



# A review on segmentation of positron emission tomography images



Brent Foster, Ulas Bagci\*, Awais Mansoor, Ziyue Xu, Daniel J. Mollura

Center for Infectious Disease Imaging, Department of Radiology and Imaging Sciences, National Institutes of Health (NIH), Bethesda, MD 20892, United States

## ARTICLE INFO

### Article history:

Received 17 August 2013

Accepted 16 April 2014

### Keywords:

Image segmentation

PET

SUV

Thresholding

PET-CT

MRI-PET

Review

## ABSTRACT

Positron Emission Tomography (PET), a non-invasive functional imaging method at the molecular level, images the distribution of biologically targeted radiotracers with high sensitivity. PET imaging provides detailed quantitative information about many diseases and is often used to evaluate inflammation, infection, and cancer by detecting emitted photons from a radiotracer localized to abnormal cells. In order to differentiate abnormal tissue from surrounding areas in PET images, image segmentation methods play a vital role; therefore, accurate image segmentation is often necessary for proper disease detection, diagnosis, treatment planning, and follow-ups. In this review paper, we present state-of-the-art PET image segmentation methods, as well as the recent advances in image segmentation techniques. In order to make this manuscript self-contained, we also briefly explain the fundamentals of PET imaging, the challenges of diagnostic PET image analysis, and the effects of these challenges on the segmentation results.

Published by Elsevier Ltd.

## Contents

1. Introduction	77
2. Background on PET imaging and segmentation	79
2.1. Radiotracers	79
2.2. Quantitative evaluation of radiotracer uptake in PET images	79
2.3. What affects SUV measurements?	79
2.4. Challenges in segmentation of PET images	80
3. Manual segmentation: ground truth construction and segmentation evaluation	80
3.1. Ground truth construction	80
3.2. Segmentation evaluation	80
3.3. Difficulties in manual segmentation	81
4. Thresholding-based methods	81
4.1. Fixed thresholding	82
4.2. Adaptive thresholding	82
4.3. Iterative thresholding method (ITM)	84
4.4. Optimal choice of thresholding, partial volume and reconstruction effects on thresholding	85
5. Stochastic and learning-based methods	85
5.1. Mixture models	85
5.2. Fuzzy locally adaptive Bayesian (FLAB) method	86
5.3. Clustering/Classification of PET image intensities	86
6. Region-based segmentation methods	87
6.1. Region growing	88
6.2. Graph-based methods	88
6.2.1. Graph-cut	88
6.2.2. Random walk	89
7. Boundary-based methods	89
7.1. Level set and active contours	89
7.2. Gradient-based methods	90

\* Corresponding author. Tel.: +1 301 402 6455.

E-mail address: [ulas.bagci@nih.gov](mailto:ulas.bagci@nih.gov) (U. Bagci).

8. Joint segmentation methods . . . . . 90  
 9. Discussion . . . . . 92  
 10. Conclusions . . . . . 92  
 Conflict of interest statement . . . . . 93  
 Acknowledgments . . . . . 93  
 References . . . . . 93

**1. Introduction**

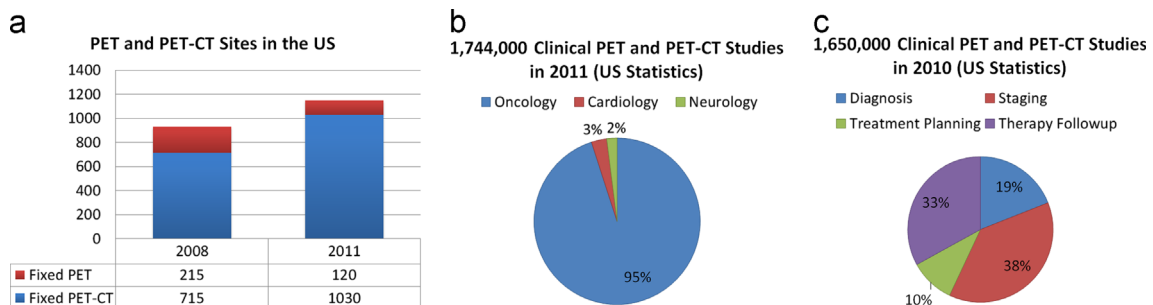
Structural imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) are widely utilized in clinical practice to examine anatomical abnormalities caused by disease. The three dimensional (3D) images produced by these techniques usually give detailed structural information about one's anatomy that can be used for diagnostic and therapeutic purposes [1]. However, structural imaging is not well suited for pathology detection applications where *cellular activity* is more significant than anatomical features [2]. The need for functional characterization leads researchers to develop PET scanners, which provide molecular information on the biology of many diseases. When combined with CT or MRI, utilizing both functional (PET) and structural information leads to a higher sensitivity and specificity than is achievable using either modality alone. Although the sensitivity of PET scans is usually much higher than conventional structural images, anatomical information from another modality (CT or MRI) is still needed to properly interpret and localize the radiotracer uptake and the PET images are somewhat limited due to low resolution. Hence, there is a frequent need for assessing functional images together with structural images in order to localize functional abnormalities and distinguish them from normal uptake of PET radiotracers, which tend to normally accumulate in the brain, heart, liver, kidneys, etc. [3–5]. PET-CT imaging and more recently MRI-PET have been used to combine complementary diagnostic information from different imaging modalities into a single imaging device, removing the need for registration [6]. Using these scanning techniques, disease can be labeled and identified such that an earlier diagnosis with more accurate staging for patients may potentially be delivered [7].

Some of the statistics for the use of PET imaging in the U.S. is summarized in Fig. 1(a). Over 1,700,000 clinical PET and PET-CT studies were reported nation-wide for 2011 only. Compared to single PET imaging, the use of PET-CT is relatively higher and continuing to increase. PET imaging is mostly used for (i) diagnosis, (ii) staging, (iii) treatment planning, and (iv) therapy follow-up, in different fields of medicine such as (1) oncology, (2) cardiology, and (3) neurology (Fig. 1(b)). PET is widely used in staging and follow-up therapy in oncology applications (Fig. 1(c)). For instance, radiation therapy, as a common cancer treatment in oncology, aims to target the boundary and volume of abnormal tissue and irradiates

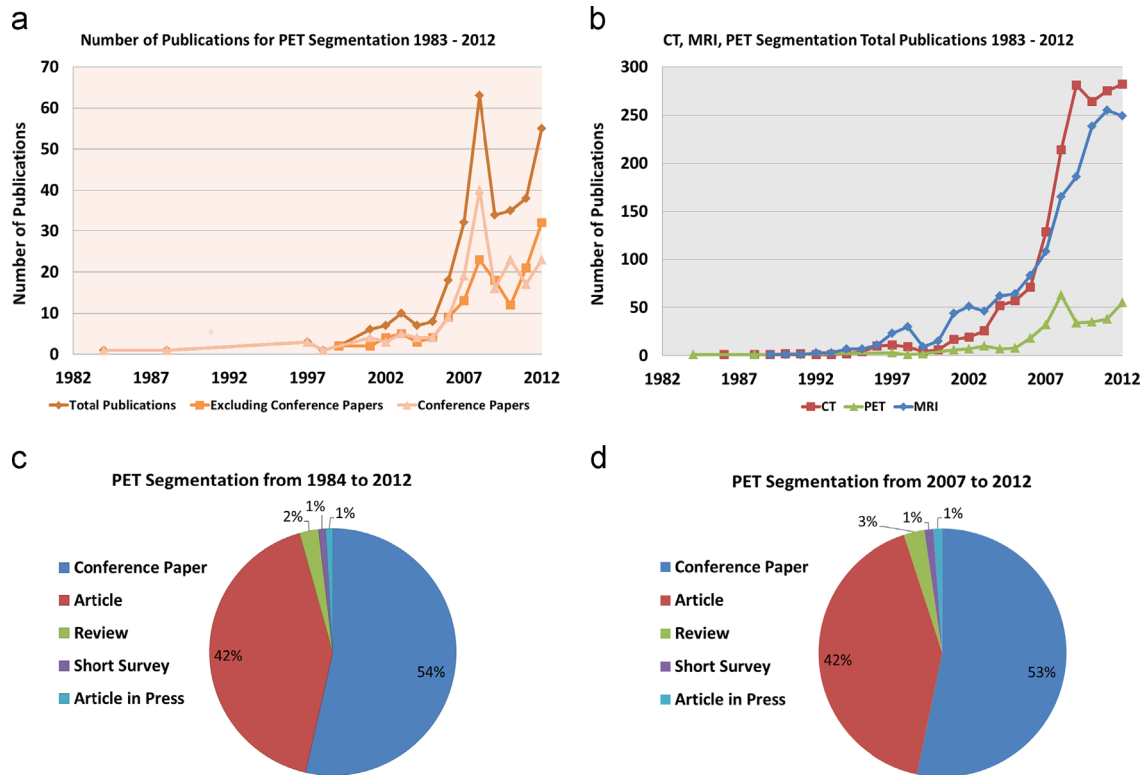
the targeted area with a high dosage of radiation, intending to eliminate all cancerous cells. In practice, the determination of this boundary (i.e., delineation) should be kept as small as possible to minimize damage to healthy tissue, but the boundary must ensure the inclusion of the entire extent of the diseased tissue [2]. PET is also used in cardiac applications such as quantifying blood flow to the heart muscle and quantifying the effects of a myocardial infarction [8]. More recently, PET has been used for imaging inflammation and infection in the lungs [9] with <sup>18</sup>F-FDG because this glucose analog localizes to activated and proliferated inflammatory cells. The new norm in clinical practice is acquiring PET-CT images instead of a single PET scan to take advantage of the functional and structural information jointly.

In pre-clinical and clinical applications, physicians and researchers use PET imaging to determine *functional characterization* of the tissues. Owing to this, clinical trials are now placing a greater reliance on imaging to provide objective measures in before, during, and after treatment processes. The functional morphology (the area, volume, geometry, texture, etc.) as well as activity measures – such as standardized uptake value (SUV) of the tissues—are of particular interest in these processes. Accurately determining quantitative measures enables physicians to assess changes in lesion biology during and after treatment; hence, it allows physicians to better evaluate tumor perfusion, permeability, blood volume, and response to therapy. Among these measures, functional volume (i.e., the volume of high uptake regions) has been proven useful for the definition of target volumes [11]. Therefore, an accurate image segmentation method, other than the conventional region of interest (ROI) analysis, is often needed for diagnostic or prognostic assessment. This functional characterization has a higher potential for proper assessment due to recent advances in PET imaging. Indeed, this higher potential has renewed interest in developing much more accurate (even globally optimal) segmentation methods to turn hybrid imaging systems into diagnostic tools [11]. Specifically, after the adoption of multi-modal imaging systems (i.e., PET-CT, MRI-PET), optimal approaches for precise segmentation and quantification of metabolic activities were crucial.

For the literature search, we used Pubmed™, IEEEExplore™, Google Scholar™, and ScienceDirect™ and listed all the relevant articles from 1983 to March 2013. Our search also included the methods specifically developed for MRI and CT for comparison



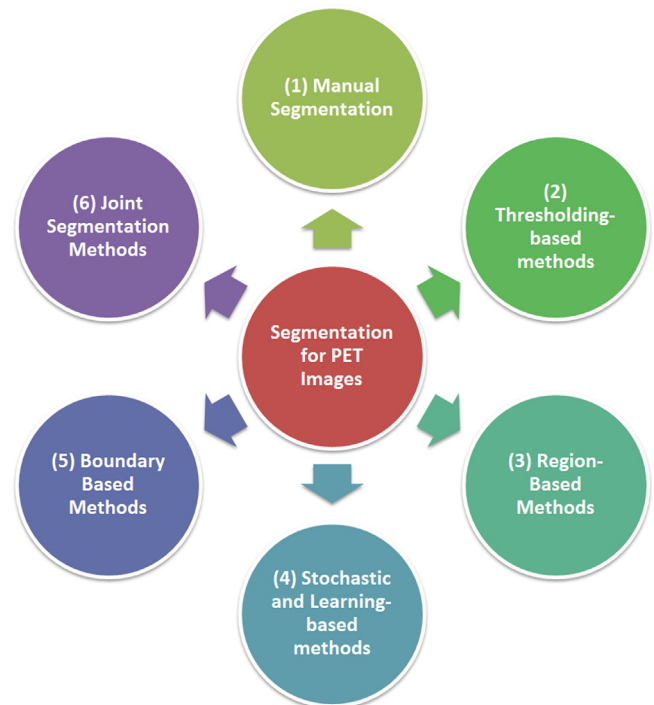
**Fig. 1.** A summary of PET technology used in the U.S. is shown in (a) [10], (b) gives the breakdown of clinical PET and PET-CT studies in 2011 by the branch of medicine and (c) demonstrates 2010 PET technology used in the U.S. for oncology applications, in which PET has been used for mostly staging and follow-up therapy.



**Fig. 2.** Analysis of publications pertaining to PET image segmentation methods and their applications is shown (from 1983 to 2012). Journal and conference publications are shown in (a). A comparison of modality dependent image segmentation methods published for MRI, CT, and PET is shown in (b). Further categorization on the published papers has been conducted in (c) and (d) from 1984 to 2012 and from 2007 to 2012, respectively.

(Fig. 2). The number of publications for PET image segmentation is further separated by publication type (conference, journal, and total) in Fig. 2(a). As a reflection of the improvements in multi-modality imaging technology (PET-CT and MRI-PET), there was a dramatic increase in the number of publications in 2008 and 2011. For a comparison, Fig. 2(b) shows how the number of publications in PET image segmentation methods compare to the number of CT and MRI based segmentation methods in the literature. Notably, the number of PET image segmentation publications has always been lower than both CT and MRI and was significantly lower before 2007. Fig. 2(c) gives the breakdown on the number of publications for segmentation techniques for PET images from 1984 to 2013. We also noted that only 2% of the articles were review papers and almost half of the total articles are journal papers (42% journal publications and 54% conference publications). For the last 6 years, Fig. 2(d) shows a snapshot of publication types from 2007 to 2013, during which the dramatic increase of PET image segmentation publications was observed. It appears that the growing interest in PET and hybrid imaging will further accelerate the methods for segmentation and quantification of lesions.

In this work, we systematically review state-of-the-art image segmentation methods for PET scans of *body images*, as well as the recent advances in PET image segmentation techniques. In order to have a complete review on the topic, the necessary knowledge of the physical principles of PET imaging are also given, along with the source of the challenges for segmentation inherent to PET images in Section 2. The state-of-the-art segmentation methods for PET images, their comparison, and recently developed advanced PET image segmentation methods are extensively analyzed in later sections, and the methods are divided into the following groups for clarity: manual segmentation and ground truth reconstruction (Section 3), thresholding-based (Section 4), stochastic and learning-based (Section 5), region-based (Section 6),



**Fig. 3.** An overview of the categories of PET segmentation methods: Manual segmentation, thresholding-based, region-based, stochastic and learning-based, boundary based, and joint segmentation methods.

boundary-based (Section 7), and multi-modality methods (Section 8). These categories are shown in Fig. 3. Due to the large number of segmentation methods, we have categorized the state-of-the-art methods into intuitive groups for easier comprehension and better

contrasting of the methods. Finally, discussions are made in Section 9, followed by conclusions in Section 10.

## 2. Background on PET imaging and segmentation

### 2.1. Radiotracers

The basic concept of PET is to label a radio-pharmaceutical compound with a biologically active ligand to form a radiotracer and inject it intravenously into a patient. The PET scanner then measures the distribution and concentration of the radiotracer accumulation throughout the patient's body as a function of time [12]. To do this, PET utilizes positron emitting radioisotopes as molecular probes so the biochemical process can be measured through imaging *in vivo* [13]. There have been many radiotracers developed and among them *FDG* ( $^{18}\text{F}$  combined with deoxyglucose) is considered the radiotracer of choice in most studies [14]. Metabolically active lesions have up regulation of glucose metabolism. For example, the rapid cell division in cancer cases and the immune response in infectious diseases require high levels of glucose. Therefore, labeling glucose with  $^{18}\text{F}$  renders these lesions detectable using PET imaging because the *FDG* accumulates in these areas [14]. Meanwhile, a large number of new compounds are also becoming prospects for PET imaging which have some advantages over *FDG* such as tracers that do not accumulate in the heart/kidney. However, *FDG* still remains the most commonly used radiotracer in the clinical routine for body imaging [15]. At the time of this writing, there is no reported study in the literature mimicking the differences of segmentation accuracy caused by using different radiotracers in PET imaging. Therefore, in this manuscript, an evaluation of image segmentation methods are assumed to be independent of the choice of radiotracer.

### 2.2. Quantitative evaluation of radiotracer uptake in PET images

A quantitative assessment of changes in *FDG* uptake in PET images is required for accurate diagnosis and assessment of treatment response, whereas a qualitative assessment of PET images is usually sufficient for the detection of lesions [16]. Qualitative assessments using PET images are often conducted visually by expert radiologists and nuclear medicine physicians [17], while various semi-quantitative and quantitative methods such as *SUV*, tumor-to-background ratio (*TBR*), nonlinear regression techniques, total lesion evaluation (*TLE*), and the Patlak-derived methods are currently undergoing extensive exploration [3]. Among these metrics, *SUV* is the most widely used quantification index for PET imaging because it gives a physiologically relevant measurement of cellular metabolism [13,16]. *SUV* standardizes the intensities of the PET images, and it is simply defined as the tissue concentration of a tracer measured by the PET image intensity at any point of time,

followed by a normalization with the injected dose, the patient's size, and a decay factor which depends on the particular radiotracer type used during the imaging [18,13]. The explicit formulation of the *SUV* computation is

$$SUV_{\chi} = \frac{C(t)}{D/\chi}, \quad (1)$$

where  $\chi$  is either body weight ( $\chi=BW$  (in g or kg)) or lean body mass ( $\chi=LBM$ ), depending on the type of *SUV* being computed [19].  $D$  is the amount of injected dose (Bq) and  $C(t)$  is the total radioactivity concentration in a given tissue at time  $t$ , and it can be directly computed from a ROI. Exploration of *SUV* and its alternative measures are outside the scope of this review; however, readers are encouraged to refer to a comprehensive review on this subject [20].

### 2.3. What affects SUV measurements?

There are physiological, physical, and procedural factors affecting the *SUV* computation. Table 1 briefly explains factors affecting *SUV* calculation, their effects, and methods proposed to balance these effects. Physiological factors include body composition (fat, weight), blood glucose concentration, and kidney function. Among these, *SUV* is quite sensitive to body weight, especially for obese patients. High blood glucose concentration can also be problematic with diabetic patients when measuring *SUV*. Therefore, many techniques were developed to reduce uncertainties in the *SUV* measurements.

Physical factors consist of the partial volume effect (PVE), reconstruction and smoothing of the images, and respiratory motion (or organ/lesion motion) artifacts. PVE and respiratory motion artifacts lead to an underestimation of *SUV* for smaller lesions. Various methods have been developed for correcting PVEs; some of these are reported in Table 1. The literature has shown that repeated *SUV* measurements of the same patient may differ by up to 30% simply from measuring and analyzing variables that influence *SUV* computation [18,36].

It is worth noting that the computation of  $SUV_{max}$  can be improved by removing statistical outliers by defining  $SUV_{max}$  differently [37,38]; however, redefining *SUV* based on post-processing steps such as a parabolic fit around the maximum intensity may only correct the *SUV* marginally. So, instead of fitting the data to a predefined curve, averaging the most intense ("hottest") voxels, which have the highest concentration of photons, together with  $SUV_{max}$  reduces the influence of a noisy outlier [36,39–42]. In [39], the repeatability of  $SUV_{max}$ ,  $SUV_{mean}$  and several ways of averaging the hottest voxels together were investigated and showed a reduced variability by a factor of 2.7 when the top 10 hottest voxels were averaged together to form an alternative way of computing  $SUV_{max}$ .

**Table 1**  
Methods for correcting physical and physiological factors influencing *SUV* Computation.

Physiological factors	Effects	Corrective measure	References
Body composition	<i>SUV</i> in obese patients overestimates <i>FDG</i>	Use of lean body mass ( $SUV_{LBM}$ ) or body surface area ( $SUV_{BSA}$ )	[19,21,22]
Blood glucose concentration	Reduced <i>FDG</i> uptake in tissues with increasing glucose levels	Control blood glucose levels before administering <i>FDG</i>	[23–25]
Uptake period	Increase of <i>SUV</i> over time in malignant tissues	Standardize the time of image acquisition	[26–28]
Respiratory motion	Reduction of $SUV_{max}$ up to 7–159%	Respiratory gating or 4D reconstruction	[29–31]
Attenuation correction and reconstruction methods	Underestimation of <i>SUV</i> with highly smoothed reconstruction by roughly 20%	Standardize reconstruction algorithm	[32,33]
PVE	Underestimates <i>SUV</i> in lesions with diameters less than 2–3 times spatial resolution of scanner	Adopt an optimal PVE factor	[34,35]



In addition to these factors, an ROI is needed for *SUV* computation, and the methodologies used for defining the ROI around a lesion can significantly affect the *SUV* based metrics. Therefore, in order to extract morphological and functional information from PET images, the ROI needs to be identified “precisely”. In other words, a precise segmentation is needed because even small errors in segmentation can distort the calculation of the *SUV* measurements by altering the region’s margins. Furthermore, inter- and intra-operator variability can be considerably high in defining these ROIs, affecting quantitation and possibly diagnostic decisions.

#### 2.4. Challenges in segmentation of PET images

Without loss of generality, image segmentation can be thought of as two related tasks: *recognition* and *delineation* [43]. Recognition is the process of determining “where” the object is and to distinguish it from other object-like entities in the image, and delineation is the act of defining the spatial extent of the object region in the image [44]. In the recognition process, high uptake regions are observed and identified by clinicians. These rough areas of where the objects are located in the image are considered as ROIs, though this process can be automated as well [45]. In delineation, the second step of segmentation, the aim is precise separation of uptake regions from the background and non-significant uptakes [43,46,47]. Some of the intrinsic and extrinsic factors that significantly affect PET image segmentation are as follows:

- (i) resolution related issues,
- (ii) large variability in the shape, texture, and location of pathologies,
- (iii) noise.

These factors increase the difficulty of segmentation in multiple ways. For example, the low resolution and high smoothing decreases the contrast between objects in the image, and boundaries between nearby objects often become unclear. Several additional factors can be counted under resolution related issues. For instance, patients are sometimes unable to hold their breath during an entire scan, and motion artifacts may occur. These artifacts from breathing blur the images severely [48]. Second, the large variability in shape or texture of the pathologies makes the segmentation problem even more challenging due to difficulty in generalizing the available PET segmentation methods for those cases. Last, noise in PET images is inherently high, leading to further difficulties in image segmentation methods that tune parameters based on the value of the  $SUV_{max}$  as well as methods that use the intensity of an initial “seed” location. As demonstrated in [49], noise affects the segmentation of PET images and is regarded as the most significant contributing factor for not having a reproducible *SUV* measurement. However, some standards and guidelines have been enacted to ensure more reproducible analyses between scans and centers [50,51].

Given the difficulties defined above and the unique challenges pertaining to PET images, there have been considerable improvements in PET image segmentation methods. These improvements are primarily due to the need of accurate and robust quantification tools that have the capability of analyzing multi-modal images in real time. An explosive growth in the use of PET-CT and more recently MRI-PET in clinics facilitates this need.

### 3. Manual segmentation: ground truth construction and segmentation evaluation

#### 3.1. Ground truth construction

An overview of the categories that PET image segmentation methods are classified into is given in Fig. 3. Before introducing the

various PET image segmentation methods as summarized in Fig. 3, it is useful and necessary to know the standard ways of evaluating the accuracy of segmentation for proper comparison. In order to evaluate an image segmentation algorithm, the true boundary of the object of interest should be identified. Unfortunately, there is no ground truth available if histopathologic samples are not available. This is the main challenge for all medical image delineation algorithms. Instead, surrogate truths (or reference standards) are used for measuring the quality of a segmentation algorithm. Using phantom images is one way to create a surrogate truth for measuring the performance of an algorithm. Phantoms have the benefit of knowing the exact dimensions of the object in the image. Additionally, a digital phantom, i.e. a synthetic image, can be constructed where the true boundary is known and imaging characteristics of a specific PET scanner can then be added [52]. However, the human anatomy is far too complex to be accurately represented via phantoms; hence, the use of phantoms is limited in terms of identifying the extent and true performance of the segmentation algorithms.

Another way, the most common one, is to use manually segmented structures and compare those structures with algorithm-generated segmentations in terms of overlap or boundary differences [53]. This strategy is currently the state-of-the-art segmentation method for the evaluation and development of medical image segmentation problems. Although it is important to incorporate as many manual segmentations as possible into the evaluation framework in order to reduce sampling error occurred due to the inherent high variations and inter-observer differences, it is often necessary to statically combine all these segmentations together to form a single ground truth for evaluation. The widely used Simultaneous Truth and Performance Level Estimation (STAPLE) method deals with this problem [54]. STAPLE estimates the ground truth segmentation by weighing each expert observer segmentation depending on an estimated performance level while also incorporating a prior model for the spatial distribution of objects being segmented [54].

#### 3.2. Segmentation evaluation

After creating a surrogate truth, there are three main categories of approaches for this task: quantifying volumetric differences, using estimators derived from the confusion matrix, shape based similarity measures, and regression based statistical methods. The volumetric difference is usually determined by simply computing the absolute percent difference in total volume between two segmentations. Since it is such an intuitive and simple metric, the percent volume difference is commonly used in the clinical literature, but this metric alone does not convey enough information to determine the similarity between two segmentations. For instance, it is highly possible for a segmentation method to produce the same volume as the volume of the ground truth, but the segmentation may still be unsatisfactory (i.e., segmentation leaks into non-object territory with some amount of volume, but still the same volume of ground truth can be obtained). Indeed, more quantitative metrics must be used along side for proper evaluation.

There are several estimators that are derived from the confusion matrix that are commonly used for segmentation evaluation. First, the dice similarity coefficient (DSC) is one of the most widely used quantitative metrics to evaluate segmentation accuracy, and this index allows for the false positives and false negatives to be combined into a single value for easy comparison [55,56]. DSC simply measures spatial overlap (in percentage) between a segmented lesion and the surrogate truth where higher DSC values indicate the *goodness* of an image segmentation. Given that the segmented volume is denoted by  $V_1$  and the surrogate truth is

shown by  $V_2$ , the DSC is computed as follows:

$$DSC(V_1, V_2) = 2 \frac{|V_1 \cap V_2|}{|V_1| + |V_2|} \quad (2)$$

where the overlap of two volumes ( $V_1 \cap V_2$ ) indicates the True Positive Volume Fraction (TPVF) (also called *sensitivity*). The amount of false positive volume segmentation is measured in the False Positive Volume Fraction (FPVF) such that  $100 - FPVF$  is the *specificity* [47,57].

For evaluating the segmentation of complex shaped lesions, boundary-based measures should be used in addition to region based metrics to quantify the shape dissimilarity between the delineated lesion and the ground truth [45,57]. Geometric metrics such as the Hausdorff distance (HD) measure how far two boundaries are from each other [58]. Thus, an accurate segmentation result would achieve a high DSC value (high regional overlap) and a low HD value (high shape similarity). The DSC and HD are commonly reported together in the literature for a more thorough evaluation.

Regression based statistical methods (i.e., Spearman and/or Pearson correlation coefficients) and the simple *mean volume difference* or *relative volume ratio* for the evaluation of segmentation methods are much more common than DSC based evaluations in clinical literature. However, readers should be aware that without having the TPVF (Sensitivity) and  $100 - FPVF$  (Specificity) pair or the DSC value, comparing statistics on the absolute volume difference does not provide complete information on segmentation accuracy. In addition to sensitivity and specificity, receiver operating characteristic (ROC) curves may be used to evaluate the performance of a delineation method by combining sensitivity and specificity of the PET segmentation algorithms for a given uncertainty level [59,60].

In summary, a precise evaluation for a segmentation algorithm should be based on sensitivity and specificity measures (or DSC) and not solely on the absolute volume based statistical evaluations. Unless otherwise specified, we assume that all research papers listed in this review completed sensitivity and specificity comparisons for the proposed segmentation methods, even though some of those studies reported only volume based evaluations. It is also worth noting that DSC or sensitivity/specificity measures are region-based evaluation criteria. For a brief review on segmentation evaluation metrics, readers are encouraged to refer to [47,55].

### 3.3. Difficulties in manual segmentation

Manually drawing a boundary around an object on the image is perhaps the most intuitive and easily implemented way of obtaining ROIs for a given image which makes it the most common method of obtaining surrogate truths, as described previously. However, it suffers from many drawbacks. Manual segmentation is highly subjective and intra- and inter-operator agreement rates are often presented in the literature to indicate the reliability of the obtained surrogate truths and the level of difficulty of the segmentation problem [61–66].

The major drawbacks of manual segmentation are that it is time consuming, labor intensive, and operator-dependent. The high intra- and inter-operator variability of the resulting delineations make the delineation process less precise and unlikely reproducible. In a recent study [72], which involved 18 physicians from 4 different departments, the agreement, defined as a volume overlap of  $\geq 70\%$ , was found only in 21.8% of radiation oncologists and 30.4% of haematologic oncologists. It appears that a partial explanation for this high intra-observer variability for manual segmentation may be attributed to the size of the lesion because smaller lesions (i.e.,  $< 4 \text{ cm}^3$ ) suffer much greater from the partial volume effect [73]. This causes the boundaries of objects to be blurred and unclear, making manual segmentation problematic. Table 2 exemplifies this high variation over a few studies reported

**Table 2**

Manual segmentation variability by expert radiologists – the gold standard: Intra- and Inter-observer percent variability is the average percent variation in the segmentation volume. The reliability coefficient is used to quantifies the consistency among multiple measurements on a scale from 0 to 1, where 1 indicates most reliable, and 0 shows unreliable.

Intra-observer (%)	Inter-observer (%)	References
42–84	44–78	[61]
5	10–18	[64]
$13.4 \pm 17.3$	$16.4 \pm 21.8$	[67]
0–44	0–66	[68]
$10.14 \pm 4.23$	$22.27 \pm 6.49$	[69]
<b>Reliability coefficient</b>		
0.91	0.85	[70]
0.90	0.82	[71]

in the literature. For seven studies, either the inter- and intra-observer variance is reported or the reliability coefficient is reported. The reliability coefficient is very similar to observer agreement rates, and it quantifies the consistency among multiple measurements on a scale from 0 to 1 [74]. Higher reliability means lower inter- and intra-variability. However, it should be noted that reliability does not imply validity. The reliability coefficient can be defined as the proportion of the total variance of interest that represents the true information being sought. More information about reliability coefficient can be found in [75]. Even among these recent highly cited studies, there is no consensus on how variable manual segmentation is as well as the experience level of the experts conducting the manual delineation.

## 4. Thresholding-based methods

Thresholding is a simple, intuitive, and popular image segmentation technique that converts a gray-level image into a binary image by defining all voxels greater than some value to be foreground and all other voxels are considered as background [76]. The thresholding-based PET image segmentation methods utilize the probability of intensities, usually by using the histogram of the image. An intuitive view on this process is that the objects of interest in the PET image, usually referred to as the highest uptake region, are much smaller than the background areas. A smaller area equates to a smaller probability of appearing on the image. Additionally, since the intensity of PET images has some physical meaning, the intensities are somewhat unique for the different tissue types and grouping specific ranges of the intensities for different objects is usually enough for a good segmentation. How to group these intensities together is the challenge, and thresholding is one approach to solve grouping problems.

Due to the nature of PET images (i.e., low resolution with high contrast), thresholding-based methods are suitable because the local or global intensity histogram usually provides a sufficient level of information for separating the foreground (object of interest) from the background. However, there is some uncertainty that cannot be avoided when using thresholding-based methods. Because of the large variability of pathologies, low resolution, inherent noise, and high uncertainties in fuzzy object boundaries, there is no general consensus on the selection of a thresholding level (especially automatic threshold selection). Therefore, an optimal threshold determination remains a challenging task. Despite all these difficulties, thresholding-based methods are still under development for improving the segmentation mechanism towards the optimal boundary extraction. Hence, it is common to see these methods both in pre-clinical and clinical studies. Here, we will review state-of-the-art thresholding-based segmentation methods for PET images and their

**Table 3**  
Example studies using fixed-threshold based PET segmentation.

Segmentation method	Anatomical area	Sample size	Accuracy	References
$T_{42\%}$ $T_{34\%}$	Static phantom Moving phantom	3 elliptical spheres 4 SBR 3 Spheres 3 Motions	Mean volume deviation (%): 8.4% Difference from ideal ranged from 3 to 94 cm <sup>3</sup> for motion volumes of 1.2–243 cm <sup>3</sup>	[80] [81]
$T_{34\%}$	Moving lollipop phantom	1 Sphere with 3 longitudinal movements	Volume deviation from ground truth (%): 1.4 ± 8.1%	[82]
$T_{50\%}$ $T_{50\%}$ $T_{42\%} : < 3 \text{ cm}$	Intact squamous cell carcinoma NSCLC NSCLC	40 Patients 101 Patients 20 Patients	Volume deviation from CT (%): 54.5% Volume deviation from CT (%): 27 ± 3% Determined threshold values such that the volumes were exactly the ground truth	[83] [84] [85]
$T_{24\%} : 3\text{--}5 \text{ cm}$ $T_{15\%} : > 5 \text{ cm}$ $T_{40\%}, SUV_{2.5}$ Manual, $T_{40\%}$ , $T_{50\%}, T_{SBR}$ Manual, $SUV_{2.5}$ $T_{40\%}, T_{SBR}$ $T_{43\%}, SUV_{2.5}$	NSCLC Oral cavity, oropharynx, hypopharynx, larynx NSCLC Rectal and anal Cancer	19 Patients 78 Lesions 25 Lesions 18 patients	Median volume deviation from CT (%): -140%, -20% Mean overlap fraction (CT): 0.61, 0.55, 0.39, 0.43 Mean GTV (cm <sup>3</sup> ): 157.7, 164.6, 53.6, 94.7 Mean radius (cm): 3.03, 3.05, 2.18, 2.52 55.4 ± 18.3, 36.7 ± 38.4 Volume difference compared to manual delineation	[86] [77] [79] [87]
$T_{SUV_{max}}, T_{iterative}, T_{Bgd}, FIT$ $T_{SBR}$	Nonspherical simulated tumors inserted into real patient PET scans NSCLC	41 Lesions 23 Tumors	Mean error volume (%): -50% ± 10%, -40% ± 40%, 4% ± 10%, 24% ± 20% Compared to histopathology: sensitivity: 66.7% and specificity: 95.0% Compared to manual segmentation: Sensitivity: 55.6% Specificity: 88.3%	[88] [89]
$T_{42\%}, T_{50\%}, FCM$ Manual, $T_{SBR}, T_{40\%}$ , $T_{50\%}, SUV_{2.5}$ Manual, $T_{Bgd_{20\%}}, T_{40\%}, T_{SBR}, SUV_{2.5}, T_{40\%}$	Sphere phantoms with diameters 13–37 (mm) High-grade gliomas Esophageal Carcinoma	6 Spheres using 4 scanners 18 patients 96 Tumors	Classification error (%): 42.6 ± 51.6, 20.3 ± 18.5, 27.8 ± 25.6 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 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	[67] [78] [90]
Gradient Based, $T_{SBR}, T_{40\%}, T_{40\%}$	Stage I–II NSCLC	10 Patients	DSC: 66%, 64%, 62%, 65%	[91]

Patients had various number of tumors  $T_{Bgd} = e^{-\alpha} T_{70\%} + \text{mean background intensity } T_{Bgd}$ ,  $= SUV_{Bgd} + \alpha(\%) * (SUV_{max} - SUV_{Bgd})$  FIT: In [88], the FIT method models the tumor as the convolution of actual tumor volume of uniform activity with a 3D Gaussian function describing the local spatial resolution in the reconstructed image. The algorithm is from [92].

comparisons using clinical data. Thresholding techniques can be further divided into several groups (Fig. 4): *Fixed Thresholding*, *Adaptive Thresholding*, and *Iterative Thresholding Method*. Also, we describe the challenges and drawbacks specifically pertaining to thresholding segmentation as well as the effects of the partial volume effect and the effects of different reconstruction algorithms.

#### 4.1. Fixed thresholding

In fixed thresholding, as its name implies, all pixels above an intensity level are assigned to a group, and everything else is considered to be background. This level may be given as an input by an expert, learned in a training set of images of the same type, or derived by analytic expression using realistic phantoms. The object boundary in any PET image is going to contain some amount of fuzziness due to the PVE, resolution related issues, and motion artifacts [76]. Thus, many thresholding methods incorporate some amount of class uncertainty, entropy criteria, between-class variance, and many other types of criterion in order to account for this fuzzy object nature [76]. In many clinical studies, a value such as a *SUV* of 2.5 is set as a pre-defined threshold level to differentiate malignant lesions from benign [48]. Similarly,  $SUV_{max}$  can also be used to separate object information from the background by using a specific percentage of  $SUV_{max}$ , which has the advantage of being normalized between patients. The most common thresholding value chosen in the clinical setting is 40–43% of the  $SUV_{max}$ , but this may not always work well and may need to be adjusted considerably for different PET images,

depending on the image properties, scanner type, reconstruction, image noise, etc. For example, several studies evaluated the commonly accepted thresholding value of 40–43% for segmenting lesions and found that this value was suitable in a broad sense. The authors suggested other thresholding values to obtain the correct boundaries such as 45% [77] and 78% [78,79]. Another drawback of this fixed thresholding approach is the tendency to overestimate the lesion boundaries, particularly for small lesions. Therefore, an adaptation of thresholding with further information or user guidance is often necessary in order to provide a clinically sound delineation. Table 3 shows some of the notable studies in the literature that used a fixed thresholding method for segmenting lesions from various body regions. It lists the thresholding value chosen, anatomical area (or disease type), sample size, and the accuracy reported in the study. As can be seen from Table 3, since there is no consensus on a fixed thresholding method for segmenting PET images, variable results were reported even for similar lesion types.

#### 4.2. Adaptive thresholding

Many fixed thresholding-based segmentation methods use digital or physical phantoms to construct and quantify the relationship between the true lesion volume and the estimated lesion volume, with respect to various image quality metrics to “adapt” the thresholding value for a particular image of interest. The source-to-background ratio (SBR or S/B), mean background intensity, estimated mean lesion intensity, and full width half maximum (FWHM)

of the scanner can be used as potential image quality metrics for this purpose.

In addition to adapting the thresholding value based on image quality metrics, it is also possible to adapt the thresholding value based on the motion artifacts of the PET image. As described previously, it lowers the contrast and mean intensity difference between the object and the background. For instance, some phantom based studies characterized the thresholding level of oscillating spheres instead to static spheres in order to mimic breathing and cardiac motions in [93]. Although adapting the threshold level to accommodate breathing and cardiac motion may have important

applications for segmenting lung cancers or inflammation in the lungs when respiratory gating is unavailable, the method presented in [93] suffers from serious shortcomings such as the requirement of a prior estimation of lesion volume from structural imaging [38]. When a prior estimation of the lesion volume does not exist (possibly due to a lack of a corresponding CT or MRI image), success rates of the delineations are restricted to lesions larger than 4 cm<sup>3</sup>. For small lesions, even with prior anatomical knowledge, the success rate is under the desired clinical accuracy.

Table 4 lists some of the state-of-the-art adaptive thresholding equations in the literature, which were validated using various

**Table 4**  
Examples of adaptive thresholding equations used in segmenting lesions from PET images.

Analytic expressions	Notes	Volume difference <sup>a</sup>	References
$T(\%) = A * e^{-C * V_{ml}}$	Parameters A, C are coefficients computed for each SBR. First to use the SBR to estimate the thresholding level.	8.4%, NA	[80]
$SUV_{CutOff} = 0.307 * SUV_{mean} + 0.588$	Iteratively selected the mean target SUV	21%, 67%	[94]
$T(\%) = 0.15 * I_{mean} + SUV_{Bgd_{mean}}$	$I_{mean}$ is approximation of mean intensity of tumor. $Bgd_{mean}$ is mean intensity in relevant background structure.	NA, 60.1%	[79]
$T(\%) = 59.1 - 18.5 * \log(V)$	Fitted a logarithmic regression curve to threshold values in PET that resulted in the same volume from CT	NA	[85] <sup>b</sup>
$T(\%) = SBR * (SUV_{max} - SUV_{Bgd_{mean}}) + SUV_{Bgd_{mean}}$	Considered influence of difference between target and background intensities	47.3%, NA	[95]
$T(\%) = SUV_{Bgd_{mean}} + (Th_{rel}) * (SUV_{max} - SUV_{Bgd_{mean}})$	$Th_{rel}$ = Relative threshold corresponding to the physical diameter of the source. Used $Th_{rel}$ to approximate the image transfer function at the interface between lesion and background.	Spheres > 15 mm accurate to $\leq \pm 1$ mm	[96]
$T(\%) = 7.8\% / V_{ml} + 61.7\% * SBR + 31.6\%$	Initialization for an iterative thresholding method	10%, 16.6%	[37]
$T(\%) = 90.787 * e^{(-0.0025 * Area)}$	Area < 448 mm <sup>2</sup>	13.8%, NA	[97]
$T(\%) = 0.00154 * Area + 28.77$	Area $\geq 448$ mm <sup>2</sup>		
$T(\%) = \frac{a * mSUV_{70} + b * BG}{SUV_{max}}$	Diameter > 30 mm, a=0.50, b=0.50, Diameter $\leq 30$ mm, a=0.67, b=0.60	4.7%, 7.5%	[98]
$T(\%) = 309.2 - 171.9 * \left(1 - \frac{1}{SBR}\right) - 9.643 * Diameter$ (mm)	Sphere inner diameter $\leq 10$ mm	NA	[99] <sup>b</sup>
$T(\%) = 151.1 - 101.8 * \left(1 - \frac{1}{SBR}\right) - 0.976 * Diameter$ (mm)	Sphere inner diameter > 10 mm		
$T(\%) = 5 + e^{3.568 + 0.197/V + -0.1069 * \log V}$	Considered scan duration and initial injected activity in model. An iterative technique based on Monte Carlo simulation studies	5.1%, 11.1%	[100]
$\ln(T(\%)) = 0.0634x^{1/3} + 0.1202y^{1/3} + 0.7327\ln(z) + 0.0597x^{1/3}\ln(z) - 0.1221y^{1/3}\ln(z) - 0.0248(xy)^{1/3} - 0.9504$	T(%) is normalized to background. x is Volume in cm <sup>3</sup> , y is the motion in mm, z is the SBR	Within 2 mm of CT volume	[101]
$T(\%) = 55.94 - 0.25 * Area$ (mm <sup>2</sup> ) + 56.96 * $\left(1 - \frac{1}{SBR}\right) + 2.82 * FWHM$ (mm)	Considered how motion in moving lung tumors affects the thresholding	Cross sectional area $\leq 133$ mm <sup>2</sup>	NA, 6.62% [102]
$T(\%) = 33.13 + 60.27 * \frac{1}{SBR} + 0.60 * FWHM$ (mm)	Cross sectional area > 133 mm <sup>2</sup>		
	Considered the resolution of the scanner for small objects where the PVE has a significant effect by including the full width half maximum (FWHM) of the point spread function of the scanner		

<sup>a</sup> Volume difference is the difference between the volume found from the adaptive thresholding segmentation and the ground truth. *Left* is the volume difference when method is applied to phantom images while *Right* is the volume difference when the method is applied to patient scans.

<sup>b</sup> These studies used spherical phantoms to determine the thresholding level that gives the true volume and then fitted these values to a curve without testing against test images.



phantoms. The studies in Table 4 are organized based on the same year of publication, with the earliest year at the top and the most recent year at the bottom. Studies in same year are in no particular order. Notably, the analytic expressions grow increasingly complex as the years advance with earlier studies only considering the estimated lesion volume while later studies also take into consideration the resolution of the scanner and various image quality metrics. The volume difference between the segmentation found using the analytic expression and the ground truth is reported. This volume difference is divided between the volume difference when the method is applied to phantom images or patient images for a better comparison, if reported in the study.

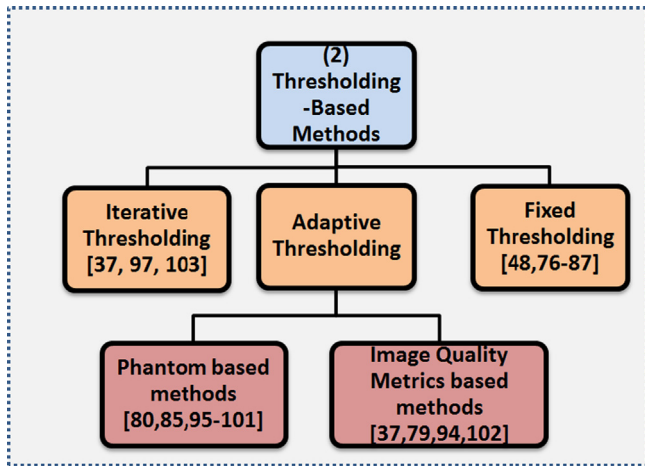


Fig. 4. An overview of intensity-based segmentation methods for PET images.

Overall, the major limitation in analytic expression based thresholding methods is that it is difficult to reproduce the same/similar segmentation results in different scanners or different patients [38]. This is because analytic expressions require precise tuning for a specific scanner, the reconstruction type, and even the patient size. Another drawback is that these expressions normally fail for lesions with a complex shape, due to an invalid analytic model for those cases (e.g. the partial volume would affect complex-shaped lesions differently than the spherical-shaped lesions). Also, it is very important to note that because of the limitations of representing anatomical structures and metabolic activities realistically, the construction and calibration of analytic expressions introduce uncertainty to the task of finding an optimal thresholding level; therefore, none of the studies listed in Table 4 are general enough to be used in clinical applications, particularly in radiotherapy planning and surgery.

4.3. Iterative thresholding method (ITM)

Commonly used adaptive thresholding methods in PET segmentation require a priori estimation of the lesion volume from anatomic images such as CT or analytic expression based on phantom geometry; however the iterative threshold method (ITM), proposed by Jentzen et al. [37] estimates the PET volumes without anatomic prior knowledge. The ITM iteratively converges to the optimum threshold to be applied to the PET image. The method is based on calibrated threshold-volume curves at varying S/B ratio acquired by phantom measurements using spheres of known volumes. The measured S/B ratios of the lesions are then estimated from PET images, and their volumes are iteratively calculated using the calibrated S/B-threshold-volume curves [97,103]. The resulting PET volumes are then compared with the known sphere volume and CT volumes of tumors that served as gold standards.

This process is illustrated in Fig. 5. The ITM begins by obtaining several curves that are observed at typical S/B ratios from the PET images. The calibrated S/B-threshold-volume curve that best fits the measured S/B ratio is then used. For the selected S/B-threshold-volume curve, there is a fixed-threshold value  $T_1$  (fixed-threshold region) for large volumes that is applied to the PET image and an initial estimation of the volume ( $V_1$ ) is made using the ellipsoid model. The volume  $V_1$  is used further to determine the second threshold value  $T_2$ . If the value  $T_2$  is significantly larger than  $T_1$ , then the threshold value  $T_2$  is applied again to the gray scale of the PET image and a second volume  $V_2$  can be calculated. If the threshold values  $T_2$  and  $T_3$  are not significantly different, the iteration stops at step 3 with an estimated volume  $V_3$ , the iteration ends at step  $n$  with an estimated volume  $V_n$  if the threshold value  $T_n$  does not deviate significantly from  $T_{n+1}$ . An example delineation process is given in Fig. 6 where optimal threshold was found in 4 iterations.

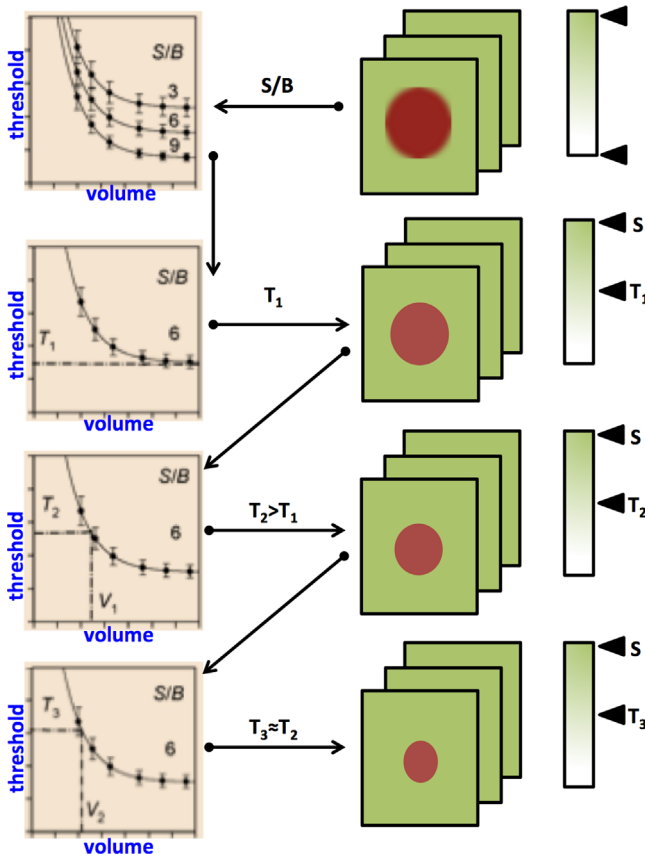


Fig. 5. Iterative thresholding method for finding optimal thresholding value.

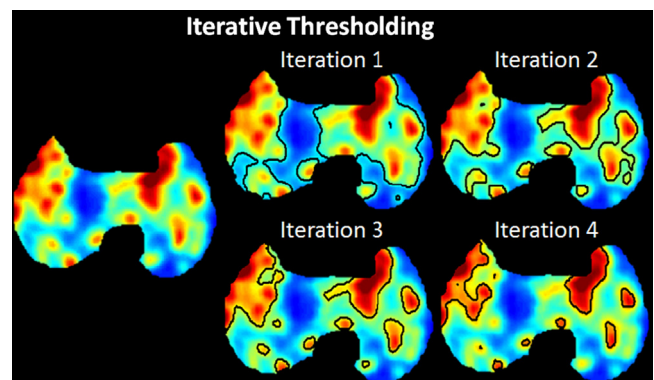


Fig. 6. The segmentation result at each iteration using the ITM is shown.

Unlike most adaptive thresholding methods, ITM does not require any prior information of the volume of interest. It only requires (a) the S/B ratio of the lesion easily taken from the PET image and (b) the S/B-threshold-volume curve to be determined once for the specific camera, reconstruction algorithm, and the radiotracer. However, it is important to note here that S/B-threshold-volume curve depends on the type of reconstruction algorithm used, especially for the small structures. Also the S/B-threshold-volume curve strongly depends on the spatial resolution of the imaging device. The ITM is further limited by the spatial resolution, implicit activity distribution, edge detection, and volume model of the lesion. For lesions with an effective diameter close to the spatial resolution of the scanner, the ITM method cannot be applied due to the PVE. Also, the ITM estimates the volume reliably only if the imaged activity distribution is homogeneous. The asymmetric activity distribution results in an underestimated volume. Additionally, the measurement of the S/B-threshold-volume curves assumed spherical lesions, and the clinical PET and CT volume calculations were performed using an ellipsoid model. Therefore, these suppositions are only an approximation of the irregularly shaped tumors. As a consequence, the clinical volume estimation may be less than accurate.

#### 4.4. Optimal choice of thresholding, partial volume and reconstruction effects on thresholding

From the conclusions reported in [103–105], the conditions to obtain an accurate (i.e., exact or very close) delineation of objects in PET images using thresholding are very strict. Considering an object whose largest diameter is less than ten times the image resolution, given as the FWHM of the point spread function (PSF), there exists a single threshold value that allows the true contours of the object to be accurately recovered *if and only if*

- (1) the object is spherical,
- (2) the uptake is uniform inside and outside the object, and
- (3) the PSF is isotropic and constant over the whole field of view.

For all other conditions, where non-spherical uptake occur or uptake is non-uniform, the choice of an optimal threshold selection is ill-posed problem; therefore, theoretical justification is not always possible [106].

Thresholding methods do not perform well with tumors that are less than 2–3 times the spatial resolution of the scanner [34,35,79,80,97,99,102,107,108]. This is because the scanner's PSF introduces the PVE. PVE comes from two sources: finite spatial resolution of imaging system and the discrete nature of PET images. Since PET is of low spatial resolution compared to CT or MRI, this factor is significant for PET imaging and most methods aim at this problem. New techniques such as digital photon counting (Philips Vereos PET/CT) is being explored to promote the resolution [109]. Furthermore, the continuous distribution of radiotracer is sampled to discrete voxels of PET images. Therefore, most voxels contains more than one tissue and the uptake value is the average of the response from all tissues within the voxel. This rule applies to all digital imaging modalities regardless of their spatial resolution. PVE is a major source of bias for PET quantification and several techniques have been proposed for PVE correction. During the reconstruction phase, spatial resolution can be enhanced by incorporating PSF information; anatomical information from CT or MRI can further be utilized as priors for better reconstruction. Deconvolution can also be applied for the purpose of resolution enhancement after reconstruction. To model the mutual influence between regions within image domain, Recovery Coefficient (single region) and Geometric Transfer Matrix (multiple regions) can be pre-calculated according to the approximate

size and shape of the target regions. Such corrections are simple but limited to mean value correction only and the assumptions of homogeneity and shape. More sophisticated methods targeting voxel level correction often make use of co-registered high resolution anatomical information from CT or MRI to define the structures and boundaries between regions. The PVE factors can subsequently be obtained by modeling the interaction between regions.

In practice, for smaller tumors, a 5% thresholding change can cause the measured volume to change by a factor of two [79,80,102,108]. It is explained in [99] that the near optimal thresholding value for small volumes depends largely on the size or diameter, due to the PVE being more influential while the optimal threshold value of larger objects has more functional dependence on the SBR. This is why many analytic thresholding expressions are piecewise linear and based on the estimated area or diameter of the lesion. Since  $SUV_{max}$  is the least affected by the PVE [34,89], there are many equations that use a percentage of the  $SUV_{max}$  to determine a reasonable threshold level (see Table 3, relative thresholding). For an in-depth review on PVE in PET imaging, a survey paper is referred [34].

Reconstruction methods also affect thresholding based segmentation algorithms. As a whole, reconstruction methods used for PET images vary on the amount of smoothing done, especially when attempting to compensate for the high noise of PET images. In addition, greater smoothing increases the difficulty of thresholding because an image with significant smoothing has a smaller absolute range of intensity values, such that a higher thresholding value is required to compensate for the decrease in contrast [102,108,110]. In addition, this in turn lowers the range of thresholding values that can be chosen to achieve an acceptable segmentation which makes it more probable that a non-optimal thresholding value is chosen. In [110], five different segmentation algorithms were compared with respect to various reconstruction algorithms, and it was found that  $SUV_{max}$  and  $SUV_{mean}$ -based fixed thresholding segmentation methods resulted in much larger volumes when high smoothing methods were used in the reconstruction process. Ideally, the method for threshold selection would compensate for the effect of reconstruction smoothing, though currently this is not considered. For detailed results on the effects of smoothing with reconstruction algorithms and threshold levels, reader may find useful information in [38,108].

Thresholding methods have proven to be simplest yet computationally the most efficient segmentation method. Their sensitivity to noise and incapability to handle intensity variations make them less than an ideal candidate for complex segmentation tasks in medical image analysis. Furthermore, with the exponential increase in the computational capabilities of hardware onboard medical devices over the past decade, the algorithmic simplicity of thresholding-based methods is becoming less attractive.

## 5. Stochastic and learning-based methods

Stochastic methods exploit differences between uptake regions and surrounding tissues statistically. Learning-based methods, similarly, use pattern recognition techniques to statistically estimate dependencies in the data. Since there are strong similarities between learning-based methods and stochastic methods, in this section we introduce the core concepts of the both groups together (Fig. 7).

### 5.1. Mixture models

The intensity distribution of objects within PET images are commonly considered to be approximately Gaussian in shape, and

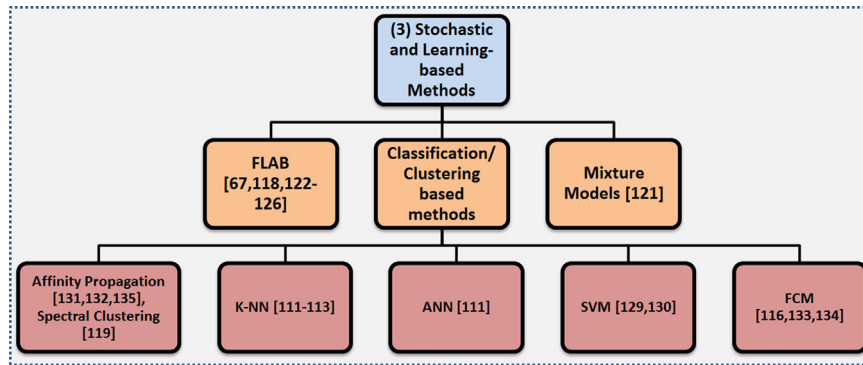


Fig. 7. An overview of the stochastic and learning-based segmentation methods.

this prior knowledge can be useful for segmentation. Gaussian Mixture Models (GMM) assume any distribution of intensities on the PET image can be approximated by a summation of Gaussian densities with the goal of identifying and separating these densities using an optimization technique such as the Expectation Maximization (EM) algorithm. There are currently several state-of-the-art GMMs in the literature that we will highlight here [118,121]. A GMM-based segmentation technique was created in [121], and three tissue classifications were considered: background, uncertain regions, and the target. A user defined ROI was required to initiate the algorithm, and the EM method was used to estimate the underlying Gaussians. Then, the voxels were assigned to one of the three classes. From these regions, the segmentation was constructed. The algorithm was evaluated on PET images of tumors from non-small cell lung cancer patients and performed favorably compared to 40%  $SUV_{max}$  and to the adaptive thresholding method.

### 5.2. Fuzzy locally adaptive Bayesian (FLAB) method

Another approach that uses Gaussian mixtures of the objects is a locally adaptive PET segmentation method based on Bayesian statistics [118]. This method is known as fuzzy locally adaptive Bayesian (FLAB) segmentation. FLAB is an unsupervised statistical method that considers the image as having two hard tissue classes and a finite level of “fuzzy levels” that includes mixtures of the hard tissue classes. Due to the fuzzy properties of the model, FLAB allows the co-existence of voxels to belong to one of two hard classes, where voxels belonging to a “fuzzy level” depend on its membership to the two hard classes. A ROI identification step is necessary to perform FLAB (like most of the other PET segmentation methods); therefore, it is not fully automated. Although FLAB has been shown to be quite robust and reproducible [118,67] for tumor volume assessments, ROI identification may be difficult when heterogeneous or high uptake regions occur [122]. Since heterogeneous or high uptake regions could be close to the object of interest, placing the ROI with these restrictions can in some cases be difficult. Moreover, a significant number of background voxels are necessary for accurate statistical modeling of tissue classes, so the number of background voxels should be small enough to avoid nearby uptake regions to be involved in model computation, other than the lesion of interest. In addition, the use of FLAB for the segmentation of heterogeneous lesions is limited to only two classes (background and foreground); therefore, FLAB may fall short when dealing with more than two classes.

Fortunately, an improved version of FLAB was recently published to deal with heterogeneous uptakes by allowing up to three tumor classes instead of just the two hard classes of the initial version of FLAB [123]. The improved version, named 3-FLAB, had a higher accuracy and robustness as compared with adaptive

threshold and FCM with a mean classification error of less than  $9\% \pm 8\%$ . Additionally, the accuracy [124], robustness [67], and reproducibility [125], as well as its clinical impact have been demonstrated in numerous papers [126–129].

### 5.3. Clustering/Classification of PET image intensities

Classification methods seek to partition a feature space derived from the image by using data with known labels [130]. Due to the requirement of training data for labeling the data, classifiers are known as supervised methods. The most common feature is image intensity itself. Classifiers can transfer the labels of the training data to new data as long as the feature space sufficiently distinguishes each label as well [130]. However, a disadvantage of supervised methods is that they generally do not incorporate spatial information into the decision of labeling, and the requirement of manual interaction to obtain training data is very labor intensive and time consuming. This eventually increases the computational complexity.

Similar to the classification methods, clustering methods can utilize the spatial information contained within the PET images but without the use of training data [131]. Since these methods do not need training data, they are termed unsupervised methods. Compared to supervised methods, clustering methods have less computational complexity; however, they are sensitive to noise and cannot integrate spatial information well due to inherent uncertainty of the data.

Examples of supervised and unsupervised methods used in PET segmentation include  $k$ -nearest neighbor ( $k$ -NN) [111–113], support vector machine (SVM) [132,133], Fuzzy C - Means (FCM) [116], artificial neural network (ANN) [111], and more recently Affinity Propagation (AP) [134,135] and spectral clustering [119]. Clustering methods aim at gathering items with similar properties (i.e., intensity values, spatial location, etc) into local groups. Similar to some advanced thresholding methods, clustering can also entail hard and soft boundaries or “fuzzy” objects [11]. These methods usually take similarities between data points as inputs and outputs a set of data points that best represent the data with corresponding labels (i.e., foreground, background). Clustering is very useful particularly when the shapes of the uptakes regions are non-convex with a heterogeneous background. Non-convex regions are quite common for different diseases, in particular pulmonary infections; hence, there is a growing interest in the use of clustering based methods to segment complex shaped uptake regions [134,135].

The spectral clustering method has been shown to have the potential to accurately delineate tumors containing inhomogeneous activities in the presence of a heterogeneous background in [119]. However, the number of tumors segmented in the study was limited, and there was no clear consensus on the choice of similarity parameters, which might not be optimal when only

intensity values are used as similarity parameters. Another common clustering method, FCM, was used in PET segmentation context first in [136], and it has been used mainly for PET brain lesion segmentation since [136,137]. The FCM algorithm classifies voxels into one of two groups, based on “fuzzy” levels where, due to the low resolution and the PVE, a particular voxel is allowed to include a mixture of multiple tissue types. FCM then decides which tissue type the voxel is most likely to belong to (i.e., it has more of one particular tissue type in the voxel over any of the other possible tissue types). Finally, the algorithm cuts the clusters into a foreground and background using a graph-based approach and converges to the global optima iteratively.

Most PET segmentation techniques are suited well for the determination of focal uptake, but generally give poor segmentation results for diffuse and multi-focal radiotracer uptake patterns such as those seen in infectious lung disease (Fig. 8). However, a recent study proposes a PET segmentation framework uniquely suited for the segmentation of these difficult uptake patterns [134,135]. This method utilized a novel similarity function that estimates the “similarity” (or affinity) of the data points on the histogram on the image(s) within the AP platform [138]. AP then uses these similarities to select the optimal thresholds to separate the image into multiple regions. In particular, [134] demonstrates the usability for quantification of small animal infectious disease models such as rabbits with tuberculosis or Ferrets with H1N1 (swine flu).

Table 5 lists some landmark studies that utilize clustering methods for PET image segmentation. The classification method used along with the other PET image segmentation methods are listed in the first column of Table 5. Also, the sample size and the

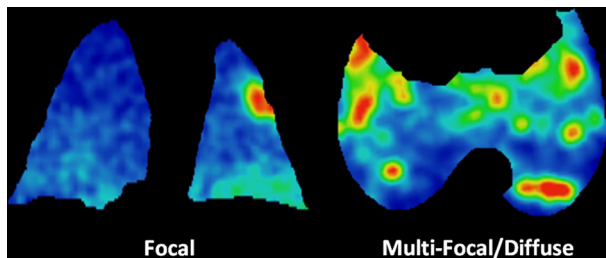


Fig. 8. Left: A representative slice (segmented to remove non-lung regions) showing focal radiotracer uptake in a small animal model while Right: demonstrates multi-focal/diffuse uptake patterns in a rabbit model infected with tuberculosis (5 weeks). Most PET segmentation techniques focus on segmenting the focal uptakes while ignoring the diffuse uptakes that occur in infectious pulmonary disease.

types of images used in the studies (such the type of phantom used or the type of disease if the studies used patient images) are listed in the second column. The quantitative results, as reported in the studies, are listed in the third column.

### 6. Region-based segmentation methods

Another distinct type of PET segmentation technique is region-based segmentation methods where the *homogeneity* of the image is the main consideration for determining object boundaries. While it is true that the region-based segmentation methods also utilize the intensities of the image, they are much more concerned about the local distribution (homogeneity) of the intensities on the image. The region-based methods are mainly divided into two subgroups when considering PET images: *Region Growing* in Section 6.1 and *Graph Cut* methods in Section 6.2. Fig. 9 shows further subgroups of these region-based segmentation methods.

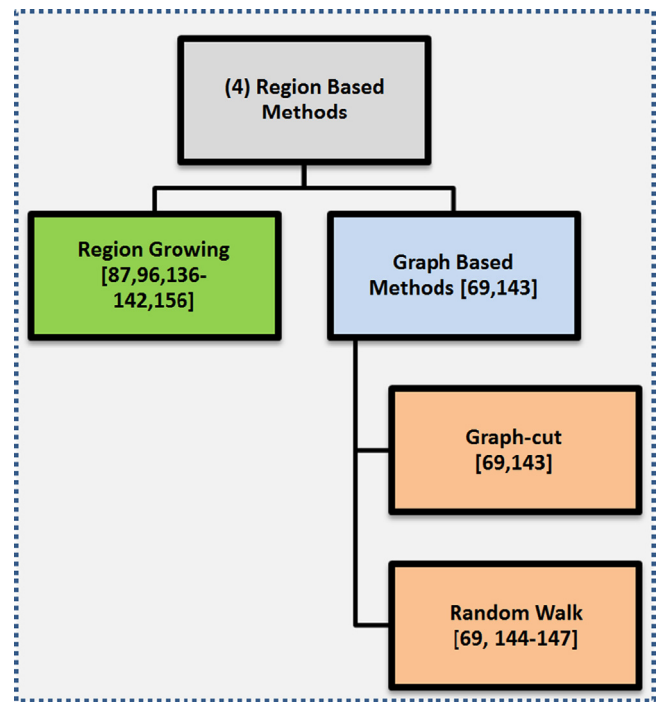


Fig. 9. An overview of the region-based segmentation methods.

Table 5 Examples of the performance of classification algorithms in PET segmentation.

Classification method	Sample size	Results	References
k-NN, ANN, Adaptive thresholding	6 Phantom spheres	Absolute relative error(%) <sup>a</sup> 6.83, 0.28, 7.61	[111]
k-NN	Monte Carlo simulation using Zubal whole-body phantom as prior	Dice similarity: ~80%–85%	[112]
k-NN, SUV <sub>2.5</sub> , T <sub>50%</sub> , T <sub>SBR</sub>	10 Head and neck cancer patients	Sensitivity, Specificity: 0.90, 0.95–0.93, 0.84–0.48, 0.98–0.68, 0.96	[113]
k-Means, MRF, Multiscale MRF	4 Lesions	Volume difference (%): 9.09, 6.97, 5.09	[114]
k-Means, MRF, Multiscale MRF	6 Spheres using the NIRMPA phantom	Volume difference (%): 42.86, 32.59, 15.36	[115]
FCM, FCM-SW <sup>a</sup>	Simulated lesions from the NCAT Phantom	Classification error (%): –10.8 ± 23.1, 0.9 ± 14.4	[116]
Standard GMM, SVFMM, CA-SVFMM, ICM, MEANF, Dirichlet Gaussian mixture model	21 NSCLC and 7 LSCC patients	Classification error (%): 21.7 ± 22.0, 8.6 ± 28.3	[117]
FLAB, FHMC, FCM, T <sub>42%</sub>	PET Image of dog lung and spherical phantoms	Misclassification ratio (%): 32.15, 12.43, 11.85, 3.52, 1.19, 0.67	[117]
Spectral clustering, adaptive thresholding	10 Spherical phantoms	Classification errors (%): 25.2, 31.2, 51.6, 55.8	[118]
	30 Simulated lesions	Dice Similarity: ~95%, 92%	[119]

<sup>a</sup> Integrated a trous wavelet transform and spatial information by first smoothing with a nonlinear anisotropic diffusion filter [120].



### 6.1. Region growing

The fundamental drawback of histogram based segmentation methods such as thresholding is that histograms provide no spatial information which is extremely valuable information not being considered. One such method that incorporates spatial information in the image along with the intensity information is Region Growing, as first presented in [139]. The algorithm starts at a user defined seed and based on the mean and standard deviation of the intensities within the local seed region, connected pixels are either included or excluded in the segmentation results. A second input, a homogeneity metric, is used to decide how different a new pixel can be from the statistics of the region already selected and can still be included in the segmentation [87]. This process is repeated until the entire region of interest has been dissected or the segmented region does not change.

Region growing is a widely applied technique for different segmentation purposes [87,139–143]. The main assumption in region growing is that regions of the object of interest have nearly constant or slowly varying intensity values to satisfy the homogeneity requirement. In region growing, it is expected that different homogeneity criterion and initial seed locations could easily affect the final segmentation results. Despite these difficulties, PET images tend to be sufficiently homogeneous in general, so region growing usually gives satisfactory results. For example, when compared to fixed thresholding techniques such as  $SUV_{2.5}$  and  $SUV_{43\%}$ , the region growing algorithm was shown to be much more accurate with a smaller standard deviation of the segmentation accuracy [87]. Furthermore, segmentation results obtained from the region growing method are highly reproducible, but, again, are strongly depend on the initialization of the segmentation. See Fig. 10 for an example of the delineation process with regard to different initializations of the seeds and homogeneity parameters. Although region growing methods have been shown to work well in homogeneous regions with appropriately set intensity homogeneity parameters, segmentation of heterogeneous structures has not been satisfactory (Fig. 10(a,b)). Region growing may fail even for sufficiently homogeneous uptake regions when the homogeneity parameter of the region growing algorithm is not appropriately set (Fig. 10(c)). Moreover, region growing methods reported in the literature for PET segmentation are not able to handle multiple object segmentation (Fig. 10(d)). Thus, homogeneity parameters for multiple lesion cases have yet to be assessed.

The main challenge in the region growing algorithm is “leakage”, that often occurs in PET images due to the high PVE, low resolution, and motion artifacts. Leakage can usually be limited using shape information [144], or be removed during an additional step after segmentation [145]. There has also been some methods proposed in the literature that discuss how to prevent or limit the leakage of region growing segmentation on PET images. Here, we highlight a

few state-of-the-art methods that aim to “constrain” region growing from leaking into the background or nearby objects. In [146], a region growing algorithm that avoids false-positive segmentation though user incorporated input was introduced. This has been done simply by ROI definition to limit the possible areas where region growing could leak into. Initially, the voxel with the highest intensity in the ROI was chosen as the starting seed. Then, an adaptive version of the conventional region growing algorithm determined the boundaries of a lesion by assessing whether there has been a sharp volume increase between iterations. After the final step, a dual-front active contour based method was applied to the segmented region using the ROI to refine the segmentation and to further reduce any background leakage that may have occurred. Similar to the final step in the previous algorithm, another region growing based method attempted to constrain the algorithm by integrating a threshold based segmentation with region growing to fine-tune an absolute thresholding level after the background signal was subtracted from the PET signals [96]. This improved the robustness of thresholding segmentation against noise by considering the homogeneity and connectedness of the segmented area of the image; however, the selection of stopping criteria and inclusion of non-target structures within the segmented areas are some of the drawbacks [96].

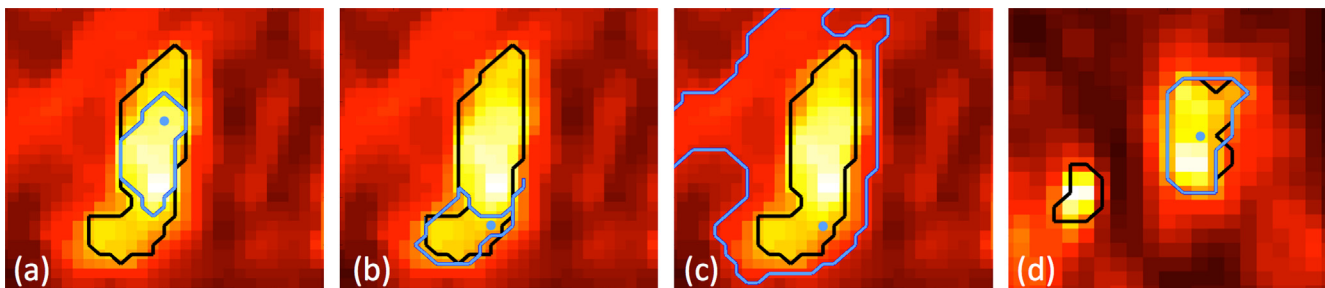
The last approach that we will describe is the condensed connected region growing (CCRG) method, which is an iterative algorithm utilizing the statistics from a user defined tumor region for segmentation. After defining the ROI, the voxel with the highest intensity is found, and the region growing begins at this location. Iteratively, the mean and standard deviation of this region is calculated, and a value derived by a formula that containing both of these metrics, is used to determine whether or not to include nearby voxels [87]. The CCRG gave significantly better segmentation results compared to thresholding-based methods; however, high false positive rates remains a challenging problem.

### 6.2. Graph-based methods

Graph-based approaches have a big advantage over other PET segmentation methods by incorporating efficient recognition into the segmentation process by using foreground and background seeds, specified by the user (or automatically) to locate the objects in the image [69]. These *seed points* act as hard constraints and combine global information with local pairwise pixel similarities [147] for optimal segmentation results. The two most common graph based methods used for PET segmentation are *Graph Cut* and *Random Walk*. We describe them in the following subsections in detail.

#### 6.2.1. Graph-cut

Initially, graph-cut constructs a graph where the *nodes* are the voxels on the image and the *edges* represent the strength of similarity between the nodes. Once the graph is constructed, the edges are cut using energy functions to minimize the cost of all



**Fig. 10.** The homogeneity metric for (a), (b)=0.1 and (c), (d)=0.3. The black outline in the images is the gold standard while the blue line is found from the region grown algorithm. The blue dot represents the location of the user defined seed. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

possible cuts. Graph-cut optimizes these partitions with respect to the energy function. Notably, graph-cut has been shown to optimally segment images using local pairwise pixel similarities [147]. However, graph-cut is not very robust and fails to give optimal results for noisy images [69].

### 6.2.2. Random walk

The random walk (RW) algorithm first appeared for computer vision applications in [148], and was used later for image segmentation in [69,149–151]. RW is robust against noise and weak boundaries, a necessary trait due to the low resolution and high noise characteristics of PET images. It was first proposed for PET image segmentation in [69]. The study compared RW with two well-known threshold based segmentation methods (described previously in this review), FLAB and FCM, and outperformed them. Authors found that RW was superior to these commonly used methods in terms of accuracy, robustness, repeatability, and computational efficiency. RW has also been used in multi-modality segmentation, as described later in Section 8. One drawback of RW is that it may not properly handle multi-focal uptake regions when distributed in large areas of the body region. Although an automated seeding process was proposed recently in [58] for this purpose, human incorporation may still be necessary for some extreme cases.

## 7. Boundary-based methods

Instead of using the statistics of the entire image or the homogeneity of the image for segmentation, boundary-based segmentation methods were designed to locate and identify the boundaries of the objects in PET images. However, locating the boundaries of the objects in PET images is challenging due to the low resolution and noise of the PET images. The boundary-based methods can be categorized into two subgroups, *Level Set/Active Contours* and *Gradient Based* methods, as shown in Fig. 11.

### 7.1. Level set and active contours

The concept of active contours, also called snakes, was first proposed in [152], where an initial contour around the object of interest deforms and moves towards the desired object's edges. 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 Caselles et al. [153]. 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 applied to PET images, a number of active contour-based segmentation techniques have been adapted in the literature. For instance, Hsu et al. [154] applied the classical active contour model to segment liver PET images. In their approach they estimated the external energy by solving a Poisson partial differential equation (PDE) and the algorithm was initialized by a Canny edge detection. Geometric active contours combined with an iterative deblurring algorithm were applied to PET images in [66] in the delineation of non-small cell lung cancer. Li et al. [146] used region growing as a pre-processing step to improve the active contour robustness for PET tumor delineation. Recently Abdoli et al. [155] combined geometric active contours with anisotropic diffusion filtering as a preprocessing step for smoothing, followed by a multi-resolution contourlet transform to segment tumors. The purpose of using the contourlet transform is to make the energy functional more effective in directing the evolving contour towards the target object.

The level set (LS) method was proposed in [156] for a way of modeling active contours by tracing interfaces between different phases of fluid flows. It has proven to be a very powerful tool for tracking moving interfaces over time and intensity. LS has been adopted in many applications involving movement of interfaces, including their widespread use in different imaging problems such as image segmentation and image registration [157,158]. Basically, LS attempts to exploit the intensity gradient information based on the concept of evolving level sets by iteratively solving Euler–Lagrange partial differential equation

$$\frac{\partial \phi}{\partial t} = V(\kappa) |\nabla \phi|, \quad (3)$$

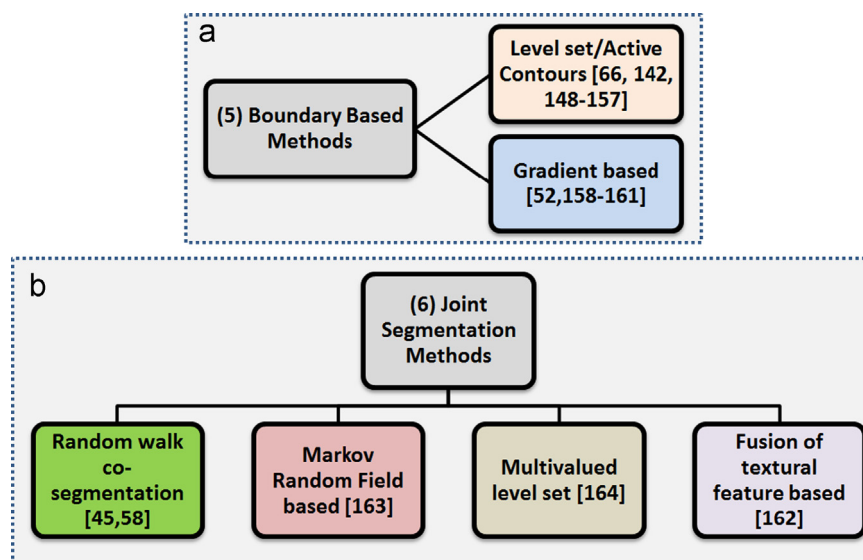


Fig. 11. An overview of the boundary-based segmentation methods.

where  $\varphi$  is an implicit function (e.g., distance) that monitors the evolving level set and  $V$  is the velocity function for controlling the expansion and shrinkage of level set that is directly proportional to the curvature  $\kappa$  and inversely proportional to the image intensity gradient. The spatial regularization imposed by LS-based methods encourages the segmentation to have a smooth boundary; therefore, the resulting segmentations have a more regular shape than obtained by other methods.

Several techniques have been proposed in the literature that employ LS with pre- or post-processing techniques to segment images in different imaging modalities including PET. As a representative method, [159] recently developed a technique that utilizes both spatial and temporal information in dynamic PET data based on a multi-phase level set. The authors explained that in a PET scan, activity contrast between organs varies from frame-to-frame because of changes in tracer distribution over time. The method defines different weighting of absolute difference in the data term to each image frame and considers the noise level and activity difference. The authors validated their segmentation method using both phantom and real mice data [159]. The method was compared with the  $k$ -means algorithm and shown to have a higher accuracy accordingly.

As described in the region growing methods, an adaptive region-growing method [143] was used as a preprocessing step, and then the LS method was used to further refine the segmentation result. The method was shown to outperform iterative threshold methods on phantom and real images. Further, a PSF-based deconvolution method was used in [160] as a preprocessing step with LS for co-registered PET/CT in order to segment lung tumors in a semi-automated way. Both methods were tested on phantom as well as clinical data and showed that it produced accurate results with high reproducibility.

LS methods have proven to be an elegant tool for tracking moving interfaces. By implicitly including curve parametrization and geometric property estimation, these methods are able to handle topological changes much better than other boundary and region-based methods. Also LS works well at segmenting multi-objects if initialization of LS has been done properly. However, depending on the energy function, the method can be computationally complex and is highly dependent on the initial condition.

## 7.2. Gradient-based methods

In general, the edges of an image usually have a sharp change in intensity values to signify the boundary of an object. To locate where these local changes of intensity occur, the gradient of the image is usually calculated between a voxel and the neighboring voxels. However, simply analyzing the abrupt changes in the PET intensity values does not often give optimal segmentation results due to several challenges that make the segmentation process less robust or not very accurate. The most significant of these challenges is the low resolution of the PET images and high PVE, which cause the boundaries to be smoothed and sometimes disconnected. PET images have also considerable noise, and this is amplified on gradient-based methods, which may result in a sub-optimal solution as well [161].

Despite all of the challenges that PET images pose for gradient-based segmentation methods, there were some attempts for compensating all of these difficulties. In [162], the images were first iteratively processed using an edge preserving filter and an iterative deconvolution algorithm to enhance the edges and reduce the effects of partial volume and smoothing that results from the reconstruction process. The deconvolution kernel involves the PSF of the PET scanner and the PSF must be estimated or known prior. Next, the watershed transform [161], which determines the edges using the gradient of an image, and

hierarchical cluster analysis were applied to the PET images. The hierarchical clustering algorithm (Wards algorithm) clustered the small patches found from the *watershed transform* together to construct the segmentation. When this method was compared to the ground truth volumes of the phantoms, a slight underestimation of approximately 10–20% was observed, and a slight radius underestimation around 0.5–1.1 mm was reported [162]. The proposed method was also evaluated on PET images of non-small cell lung cancer and compared to volumes derived from CT, other PET image threshold-based techniques, as well as histopathology slices of surgical specimens that were inated with gelatin, frozen, and then sliced [91]. The segmentation method was shown to have a high accuracy with respect to 3D tumor volumes derived from histopathology and had the best estimation of true tumor volume as compared to the various PET segmentation methods.

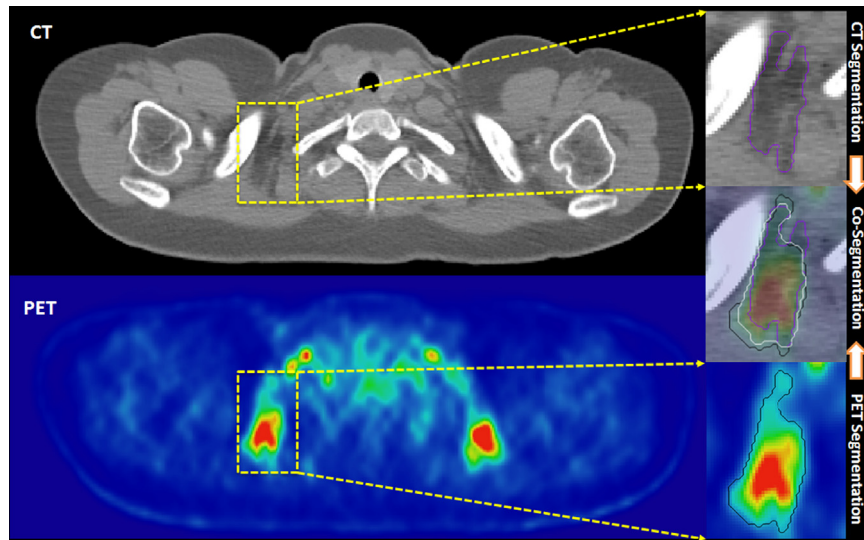
Another gradient based PET image segmentation method has been developed named GRADIENT (MIM Software, Cleveland, OH) and was validated in several studies demonstrating higher performance than manual and constant threshold methods in a phantom study [52,163]. GRADIENT requires a user defined initial starting point and a user defined ellipse which is then used for the initial bounding region for gradient detection [52]. In addition, another method, [164], used the watershed algorithm for segmenting noisy PET transmission images by utilizing a multi-resolution approach to deal with the PVE and the excessive noise in the data. However, the amount of smoothing used in the pre-processing and post-processing steps to fuse the over-segmented regions together, and the noise in PET images are some of the unsolved issues in PET image segmentation problems.

## 8. Joint segmentation methods

Image fusion involves combining two or more images of differing modalities to create a composite that contains complementary information from the inputs. Before PET-CT and MRI-PET hybrid scanners were developed, image registration techniques were being used to align images. It is evident that fused images are more suitable for visual perception, particularly for radiologists as analyzing fused images reduces uncertainty and minimizes redundancy in the output while maximizing relevant information rather than analyzing the images alone. In parallel to the developments of multi-modal scanners (PET-CT and MRI-PET), there have been recent attempts in the literature to bring the usefulness of integrating anatomical information (from CT and MRI) with functional data (from PET) for joint delineation of tumors [45,165–168]. Fig. 12 shows an example segmentation that incorporates information from PET and CT images where the individual segmentations from PET and CT are shown in white and pink, respectively, and the resulting joint-segmentation (i.e., co-segmentation) is in black. There are multiple-benefits of using a co-segmentation algorithm other than the conventional advantages of image fusion. First, co-segmentation algorithms bring increased robustness into the lesion delineation process due to unified information. Second, they provide wider spatial and temporal coverage of the tissues. Third, there is less uncertainty and more reliability in results of co-segmentation algorithms. All these benefits are due to the fact that co-segmentation algorithms try to mimic human visual image recognition performance which depends on the amount of informative features such as corners, texture, edges, and lines available in the images [169]. Co-segmentation algorithms optimize the number of these details by unifying two or more images into the same platform.

Currently, there are four main ideas on how to incorporate anatomical and PET information into the same space. First, a multi-valued LS deformable model was developed in [167] for integrating





**Fig. 12.** Here is an example of a segmentation that incorporates anatomical and functional information from multi-modalities (PET and CT). The original images are shown on the left while a zoom in view showing the segmentation (using the information only from the respective image) is provided on the right. The resulting co-segmentation is in the middle image on the right in white.

individual segmentations from PET and CT together to incorporate the information from both. The individual segmentations were combined together using the multi-valued LS method. Second, textural features from CT images were used to distinguish cancerous tissue types, and PET information was incorporated into this knowledge [165,168]. However, these efforts have some drawbacks when utilizing PET and CT information simultaneously such as potentially unrealistic assumption of one-to-one correspondences between anatomical and functional images of lesions and lack of standard for combining feature sets from different imaging modalities, as well as long execution times and sub-optimal solutions to individual segmentation problems. As a possible solution, third, a joint PET-CT image segmentation method was proposed in [166] and extended in [170], where a MRF algorithm was formulated on a graph. The approach formulates the segmentation problem as a minimization problem of a Markov random field model, which encodes the information from both modalities. Furthermore, the optimization task of the MRF minimization is solved using a graph-cut based method. Two sub-graphs are constructed for the segmentation of the PET and the CT images, respectively. Later, authors extended their method in [170] to achieve consistent results in two modalities by adding an adaptive context cost between the two sub-graphs. Although an optimal solution can be obtained by solving a single maximum flow problem, which leads to simultaneous segmentation of the tumor volumes in both modalities, the method itself requires user interaction, and it was only used with images from the head-neck with large tumors. Its performance in small uptake regions was not assessed. Another shortcoming of the proposed approach was due to the potentially unrealistic assumption that there is a one-to-one correspondence between PET and CT delineations. However, it is the first attempt in the literature showing full interaction between anatomical and functional information, and it has been successfully applied for simultaneous segmentation of head and neck tumors from PET-CT images. Furthermore, authors showed that graph-based PET-CT co-segmentation algorithm concurrently segment tumor from both modalities, achieving globally optimal solutions in low-order polynomial time.

Forth, the assumption of one-to-one-correspondence was recently relaxed in [45], where a fully automated co-segmentation method was driven by the uptake regions from PET, and then the correct anatomical boundaries were found in the corresponding CT images.

The proposed method was based on the *random walk* image co-segmentation, and more importantly, it did not have the assumption of one-to-one correspondence between PET and CT images or PET and MR (i.e., the lesions may have had a smaller or larger uptake region on PET compared to the anatomical abnormality in the CT or MR image). Therefore, this method is more realistic for clinical and pre-clinical applications, where relative differences may occur in structural and functional regions. Recently, the method from [45] was applied to quantify lesions from PET, PET-CT, PET-MRI, and MRI-PET-CT images in [58]. For an example of their findings, when considering only the information from the PET image, the co-segmentation method had a DSC of  $83.23 \pm 1.87\%$ . But when considering the information from CT and the information from PET, the DSC increased statistically significantly to  $91.44 \pm 1.71\%$ . The authors also compared their method with the state-of-the-art co-segmentation method [166] (described above) and it outperformed that method which has a high DSC of  $89.34 \pm 1.95\%$ . In both studies, evaluation of the algorithms are based on the ground-truth annotations manually obtained from multiple observers.

Radiologists visually judge spatial relationships between images better when images are fused. Appropriately and jointly displaying the PET-CT and MRI-PET images is important in many diagnostic tasks. For example, it has been shown that fusion of abdominal images from different modalities can improve diagnosis and monitoring of disease progression [171]. Indeed, hybrid imaging techniques have proven useful for the evaluation of patients with cancer including diagnosis, staging, treatment planning, and monitoring the response to therapy including disease progression [172]. Although it was widely accepted that the process of combining relevant information from two or more images into a single image carries more information than the single image alone, segmentation of the lesions was being conducted in single images until the co-segmentation algorithms came into play. Similar to the fusion process that radiologists use for qualitative evaluation of the lesions, the co-segmentation process combines the strength of multi-modal images to facilitate a globally optimal lesion boundary.

Future research into the co-segmentation of functional and anatomical images will strive towards improving the efficiency of the methods so that they can be used in clinical routine readily. Although available methods are shown to be efficient enough, real time processing of images will require several post-processing and



interactive methods to be conducted prior to the delineation process. Therefore, computational efficiency and even the use of the co-segmentation algorithms within the hardware (i.e., scanner) will be the potential directions.

## 9. Discussion

This review gives an overview of the current image segmentation techniques for PET images. The similarities, main ideas, assumptions, and quantitative comparisons between the many PET segmentation methods have been outlined to give researchers and clinicians an idea of which method is applicable for most situations. When choosing which method is appropriate for a specific quantification application, it should be noted that there are some important details outside the scope of this manuscript. For instance, application specific algorithms for particular radiotracers or diseases were not included in this manuscript. Different radiotracers can also have impact on the performance of individual segmentation algorithms. Specifically, both location and degree of concentration will change using different radiotracers. Therefore, for basic threshold-based methods, it will lead to considerable change in the appropriate thresholds. On the other hand, for more advanced methods using region information and edge information, it will influence less on the final segmentation result. Furthermore, the use of contrast media and metallic implants have been associated with focal radiotracer uptake, which may affect the accuracy of the image segmentation [173–175].

Segmentation methods developed for dynamic PET imaging were not included in this review either. Dynamic PET images are primarily used in research and drug development applications, and the long imaging times makes dynamic PET imaging extremely impractical for use in routine clinics. Interested readers should refer to [176–182] for examples of current research on this topic.

To the best of our knowledge, there is currently no study in the literature that directly compares the computational time of the PET image segmentation methods utilizing standardized images and hardware. Nevertheless, most of the PET segmentation algorithms can be considered efficient enough to be used in clinics due to the advancement of parallel computing and powerful workstations. However, we should also note that the magnitude of user interaction varies significantly depending on the pre-clinical/clinical applications and this may greatly affect the time for obtaining the segmentation results. For instance, supervised clustering methods require time-consuming training and labeling of the data while region-based methods often require a single seed location within the object of interest.

From the relevant publications that we reviewed in this work, it appears that methods for PET image segmentation are advancing towards fully automatic approaches that are clinically more feasible: (a) in the efficiency sense, and (b) the decrease in the variability from the lack of required input. Regarding the research direction (b), the choice of segmentation method for the particular application can be very specific. For instance, the AP-based segmentation method [135] was shown to be superior to other methods when uptake are diffuse and multi-focal; therefore, clustering based algorithms are preferred to other algorithms when uptake has non-spherical shapes and is multi-focal, as common in infectious lung diseases. In tumor segmentation, for another example, if manual ROI definition is not time consuming, then preferably FLAB method [135] or other fuzzy clustering methods [135] can be used to delineate tumors successfully. When there are multiple lesions and it is difficult to define individual ROIs for every lesion, then graph-cut, random walk, and region growing type algorithms are preferred due to the fast convergence,

accurate results, and more automated nature. In particular, the random walk algorithm is preferred when noise is a concern, and the region growing algorithm is less preferred due to lack of a leakage control system. Last, but not least, when lesions are spherical or near-spherical, thresholding methods, in particular ITM or adaptive thresholding, are satisfactory enough and the segmentation results are not statistically significantly different from each other [135].

Research in the PET image segmentation area will continue, perhaps indefinitely, as there is no single general optimal solution framework for all clinical problems in terms of accuracy, precision, and efficiency. Indeed, it is important to note that image segmentation has an infinite dimensional solution space which encourages researchers to distinguish image segmentation methods based on the application domain specific solution starting from general solution frameworks. In this way, even though the solution space is constrained to certain methods and the methods developed for specific clinical application will only have subtle differences in their accuracy measurements, there will still be improvements in terms of efficiency of the methods. Improving the efficiency of the image segmentation methods pertaining to application specific domains in PET imaging will shed new light on how to deal with a large number of data in a shorter time interval.

From reviewing a vast amount of research in PET segmentation methods, it is clear to the authors that one of the future directions of this community is an immediate need for standardization between the many different segmentation methods such as a publicly available database of PET images for evaluating new and old methods against. This database should consist of phantom images (such as the excellent PET simulation images from the Monte Carlo-based Geant4 application for emission tomography (GATE) software [183]) as well as small animal and human PET images with multiple manual delineations. Currently, beyond taking the time for implementing and conducting research on the optimal parameters for a new segmentation method against a common image database, there is no way to fairly and thoroughly evaluate one segmentation algorithm or framework versus another. In this review, the accuracy of the many methods that was reported in the respect studies was given, but when comparing and considering these methods using the reported accuracy there should be a reasonable amount of skepticism. It is apparent that some studies may have done more extensive and through testing versus another and without a common database of images and standardized testing these numbers are less significant. There are similar databases in the computer vision, machine learning, and many other fields of image processing research and this common database would aid in advancing the PET segmentation community significantly. Notably, there is an ongoing study proposing a protocol for evaluation of current and future segmentation methods based on a general framework, which will be expanded and adapted to PET imaging [184,185]. With the completion of that study, a benchmark database for validation and comparison of segmentation methods will be available and very beneficial for PET imaging and image processing applications.

## 10. Conclusions

PET imaging provides quantitative functional information on diseases, and image segmentation is of great importance for extracting this information. In this paper, we presented the state-of-the-art image segmentation methods that are commonly used for PET imaging, as well as the recent advances in techniques applicable to PET, PET-CT, and MRI-PET images. We investigated different segmentation methods in detail; results were listed and

compared throughout this review. Given the vast number and wide variety of methods for approaching the segmentation task, this review compares and contrasts the state-of-the-art methods and provides researchers and clinicians with detailed segmentation methods that are well suited for any particular application. We noted that although there is no PET image segmentation method that is optimal for all applications or can compensate for all of the difficulties inherent to PET images, development of trending image segmentation techniques which combine anatomical information and metabolic activities in the same hybrid frameworks (PET-CT, PET-MR, and MRI-PET-CT) is encouraging and open to further investigations. Continued refinements of PET image segmentation methods in parallel to these advances in imaging instrumentation will provide basis for improved evaluation of prognosis.

### Conflict of interest statement

None declared.

### Acknowledgments

This research is supported by the Center for Infectious Disease Imaging (CIDI), the Intramural Program of the National Institutes of Allergy and Infectious Diseases (NIAID), and the National Institutes of Bio-imaging and Bioengineering (NIBIB) at the National Institutes of Health (NIH). We thank Dr. Sanjay Jain for kindly providing the rabbit TB images.

### References

- [1] T. Seute, P. Leffers, G. ten Velde, A. Twijnstra, Detection of brain metastases from small cell lung cancer, *Cancer* 112 (8) (2008) 1827–1834.
- [2] M. MacManus, U. Nestle, K. Rosenzweig, I. Carrio, C. Messa, O. Belohlavek, M. Danna, T. Inoue, E. Deniaud-Alexandre, S. Schipani, et al., Use of PET and PET/CT for radiation therapy planning: laea expert report 2006–2007, *Radiother. Oncol.* 91 (1) (2009) 85–94.
- [3] S. Basu, T. Kwee, S. Surti, E. Akin, D. Yoo, A. Alavi, Fundamentals of PET and PET/CT imaging, *Ann. NY Acad. Sci.* 1228 (1) (2011) 1–18.
- [4] D. Lardinois, W. Weder, T. Hany, E. Kamel, S. Korom, B. Seifert, G. von Schulthess, H. Steinert, Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography, *New Engl. J. Med.* 348 (25) (2003) 2500–2507.
- [5] L. Kostakoglu, H. Agress Jr., S. Goldsmith, Clinical role of FDG PET in evaluation of cancer patients, *Radiographics* 23 (2) (2003) 315–340.
- [6] M. Judenhofer, Simultaneous PET-MRI: a new approach for functional and morphological imaging, *Nat. Med.* 14 (4) (2008) 459–465.
- [7] D. Evanko, Two pictures are better than one, *Nat. Methods* 5 (5) (2008) 377.
- [8] P. Kaufmann, P. Camici, Myocardial blood flow measurement by PET: technical aspects and clinical applications, *J. Nucl. Med.* 46 (1) (2005) 75–88.
- [9] L.S. Zhao, B., L.H. Schwartz, Imaging surrogates of tumor response to therapy: anatomic and functional biomarkers, *J. Nucl. Med.* 50 (2) (2009) 239–249.
- [10] I.M.I. Division, Pet Market Summary Report, Technical Report, 2011.
- [11] V. Gregoire, K. Haustermans, X. Geets, S. Roels, M. Lonnew, Pet based treatment planning in radiotherapy: a new standard? *J. Nucl. Med.* 48 (2007) 685–775.
- [12] J. Votaw, The aapm/rsna physics tutorial for residents. Physics of PET, *Radiographics* 15 (5) (1995) 1179–1190.
- [13] S. Basu, H. Zaidi, S. Holm, A. Alavi, Quantitative techniques in PET-CT imaging, *Curr. Med. Imaging Rev.* 7 (3) (2011) 216–233.
- [14] M. Evelina, S. Gian, T. Federica, Z. Angelo, D. Filippo, P. Giulia, R. Luigi, Z. Federica, T. Silverio, 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), *J. Exp. Clin. Cancer Res.* 27 (2008) 52.
- [15] G. Kramer-Marek, M. Bernardo, D. Kiesewetter, U. Bagci, M. Kuban, A. Omer, R. Zielinski, J. Seidel, P. Choyke, J. Capala, Pet of her2-positive pulmonary metastases with 18F-zher2: 342 affibody in a murine model of breast cancer: comparison with 18F-FDG, *J. Nucl. Med.* 53 (6) (2012) 939–946.
- [16] R. Wahl, H. Jacene, Y. Kasamon, M. Lodge, From recist to perclist: evolving considerations for PET response criteria in solid tumors, *J. Nucl. Med.* 50 (Suppl 1) (2009) 122S–150S.
- [17] V. Lowe, J. Hoffman, D. DeLong, E. Patz, R. Coleman, et al., Semiquantitative and visual analysis of FDG-PET images in pulmonary abnormalities, *J. Nucl. Med.* 35 (11) (1994) 1771–1776.
- [18] M. Lodge, M. Chaudhry, R. Wahl, Noise considerations for PET quantification using maximum and peak standardized uptake value, *J. Nucl. Med.* 53 (7) (2012) 1041–1047.
- [19] C. Kim, N. Gupta, B. Chandramouli, A. Alavi, Standardized uptake values of FDG: body surface area correction is preferable to body weight correction, *J. Nucl. Med.* 35 (1) (1994) 164–167.
- [20] M. Kelly, SUV: advancing comparability and accuracy, White Paper Siemens.
- [21] K. Zasadny, R. Wahl, Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-d-glucose: variations with body weight and a method for correction, *Radiology* 189 (3) (1993) 847–850.
- [22] Y. Sugawara, K. Zasadny, A. Neuhoff, R. Wahl, Re-evaluation of the standardized uptake value for FDG: variations with body weight and methods for correction, *Radiology* 213 (2) (1999) 521–525.
- [23] P. Lindholm, H. Minn, S. Leskinen-Kallio, J. Bergman, U. Ruotsalainen, H. Joensuu, et al., Influence of the blood glucose concentration on FDG uptake in cancer—a PET study, *J. Nucl. Med.* 34 (1) (1993) 1–6.
- [24] F. Crippa, C. Gavazzi, F. Bozzetti, C. Chiesa, C. Pascali, A. Bogni, V. De Sanctis, D. Decise, M. Schiavini, G. Cucchetti, E. Bombardieri, The influence of blood glucose levels on [18F] fluorodeoxyglucose (FDG) uptake in cancer: a PET study in liver metastases from colorectal carcinomas, *Tumori* 83 (4) (1997) 748–752.
- [25] L. Shankar, J. Hoffman, S. Bacharach, M. Graham, J. Karp, A. Lammertsma, S. Larson, D. Mankoff, B. Siegel, A. Van den Abbeele, et al., Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in national cancer institute trials, *J. Nucl. Med.* 47 (6) (2006) 1059–1066.
- [26] M. Lodge, J. Lucas, P. Marsden, B. Cronin, M. O'Doherty, M. Smith, A PET study of 18 FDG uptake in soft tissue masses, *Eur. J. Nucl. Med. Mol. Imaging* 26 (1) (1999) 22–30.
- [27] J. Keyes Jr., Suv: Standard uptake or silly useless value? *J. Nucl. Med.* 36 (10) (1995) 1836–1839.
- [28] S. Beaulieu, P. Kinahan, J. Tseng, L. Dunnwald, E. Schubert, P. Pham, B. Lewellen, D. Mankoff, Suv varies with time after injection in 18F-FDG PET of breast cancer: characterization and method to adjust for time differences, *J. Nucl. Med.* 44 (7) (2003) 1044–1050.
- [29] S.A. Nehmeh, Y.E. Erdi, C.C. Ling, K.E. Rosenzweig, H. Schoder, S.M. Larson, H.A. Macapinlac, O.D. Squire, J.L. Humm, Effect of respiratory gating on quantifying PET images of lung cancer, *J. Nucl. Med.* 43 (7) (2002) 876–881.
- [30] S. Nehmeh, Y. Erdi, T. Pan, A. Pevsner, K. Rosenzweig, E. Yorke, G. Mageras, H. Schoder, P. Vernon, O. Squire, et al., Four-dimensional (4d) PET/ct imaging of the thorax, *Med. Phys.* 31 (2004) 3179–3186.
- [31] Y. Erdi, S. Nehmeh, T. Pan, A. Pevsner, K. Rosenzweig, G. Mageras, E. Yorke, H. Schoder, W. Hsiao, O. Squire, et al., The CT motion quantification of lung lesions and its impact on PET-measured SUVs, *J. Nucl. Med.* 45 (8) (2004) 1287–1292.
- [32] T. Pan, O. Mawlawi, S. Nehmeh, Y. Erdi, D. Luo, H. Liu, R. Castillo, R. Mohan, Z. Liao, H. Macapinlac, Attenuation correction of PET images with respiration-averaged CT images in PET/CT, *J. Nucl. Med.* 46 (9) (2005) 1481–1487.
- [33] C. Ramos, Y. Erdi, M. Gonen, E. Riedel, H. Yeung, H. Macapinlac, R. Chisin, S. Larson, FDG-PET standardized uptake values in normal anatomical structures using iterative reconstruction segmented attenuation correction and filtered back-projection, *Eur. J. Nucl. Med. Mol. Imaging* 28 (2) (2001) 155–164.
- [34] M. Soret, S. Bacharach, I. Buvat, Partial-volume effect in PET tumor imaging, *J. Nucl. Med.* 48 (6) (2007) 932–945.
- [35] S. Srinivas, T. Dhurairaj, S. Basu, G. Bural, S. Surti, A. Alavi, A recovery coefficient method for partial volume correction of PET images, *Annu. Nucl. Med.* 23 (4) (2009) 341–348.
- [36] F. Fahey, P. Kinahan, R. Doot, M. Kocak, H. Thurston, T. Poussaint, Variability in PET quantitation within a multicenter consortium, *Med. Phys.* 37 (7) (2010) 3660–3666.
- [37] W. Jentzen, L. Freudenberg, E. Eising, M. Heinze, W. Brandau, A. Bockisch, Segmentation of PET volumes by iterative image thresholding, *J. Nucl. Med.* 48 (1) (2007) 108–114.
- [38] J. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonnew, V. Grégoire, Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms, *Radiother. Oncol.* 69 (3) (2003) 247–250.
- [39] I. Burger, D. Huser, C. Burger, G. von Schulthess, A. Buck, Repeatability of FDG quantification in tumor imaging: averaged SUVs are superior to SUV (max), *Nucl. Med. Biol.* 39 (5) (2012) 666–670.
- [40] Y. Nakamoto, K. Zasadny, H. Minn, R. Wahl, Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18F] fluoro-d-glucose, *Mol. Imaging Biol.* 4 (2) (2002) 171–178.
- [41] L. Velasquez, R. Boellaard, G. Kollia, W. Hayes, O. Hoekstra, A. Lammertsma, S. Galbraith, Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies, *J. Nucl. Med.* 50 (10) (2009) 1646–1654.
- [42] N. Krak, R. Boellaard, O. Hoekstra, J. Twisk, C. Hoekstra, A. Lammertsma, Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial, *Eur. J. Nucl. Med. Mol. Imaging* 32 (3) (2005) 294–301.

- [43] U. Bagci, X. Chen, J. Udupa, Hierarchical scale-based multi-object recognition of 3d anatomical structures, *IEEE Trans. Med. Imaging* 31 (3) (2012) 777–789.
- [44] P. Saha, J. Udupa, Scale-based diffusive image filtering preserving boundary sharpness and fine structures, *IEEE Trans. Med. Imaging* 20 (11) (2001) 1140–1155.
- [45] U. Bagci, J.K. Udupa, J. Yao, D.J. Mollura, Co-segmentation of functional and anatomical images, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, Springer, 2012, pp. 459–467.
- [46] U. Bagci, M. Bray, J. Caban, J. Yao, D. Mollura, Computer-assisted detection of infectious lung diseases: a review, *Comput. Med. Imaging Graphics* 36 (1) (2012) 72–84.
- [47] J. Udupa, V. Leblanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. Currie, B. Hirsch, J. Woodburn, A framework for evaluating image segmentation algorithms, *Comput. Med. Imaging Graph.* 30 (2) (2006) 75–87.
- [48] U. Nestle, S. Kremp, A. Grosu, Practical integration of 18F-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (nslc): the technical basis, icru-target volumes, problems, perspectives, *Radiother. Oncol.* 81 (2) (2006) 209–225.
- [49] R. Boellaard, N. Krak, O. Hoekstra, A. Lammertsma, Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study, *J. Nucl. Med.* 45 (9) (2004) 1519–1527.
- [50] D. Delbeke, R.E. Coleman, M.J. Guibertau, M.L. Brown, H.D. Royal, B.A. Siegel, D.W. Townsend, L.L. Berland, J.A. Parker, K. Hubner, et al., Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0, *J. Nucl. Med.* 47 (5) (2006) 885–895.
- [51] R. Boellaard, Standards for PET image acquisition and quantitative data analysis, *J. Nucl. Med.* 50 (Suppl 1) (2009) 11S–20S.
- [52] 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, *Int. J. Radiat. Oncol. Biol. Phys.* 82 (3) (2012) 1164–1171.
- [53] U. Bagci, B. Foster, K. Miller-Jaster, B. Luna, B. Dey, W.R. Bishai, C.B. Jonsson, S. Jain, D.J. Mollura, A computational pipeline for quantification of pulmonary infections in small animal models using serial PET-CT imaging, *EJNMMI Res.* 3 (1) (2013) 1–20.
- [54] S.K. Warfield, K.H. Zou, W.M. Wells, Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation, *IEEE Trans. Med. Imaging* 23 (7) (2004) 903–921.
- [55] D. Dogra, A. Majumdar, S. Sural, Evaluation of segmentation techniques using region area and boundary matching information, *J. Vis. Commun. Image Represent.* 23 (2012) 150–160.
- [56] J.A. Lee, Segmentation of positron emission tomography images: some recommendations for target delineation in radiation oncology, *Radiother. Oncol.* 96 (3) (2010) 302–307.
- [57] U. Bagci, J. Yao, K. Miller-Jaster, X. Chen, D. Mollura, Predicting future morphological changes of lesions from radiotracer uptake in 18F-FDG-PET images, *Plos One* 8 (2) (2013) e57105.
- [58] U. Bagci, J. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, D.J. Mollura, Joint segmentation of functional and anatomical images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images, *Med. Image Anal.* 17 (8) (2013) 929–945.
- [59] C. Lin, E. Itti, C. Haioun, Y. Petegnief, A. Luciani, J. Dupuis, G. Paone, J.-N. Talbot, A. Rahmouni, M. Meignan, Early 18F-FDG PET for prediction of prognosis in patients with diffuse large b-cell lymphoma: SUV-based assessment versus visual analysis, *J. Nucl. Med.* 48 (10) (2007) 1626–1632.
- [60] D. Hellwig, T. Graeter, D. Ukena, A. Groeschel, G. Sybrecht, H. Schaefers, C. Kirsch, 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? *J. Nucl. Med.* 48 (11) (2007) 1761–1766.
- [61] J. Fox, R. Rengan, W. O'Meara, E. Yorke, Y. Erdi, S. Nehmeh, S. Leibel, K. Rosenzweig, Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int. J. Radiat. Oncol. Biol. Phys.* 62 (1) (2005) 70–75.
- [62] C. Caldwell, K. Mah, Y. Ung, C. Danjoux, J. Balogh, S. Ganguli, L. 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, *Int. J. Radiat. Oncol. Biol. Phys.* 51 (4) (2001) 923–931.
- [63] R. Steenbakkers, J. Duppen, I. Fitton, K. Deurloo, L. Zijp, E. Comans, A. Uitterhoeve, P. Rodrigues, G. Kramer, J. Bussink, et al., Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis, *Int. J. Radiat. Oncol. Biol. Phys.* 64 (2) (2006) 435–448.
- [64] C. Fiorino, M. Reni, A. Bolognesi, G. Cattaneo, R. Calandrino, Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning, *Radiother. Oncol.* 47 (3) (1998) 285–292.
- [65] P. Giraud, S. Elles, S. Helfre, Y. De Rycke, V. Servois, M. Carette, C. Alzieu, P. 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, *Radiother. Oncol.* 62 (1) (2002) 27–36.
- [66] I. El Naqa, J. Bradley, J. Deasy, K. Biehl, R. Laforest, D. Low, Improved analysis of PET images for radiation therapy, in: *International Conference on the Use of Computers in Radiation Therapy*, Seoul Korea, 2004.
- [67] M. Hatt, C. Cheze Le Rest, N. Albarghach, O. Pradier, D. Visvikis, PET functional volume delineation: a robustness and repeatability study, *Eur. J. Nucl. Med. Mol. Imaging* 38 (4) (2011) 663–672.
- [68] J. Erasmus, G. Gladish, L. Broemeling, B. Sabloff, M. Truong, R. Herbst, R. Munden, Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response, *J. Clin. Oncol.* 21 (13) (2003) 2574–2582.
- [69] U. Bagci, J. Yao, J. Caban, E. Turkbey, O. Aras, D. Mollura, A graph-theoretic approach for segmentation of PET images, in: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2011*, pp. 8479–8482.
- [70] S. Breen, J. Publicover, S. De Silva, G. Pond, K. Brock, B. O'Sullivan, B. Cummings, L. Dawson, A. Keller, J. Kim, et al., Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers, *Int. J. Radiat. Oncol. Biol. Phys.* 68 (3) (2007) 763–770.
- [71] C. Beckers, C. Ribbens, B. André, S. Marcelis, O. Kaye, L. Mathy, M. Kaiser, R. Hustinx, J. Foidart, M. Malaise, Assessment of disease activity in rheumatoid arthritis with 18F-FDG PET, *J. Nucl. Med.* 45 (6) (2004) 956–964.
- [72] H. Vorwerk, G. Beckmann, M. Bremer, M. Degen, B. Dietl, R. Fietkau, T. Gsänger, R. Hermann, M. Alfred Herrmann, U. Holler, et al., The delineation of target volumes for radiotherapy of lung cancer patients, *Radiother. Oncol.* 91 (3) (2009) 455–460.
- [73] B. Shah, N. Srivastava, A. Hirsch, G. Mercier, R. Subramaniam, Intra-reader reliability of FDG PET volumetric tumor parameters: effects of primary tumor size and segmentation methods, *Annu. Nucl. Med.* 26 (9) (2012) 707–714.
- [74] N. Webb, R. Shavelson, E. Haertel, Reliability coefficients and generalizability theory, *Handbook of Statistics* 26 (2006) 81–124.
- [75] K.H. Zou, S.K. Warfield, A. Bharatha, C. Tempany, M.R. Kaus, S.J. Haker, W.M. Wells III, F.A. Jolesz, R. Kikinis, Statistical validation of image segmentation quality based on a spatial overlap index sup 1 sup: scientific reports, *Acad. Radiol.* 11 (2) (2004) 178–189.
- [76] P. Saha, J. Udupa, Optimum image thresholding via class uncertainty and region homogeneity, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (7) (2001) 689–706.
- [77] D. Schinagl, W. Vogel, A. Hoffmann, J. van Dalen, W. Oyen, J. Kaanders, Comparison of five segmentation tools for 18F-fluoro-deoxy-glucose-positron emission tomography-based target volume definition in head and neck cancer, *Int. J. Radiat. Oncol. Biol. Phys.* 69 (4) (2007) 1282–1289.
- [78] H. Vees, S. Senthamizhchelvan, R. Miralbell, D. Weber, O. Ratib, H. Zaidi, Assessment of various strategies for 18 F-FET PET-guided delineation of target volumes in high-grade glioma patients, *Eur. J. Nucl. Med. Mol. Imaging* 36 (2) (2009) 182–193.
- [79] U. Nestle, S. Kremp, A. Schaefer-Schuler, C. Sebastian-Welsch, D. Hellwig, C. Rube, C. 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, *J. Nucl. Med.* 46 (8) (2005) 1342–1348.
- [80] Y. Erdi, O. Mawlawi, S. Larson, M. Imbriaco, H. Yeung, R. Finn, J. Humm, Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding, *Cancer* 80 (S12) (1997) 2505–2509.
- [81] C. Caldwell, K. Mah, M. Skinner, C. 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, *Int. J. Radiat. Oncol. Biol. Phys.* 55 (5) (2003) 1381–1393.
- [82] C. Nagel, G. Bosmans, A. Dekker, M. Öllers, D. De Ruyscher, P. Lambin, A. Minken, N. Lang, K. Schäfers, Phased attenuation correction in respiration correlated computed tomography/positron emitted tomography, *Med. Phys.* 33 (6) (2006) 1840–1847.
- [83] A. Paulino, M. Koshy, R. Howell, D. Schuster, L. Davis, Comparison of CT-and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer, *Int. J. Radiat. Oncol. Biol. Phys.* 61 (5) (2005) 1385–1392.
- [84] E. Deniaud-Alexandre, E. Touboul, D. Lerouge, D. Grahek, J. Foulquier, Y. Petegnief, B. Grès, H. El Balaa, K. Kerrou, et al., Impact of computed tomography and 18F-deoxyglucose coincidence detection emission tomography image fusion for optimization of conformal radiotherapy in non-small-cell lung cancer, *Int. J. Radiat. Oncol. Biol. Phys.* 63 (5) (2005) 1432–1441.
- [85] K. Biehl, F. Kong, F. Dehdashti, J. Jin, S. Mutic, I. ElNaqa, B. Siegel, J. 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? *J. Nucl. Med.* 47 (11) (2006) 1808–1812.
- [86] R. Hong, J. Halama, D. Bova, A. Sethi, B. Emami, Correlation of PET standard uptake value and ct window-level thresholds for target delineation in CT-based radiation treatment planning, *Int. J. Radiat. Oncol. Biol. Phys.* 67 (3) (2007) 720–726.
- [87] E. Day, J. Betler, D. Parda, B. Reitz, A. Kirichenko, S. Mohammadi, M. Miften, A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients, *Med. Phys.* 36 (10) (2009) 4349–4358.
- [88] P. Tylski, S. Stute, N. Grotus, K. Doyeux, S. Hapdey, I. Gardin, B. Vanderlinden, I. Buvat, Comparative assessment of methods for estimating tumor volume and standardized uptake value in 18F-FDG PET, *J. Nucl. Med.* 51 (2) (2010) 268–276.
- [89] A. Van Baardwijk, G. Bosmans, L. Boersma, J. Buijsen, S. Wanders, M. Hochstenbag, R. 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



- the primary tumor and involved nodal volumes, *Int. J. Radiat. Oncol. Biol. Phys.* 68 (3) (2007) 771–778.
- [90] W. Yu, X. Fu, Y. Zhang, J. Xiang, L. Shen, G. Jiang, J. Chang, Gtv spatial conformity between different delineation methods by 18 FDG PET/CT and pathology in esophageal cancer, *Radiother. Oncol.* 93 (3) (2009) 441–446.
- [91] M. Wanet, J.A. Lee, B. Weynand, M. De Bast, A. Poncelet, V. Lacroix, E. Coche, V. Grégoire, 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, *Radiother. Oncol.* 98 (1) (2011) 117–125.
- [92] C. Chen, R. Muzic Jr., A. Nelson, L. Adler, Simultaneous recovery of size and radioactivity concentration of small spheroids with PET data, *J. Nucl. Med.* 40 (1) (1999) 118–130.
- [93] B. Yaremko, T. Riauka, D. Robinson, B. Murray, A. Alexander, A. McEwan, W. Roa, Thresholding in PET images of static and moving targets, *Phys. Med. Biol.* 50 (24) (2005) 5969–5982.
- [94] Q. Black, I. Grills, L. Kestin, C. Wong, J. Wong, A. Martinez, D. Yan, Defining a radiotherapy target with positron emission tomography, *Int. J. Radiat. Oncol. Biol. Phys.* 60 (4) (2004) 1272–1282.
- [95] L. Drever, D. Robinson, A. McEwan, W. Roa, A local contrast based approach to threshold segmentation for PET target volume delineation, *Med. Phys.* 33 (6) (2006) 1583–1594.
- [96] J. Davis, B. Reiner, M. Huser, C. Burger, G. Szekely, I. Ciernik, Assessment of 18f PET signals for automatic target volume definition in radiotherapy treatment planning, *Radiother. Oncol.* 80 (2006) 43–50.
- [97] L. Drever, W. Roa, A. McEwan, D. Robinson, Iterative threshold segmentation for PET target volume delineation, *Med. Phys.* 34 (4) (2007) 1253–1265.
- [98] A. Schaefer, S. Kremp, D. Hellwig, C. Rübke, C. Kirsch, 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, *Eur. J. Nucl. Med. Mol. Imaging* 35 (11) (2008) 1989–1999.
- [99] M. Brambilla, R. Matheoud, C. Secco, G. Loi, M. Krengli, E. Inglese, Threshold segmentation for PET target volume delineation in radiation treatment planning: the role of target-to-background ratio and target size, *Med. Phys.* 35 (4) (2008) 1207–1213.
- [100] S. Nehmeh, H. El-Zeftawy, C. Greco, J. Schwartz, Y. Erdi, A. Kirov, C. Schmidlein, A. Gyau, S. Larson, J. Humm, An iterative technique to segment PET lesions using a monte carlo based mathematical model, *Med. Phys.* 36 (10) (2009) 4803–4809.
- [101] A. Riegel, M. Bucci, O. Mawlawi, V. Johnson, M. Ahmad, X. Sun, D. Luo, A. Chandler, 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, *Med. Phys.* 37 (4) (2010) 1742–1752.
- [102] R. Matheoud, P. Della Monica, C. Secco, G. Loi, M. Krengli, E. Inglese, M. Brambilla, Influence of different contributions of scatter and attenuation on the threshold values in contrast-based algorithms for volume segmentation, *Phys. Medica* 27 (1) (2011) 44–51.
- [103] J. van Dalen, A. Hoffmann, V. Dicken, W. Vogel, B. Wiering, T. Ruers, N. Karssemeijer, W. Oyen, A novel iterative method for lesion delineation and volumetric quantification with FDG PET, *Nucl. Med. Commun.* 28 (6) (2007) 293–485.
- [104] M. King, D. Long, A. Brill, Spect volume quantitation: influence of spatial resolution, source size and shape, and voxel size, *Med. Phys.* 18 (5) (1997) 1016–1024.
- [105] F. Hofheinz, S. Dittrich, C. Potzsch, J. van den Hoff, Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold, *Phys. Med. Biol.* 55 (4) (2010) 1099–1113.
- [106] J. Lee, Segmentation of positron emission tomography images: some recommendations for target delineation in radiation oncology, *Radiother. Oncol.* 96 (3) (2010) 302–307.
- [107] L. Geworski, B. Knoop, M. de Cabrejas, W. Knapp, D. Munz, Recovery correction for quantitation in emission tomography: a feasibility study, *Eur. J. Nucl. Med. Mol. Imaging* 27 (2) (2000) 161–169.
- [108] E. Ford, P. Kinahan, L. Hanlon, A. Alessio, J. Rajendran, D. Schwartz, M. Phillips, Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes, *Med. Phys.* 33 (2006) 4280–4288.
- [109] Philips, Vereos PET/CT, ([www.healthcare.philips.com/us\\_en/clinicalspecialties/radiology/solutions/vereos](http://www.healthcare.philips.com/us_en/clinicalspecialties/radiology/solutions/vereos)), [Online; accessed 19-December-2013] (2013).
- [110] C. Ballangan, C. Chan, X. Wang, D. Feng, The impact of reconstruction algorithms on semi-automatic small lesion segmentation for PET: a phantom study, in: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011*, pp. 8436–8483.
- [111] S. Mhd Saeed, A. Maysam, A. Abbes, Z. Habib, Artificial neural network-based system for PET volume segmentation, *Int. J. Biomed. Imaging* (2010).
- [112] J. Kim, L. Wen, S. Eberl, R. Fulton, 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, 2007*, pp. 4242–4245.
- [113] H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, R. Tirona, 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) (2009) 618–625.
- [114] D. Montgomery, A. Amira, H. Zaidi, Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model, *Med. Phys.* 34 (2) (2007) 722–736.
- [115] A. Amira, S. Chandrasekaran, D. Montgomery, I. Servan Uzun, A segmentation concept for positron emission tomography imaging using multiresolution analysis, *Neurocomputing* 71 (10) (2008) 1954–1965.
- [116] S. Belhassen, H. Zaidi, A novel fuzzy c-means algorithm for unsupervised heterogeneous tumor quantification in PET, *Med. Phys.* 37 (3) (2010) 1309–1324.
- [117] T. Nguyen, J. Wu, Dirichlet gaussian mixture model: application to image segmentation, *Image Vis. Comput.* 29 (12) (2011) 818–828.
- [118] M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, D. Visvikis, A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET, *IEEE Trans. Med. Imaging* 28 (6) (2009) 881–893.
- [119] F. Yang, P. Grigsby, Delineation of FDG-PET tumors from heterogeneous background using spectral clustering, *Eur. J. Radiol.* 81 (11) (2012) 3535–3541.
- [120] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (7) (1990) 629–639.
- [121] M. Aristophanous, B.C. Penney, M.K. Martel, C.A. Pelizzari, A gaussian mixture model for definition of lung tumor volumes in positron emission tomography, *Med. Phys.* 34 (2007) 4223–4235.
- [122] F. Hofheinz, J. Langner, J. Petr, B. Beuthien-Baumann, J. Steinbach, J. Kotzerke, J. van den Hoff, An automatic method for accurate volume delineation of heterogeneous tumors in PET, *Med. Phys.* 40 (8).
- [123] M. Hatt, C. Cheze le Rest, P. Descourt, A. Dekker, D. De Ruyscher, M. Oellers, P. Lambin, O. Pradier, D. Visvikis, Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications, *Int. J. Radiat. Oncol. Biol. Phys.* 77 (1) (2010) 301–308.
- [124] M. Hatt, C. Cheze-le Rest, A. Van Baardwijk, P. Lambin, O. Pradier, D. Visvikis, Impact of tumor size and tracer uptake heterogeneity in 18F-FDG PET and CT non-small cell lung cancer tumor delineation, *J. Nucl. Med.* 52 (11) (2011) 1690–1697.
- [125] M. Hatt, C. Cheze-Le Rest, E.O. Aboagye, L.M. Kenny, L. Rosso, F.E. Turkheimer, N.M. Albarghach, J.-P. Metges, O. Pradier, D. Visvikis, Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements, *J. Nucl. Med.* 51 (9) (2010) 1368–1376.
- [126] A. Le Maitre, M. Hatt, O. Pradier, C. Cheze-le Rest, D. Visvikis, Impact of the accuracy of automatic tumour functional volume delineation on radiotherapy treatment planning, *Phys. Med. Biol.* 57 (17) (2012) 5381.
- [127] B.H. de Figueiredo, T. Merlin, H. de Clermont-Gallerande, M. Hatt, D. Vimont, P. Fernandez, F. Lamare, Potential of [18f]-fluoromisonidazole positron-emission tomography for radiotherapy planning in head and neck squamous cell carcinomas, *Strahlentherapie Onkologie* 189 (12) (2013) 1015–1019.
- [128] F. Hofheinz, C. Potzsch, L. Oehme, B. Beuthien-Baumann, J. Steinbach, J. Kotzerke, J. van den Hoff, Automatic volume delineation in oncological PET. Evaluation of a dedicated software tool and comparison with manual delineation in clinical data sets, *Nuklearmedizin* 51 (1) (2012) 9–16.
- [129] D. Torigian, L.R.F. S. Alapati, G. Bodapati, F. Hofheinz, J. van den Hoff, B. Saboury, A. Alavi, Feasibility and performance of novel software to quantify metabolically active volumes and 3d partial volume corrected suv and metabolic volumetric products of spinal bone marrow metastases on 18F-FDG-PET/ct, *Hell J. Nucl. Med.* 14 (1) (2011) 8–14.
- [130] D.L. Pham, C. Xu, J.L. Prince, Current methods in medical image segmentation 1, *Annu. Rev. Biomed. Eng.* 2 (1) (2000) 315–337.
- [131] Z. Ma, J. Tavares, R. Jorge, Segmentation of structures in medical images: review and a new computational framework, in: *Proceedings of 8th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering, 2008*.
- [132] A. Kerhet, C. Small, H. Quon, T. Riauka, R. Greiner, A. McEwan, W. Roa, Segmentation of lung tumours in positron emission tomography scans: a machine learning approach, *Artif. Intell. Med.* (2009) 146–155.
- [133] E. Yoshida, K. Kitamura, Y. Kimura, F. Nishikido, K. Shibuya, T. Yamaya, H. Murayama, Inter-crystal scatter identification for a depth-sensitive detector using support vector machine for small animal positron emission tomography, *Nucl. Instrum. Methods Phys. Res. A: Accelerat. Spectrom. Detect. Assoc. Equip.* 571 (1) (2007) 243–246.
- [134] B. Foster, U. Bagci, B. Luna, B. Dey, W. Bishai, S. Jain, Z. Xu, D.J. Mollura, Robust segmentation and accurate target definition for positron emission tomography images using affinity propagation, *ISBI* (2013) 1461–1464.
- [135] B. Foster, U. Bagci, Z. Xu, B. Luna, W. Bishai, S. Jain, D.J. Mollura, Robust segmentation and accurate target definition for positron emission tomography images using affinity propagation, *IEEE Trans. Biomed. Eng.* 61 (3) (2014) 711–724.
- [136] W. Zhu, T. Jiang, Automation segmentation of PET image for brain tumors, in: *Nuclear Science Symposium Conference Record, 2003 IEEE, vol. 4, 2003*, pp. 2627–2629.
- [137] A. Boudraa, J. Champier, L. Cinotti, J. Bordet, F. Lavenne, J. Mallet, Delineation and quantitation of brain lesions by fuzzy clustering in positron emission tomography, *Comput. Med. Imaging Graph.* 20 (1) (1996) 31–41.
- [138] B. Frey, D. Dueck, Clustering by passing messages between data points, *Science* 315 (5814) (2007) 972–976.
- [139] R. Adams, L. Bischof, Seeded region growing, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (6) (1994) 641–647.
- [140] Z. Xu, Z. Gao, E. Hoffman, P. Saha, Tensor scale-based anisotropic region growing for segmentation of elongated biological structures, in: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), 2012*, pp. 1032–1035.



- [141] Z. Xu, U. Bagci, A. Kubler, B. Luna, S. Jain, W.R. Bishai, D.J. Mollura, Computer-aided detection and quantification of cavitary tuberculosis from CT scans, *Med. Phys.* 40 (11).
- [142] Z. Xu, U. Bagci, B. Foster, D.J. Mollura, A hybrid multi-scale approach to automatic airway tree segmentation from CT scans, in: 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI), 2013, pp. 1308–1311.
- [143] H. Li, W.L. Thorstad, K.J. Biehl, R. Laforest, Y. Su, K.I. Shoghi, E.D. Donnelly, D.A. Low, W. Lu, A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours, *Med. Phys.* 35 (8) (2008) 3711–3721.
- [144] Z. Xu, P.K. Saha, S. Dasgupta, Tensor scale: an analytic approach with efficient computation and applications, *Comput. Vis. Image Underst.* 116 (10) (2012) 1060–1075.
- [145] Z. Xu, F. Zhao, R. Bhagalia, B. Das, Generic rebooting scheme and model-based probabilistic pruning algorithm for tree-like structure tracking, in: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), 2012, pp. 796–799.
- [146] H. Li, W. Thorstad, K. Biehl, R. Laforest, Y. Su, K. Shoghi, E. Donnelly, D. Low, W. Lu, A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours, *Med. Phys.* 35 (8) (2008) 3711–3721.
- [147] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (11) (2001) 1222–1239.
- [148] H. Wechsler, M. Kidode, A random walk procedure for texture discrimination, *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* 1 (3) (1979) 272–280.
- [149] S. Andrews, G. Hamarneh, A. Saad, Fast random walker with priors using precomputation for interactive medical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, 2010, pp. 9–16.
- [150] L. Grady, Random walks for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (11) (2006) 1768–1783.
- [151] Z. Xu, U. Bagci, B. Foster, A. Mansoor, D.J. Mollura, Spatially constrained random walk approach for accurate estimation of airway wall surfaces, in: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*, Lecture Notes in Computer Science, vol. 8150, Springer Berlin Heidelberg, 2013, pp. 559–566.
- [152] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, *Int. J. Comput. Vis.* 1 (4) (1988) 321–331.
- [153] V. Caselles, F. Catté, T. Coll, F. Dibos, A geometric model for active contours in image processing, *Numer. Math.* 66 (1) (1993) 1–31.
- [154] C.-Y. Hsu, C.-Y. Liu, C.-M. Chen, Automatic segmentation of liver PET images, *Comput. Med. Imaging Graph.* 32 (7) (2008) 601–610.
- [155] M. Abdoli, R. Dierckx, H. Zaidi, Contourlet-based active contour model for PET image segmentation, *Med. Phys.* 40 (8) (2013) 082507.
- [156] 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, Cambridge, United Kingdom, 1999.
- [157] B. Vemuri, J. Ye, Y. Chen, C. Leonard, A level-set based approach to image registration, in: *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2000, pp. 86–93.
- [158] B.C. Vemuri, J. Ye, Y. Chen, C. Leonard, et al., Image registration via level-set motion: applications to atlas-based segmentation, *Med. Image Anal.* 7 (1) (2003) 1–20.
- [159] J. Qi, Pet image segmentation and reconstruction using level set method (Ph.D. thesis), University of Toronto, 2011.
- [160] I. El Naqa, J. Bradley, J. Deasy, K. Biehl, R. Laforest, D. Low, Improved analysis of PET images for radiation therapy, in: *Proceedings of the 14th International Conference on the Use of Computers in Radiation Therapy*, 2004, pp. 361–363.
- [161] Y.C. Lin, Y.P. Tsai, Y.P. Hung, Z.C. Shih, Comparison between immersion-based and toboggan-based watershed image segmentation, *IEEE Trans. Image Process.* 15 (3) (2006) 632–640.
- [162] X. Geets, J.A. Lee, A. Bol, M. Lonneux, V. Grégoire, A gradient-based method for segmenting FDG-PET images: methodology and validation, *Eur. J. Nucl. Med. Mol. Imaging* 34 (9) (2007) 1427–1438.
- [163] S. Liao, B.C. Penney, H. Zhang, K. Suzuki, Y. Pu, Prognostic value of the quantitative metabolic volumetric measurement on 18F-FDG PET/ct in stage iv nonsurgical small-cell lung cancer, *Acad. Radiol.* 19 (1) (2012) 69–77.
- [164] C. Riddell, P. Brigger, R. Carson, S. Bacharach, The watershed algorithm: a method to segment noisy PET transmission images, *IEEE Trans. Nucl. Sci.* 46 (3) (1999) 713–719.
- [165] H. Yu, C. Caldwell, K. Mah, D. Mozeg, Coregistered FDG PET/ct-based textural characterization of head and neck cancer for radiation treatment planning, *IEEE Trans. Med. Imaging* 28 (3) (2009) 374–383.
- [166] D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, X. Wu, Globally optimal tumor segmentation in PET-ct images: A graph-based co-segmentation method, in: *Information Processing in Medical Imaging*, 2011, pp. 245–256.
- [167] I. El Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J. Bradley, P. Grigsby, J. Deasy, Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning, *Med. Phys.* 34 (12) (2007) 4738–4749.
- [168] D. Markel, C. Caldwell, H. Alasti, H. Soliman, Y. Ung, J. Lee, A. Sun, Automatic segmentation of lung carcinoma using 3d texture features in 18-FDG PET/ct, *Int. J. Mol. Imaging*, <http://dx.doi.org/10.1155/2013/980769>.
- [169] A. Toet, H.M.A., S. Nikolov, J. Lewis, T. Dixon, D. Bull, C. Canagarajah, Towards cognitive image fusion, *Inf. Fusion* 11 (2) (2010) 95–103.
- [170] Q. Song, J. Bai, D. Han, S. Bhatia, W. Sun, W. Rockey, J. Bayouth, J. Buatti, X. Wu, Optimal co-segmentation of tumor in PET-CT images with context information, *IEEE Trans. Med. Imaging* 32 (2013) 1685–1697.
- [171] F. Giesel, A. Mehndiratta, J. Locklin, M. McAuliffe, S. White, P. Choyke, M. Knopp, B. Wood, U. Haberkorn, H. von Tengg-Koblighk, Image fusion using CT, MRI and PET for treatment planning, navigation and follow up in percutaneous RFA, *Exp. Oncol.* 31 (2) (2009) 106–114.
- [172] O. Israel, Z. Keidar, G. Iosilevsky, L. Bettman, J. Sachs, A. Frenkel, Review the fusion of anatomic and physiologic imaging in the management of patients with cancer, *Semin. Nucl. Med.* 31 (3) (2001) 191–205.
- [173] A. Bockisch, T. Beyer, G. Antoch, L.S. Freudenberg, H. Kühl, J.F. Debatin, S.P. Müller, Positron emission tomography/computed tomography – imaging protocols, artifacts, and pitfalls, *Mol. Imaging Biol.* 6 (4) (2004) 188–199.
- [174] W. Sureshbabu, O. Mawlawi, PET/CT imaging artifacts, *J. Nucl. Med. Technol.* 33 (3) (2005) 156–161.
- [175] V. Keereman, R. Holen, P. Mollet, S. Vandenberghe, The effect of errors in segmented attenuation maps on PET quantification, *Med. Phys.* 38 (11) (2011) 6010–6019.
- [176] J. Kim, W. Cai, D. Feng, S. Eberl, Segmentation of voi from multidimensional dynamic PET images by integrating spatial and temporal features, *IEEE Trans. Inf. Technol. Biomed.* 10 (4) (2006) 637–646.
- [177] F. Turkheimer, P. Edison, N. Pavese, F. Roncaroli, A. Anderson, A. Hammers, A. Gerhard, R. Hinz, Y. Tai, D. Brooks, Reference and target region modeling of [11c]-(-)-pk11195 brain studies, *J. Nucl. Med.* 48 (1) (2007) 158–167.
- [178] Y. Kimura, M. Senda, N. Alpert, Fast formation of statistically reliable FDG parametric images based on clustering and principal components, *Phys. Med. Biol.* 47 (3) (2002) 455–468.
- [179] R. Maroy, R. Boisgard, C. Comtat, V. Frouin, P. Cathier, E. Duchesnay, F. Dollé, P. Nielsen, R. Trébossen, B. Tavitian, Segmentation of rodent whole-body dynamic PET images: an unsupervised method based on voxel dynamics, *IEEE Trans. Med. Imaging* 27 (3) (2008) 342–354.
- [180] H. Guo, R. Renaut, K. Chen, E. Reiman, Clustering huge data sets for parametric PET imaging, *Biosystems* 71 (1) (2003) 81–92.
- [181] K. Wong, D. Feng, S. Meikle, M. Fulham, Segmentation of dynamic PET images using cluster analysis, *IEEE Trans. Nucl. Sci.* 49 (1) (2002) 200–207.
- [182] T. Shepherd, R. Owenius, Gaussian process models of dynamic PET for functional volume definition in radiation oncology, *IEEE Trans. Med. Imaging* 31 (8) (2012) 1542–1546.
- [183] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assie, D. Autret, S. Avner, R. Barbier, M. Bardies, P. Bloomfield, et al., Gate: a simulation toolkit for PET and spect, *Phys. Med. Biol.* 49 (19) (2004) 4543.
- [184] T. Shepherd, B. Berthon, P. Galavis, E. Spezi, A. Apte, J. Lee, D. Visvikis, M. Hatt, E. de Bernardi, S. Das, I.E. Naqa, U. Nestle, C. Schmidlein, H. Zaidi, A. Kirov, Design of a benchmark platform for evaluating PET-based contouring accuracy in oncology applications, in: *European Association for Nuclear Medicine Annual Meeting*, 2012.
- [185] B. Berthon, E. Spezi, C. Schmidlein, A. Apte, P. Galavis, H. Zaidi, E.D. Bernardi, J. Lee, A. Kirov, Development of a software platform for evaluating automatic PET segmentation methods, in: *European Society for Radiotherapy and Oncology Meeting*, 2013.