment Response." The Journal of Nuclear Medicine **50**(1): 88-99.

Tylski, P., G. Bonniaud, E. Decenciere, J. Stawiaski, J. Coulot, D. Lefkopoulos and M. Ricard (2006). (18)F-FDG PET images segmentation using morphological watershed: a phantom study. 2006 IEEE Nuclear Science Symposium Conference Record**:** 2063-2067.

Vanderhoek, M., S. B. Perlman and R. Jeraj (2012). "Impact of the Definition of Peak Standardized Uptake Value on Quantification of Treatment Response." Journal of Nuclear Medicine **53**(1): 4-11.

Wang, X., C. Ballangan, H. Cui, M. Fulham, S. Eberl, Y. Yin and D. Feng (2014). "Lung Tumor Delineation Based on Novel Tumor-Background Likelihood Models in PET-CT Images." Nuclear Science, IEEE Transactions on **61**(1): 218-224.

Wang, X., H. Cui, C. Ballangan and D. D. Feng (2012). Lung tumor segmentation and separation from PET volumes based on Tumor-Customized Downhill. IEEE-EMBS International Conference on Biomedical and Health Informatics**:** 820-823.

Yang, W., J. Cai, J. Zheng and J. Luo (2010). "User-Friendly Interactive Image Segmentation Through Unified Combinatorial User Inputs." Image Processing, IEEE Transactions on **19**(9): 2470-2479.

Zhou, J. (2012). Gaining Insights Into Volumetric Data Visualization: A Semi-Automatic Transfer Function Generation Approach Using Contour Tree Analyses. Germany, LAP LAMBERT Academic Publishing.

Zhou, J., C. Xiao and M. Takatsuka (2013). "A multi-dimensional importance metric for contour tree simplification." Journal of Visualization **16**(4): 341-349.