Tumor Lesion Segmentation from 3D PET using a Machine Learning driven Active Surface

Payam Ahmadvand[1], Nóirín Duggan[1], François Bénard[2,3], and Ghassan Hamarneh[1]

[1] Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, Canada
{pahmadva,nduggan,hamarneh}@sfu.ca
[2] BC Cancer Agency, Vancouver, BC, Canada
fbenard@bccrc.ca
[3] Department of Radiology, University of British Columbia, Vancouver, BC, Canada

Abstract. One of the key challenges facing wider adoption of positron emission tomography (PET) as an imaging biomarker of disease is the development of reproducible quantitative image interpretation tools. Quantifying changes in tumor tissue, due to disease progression or treatment regimen, often requires accurate and reproducible delineation of lesions. Lesion segmentation is necessary for measuring tumor proliferation/shrinkage and radiotracer-uptake to quantify tumor metabolism. In this paper, we develop a fully automatic method for lesion delineation, which does not require user-initialization or parameter-tweaking, to segment novel PET images. To achieve this, we train a machine learning system on anatomically and physiologically meaningful imaging cues, to distinguish normal organ activity from tumorous lesion activity. The inferred lesion likelihoods are then used to guide a convex segmentation model, guaranteeing reproducible results. We evaluate our approach on datasets from The Cancer Imaging Archive trained on data from the Quantitative Imaging Network challenge that were delineated by multiple users. Our method not only produces more accurate segmentation than state-of-the art segmentation results, but does so without any user interaction.

1 Introduction

Positron emission tomography (PET) is a medical imaging modality that captures functional processes (e.g. metabolism) in the body. Quantitative analysis of tumor tissues in PET is a crucial step towards precise dosimetry and radiation therapy treatment planning. A critical challenge towards quantitative imaging is the ability to distinguish normal activity in the heart, brain and kidneys from abnormal activity due to the presence of malignant lesions. The aim of this paper is to outline a method for automatic localization and segmentation of lesions.

While popular methods for PET segmentation continue to include variants of thresholding [20], more recently sophisticated approaches based on Bayesian-based classification [13], belief function theory [18] and possibility theory [8] have

been proposed. Graph based methods based on the random walker algorithm as well as the maximum-flow method have also been reported [7, 22, 14]. Table 1 summarizes recent methods for PET and PET-CT segmentation according to whether they handle normal activity, are reproducible[4], the level of user interaction required, as well as the number of parameters which need to be set and whether they utilize CT. The reader is also referred to [11] for a recent survey.

In this paper, we propose a new fully automatic approach to lesion delineation in PET that in contrast with previous works in Table 1, is the only method that is (i) fully automatic (i.e. does not require user-initialization or parameter-tweaking when segmenting novel images); (ii) does not require registered CT scans; and (iii) is able to distinguish between radiotracer activity levels in normal tissue vs. the target tumorous lesions. We focus on 18F-FDG, which is by far the dominant radiotracer used in oncological PET imaging studies. This radiotracer provides unique challenges due to physiological uptake in structures such as the brain, kidneys, and other normal organs. Our method is a hybrid machine learning - active surface formulation, which utilizes physiologically meaningful cues (metabolic activity, anatomical position and regional appearance statistics) and a convex optimization guaranteeing reproducible results.

Table 1: Methods proposed for PET/PET-CT segmentation

Method	Technique	Handles * Normal Activity	Reproducible$^{\mp}$	User Interaction	Parameters†	Modality
Song [22] (2013)	Graph based	✗	✗	Seed points and radii for each tumor	7	PET/CT
Ju [14] (2015)	Graph based	✗	✗	Seed points	15	PET/CT
Foster [10] (2014)	Affinity Propagation clust.	✗	◇	Manual correction of registration	7	PET/CT
Lelandais [18] (2012)	Belief-theory with FCM	✗	○	User defined ROI	1	PET/CT
Hatt [13] (2010)	Bayesian based class.	✗	○	User defined ROI	1	PET
Dewalle-Vignion [8] (2011)	Maximum of Intensity Propagations	✗	○	ROI*N‡	1	PET
Abdoli [1] (2013)	Active contours	**	✗	Initial curve	3	PET
Layer [17] (2015)	EM-based GMM	✗	✗	ROI & Seed points	5	PET
Bagci [2] (2013)	Graph based	✗	✗	-	5	PET/CT
Bi [3] (2014)	SVM class.	✓	✗	-	5	PET/CT
Cui [7] (2015)	Graph based	✓	✗	-	3	PET/CT
Lapuyade-Lahorgue [16] (2015)	Fuzzy C-means clust.	✓	✗	-	5	PET
Zeng [24] (2013)	Active surface modeling	✓	✗	-	8	PET
Yu [23] (2009)	kNN class.	✓	✓	-	3	PET/CT
Our Method	Machine Learning & Convex seg.	✓	✓	-	0	PET

*Handles normal activity automatically *i.e.* without user seed points $^{\mp}$Accuracy dependent on parameter choice, which includes size of ROI. †Parameters which were empirically set rather than learned. $^{\circ}$Sensitivity to ROI selection not tested. ‡N is the number of projection directions which was set at 3 in [8]. **Discussion of initial curve placement not included. $^{\circ}$Method is reproducible on the condition that a registered CT is provided and that the method is applied to lung data only.

[4] A method is described as non-reproducible when its results are dependent on image-specific parameter tuning/initialization or other user interaction.

2 Method

At a high level, our method consists of the following steps: Firstly, given our emphasis on head-and-neck (H&N) cancer, the bladder is detected and cropped out by removing all transverse slices inferior to the bladder. To reduce inter-subject image variability, the intensity of all volumes is normalized to zero mean and unit variance. In a one-time training stage, seeds from lesions, background, and regions of normal activity are selected manually, along with the associated class labels. The random forest is trained on features extracted from these labeled seed points. For each novel image, a probability map is then generated based on the output of the classifier, which predicts the label likelihood of each voxel. Finally, an active surface is initialized automatically and is optimized to evolve a regularized volume segmentation delineating only the boundaries of tumor lesions.

2.1 Machine Learning: Features and Classification

Most automatic feature learning methods ignore expert domain knowledge and, to avoid overfitting, require a large number of datasets, which are difficult to acquire in many medical imaging domains. We therefore resorted to designing specific features based on the following problem-specific imaging cues:

Radiotracer Uptake (5 features): The first group of features is based on the standardized uptake value (SUV), which plays an important role in locating tumors. SUV is computed as the ratio of the image-derived radioactivity concentration to the whole body concentration of the injected radioactivity. The activity value of each voxel along with max, min, mean and standard deviation of SUV in a window size 3×3×3 around each voxel are added to the features vector to encode activity information.

Spatially-Normalized Anatomical Position (1 feature): Using the approximate position of tumors as a feature is useful since in each type of nonmetastatic cancer, lesions are located in a specific organ (e.g. H&N in our dataset). However, position coordinates must first be described within a standardized frame-of-reference over all the images. The field-of-view of PET images usually spans from the brain to the middle of the femur; however, this is not always the case. To deal with this variability in scans and obtain a common frame of reference across all images, it is necessary to first ensure anatomical correspondence. The most superior point of the bladder, which is a high-uptake organ, is used as the first anatomical landmark. The second landmark used is the most superior part of the image (i.e. top of the brain). Each image is then spatially normalized along the axial direction based on these two landmarks to obtain a new normalized axial position feature with values ranging between 0 (most inferior) to 1 (most superior). No normalization was needed within the transverse plane itself.

PET Image Texture (8 features): First, four standard 3-D Haar-like features (edge, line, rectangle, and center-surround) are used to capture the general texture pattern of a 10×10×10 region around each voxel in an image [12]. Second,

the following four texture statistics are calculated for each transverse plane: cluster prominence [21], homogeneity [21], difference variance, and inverse difference normalized (INN) [6]. These were found to be particularly useful for distinguishing normal brain from H&N lesion activity along the anteroposterior dimension. The four feature values for a given plane were assigned to each voxel in that plane.

Radiotracer Activity Homogeneity (2 features): Tumor homogeneity is a measure of the uniformity of tumor pixel intensities [15]. The tumor homogeneity is on a vector of all activity values of 3D neighbors centered around a given voxel. We used two homogeneity features at two scales, using window sizes 3×3×3 and 4×4×4, i.e. n=27 and 64, respectively.

Classification In total, the feature vector assigned to each voxel was of length 16. For training, labeled samples (i.e. feature vectors at seed pixels) were collected from the lesion, body, air background, kidneys, heart, and brain regions. These can be considered as six distinguishable classes; however, we consider the heart and kidneys as one class since they are close to each other in terms of location and far from H&N lesions. Having the brain as a separate class was needed to facilitate discrimination between H&N lesions since the brain and lesions are in close proximity. The values of each feature are normalized to zero mean and unit variance over the training samples. A random forest classifier is used to predict the label probability on a voxel-by-voxel basis. The parameters of the classifier were trained using leave-one-image-out cross-validation on the training set. Having more than 50 decision trees did not increase the classification accuracy, and the value for the number of variables randomly sampled at each split was set to the square root of the number of features, as it was automatically set and found to be robust. Although training on thousands of samples using random forests is fast (0.15 sec), predicting the labels of millions of voxels using a MATLAB implementation was time-consuming (up to 20 hours per volume). Instead, we used a fast C implementation [19], with parallel processing distributed on 4 CPU cores (one slice per core). This resulted in a reduction in prediction time by a factor of 50 to reach the average running time of ~15 minutes. Fig. 1 shows sample probability maps generated by our trained random forest.

2.2 Convex Segmentation with Learned Likelihoods

To produce a final segmentation of the lesion tissue, the posterior model produced by the random forest classifier is included as a data (likelihood) term into the convex segmentation formulation of Bresson et al. [4]:

$$E(u) = \int_{\Omega} |\nabla u(x)|\, dx + \int_{\Omega} u(x)(P_{obj}(x) - P_{bg}(x) + C_A)dx \qquad (1)$$

where $u(x)$ in [0,1] is the segmentation label field, Ω is the image domain, the first term is the boundary regularization term, and the second is the data term. P_{obj} is the probability of a lesion and $P_{bg} = 1 - P_{obj}$ (i.e. P_{bg} groups together the likelihood probabilities of all non-lesion classes), and C_A is a constant that

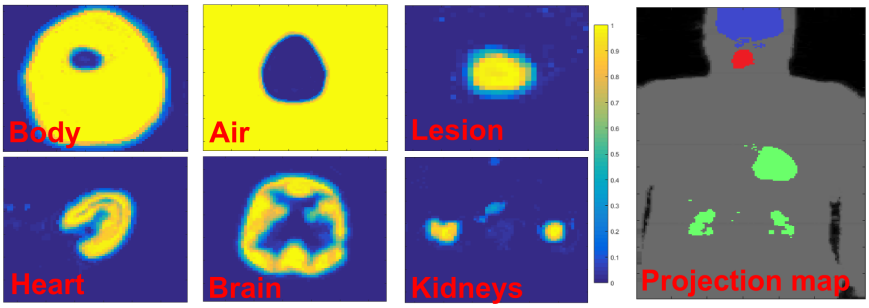


Fig. 1: Example of class probabilities. Images on the left: Probability maps over different classes in the transverse plane (different scales used for clarity). Right: Maximum probability projection.

penalizes the area of the segmentation surface and is used to constrain the size of small unconnected components (e.g. small lymph nodes). The contour is automatically initialized as a "box-shell" around the border of the 3D volume. Given the convex formulation, the algorithm converges to the same global solution with any initialization. One hundred (100) segmentation update iterations were enough for convergence for all our experiments.

3 Results

Experimental Setup: We evaluated our approach on the H&N cancer collection provided by The Cancer Imaging Archive (TCIA) [5, 9]. Ten images from this collection were selected by the Quantitative Imaging Network (QIN) team for the QIN H&N challenge, in which all lesions and lymph nodes with high activity were delineated by experts. We chose these 10 images for training and tested on 15 new images with one or two lesions in each image from the same collection. The training data came from Siemens Biograph Duo scanner. The reconstruction parameters were 2 iterations/ 8 subsets (iterative reconstruction algorithm: OSEM) with a 5 mm smoothing Gaussian filter and voxel size 0.354×0.354×3.37 mm. For testing, data from the original scanner as well as additional data from Siemens Biograph 40 scanner was used (also with OSEM reconstruction algorithm, parameters 4 iterations, 8 subsets) with a 7 mm smoothing Gaussian filter and voxel size 0.339×0.339×5 mm. Six manual segmentations (three users, two trials each) are provided by TCIA for all 25 images. Training seed voxels were collected from the lesions, background, and regions of normal activity, for a total of 1108 seeds. The seed voxels and the manual segmentations of the training data (only) were used to tune hyperparameters and select features; these hyper-parameters were fixed across all test images.
Quantitative Segmentation Results: Our segmentation method was validated against manual segmentations using the Dice similarity coefficient (DSC), Jaccard index, false positive (FP), and false negative (FN) rates. These results

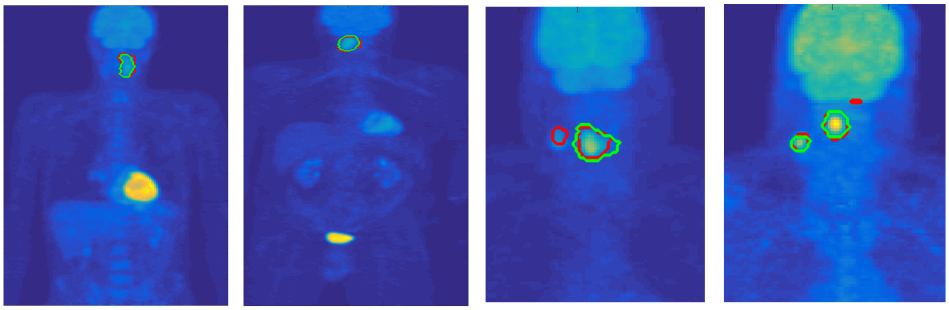
are presented in Table 2. Different combinations of classes for training the classifier prior to segmentation were evaluated (Rows 1-5). Note that Row 6 of Table 2 reports the result of the performance of our method without the final segmentation step (i.e. a voxel is assigned to the class with highest random forest probability). Note also that DSC, Jaccard, FP, and FN are the average values obtained from comparing the automatic segmentation with the 6 manual delineations (3 users, 2 delineations each). We compared our method with the state-of-the-art work of Foster et al. [10] using their publicly available PET-only segmentation software (Row 7). In Table 2 we observe that our proposed method, with five different classes and convex segmentation, outperforms [10]. Not only did our method achieve higher DSC and Jaccard and lower FP and FN, but our method is also fully automatic and reproducible. On the other hand, the software of Foster et al. required selecting a 3D region of interest around the lesions and tweaking parameters until the best performing result is obtained, an extra burden that took an experienced user (knowledgeable in PET imaging and understands the meaning of the underlying parameters) on average 5 minutes per image. Row 8 is the average agreement computed over all scans between manual segmentations produced by the three expert users where each combination of users was included i.e. (user1,user2), (user1,user3) etc.

Segmentation Agreement: We evaluated the inter- and intra-user agreement in lesion delineation and compared it to results from our automated method and those obtained from expert users. We found that user2 has the highest intra-user variability with an average DSC of 0.844, and that user3 agreed the least with the two other users with an average DSC of 0.787. Our fully automated method had an average DSC agreement of 0.741 with other users, i.e. our method falls short by only 4% from performing as well as expert user3.

Quantitative Segmentation Results: Fig. 2 shows examples of segmentation results from four cases. In the third case, our method segmented a lesion that was missed in the manual segmentation. In the fourth case, a small segmentation leakage into the inferior part of brain is observed.

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

	Method	DSC	Jaccard	FP	FN
1	[Air + Body + Kidney + Heart + Brain] [Lesion]	0.71	0.57	0.66	0.03
2	[Air + Body] [Kidney + Heart + Brain] [Lesion]	0.71	0.56	0.60	0.04
3	[Air] [Body] [Kidney + Heart + Brain] [Lesion]	0.73	0.59	0.41	0.05
4	[Air + Body] [Kidney + Heart] [Brain] [Lesion]	0.72	0.57	0.57	0.04
5	[Air] [Body] [Kidney + Heart] [Brain] [Lesion]	0.74	0.60	0.46	0.04
6	[Air] [Body] [Kidney + Heart] [Brain] [Lesion] w/o segmentation	0.71	0.56	0.70	0.03
7	Foster et al. (2014) [10]	0.71	0.58	1.18	0.05
8	Average pairwise agreement between 3 expert segmentations	0.80	0.67	0.27	0.17



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

Fig. 2: Qualitative segmentation results. The PET image is rendered using maximum intensity projection. Our proposed segmentation is shown as red contours, while an example manual segmentation is shown in green. Note in (c) that our method captures a valid lesion missed by the user. In (d), we see an example of segmentation leakage into the inferior part of the brain.

4 Conclusion

We present the first work that fully automates the segmentation of lesions relying solely on PET. Our method is able to isolate abnormal lesion activity from the background and from other tissue regions of normal radiotracer uptake. Our method is a convex segmentation technique that is guided by learned likelihood terms. The learning is based on a classification model trained on anatomically and physiologically meaningful cues. The use of convex formulation together with a trained classifier to learn all parameters precludes the need for human interaction (i.e. initialization and parameter tuning) and results in fully reproducible results on new test images (obtained from a different scanner with different reconstruction parameters). Our approach outperforms a recently published state-of-the-art method for this application and differed in average Dice similarity coefficient DSC by just 4% compared with an expert user. In future work, the approach will be extended to other radiotracers of interest in oncology. It will also be extended to incorporate anatomical information from CT, when available.

Acknowledgements. Funding provided by the Canadian Institutes of Health Research (OQI-137993).

References

1. Abdoli, M., et al.: Contourlet-based active contour model for PET image segmentation. Med. Phys. 40(8), 082507:1–12 (2013)
2. Bagci, U., et al.: Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. Med. Image Anal. 17(8), 929–945 (2013)

3. Bi, L., et al.: Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies. In: MICCAI. pp. 569–576 (2014)
4. Bresson, X., et al.: Fast global minimization of the active contour/snake model. J. Math. Imaging Vision 28(2), 151–167 (2007)
5. Clark, K., et al.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging 26(6), 1045–1057 (2013)
6. Clausi, D.A.: An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Remote Sens 28(1), 45–62 (2002)
7. Cui, H., et al.: Primary lung tumor segmentation from PET–CT volumes with spatial–topological constraint. Int J Comput Assist Radiol Surg 11(1), 19–29 (2015)
8. Dewalle-Vignion, A., et al.: A new method for volume segmentation of PET images, based on possibility theory. IEEE Trans. Med. Imag. 30(2), 409–423 (Feb 2011)
9. Fedorov, A., et al.: DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. PeerJ 4, e2057 (2016)
10. Foster, B., et al.: Segmentation of PET images for computer-aided functional quantification of tuberculosis in small animal models. IEEE Trans. Biomed. Eng. 61(3), 711–724 (March 2014)
11. Foster, B., et al.: A review on segmentation of positron emission tomography images. Comput. Biol. Med. 50, 76–96 (2014)
12. Haralick, R.M., et al.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. (6), 610–621 (1973)
13. Hatt, M., et al.: Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. Int J Radiat Oncol Biol Phys 77(1), 301–308 (2010)
14. Ju, W., et al.: Random walk and graph cut for co-segmentation of lung tumor on PET-CT images. IEEE Trans Image Processing 24(12), 5854–5867 (Dec 2015)
15. Kumar, A., et al.: A graph-based approach for the retrieval of multi-modality medical images. Med. Image Anal. 18(2), 330–342 (2014)
16. Lapuyade-Lahorgue, J., et al.: Speqtacle: An automated generalized fuzzy c-means algorithm for tumor delineation in PET. Med. Phys. 42(10), 5720–5734 (2015)
17. Layer, T., et al.: PET image segmentation using a gaussian mixture model and markov random fields. EJNMMI Physics 2(1), 1–15 (2015)
18. Lelandais, B., et al.: Segmentation of biological target volumes on multi-tracer PET images based on information fusion for achieving dose painting in radiotherapy. In: MICCAI. pp. 545–552 (2012)
19. Liaw, A., et al.: Classification and regression by randomForest. R news 2(3), 18–22 (2002)
20. Nestle, U., et al.: 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. J. Nucl. Med. 46(8), 1342–1348 (2005)
21. Soh, L.K., et al.: Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Trans. Geosci. Remote Sens. 37(2), 780–795 (1999)
22. Song, Q., et al.: Optimal co-segmentation of tumor in PET-CT images with context information. IEEE Trans. Med. Imag. 32(9), 1685–1697 (Sept 2013)
23. Yu, H., et al.: Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. Int J Radiat Oncol Biol Phys 75(2), 618–625 (2009)
24. Zeng, Z., et al.: Unsupervised tumour segmentation in PET using local and global intensity-fitting active surface and alpha matting. Comput Biol Med 43(10), 1530–1544 (Oct 2013)