Medical Image Analysis 51 (2019) 169–183

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Disease quantification on PET/CT images without explicit object delineation

Yubing Tong[a], Jayaram K. Udupa[a,*], Dewey Odhner[a], Caiyun Wu[a], Stephen J. Schuster[b], Drew A. Torigian[a,b]

[a] Medical Image Processing group, Department of Radiology, 3710 Hamilton Walk, Goddard Building, 6th Floor, Philadelphia, PA 19104, United States
[b] Abramson Cancer Center, Perelman Center for Advanced Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States

ARTICLE INFO

Article history:
Received 30 July 2018
Revised 17 October 2018
Accepted 9 November 2018
Available online 10 November 2018

Keywords:
Image segmentation
Object recognition
PET/CT
Total lesion glycolysis (*TLG*)
Cancer
Disease quantification
Quantitative radiology

ABSTRACT

Purpose: The derivation of quantitative information from images in a clinically practical way continues to face a major hurdle because of image segmentation challenges. This paper presents a novel approach, called automatic anatomy recognition-disease quantification (AAR-DQ), for disease quantification (DQ) on positron emission tomography/computed tomography (PET/CT) images. This approach explores how to decouple DQ methods from explicit dependence on object (e.g., organ) delineation through the use of only object recognition results from our recently developed automatic anatomy recognition (AAR) method to quantify disease burden.

Method: The AAR-DQ process starts off with the AAR approach for modeling anatomy and automatically recognizing objects on low-dose CT images of PET/CT acquisitions. It incorporates novel aspects of model building that relate to finding an optimal disease map for each organ. The parameters of the disease map are estimated from a set of training image data sets including normal subjects and patients with metastatic cancer. The result of recognition for an object on a patient image is the location of a fuzzy model for the object which is optimally adjusted for the image. The model is used as a fuzzy mask on the PET image for estimating a fuzzy disease map for the specific patient and subsequently for quantifying disease based on this map. This process handles blur arising in PET images from partial volume effect entirely through accurate fuzzy mapping to account for heterogeneity and gradation of disease content at the voxel level without explicitly performing correction for the partial volume effect. Disease quantification is performed from the fuzzy disease map in terms of total lesion glycolysis (TLG) and standardized uptake value (SUV) statistics. We also demonstrate that the method of disease quantification is applicable even when the "object" of interest is recognized manually with a simple and quick action such as interactively specifying a 3D box ROI. Depending on the degree of automaticity for object and lesion recognition on PET/CT, DQ can be performed at the object level either semi-automatically (DQ-MO) or automatically (DQ-AO), or at the lesion level either semi-automatically (DQ-ML) or automatically.

Results: We utilized 67 data sets in total: 16 normal data sets used for model building, and 20 phantom data sets plus 31 patient data sets (with various types of metastatic cancer) used for testing the three methods DQ-AO, DQ-MO, and DQ-ML. The parameters of the disease map were estimated using the leave-one-out strategy. The organs of focus were left and right lungs and liver, and the disease quantities measured were *TLG*, SUV_{Mean}, and SUV_{Max}. On phantom data sets, overall error for the three parameters were approximately 6%, 3%, and 0%, respectively, with *TLG* error varying from 2% for large "lesions" (37 mm diameter) to 37% for small "lesions" (10 mm diameter). On patient data sets, for non-conspicuous lesions, those overall errors were approximately 19%, 14% and 0%; for conspicuous lesions, these overall errors were approximately 9%, 7%, 0%, respectively, with errors in estimation being generally smaller for liver than for lungs, although without statistical significance.

Conclusions: Accurate disease quantification on PET/CT images without performing explicit delineation of lesions is feasible following object recognition. Method DQ-MO generally yields more accurate results than DQ-AO although the difference is statistically not significant. Compared to current methods from the literature, almost all of which focus only on lesion-level DQ and not organ-level DQ, our results were

* Corresponding author.
E-mail addresses: jay@pennmedicine.upenn.edu, jay@mail.med.upenn.edu (J.K. Udupa).

https://doi.org/10.1016/j.media.2018.11.002

comparable for large lesions and were superior for smaller lesions, with less demand on training data and computational resources. DQ-AO and even DQ-MO seem to have the potential for quantifying disease burden body-wide routinely via the AAR-DQ approach.

1. Introduction

1.1. Background

It is now generally believed that quantitative radiology (QR), when brought to routine clinical practice, will bring about significant enhancement of the role of radiology in the medical milieu, potentially spawning numerous new advances in medicine. The derivation of quantitative information from images, however, continues to face a major image analysis hurdle, namely the identification and delineation of "objects" of interest in the image. The "object" may be an anatomic organ, a sub-organ, a tissue region, a pathological region, or an anatomic zone such as a well-defined lymph node station. Called *image segmentation*, this process has a rich and long history spanning nearly 5 decades (Doyle, 2012; Narasimhan and Fornango, 1963) in the general area of image processing. Image segmentation has, however, remained the toughest challenge in image analysis and an essential roadblock to the practical clinical implementation of QR.

Perhaps owing to the above challenge, most past efforts on the development of segmentation algorithms have focused on specific organs and image modalities. The literature on algorithms for segmenting each important organ in the body such as brain (Ashburner and Friston, 2009; Zhou and Rajapakse, 2005), heart (Zhuang et al., 2015, Zhuang and Shen, 2016), lungs (Mansoor et al, 2014, Kohlmann et al., 2015), and liver (Priyadarsini et al., 2012; Goceri et al., 2014) individually on magnetic resonance imaging (MRI) and computed tomography (CT) images is vast. While these algorithms have brought about many advances in the study of diseases pertaining to these specific organs, they are not generalizable to body-region-wide or body-wide applications that require the segmentation of all or all major objects/organs in the region under consideration. A new breed of methodologies is now evolving to address this issue of segmenting multitudes of organs situated body-wide or in an entire body region (Udupa et al., 2014; Kashyap et al., 2018; Lee et al., 2016; Oliveira et al., 2018; Hu et al., 2017; Okada et al., 2015; Tong et al., 2015; Namias et al., 2016).

While it remains to be seen how these advances will influence the practice of QR in the future, it is time to think about how to decouple methods of disease quantification from explicit dependence on image segmentation when (and to the extent) possible. In this regard, as formulated in all of our previous segmentation work, it is helpful to formulate image segmentation as being comprised of two related processes – object *recognition* and object *delineation. Recognition* is the high-level process of determining the whereabouts of the object or locating the object in an image. *Delineation* is the low-level process of precisely demarcating the boundary of or the region occupied by the object in the image. Note that, in a segmentation method, each process can be implemented to operate manually or fully automatically or at different levels of automation by using different strategies.[1] For example, in fully manual segmentation, both processes are manually executed. In the Live Wire method (Falcao et al., 1998) (where this two-tier division of segmentation was first introduced), automatic delineation occurs in real time in response to manual recognition where the two processes are coupled tightly and synergistically so that segmentation becomes accurate, agreeable to the human operator in real time "on the fly", and efficient with no requirement for post-hoc correction.

Generally, the recognition process can be automated to perform much more robustly than delineation. On many occasions, delineation becomes ill-defined due to artifacts such as noise, beam hardening effects, blur, image non-uniformity, and intensity non-standardness, and the presence of pathology and its variations, even though recognition can be performed quite effectively. The above dichotomization of segmentation is useful for determining when recognition alone may be sufficient and when delineation may also be needed for the image analysis task at hand. This allows making the process of quantification more robust and to a large extent independent of the vagaries of the segmentation (delineation) process if image analysis can be performed by just recognition alone, irrespective of whether it is accomplished manually or automatically or with a variable degree of automation. While some image analysis applications such as radiation treatment planning may require generation of object contours, and hence explicit delineation, there is no need to tie up many other tasks to the success of delineation. In this manuscript, following this tenet, we will demonstrate that automatic and accurate quantification of disease in different organs in patients with cancer is feasible immediately following object recognition via whole-body positron emission tomography/computed tomography (PET/CT) image acquisitions.

1.2. Related work and scientific gaps

Currently, body-wide ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG)-PET/CT is the most commonly used modality for molecular imaging of patients with cancer. FDG-PET/CT improves the sensitivity for detection of pathology at the molecular, subcellular, or cellular level well before gross anatomic changes manifest, and simultaneously improves the specificity of diagnosis to distinguish whether macroscopic abnormalities are benign or malignant in nature (Kwee et al., 2013a,b). As such, it changes management in up to 40% of patients with cancer prior to implementation of treatment, often due to improved detection of regional lymph node metastases and distant metastases in the body, and improves the diagnostic performance of post-treatment assessment compared to structural imaging with CT or MRI alone (Kwee et al., 2013a,b; Hillner et al, 2008). Importantly, FDG-PET/CT provides image data that are quantifiable prior to and following treatment, allowing for individualized regional and global disease assessment of patients with cancer (Kwee et al., 2013a,b).

To appreciate the importance of disease quantification body-region-wide or body-wide in patients with cancer and other disease conditions, let us examine the primary clinical tasks for which medical imaging is used.

1. *Screening*: The goal is to detect disease that would eventually become clinically significant if untreated prior to the onset of clinical symptoms or signs in order to improve patient outcome.

[1] The word *segmentation* is often used to refer to delineation where recognition is implicitly assumed but often manually performed. However, methods with different levels of automation, varying from purely manual to fully automatic in the two-dimensional space of the degree of automation for recognition and delineation, are also available.

2. *Detection/diagnosis*: The goal is to determine if pathology is present or not, and to determine what the pathology is caused by (e.g., neoplastic disease, inflammatory/infectious disease, traumatic injury, congenital disorder, etc.) in patients who have symptoms or signs of disease. If the pathology is neoplastic in nature, other related tasks are to determine whether it is benign or malignant in nature as well as indolent or aggressive in biologic behavior, and to determine its specific histopathologic subtype when possible.
3. *Staging*: The goal is to determine the spatial extent and burden (i.e., stage) of disease, so that optimal approaches to manage the patient may be devised.
4. *Prognosis assessment*: The goal is to predict clinical outcome (i.e., prognosis) of the patient (e.g., overall survival, quality of life, etc.) prior to the implementation of therapy.
5. *Treatment planning*: The goal is to arrive at an optimal plan for patient treatment by utilizing available information in order to maximize patient clinical outcomes by treating the disease while simultaneously minimizing adverse events related to the treatment itself.
6. *Treatment prediction assessment*: The goal is to predict if and how well a particular disease will respond to a particular therapy prior to or during early implementation of treatment.
7. *Treatment response assessment*: The goal is to assess if and how well a particular disease has already responded to treatment (e.g., complete response, partial response, stable disease, progressive disease) during or following implementation of treatment.
8. *Restaging/ surveillance*: The goal is to monitor for disease relapse after a disease condition has been fully treated, and to again determine (i.e., restage) the spatial extent and burden of recurrent disease when present.

Medical images are currently acquired for most of the above tasks and subsequently interpreted mainly qualitatively or at best semi-quantitatively because of the bottleneck of image segmentation. The ability to routinely quantify disease body-wide in a production mode can potentially improve many of these tasks significantly, particularly Tasks 3–7, since manual methods of quantification are error-prone, are subject to intra- and inter-observer variations, are labor-intensive, result in suboptimal throughput in clinical practice, and are just impractical when the disease is extensive.

Several commercial vendors currently offer software for disease measurement (Withofs et al., 2014; Geworski et al., 2010; Hofheinz et al., 2007,2010). They all operate under the paradigm of first manually performing recognition of diseased tissue regions by manually specifying a *region of interest* (ROI), subsequently automatically delineating lesions by making use of information from PET alone or from both PET and CT, and finally measuring disease burden in the form of tumor volumes and PET standardized uptake value (SUV) statistics within tumor lesions. Numerous papers have also been published (Bagci, et al., 2013; Fagundes et al., 2018; Piert et al., 2018; Xu et al., 2017; Mena et al., 2017; Ju et al., 2015; Cui et al., 2016; Tan et al., 2017) whose focus has been accurate segmentation (delineation) of tumors from combined PET and CT information once adequate recognition help is given manually (such as via placement of an ROI). For body-wide applications, these current methods leave two key gaps which prevent their routine clinical use. (1) Manual perusal of image slices and specification of ROIs for each tumor site in such a manner that lesion delineation will work (and hence disease quantification will be accurate) are labor-intensive and impractical when the disease is extensive, even if confined to a single organ, and especially when involving organs and tissues body-wide. There are also associated issues of intra- and inter-operator variations in disease measurements. For accurate measurement, ROIs need to be selected carefully to properly enclose tumor sites but not too much of other tissue regions. Except when disease is discrete and focal, selecting an appropriate ROI itself becomes challenging, making the quantification of multifocal or diffuse disease quite difficult. (2) Currently available automated methods are organ-specific and not generalizable, leaving disease measurement by organ and by body region an open problem. The novel methodology presented in this paper, which we will refer to as *AAR-DQ* (short for *automatic anatomy recognition-disease quantification*) is an attempt to overcome these hurdles.[2]

1.3. Outline of proposed approach

The AAR-DQ methodology, schematically illustrated in Fig. 1, starts off with the AAR approach (Udupa et al., 2014; Wang et al., 2016) for modeling anatomy and automatically recognizing objects on low-dose CT images of PET/CT acquisitions. It incorporates novel aspects of model building that relate to finding an optimal disease map for each organ. In Section 2, we describe these aspects and summarize the other previously published model building and recognition steps of AAR for completeness. The result of recognition for an object on a patient image I is the location of a fuzzy model for the object which is optimally adjusted for I. The model is used as a fuzzy mask on the PET image for estimating a fuzzy disease map for the specific patient and subsequently for quantifying disease based on this map. This process, described in Section 3, handles blur arising in PET images from partial volume effect entirely through accurate fuzzy mapping to account for heterogeneity and gradation of disease content at the voxel level without explicitly performing correction for the partial volume effect. Disease quantification is performed from the fuzzy disease map in terms of volumes and SUV statistics. We also demonstrate that the method of disease quantification is applicable even when the "object" of interest is recognized manually with a simple and quick action such as interactively specifying a 3D box ROI. Experimental evaluation studies are carried out, as described in Section 4, on phantoms with known radiotracer concentrations as well as on patient PET/CT images where manual disease measurement is taken as the reference standard ground truth for comparison with results from the AAR-DQ approach. Our results and concluding remarks are summarized in Sections 4 and 5, respectively.

The AAR-DQ approach has the following unique features: (1) It decouples dependence on explicit segmentation (delineation) of the organ and diseased tissue regions and performs DQ directly from object location information found automatically. This makes the disease quantification process robust, efficient, and practical. (2) It takes a fuzzy approach for handling uncertainties for object modeling, object recognition, disease mapping, and disease quantification, which obviates the need for explicitly correcting for phenomena such as the partial volume effect. (3) By the characteristics of the AAR approach, AAR-DQ is not tied to any specific object, and hence is applicable body-wide.

A preliminary version of this paper appeared in the proceedings of the 2017 SPIE Medical Imaging Conference (Tong et al., 2017). The present paper includes the following significant enhancements over the conference paper: (i) The concept of conspicuous and non-conspicuous lesion level disease quantification as well as corresponding results. (ii) A more comprehensive Introduction with a general and deeper literature review of disease quantification on PET/CT. (iii) A detailed description of the complete framework of AAR for DQ, a description of the complete family of AAR-DQ approaches including DQ-AO, DQ-MO, and DQ-ML strategies. (iv) Extensive experimental results at object level and lesion level, includ-

[2] In this paper, we focus on disease quantification in organs. Disease quantification in lymph node zones will be dealt with in a separate paper.

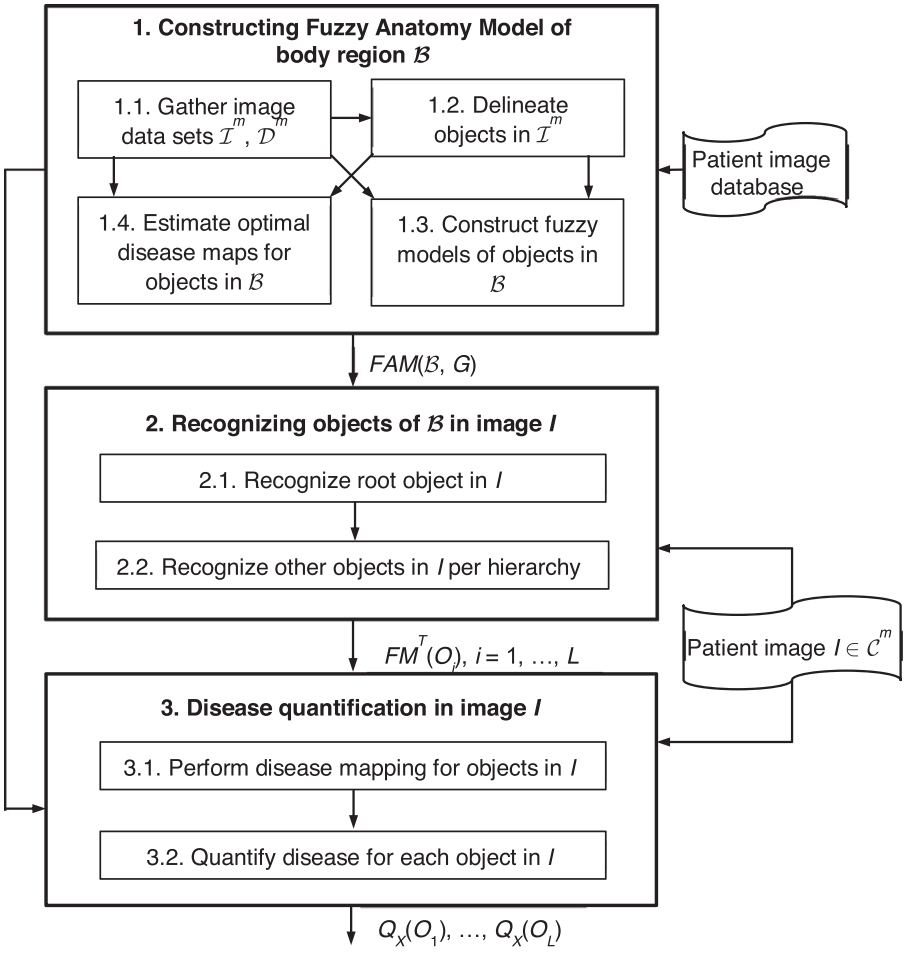

Fig. 1. A schematic representation of the AAR-DQ approach.

ing conspicuous and non-conspicuous lesions. (v) A comprehensive comparison with other current methods.

2. Anatomy recognition

We will follow the notation used in the previous AAR publications (Udupa et al., 2014; Wang et al., 2016) closely, but will need some new terminology as well.

G: The patient *population group* under consideration.
B: The *body region* of focus, in our case the upper body torso (i.e., thorax and abdomen combined).
O_1, ..., O_L: L *objects* or organs of B.
$\mathcal{I}^m = \{I_1^m, ..., I_N^m\}$: A set of *training images* in modality m of body region B from N subjects in group G which are used for constructing models. In our case, $m \in$ {CT, PET}, where CT refers to the low-dose CT image of PET/CT acquisitions. We assume that the images in $\mathcal{I}^m$ are near normal and that the CT and PET images of the same PET/CT acquisition are in registration.
$\mathcal{I}_b = \{I_{n,\ell} : 1 \le n \le N \& 1 \le \ell \le L\}$: The set of all *binary images* used for model building, $I_{n,\ell}$ being the binary image representing object O_l in image I_n^m. Since CT and PET images are in registration, binary image $I_{n,\ell}$ is applicable as a mask for object O_l in both I_n^{CT} and I_n^{PET}.
$FM(O_l)$: *Fuzzy model* of object O_l derived from the set of all binary images of O_l.
$d_O(x)$: *Disease map* associated with object O. It maps SUV x at a voxel v within O to disease severity at v on a [0, 1] scale.
$\mathcal{D}^m = \{D_1^m, ..., D_K^m\}$: A set of *training images* in modality $m \in$ {CT, PET} of patients in group G with disease. These data will be used for estimating (training) the parameters of the disease map $d_O(x)$ for each object O.
$\mathcal{C}^m = \{C_1^m, ..., C_M^m\}$: A set of *test images* in modality $m \in$ {CT, PET} of patients in group G with disease. These data will be used for testing the AAR-DQ approach.
FAM(B, G): *Fuzzy anatomy model* of the whole object assembly in B which includes all prior information gathered about objects such as the hierarchical arrangement of objects, their SUV properties, disease maps, object relationships, fuzzy models, etc.
$FM^T(O)$: Transformed $FM(O)$ corresponding to a state when O is recognized in a patient image. $Q_X(O)$: A set of quantitative measures[3] describing the disease of O.

We will follow the schematic in Fig. 1 for describing the AAR-DQ approach. Since Steps 1.4, 3.1, and 3.2 are novel additions to the previous AAR approach, they will be described in detail while the other steps will be briefly summarized for completeness.

2.1. Constructing anatomy model of the body region

For completeness, we will briefly summarize Steps 1.1 to 1.3 (Fig. 1) of the model building process first, which are introduced in

[3] Notation Q_X is fashioned after notations D_X and R_X commonly used for diagnostics and therapeutics, and is intended to denote quantitative disease analytics.

the following sections, respectively, including gathering images, delineating objects, and constructing fuzzy models, see (Udupa et al., 2014, Section 2) and (Wang et al., 2016, Section 2.1) for details. This will be followed by a detailed description of Step 1.4, which is for estimating the optimal parameters of disease maps.

2.1.1. Gathering images

All images were selected from our health system patient image database by a board-certified radiologist (co-author DAT) following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act waiver. For the normal set $\mathcal{I}^m$, the whole-body FDG-PET/CT images selected were near normal with exception of minimal incidental focal abnormalities such as cysts, small pulmonary nodules, etc. For abnormal sets $\mathcal{D}^m$ and $\mathcal{C}^m$, the same radiologist selected whole-body FDG-PET/CT images of patients with various types of metastatic cancer involving multiple organ systems. The combined patient data set $\mathcal{D}^m \cup \mathcal{C}^m$ included 31 patients in total with age 60.8 ± 9.8 yrs and normal subject data sets ($\mathcal{I}^m$) included 16 subjects in total with age 44.7 ± 10.2. All data sets included both CT and PET images.

2.1.2. Delineating objects

As per AAR methodology, anatomic body regions and the organs in them are precisely defined first (see Table 1 in Udupa et al., 2014 and Table 2 in Wang et al., 2016). All objects are then delineated following these definitions, strict tracing protocols, and scrutiny of delineations as described in Udupa et al. (2014) and Wang et al. (2016). This step generates the set of binary images $\mathcal{I}_b = \{I_{n,\ell} : 1 \le n \le N \& 1 \le \ell \le L\}$ from the input set of images $\mathcal{I}^m$. The tracings are done on the CT images of this set.

2.1.3. Constructing fuzzy models

We will focus on the following objects in the body torso in this paper for demonstrating the new ideas underlying disease quantification. The AAR-DQ approach is applicable to other and any number of objects if they can be recognized with adequate accuracy. BT: Upper body torso, which is made up of thoracic and abdominal body regions. BTSkn: The outer boundary of the body torso skin, the interior of which constitutes the upper body torso region. TSkn and ASkn: Similar to BTSkn but defined for the thoracic and abdominal body regions, respectively. LPS, RPS: Left and right pleural spaces including lungs, respectively. PS: Pleural spaces including lungs = LPS+RPS. Lvr: Liver. As in previous AAR methods, all objects considered in this work include their interior 3D region and not just the boundary.

The Fuzzy Anatomy Model *FAM*(B, *G*) is defined by 5 entities:

$$FAM(B, G) = (H, M, \rho, \lambda, \eta). \quad (1)$$

For a detailed description of these parameters, see Udupa et al. (2014). Briefly, H is a hierarchy of objects in B, represented as a tree. This tree structure permits imposing an order among objects and allows encoding non-linear and very detailed anatomic information about the population into the model. M is a set of fuzzy models, one model for each of the L objects in B, $M = \{FM(O_k): k = 1, ..., L\}$. ρ describes the parent-to-offspring relationship in H over the population. λ is a family of scale factor ranges. η denotes a set of measurements pertaining to the object assembly in B including intensity properties and all learned parameters that are needed for object recognition and disease quantification.

We will choose the object hierarchy H depicted in Fig. 2 for constructing *FAM*(B, *G*), where 3D renderings for different parts (object models) are illustrated. Fuzzy object model building will follow the hierarchy H by starting from root object of BTSkn, and then to other offspring objects. Recognition in the following

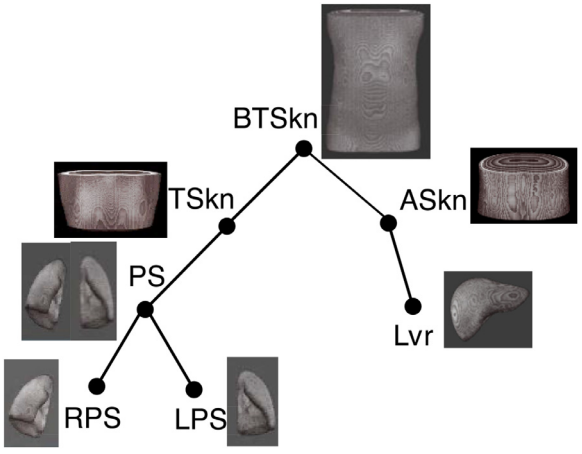

Fig. 2. Hierarchy chosen for the objects. Object abbreviations are described in the text.

Section 2.2 will also follow the same hierarchy. The fuzzy model set M is built from training binary images in the set $\mathcal{I}_b$ as described in (Udupa et al., 2014). This process consists of estimating the mean shape length and mean geometric center over all samples in $\mathcal{I}_b$ of each object O_l, repositioning all samples of O_l to this mean position, and rescaling them to mean shape length. Subsequently, a distance transform is applied to each resulting sample, and the average distance of the samples is computed and transformed to a fuzzy object membership value. From the repositioned and resized samples, the parent-to-offspring relationship ρ_l of O_l with respect to its unique parent in the hierarchy is estimated. Similarly, the size variation bounds $\lambda = \{\lambda_l: 1 \le l \le L\}$ over all samples are estimated from the same samples using the shape length of each O_l.

The fifth element η of *FAM* (B, *G*) stores values of parameters needed for *object recognition* and *disease quantification*. The parameters for *object recognition* are estimated as described in Udupa et al. (2014) and Wang et al. (2016). Briefly, apart from hierarchy H, fuzzy model set M, object relationship ρ, and scale variation λ, the only additional parameter needed is the optimal threshold Th_l for each object O_l. These parameters are estimated from image sets $\mathcal{I}^m$ and $\mathcal{I}_b$ by using Algorithm OTE described in Wang et al. (2016). Th_l is estimated by searching for a threshold interval that maximally separates the histogram of O_l from the histogram of the complement of O_l over all images in $\mathcal{I}^m$. In our case, m = CT, that is, we use the CT image for object recognition.

2.1.4. Estimating optimal disease maps

For any PET image I, let I_S denote the corresponding SUV image as defined by (Torigian et al., 2011)

$$I_S(v) = \frac{I^c(v)}{ID/BW}, \quad (2)$$

where ID is the injected dose of the radiotracer (expressed in MBq), BW is the body weight of the patient (expressed in g) whose acquired PET image is I, and $I^c(v)$ denotes the radioactivity concentration (expressed in MBq/cc where we assume 1 cc of tissue weighs 1 g) measured at voxel v of I which is corrected for decay from the time of injection to the time of image acquisition.

For *disease quantification*, we will employ a parametric function called *disease map*, denoted $d_O(x)$ which maps SUV value x at any voxel within object O to disease severity value on a [0, 1] scale.

This map is intended to be specific to object[4] O. The parameters of $d_O(x)$ will be estimated as explained below. Let μ_n^O and σ_n^O be the mean and standard deviation of SUV within *normal* object O as determined from normal image set $\mathcal{I}^{PET}$, binary image set $\mathcal{I}_b$, and the corresponding SUV images I_S.

Disease map $d_O(x)$ is modeled as $d_O(x) = \max[0, g_d(x) - g_n(x)]$, where $g_d(x)$ and $g_n(x)$ are half and full Gaussians with parameters (μ_d^O,σ_d^O) and (μ_n^O,σ_n^O), respectively. Our intent is that $g_n(x)$ describes the SUV distribution within the *normal* tissues of object O, and $g_d(x)$ expresses SUV-to-degree-of-disease relationship for O.

$$g_d(x) = \begin{cases} \exp[-(x-\mu_d)^2/2\sigma_d^2], & if\ x < \mu_d, \quad g_n(x) = \exp[-(x-\mu_n)^2/2\sigma_n^2]. \\ 1, & if\ x \geq \mu_d \end{cases} \tag{3}$$

The disease map $d_O(x)$ removes any contributions from normal tissue to the "degree of disease". When $g_d(x) < g_n(x)$, $d_O(x)$ is forced to be equal to zero to guarantee that the value of $d_O(x)$ is within [0, 1]. Assume for now that parameters (μ_d^O,σ_d^O) and (μ_n^O,σ_n^O) have been determined (see below for estimation method). The total disease burden within O is described in terms of *total lesion glycolysis* (*TLG*) defined as follows.

For any object O in any PET image I in $\mathcal{D}^m$, let $TLG_O(I)$ denote the (true) total lesion glycolysis of all lesions of O in I as determined by a reference method; that is, $TLG_O(I)$ is the sum over all lesions of the product of the lesion volume and its mean SUV. Our goal is to arrive through $d_O(x)$ at an estimate of the total disease burden of O that is as close as possible to $TLG_O(I)$. The disease quantity estimated by the AAR-DQ approach will be called *fuzzy total lesion glycolysis* of O in I, denoted $fTLG_O(I)$. The definition of $fTLG_O(I)$ requires the specification of a region within I. This region may be the whole image domain, an entire body region, or any specified ROI (binary or fuzzy), including in particular the fuzzy model $FM^T(O)$ of O localized in I by the AAR recognition step. Keeping these possibilities in mind, we denote such a general "ROI" specified in I by A and use the notation $A(v)$ to denote the membership of voxel v of I in the ROI to allow for A to be also a fuzzy ROI such as $FM^T(O)$. Thus, when A is a binary mask, $A(v) \in \{0, 1\}$, and when A is fuzzy, $A(v) \in [0, 1]$. The traditional approach for calculating TLG is first to obtain a binary lesion mask by applying a thresholding or hard segmentation operation on every voxel where partial volume effect occurs, and then utilizing the binary mask, which usually covers a smaller region than the original region used, for TLG estimation. In this paper, we consider a fuzzy volume instead. Some voxels would be discarded in the traditional approach after segmentation such that there can be no contribution from them to TLG. Yet, those voxels may still contribute to TLG and hence should not be discarded, and so their contribution can instead be described by using fuzzy membership followed by disease map as we describe in this paper.

With this generality, for the disease under consideration, we define the *fuzzy total lesion glycolysis within an ROI A*, denoted $fTLG_A(I)$, in a PET image I by

$$fTLG_A(I) = \upsilon \sum_{all\ v\ in\ A} d_O(I_S(v))I_S(v)A(v). \tag{4}$$

In words, $fTLG_A(I)$ (expressed in cc) is a weighted sum of the SUV values of voxels within mask A multiplied by the voxel volume υ (expressed in cc), assuming all voxels are of the same size. There are two weights for each voxel – $A(v)$, which is the mask weight, and $d_O(x)$, which is the disease weight based on the SUV $I_S(v)$ at v. Of course, $d_O(I_S(v))$ is unknown at this point at v. Along similar lines, to accommodate a general (hard) mask A, we modify the previously defined true total lesion glycolysis $TLG_O(I)$ within object O to $TLG_A(I)$ such that the latter denotes the sum of the true total lesion glycolysis of all lesions within binary mask A.

The map $d_O(x)$ is completely determined by μ_d^O and σ_d^O. Our idea is to estimate the parameters of $d_O(x)$ optimally so that the disease weight expressed by $d_O(x)$ brings our estimate $fTLG_A(I)$ of the disease burden as close as possible to the true estimate $TLG_A(I)$. We perform this estimation by optimizing the following function.

$$(m_d^O, s_d^O) \in^{argmin}_{\mu_d^O,\sigma_d^O} [\sum_{I \in D^m} (TLG_A(I) - fTLG_A(I))^2]. \tag{5}$$

The optimal parameters of $d_O(x)$ are denoted by (m_d^O, s_d^O). We find the optimal parameters using Powell's NEWUOA software (Powell, 2006).

The fuzzy treatment in disease quantification allows for handling both the segmentation issue of deciding whether or not a voxel belongs to a lesion as well as the determination of the SUV measurement at each voxel without explicit partial volume correction and binary segmentation commitment.

Estimation of the disease map $d_O(x)$ requires parameters μ_n^O and σ_n^O, which in turn need data sets $\mathcal{I}^{CT}$ and $\mathcal{I}^{PET}$. We estimateμ_n^O and σ_n^O directly from sets $\mathcal{I}^{PET}$ and $\mathcal{I}_b$. Estimating $TLG_A(I)$ requires data sets $\mathcal{D}^m$, $m \in$ {CT, PET}, and is challenging at the lesion level, and hence at organ level, mainly because of the extreme variability of the fuzziness of the lesions. We take the approach described below to establish $TLG_A(I)$.

2.1.5. Establishing true disease measurements

Commercial clinical software systems generally require a human operator to specify an ROI manually corresponding to each lesion to be quantified on the PET image. As illustrated in Fig. 3, the ROI should be specified fairly close to the lesion boundary on a slice, and the extent of the ROI in the third dimension orthogonal to the slice plane should also be indicated. The software then generally performs an iterative thresholding operation, sometimes with partial volume correction depending on the particular software system, and outputs the volume of the lesion, commonly known as *metabolic lesion volume* (MLV) (expressed in cc), mean and maximum SUV of the lesion, and then a product of MLV and mean SUV of the lesion called *total lesion glycolysis* (TLG) (expressed in cc).

We used one such software system called ROVER (Hofheinz et al., 2007,2010) for generating reference true measurements. We found this software to be adequate for use in the above manner for large, well-defined, and focal lesions. We refer to such lesions which can be delineated automatically by the clinical software via an ROI without requiring parameter adjustment and whose segmentations seem visually accurate as *conspicuous* lesions. However, the behavior of this software was generally not stable for lesions that are not well-defined, large, or focal. Accordingly, those lesions whose delineation by the clinical software requires manual adjustment of parameters or delineations on a per-lesion basis will be referred to as *non-conspicuous* lesions. For conspicuous lesions, we generated true measurements by using the clinical software. For non-conspicuous lesions, we created reference measurements by individually thresholding each lesion on the PET image to produce visually optimal results under the guidance of an expert radiologist (DAT) who has over 10 years of experience in making such measurements clinically. Examples of conspicuous and non-conspicuous lesions are shown in Fig. 3. As shown in the two bottom rows, non-conspicuous lesions labeled via arrows in the ROIs (circles/ ovals) may be over-segmented or under-segmented or missed when using clinical software, subsequently requiring manual adjustment.

[4] Implicit in our assumption is the fact that the map is specific to a particular disease of O, which in the current study is cancer. Multiple other disease conditions will be considered in our future work.

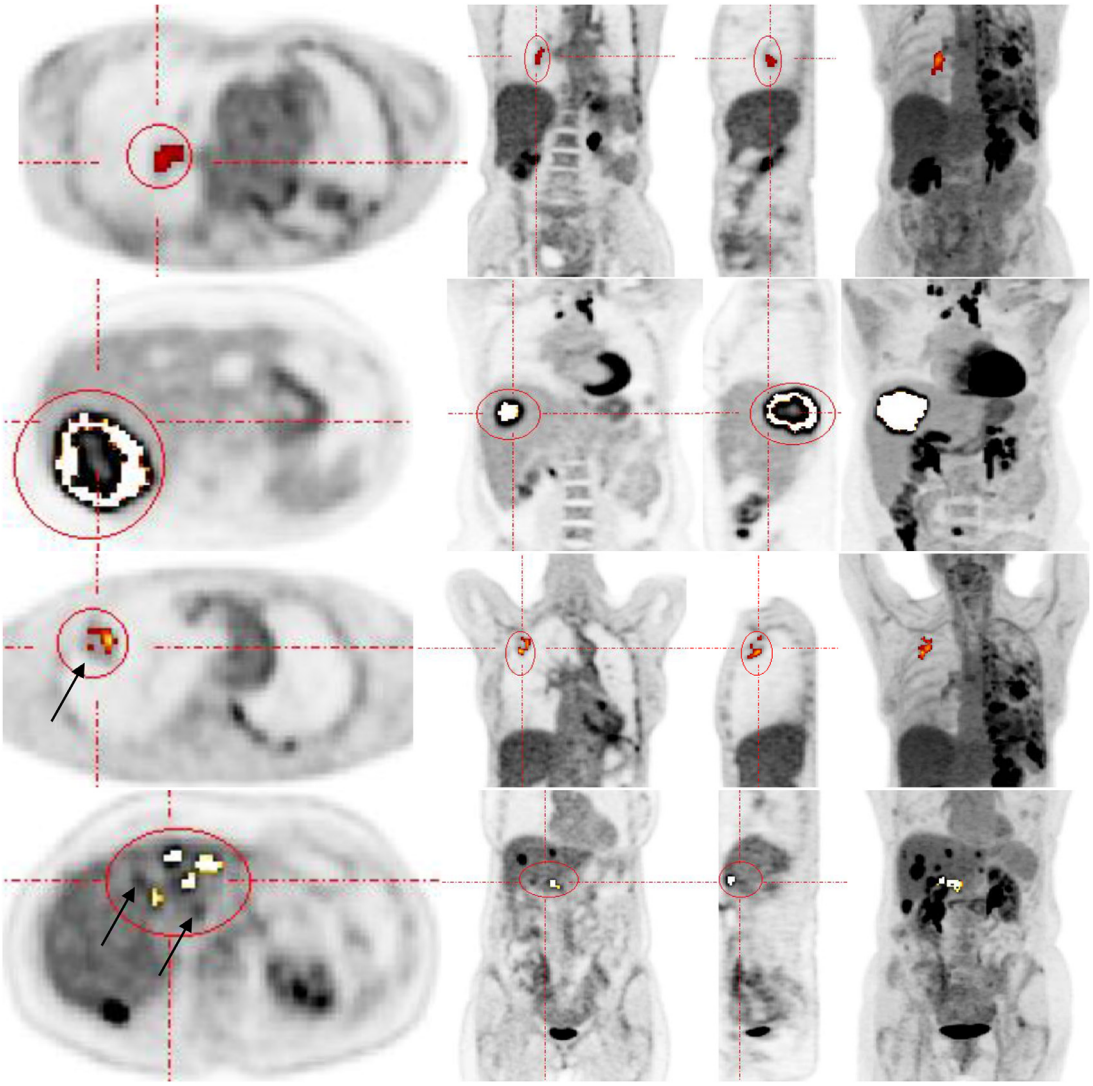

Fig. 3. Illustration of conspicuous and non-conspicuous lesions of the lung and liver on PET images. The rows correspond to 4 different patients with metastatic cancer, and the columns represent (left to right) axial, coronal, and sagittal slices and Maximum Intensity Projection in the coronal plane. ROIs (circles/ ovals) required by a commercial clinical software system and its output segmented tumors are shown as overlay. The top two rows show examples of adequate segmentation of conspicuous lesions by the software, whereas the bottom two rows show examples of inadequate segmentation of non-conspicuous lesions (arrows). Note that, even for conspicuous lesions, their fuzzy regions and boundaries are left uncovered.

In summary, at the end of the model building step (see Fig. 1), we have a fuzzy anatomy model $FAM(B, G)$ of the body region B complete with anatomic information encapsulated in H, M, ρ, and λ, and disease map information included in η for the particular disease being studied for each object of B.

2.2. Recognizing objects

We will use both manual and automatic recognition strategies. In the manual mode, we will present two methods, one at the object level and another at the lesion level, and in both, one ROI A will be specified by using rectangular boxes[5] to keep the manual recognition and specification of A simple and efficient. For further reference, we will denote these manual methods by MO and ML, respectively; in MO, the object is specified by a given mask, such as rectangular box just enclosing the object, and in ML, the box specified just encloses each lesion. The goal of the MO method is to demonstrate that, albeit manual, it can be used to efficiently and accurately quantify the total disease burden within an object via the proposed approach. This is currently not feasible by employing clinically available software systems. The current manual way of quantifying each lesion on its own is not practical when lesions and/or involved organs are numerous. The goal of the ML method is to demonstrate that even when each lesion is identified manually, the proposed approach performs quantification accurately. In both manual methods, the ROI A for disease quantification is a binary mask.

The real thrust of this manuscript is on the automatic mode at the object level, which we will denote by AO, wherein objects are recognized automatically by the AAR approach. We are not introducing any new concepts in this paper over those in Udupa et al. (2014) and Wang et al. (2016) for organ recognition per se. The process follows the recognition algorithms of Udupa et al. (2014), Wang et al. (2016) and starts off by first recognizing the root object and then follows the hierarchy displayed in Fig. 2 to locate other objects. The output of the automatic recog-

[5] The ROI mask does not need to be rectangular; in fact, it can be any given shape to localize the target object. CAVASS software (Grevera et al., 2007) supports an efficient way for generating such a mask, which involves manually drawing an ROI on a single slice and then propagating it to other slices automatically.

nition step is the modified fuzzy model $FM^T(O)$ for each object O optimally adjusted to the manifestation of O in the CT image of the given PET/CT pair. In other words, in this case, the ROI for object O for disease quantification is a fuzzy mask $A = FM^T(O)$.

In summary, we will utilize three methods of recognition – manual at object level (MO), manual at lesion level (ML), and automatic at object level (AO) to define the mask A needed for disease quantification.

3. Disease quantification

We will denote the disease quantification procedures for the three recognition methods AO, MO, and ML by DQ-AO, DQ-MO, and DQ-ML, respectively. The disease quantification procedure DQ-AO summarized in the box for recognition method AO follows directly Eq. (4). For recognition method MO, the procedure DQ-MO is the same as DQ-AO except that in Step 1, objects are identified manually, dilation in Step 3 is not performed, and the specified ROI is to be considered as a binary mask A. For DQ-ML, "object" is to be interpreted as a lesion and then the procedure DQ-MO is to be followed. In this case, L indicates the number of lesions quantified.

In words, in procedure DQ-AO, each anatomic object O is first recognized or localized in the CT image automatically. The fuzzy model $FM^T(O)$ found in this process is then dilated.[6] The resulting fuzzy mask, called A, is then applied to the SUV image derived from the input PET image. The disease map $d_O(x)$ of O is retrieved from the element η of $FAM(B, G)$ and the total disease burden $fMTV_O$ of O is computed via Eq. (4). The contribution from each voxel within the localized object is weighted by the level of disease at the voxel and the level of certainty for the voxel to belong to O. For manual methods MO and ML, the weight of belongingness to the object is binary but the weight of disease severity coming from $d_O(x)$ is fuzzy. Note that, for methods MO and ML, there may be voxels in the binary mask A that are outside the object proper which will be weighted by 1. However, if the disease map is accurate for O, such voxels will receive negligible disease weight from $d_O(x)$. Finally, the output of the procedure for each object consists of the object's SUV_{Mean} and SUV_{Max} and its estimated $fTLG_A$ value with the appropriate interpretation of the meaning of these entities as explained above for the cases of DQ-AO, DQ-MO and DQ-ML.

Procedure DQ-AO

Input: An image pair (I^{CT}, I^{PET}) of $\mathcal{B}$ from the set $\mathcal{C}^m$ and $FAM(\mathcal{B}, G)$.
Output: $Q_X(O_i)$, $i = 1, \ldots, L$.
Begin
1. Recognize objects O_i, $i = 1, \ldots, L$, in I^{CT};
2. *For* each object O_i, *do*
3. Perform gray-level dilation of $FM^T(O_i)$ and let the result be A;
4. Compute SUV_{mean}, SUV_{max}, and $fTLG_A(I^{PET}) = \upsilon \sum_{all\ v\ in\ A} d_O(I_S^{PET}(v)) I_S^{PET}(v) A(v)$, where υ denotes the volume of a voxel in A;
5. $Q_X(O_i) = [SUV_{mean}(A), SUV_{max}(A), fTLG_A(I^{PET})]$;
6. *EndFor*;
7. Output $Q_X(O_i)$, $i = 1, \ldots, L$;
End

In summary, we have described three methods for disease quantification in a single general framework: DQ-AO, DQ-MO, and DQ-ML. DQ-AO and DQ-MO perform disease quantification at the whole object (organ) level, with DQ-AO recognizing objects automatically and DQ-MO localizing objects manually. DQ-ML performs quantification at the lesion level, after an ROI is specified manually for each lesion for its recognition. DQ-ML is not a practical method, but it is included since it is similar to current clinically used software in terms of the manual labor required. DQ-AO is an automatic production-mode strategy, whereas DQ-MO is a less automated but yet practical method.

[6] The purpose of the dilation operation is to make sure that the object is fully covered by the model fuzzy mask. We perform dilation by 10 mm which roughly corresponds to AAR's object localization error.

Table 1
NEMA phantom and PET/CT image acquisition details.

Phantom weight	**10.9kg**
Isotope	^{68}Ge
Pre-injection amount of radioactivity	4.54 mCi
Activity target-to-background (T/B) ratio	4:1
Scanner	GE Discovery STE-16 PET/CT scanner
PET scan duration	5 min
PET voxel size / scene size	2.73*2.73*3.27 mm /128*128*47
CT voxel size / scene size	0.68*0.68*2.50 mm /512*512*63
CT tube voltage	120 kV
Field-of-view (PET and CT)	350 mm

4. Experiments, results, discussion

We conducted experiments on phantom data, where the true quantity of "disease" is known, as well as on patient data, where "true" disease is established by employing a clinically used commercial software system as described in Section 2.1.

4.1. Phantom data

We have utilized the publicly available National Electrical Manufacturers Association (NEMA) PET phantom data sets for studying the behavior of our proposed DQ strategies whose specifications are as follows.

The phantom data sets contain 20 scans previously acquired on a PET/CT scanner (Discovery STE, General Electric, Waukesha, WI) using a NEMA NU-2 IQ phantom (Mansoor et al., 2014; Kohlmann et al., 2015) (Data Spectrum, Durham NC). The central 5 cm diameter "lung" cylinder had been removed, the initial background activity level had been set to be equivalent to 15 mCi in a 70 kg patient, and the background activity level was approximately 9.5 mCi after 6 months (given the 271-day half-life of ^{68}Ge). Six hollow spherical inserts (with diameters of 37, 28, 22, 17, 13, and 10 mm to simulate lesions of different sizes) were used, all of which had an activity target-to-background ratio of 4:1. Additional details pertaining to the phantom and PET/CT image acquisition are summarized in Table 1.

Within the phantom, we consider each separate spherical insert (with its radioactive contents) as a "lesion", and different groups of spherical inserts (with their radioactive contents) including portions of the background as "objects" (i.e., "organs") for the purpose of testing lesion-level and object-level methods. Representative PET and CT images of the phantom are shown in different imaging planes in Fig. 4. Since automatic recognition by the AAR process is not relevant for phantoms, we tested methods DQ-MO and DQ-ML only. For DQ-MO, we specified one circular ROI to enclose all individual "lesions" (i.e., spherical inserts) going through all slices that encompassed the spherical inserts, the idea being that the ROI would encompass an "organ" (including background and spherical inserts) to emulate the process of quantifying all "lesions" within an organ collectively. For DQ-ML, a rectangular ROI was specified around each of the "lesions" (i.e., spherical inserts). The processes of specifying ROIs at the "organ" and "lesion" level for the phantom data set are illustrated in Fig. 5 for methods DQ-MO and DQ-ML. This figure also demonstrates the fuzzy disease maps obtained at the organ level and lesion level by these methods.

Some discussion is in order regarding how to establish true Q_X values for phantom data. For these data, the actual radiotracer activity is known and so also the volume of every sphere. Thus, it

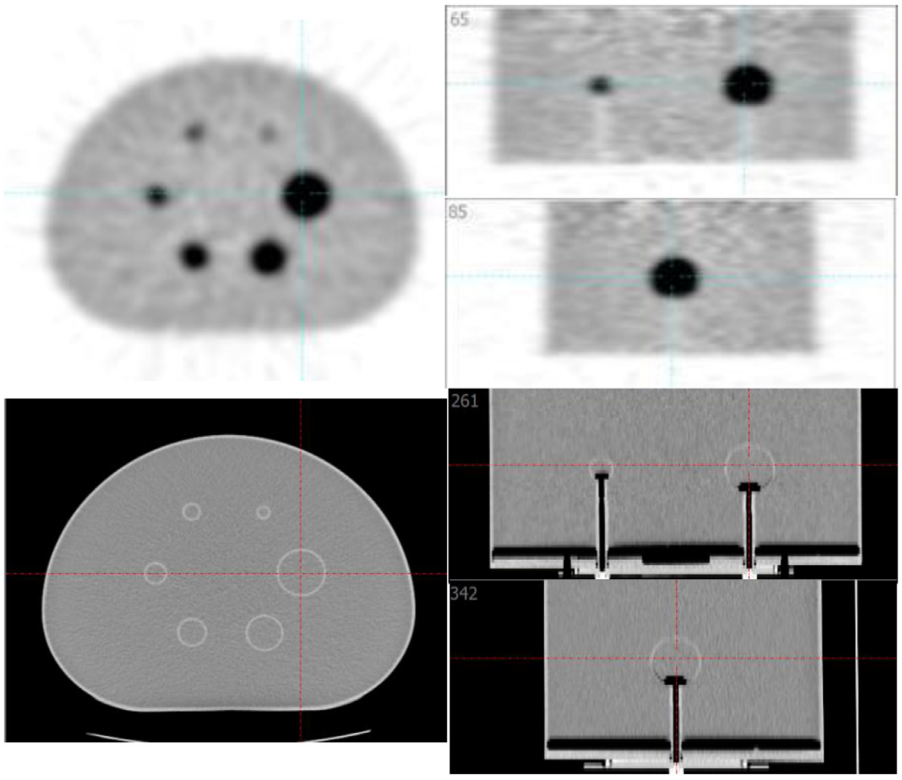

Fig. 4. PET images (top row) and CT images (bottom row) of a NEMA phantom showing 6 spherical inserts displayed in axial (left column), sagittal (right column top), and coronal (right column bottom) planes.

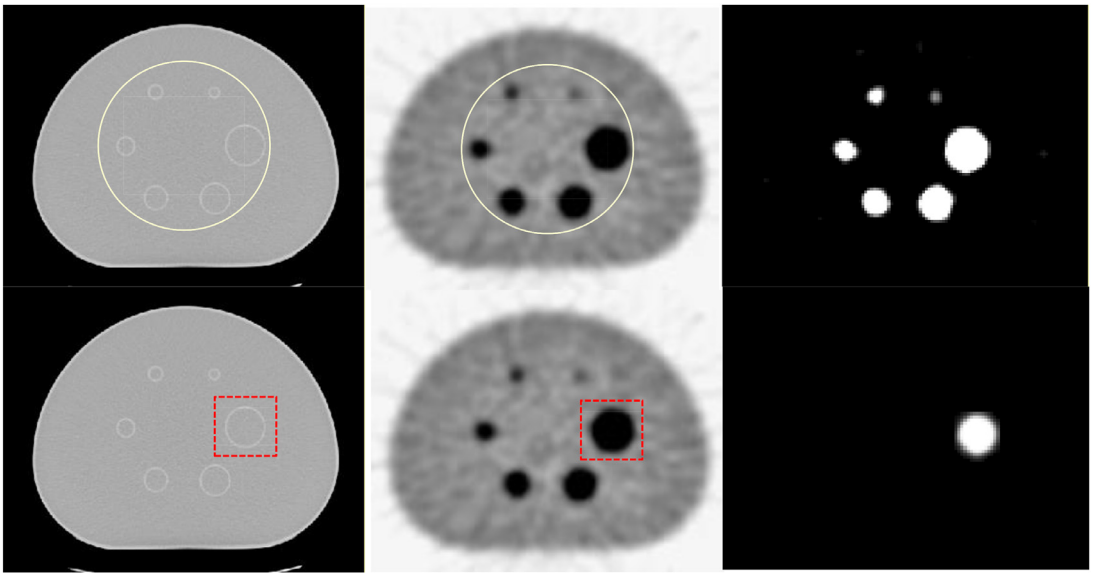

Fig. 5. Examples of the manual recognition process and the resulting fuzzy disease maps for methods DQ-MO and DQ-ML for phantom data sets. (a) Method DQ-MO (top row): A CT (I^{CT}) image (left) and a PET (I^{PET}) image (middle) with ROI placed at the "organ" level on I^{PET}, and the resulting disease map (right). (b) Method DQ-ML (bottom row) similar to (a) but with ROI placed at the "lesion" level.

is possible to calculate the theoretical absolute true TLG value for each "lesion". However, since we thought it is prudent to estimate disease maps in exactly the same way for phantoms as for patient data sets, the disease maps for phantoms were "trained" on PET images as in the case of patient studies. One consequence of training on PET images is that TLG estimated by our method will always be (much) lower than the theoretical true value since the activity level in the "lesions" will be always lower than the actual value, especially much lower for small "lesions". An alternative approach would be to "train" our method on true theoretical activity rather than on the activity observed in the PET images. This can be accomplished by making appropriate changes to Eqs. (3) and (5). We chose not to pursue this direction for two reasons. Firstly, it would be impossible to train on the theoretical true activity values in patient studies since these values are impossible to establish at lesion level and even at object level. Secondly, this would have made the actual process of DQ itself and its evaluation different for phantoms and patient cases. Therefore, we decided to establish "true"

Table 2
Mean and standard deviation of % errors in TLG estimation on phantom PET/CT scans for "lesions" and "organs" using the DQ-ML and DQ-MO methods.

Method		DQ-ML						DQ-MO
Spherical insert diameter (mm)		37	28	22	17	13	10	All spherical inserts included
TLG error	Absolute (cc) mean (SD)	4.66 (2.98)	8.70 (3.09)	7.99 (1.75)	4.53 (0.81)	1.69 (0.29)	0.45 (0.36)	21.99 (7.70)
	% mean (SD)	1.93 (1.24)	9.10 (2.99)	19.18 (3.52)	32.15 (4.68)	34.43 (6.59)	36.96 (7.38)	5.82 (1.86)
SUV_{Mean} error	Absolute mean (SD)	0.39 (0.09)	0.64 (0.08)	0.75 (0.09)	0.56 (0.08)	0.23 (0.05)	0.09 (0.09)	0.21 (0.11)
	% mean (SD)	3.86 (0.89)	7.11 (0.97)	9.37 (1.28)	8.24 (1.19)	3.89 (0.98)	2.04 (2.00)	2.22 (1.20)
SUV_{Max} error	Absolute mean (SD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	% mean (SD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Q_X values for phantom cases by using the activity values actually observed in the PET images but estimated by using the known true "masks" of the spheres in PET images in the spirit of avoiding the big hurdle of lesion segmentation.

We used a leave-one-out (LOO) strategy to evaluate the performance of our DQ approaches. In particular, there were 20 experiments in total where for each experiment, spherical inserts from 19 phantom PET/CT data sets were used to estimate the parameters of $d_O(x)$ and the remaining 1 data set was used to test the accuracy of quantitative parameters Q_X by the proposed methods. Table 2 lists the mean and standard deviation of errors in estimating TLG, SUV_{Mean}, and SUV_{Max} for "lesions" and "organs" within phantom data sets using DQ-ML and DQ-MO methods, where average error % is calculated as [(estimated value − ground truth value) / (ground truth value)] × 100 over all tested samples for each method. For the DQ-MO method, where all "lesions" together are considered within an "organ", the average TLG error was ∼6%. For the DQ-ML method, we observed that the estimates of TLG for small "lesions" had larger errors than those of larger "lesions", with average error ranging from 2% to 37%. At the lesion level, SUV_{Mean} achieves error within 2%–9% and SUV_{Max} achieves error close to zero for all lesions.

Lesions with small diameter have strong partial volume effect (compared to their size) and TLG error increases with decreasing sphere size. This has been previously observed by other methods tested on these phantoms (Kinahan and Fletcher, 2010; Doot et al., 2010), where the ratio defined by measured value/true value was found to be ∼0.3 for small lesions. Expressed as *%TLG* error as we defined above, this error is ∼70%. %TLG error is ∼50% for lesions with diameter 13 mm, ∼40% and 25% for larger lesions with diameter 17 mm and 22 mm, and ∼3% for lesions with the maximum diameter 37 mm (Kinahan and Fletcher, 2010; Doot et al., 2010). Thus, although our method's performance is much poorer for small lesions than for large lesions, it still outperforms current approaches, especially on small lesions. We note that there is an issue in properly expressing error in measuring small lesions. Even small absolute errors (in cc) become exaggerated when expressed as a percent of the total lesion activity when lesions are small. A somewhat similar phenomenon occurs also for other metrics. This is the reason that we also listed absolute errors as well for all three measures in Table 2.

4.2. Patient data

Body-wide FDG-PET/CT scans from human subjects were utilized in this study. Details about data sets $\mathcal{I}^m$, $\mathcal{C}^m$, and $\mathcal{D}^m$ are summarized in Table 3. We used 16 normal data sets in $\mathcal{I}^m$ for building the AAR model. For estimating the parameters of the disease map, we again employed the LOO strategy – of the 31 patient data sets, 30 were used for training, one was set aside for testing, and the process was repeated 31 times.

All scans had previously been acquired on PET/CT scanners with time-of-flight capabilities (Gemini TF, Philips Medical Systems, Bothell, WA). 3D PET data had been acquired either from the skull vertex to the toes or from the skull base to the proximal thighs ∼60 min after intravenous administration of ∼555 MBq of FDG for ∼3 min per bed position. Image reconstruction had been performed at 4 mm nominal slice thickness in the axial plane using a list-mode maximum-likelihood expectation-maximization (ML-EM) algorithm with 33 ordered subsets and 3 iterations, and the system model included time-of-flight as well as normalization, attenuation, randoms, and scatter corrections, where rescaled low-dose CT images were utilized for attenuation correction of PET images. All patient PET/CT images were selected from our health system patient image database by a board-certified radiologist (DAT) following approval for this study from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act waiver. 20 conspicuous lung lesions and 20 conspicuous liver lesions were assessed to illustrate the disease quantification approach. We also assessed 10 non-conspicuous lesions (3 lung lesions and 7 liver lesions) for comparison of the disease quantification approach performance with that for conspicuous lesions. An LOO strategy was used for evaluation.

The objects (i.e., organs) considered were PS (LPS+RPS) (i.e., the lungs) and Lvr (i.e., the liver) in this initial study. Other objects shown in the hierarchy of Fig. 2 are needed for accurate recognition of the objects for which DQ is performed. Organ-level reference standard measurements were obtained by aggregating lesion-level reference standard measurements from all lesions within the organ of interest, with SUV_{Mean} calculated as the mean of all lesion SUV_{Mean} values, SUV_{Max} as the maximum of all lesion SUV_{Max} values, and TLG_O as the sum of TLG values of all lesions within the organ.

Representative estimated disease maps of metastatic cancer lesions for liver and right lung are displayed in Fig. 6 for methods DQ-AO and DQ-MO. The recognized object fuzzy mask for DQ-AO and the specified binary object mask for DQ-MO are shown overlaid on the underlying CT and PET images in the top two rows. Similarly, Fig. 7 shows representative estimated fuzzy disease maps for method DQ-ML. As one can see from these examples, given appropriate ROI placement by AAR or by the manual approach for objects or lesions, the disease map within the ROIs can correctly capture conspicuous lesions within the ROIs.

Table 4 lists the mean and standard deviation of errors (absolute and %) in estimated disease quantities Q_X (TLG, SUV_{Mean}, and SUV_{Max}) for patient lesions (including conspicuous and non- conspicuous types) within the liver and lungs using the DQ-AO and

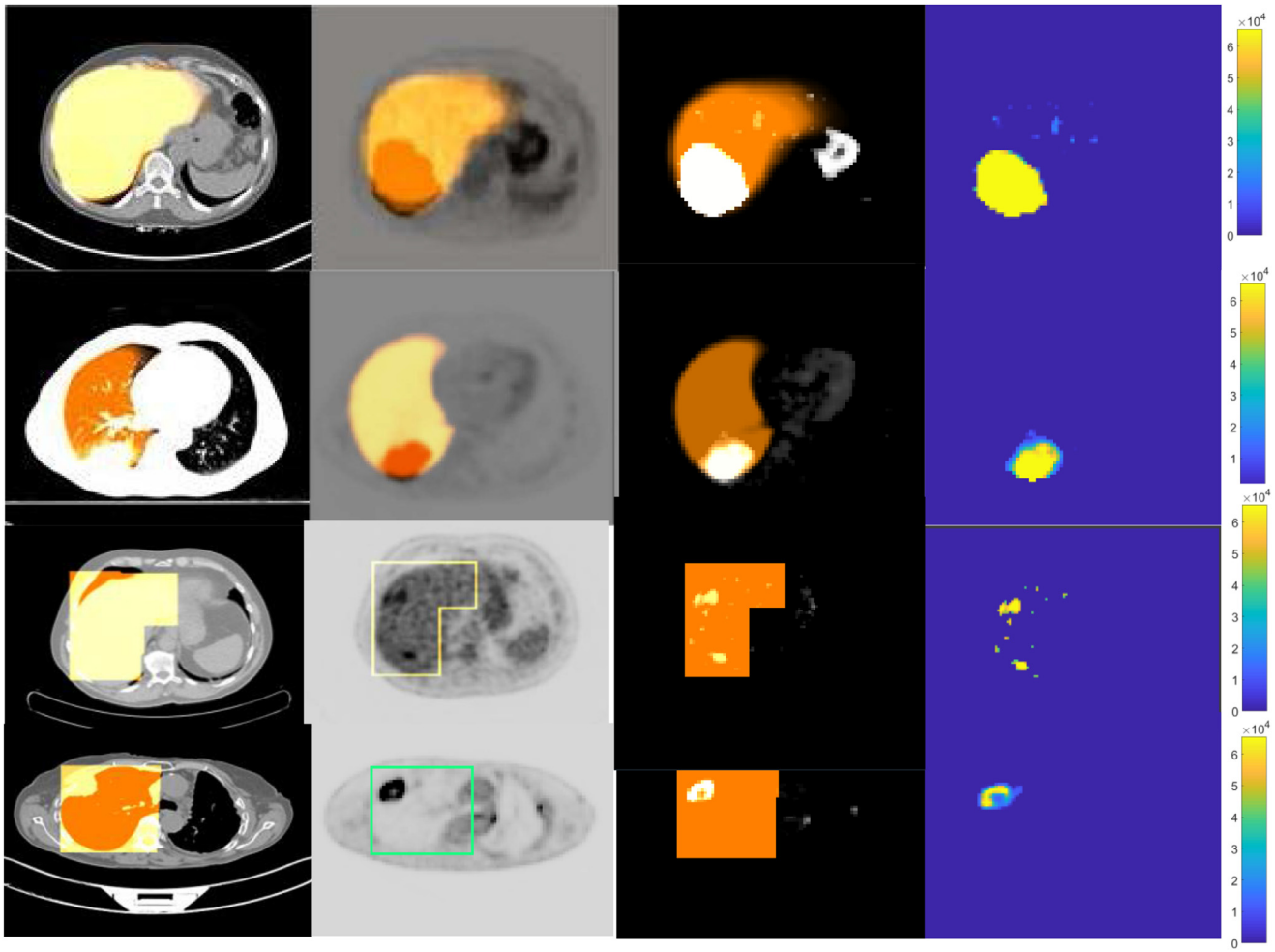

Fig. 6. Examples of disease maps of metastatic cancer lesions estimated at the organ level by methods DQ-AO (top two rows) and DQ-MO (bottom two rows) for Lvr (liver) and RPS (right pleural space including right lung). Only single representative axial slices of the CT, PET, and associated disease map images (left to right) are shown from these two different patient data sets in different organs.

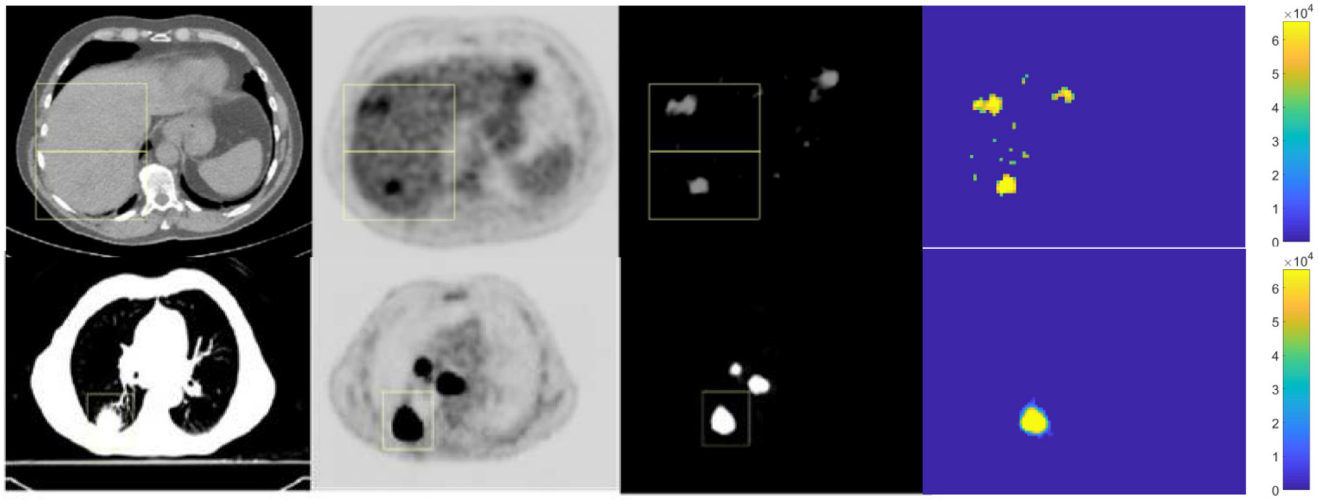

Fig. 7. Examples of disease maps of metastatic cancer lesions estimated at the lesion level by method DQ-ML for liver lesions (top row) and lung lesions (bottom row). In the top row, the other prominent hot spot is not a liver lesion but comes from the uptake in heart. Only single representative axial slices of the CT, PET, and associated disease map images (left to right) are shown.

Table 3
Summary of normal and patient PET/CT data sets used in this study. Conspicuous lesions (40 in total) and non-conspicuous lesions (10 in total) are both included.

Data set	Number of subjects, number & type of lesions	Scan details	Image size/ voxel size
$\mathcal{I}^m$ Normal	16 (no lesions)	Whole-body, unenhanced, axial	PET: 144 × 144 × 338–443, 4 × 4 × 4 mm^3 CT: 512 × 512 × 338–443, 1.2 × 1.2 × 4 mm^3
	7 (10 non-conspicuous lesions: 7 liver + 3 lung)	Whole-body or from skull base to proximal thighs, unenhanced, axial	PET: 200 × 200 × 326–440, 4.07 × 4.07 × 3.00 mm^3 CT: 512 × 512 × 326–440, 0.98 × 0.98 × 3.00 mm^3
$\mathcal{D}^m$, $\mathcal{C}^m$ Pathological	14 (20 conspicuous lung lesions) 10 (20 conspicuous liver lesions)	Whole-body or from skull base to proximal thighs, unenhanced, axial	PET: 200 × 200 × 326–440, 4.07 × 4.07 × 3.00 mm^3 CT: 512 × 512 × 326–440, 0.98 × 0.98 × 3.00 mm^3

Table 4
Mean and standard deviation of % errors and absolute errors in estimating TLG, SUV_{Mean}, and SUV_{Max} on patient PET/CT scans for all lesions within the liver and lungs via DQ-MO and DQ-AO.

Method		Liver lesions		Lung lesions		All lesions	
		DQ-MO	DQ-AO	DQ-MO	DQ-AO	DQ-MO	DQ-AO
TLG error	Absolute (cc) mean (SD)	78.74 (93.92)	105.70 (127.74)	12.23 (22.60)	21.65 (41.6)	46.06 (66.67)	67.49 (94.55)
	% mean (SD)	11.30 (7.22)	12.27 (8.85)	9.55 (4.05)	14.67 (9.24)	9.81 (9.77)	13.65 (11.75)
SUV_{Mean} error	Absolute mean (SD)	0.52 (0.41)	0.42 (0.26)	0.83 (1.01)	0.69 (0.97)	0.66 (0.57)	0.62 (0.65)
	% mean (SD)	12.62 (6.60)	11.18 (6.59)	18.59 (14.19)	17.48 (9.57)	16.23 (13.47)	14.97 (13.79)
SUV_{Max} error	Absolute mean (SD)	0.50 (1.47)	0.46 (1.37)	0.56 (1.00)	0.56 (1.00)	0.47 (1.28)	0.45 (1.22)
	% mean (SD)	0.39 (0.49)	0.39 (0.49)	0.07 (0.14)	0.07 (0.14)	0.26 (0.81)	0.27 (0.80)

Table 5
Mean and standard deviation of % errors and absolute errors in TLG, SUV_{Mean}, and SUV_{Max} estimation on patient PET/CT scans for individual conspicuous lesions within liver and lung via DQ-ML.

		Liver lesions	Lung lesions	All lesions
TLG error	Absolute (cc) mean (SD)	5.06 (6.01)	11.26 (23.07)	8.23 (17.14)
	% mean (SD)	6.63 (3.78)	11.83 (3.92)	9.07 (3.86)
SUV_{Mean} error	Absolute mean (SD)	0.14 (0.11)	0.58 (0.33)	0.34 (0.33)
	% mean (SD)	3.80 (3.21)	9.68 (7.13)	6.71 (5.17)
SUV_{Max} error	Absolute mean (SD)	0.001 (0.003)	0.00 (0.00)	0.0001 (0.0002)
	% mean (SD)	0.04 (0.02)	0.00 (0.00)	0.02 (0.01)

DQ-MO methods with respect to the quantities from the reference method over test data sets $\mathcal{C}^m$.[7] Table 5 lists comparable results using the DQ-ML method on conspicuous lesions, and Table 6 shows results for the DQ-ML method separately for conspicuous and non-conspicuous lesions. As in Table 2, we list the absolute errors and % errors in Tables 4–6.

From Tables 4–6, we observe the following. Disease quantification at the lesion level via the DQ-ML method generally had lower %errors than at the organ level via the DQ-MO or DQ-AO methods, and the errors were generally lower for liver lesions compared to lung lesions at the lesion level via the DQ-ML method. TLG estimation from the proposed approach achieved lower %error than for SUV_{Mean} estimation. For all methods, the estimation of SUV_{Max} was most accurate with less than 0.5% error.

For disease quantification at the organ level, TLG estimation via the DQ-MO method had lower %error than that via the DQ-AO method for liver lesions and higher error for lung lesions. SUV_{Mean} estimation via the DQ-MO method had lower %error than via the DQ-AO method for both liver and lung lesions, whereas SUV_{Max} estimation via the DQ-MO and DQ-AO methods had similar levels of error. However, none of the differences between DQ-AO and DQ-MO methods is statistically significant.

Table 6 shows that the proposed DQ-ML method has higher accuracy for conspicuous lesions than for non-conspicuous lesions, which is not surprising given the challenges in establishing ground

Table 6
Mean and standard deviation of % errors and absolute errors in TLG, SUV_{Mean}, and SUV_{Max} estimation for patient conspicuous vs. non-conspicuous (liver and lung) lesions using DQ-ML method.

Lesion type No. of lesions		Conspicuous	Non-conspicuous
		40	10
TLG error	Absolute (cc) mean (SD)	8.23 (17.14)	15.22 (15.92)
	% mean (SD)	9.07 (3.86)	18.60 (14.28)
SUV_{Mean} error	Absolute mean (SD)	0.34 (0.33)	1.71 (0.65)
	% mean (SD)	6.71 (5.17)	13.84 (10.34)
SUV_{Max} error	Absolute mean (SD)	0.0001 (0.0002)	0.001 (0.001)
	% mean (SD)	0.02 (0.01)	0.04 (0.01)

[7] The absolute errors for *TLG* may appear large in Table 4, but note that true *TLG* itself is very large at the organ level (in the 1000 s) because of contribution from all lesions in the object region, and more importantly, due to multiplication of volume by SUV.

Table 7
Comparison of the AAR-DQ approach with recent literature for disease burden estimation on PET/CT images considering TLG error, SUV_{Mean} and SUV_{Max}.

	Methods	Number of Training, testing samples, cross validation	Phantom and number	Error in disease burden estimation
Med. Phys., (Ford et al., 2006)	Threshold based segmentation for tumors on PET	–	NEMA, 20 ^{18}F	SUV_{Max} errors within 0.0%-60% for spheres from the largest to the smallest, no TLG error reported
Semin CT MR, (Kinahan, and Fletcher, 2010); Doot et al. Med. Phys. (Doot et al., 2010)	Manually draw circles with diameter 10 mm at sphere centers for all spheres	–	NEMA, 20, ^{18}F/Ge^{68}	SUV_{Max}, SUV_{Mean} errors within 0.0%-60% for spheres from largest diameter (37 mm) to the smallest (10 mm)
EJNMMI Phys, (Ziegler et al., 2015)	Manually draw circles (only for 4 spheres) around lesions on PET	–	NEMA, 20 ^{18}F	SUV_{Mean} error within 19.8%–63.2%, no TLG error reported
CMIG, (Taghanaki et al., 2018)	Multiple random forest machine learning approach to predict disease burden	55 patients, each patient with one (conspicuous) tumor, LOO cross validation	NEMA, 20 ^{18}F	For phantom, 13.03% for TLG error (no SD for TLG reported); *activity* (SUV_{Mean}) error 5.7 ± 5.25%, no SUV_{Max} error reported For patient, 13.83% ± 21.47% for SUV_{Mean} error and 12.17% ± 5.34% for TLG error
Proposed approach	Disease-map-based approach for object/lesion level DQ	40 conspicuous lesions, 10 non-conspicuous lesions, LOO cross validation	NEMA, 20 Ge^{68}	For phantom, at object level DQ, TLG error 5.82% ± 1.86% with SUV_{Mean} error < 3%, SUV_{Max} error close to zero; For patient, lesion level DQ, TLG error 9.1% ± 3.9% with SUV_{Mean} error < 7%, SUV_{Max} error < 1%

truth measurements for non-conspicuous lesions. Yet, even for non-conspicuous lesions, the proposed approach for TLG, SUV_{Mean}, and SUV_{Max} estimation was accurate in approximately 81%, 86%, and 100% of lesions in terms of %error.

To our knowledge, disease quantification at the object (organ) level without explicit object delineation on PET/CT images has not been previously reported in the literature, whereas almost all literature exists regarding PET/CT-based disease quantification at the lesion level. In our study, object level TLG estimation had a %error of 5.82% ± 1.86% via the DQ-MO method on PET/CT phantom data sets, and 9.81% ± 9.77% and 13.65% ± 11.75% via the DQ-MO and DQ-AO methods, respectively, on PET/CT patient data sets. At the lesion level, TLG estimation had an error ranging from 2 to 37% via the DQ-ML method on PET/CT phantom data sets depending on the "lesion" size, and an error of 9.07% ± 3.86% on PET/CT patient data sets. These lesion level results are comparable to those from a most recent study by Taghanaki et al. on PET/CT phantom data sets which utilized a multiple layer random forest tree method based on features extracted from 3D patches (very similar to deep learning approaches Taghanaki et al., 2018). This method reported an average relative absolute error of 12.17% ± 5.34% for lesion level TLG estimation, a relative absolute error of 13.83% ± 21.47% for estimating SUV_{Mean} of lesions smaller than 2 mL in volume for patient data, as well as an error of 5.70 ± 5.25% for estimating SUV_{Mean} on NEMA phantom data. From our approach, SUV_{Mean} error (2.22% ± 1.20%) and SUV_{Max} error (close to zero) on the same phantom data are better than results in Kinahan and Fletcher (2010), Doot et al. (2010), and Taghanaki et al. (2018). More general comparison between our approach and related research in the literature is summarized in Table 7.

We must note that a quantitative understanding/grading of the reported methods is almost impossible since the data sets used, acquisition protocols and resolutions, considered objects, training and test data set subdivisions, and cross validation strategies are all different in these methods. Table 7 summarizes recent literature (Kinahan and Fletcher, 2010; Doot et al., 2010; Taghanaki et al., 2018; Ford et al., 2006; Ziegler et al., 2015) dealing with disease burden estimation on PET/CT. Some methods have been tested on the NEMA phantoms. Methods based on deep learning techniques generally require a large number of training data sets.

Among all methods listed in Table 7, the one in Taghanaki et al. (2018) comes close in spirit to our approach. There are several advantages of our approach over that in Taghanaki et al. (2018). (i) We demonstrate the generality of AAR-DQ at object level and lesion level, with both manual and automatic recognition steps, while the referred work seems to operate only at lesion level. It is not clear in Taghanaki et al. (2018) how the lesion-level ROIs are generated with the only information provided stating that ROIs were drawn around lesions by an expert on PET. We must also note that AAR-DQ is set up in a general manner so it can make use of existing object models and anatomy models constructed for other applications involving object segmentation, etc. The only additional component required, namely disease map, is easily encoded into the anatomy model. (ii) AAR-DQ requires much smaller number of training samples than Taghanaki et al. (2018). (iii) The performance of our approach was comparable to Taghanaki et al. (2018) on patient (lesion-level) data sets but better than that reported in (Taghanaki et al., 2018) on phantom data sets. (iv) Our disease map estimation (training) step takes ∼2 min on a desktop computer with 4 Intel i7-core CPUs, 64GRAM, and under Ubuntu 16.04 OS. Computational timing is not reported in (Taghanaki et al., 2018) and we suspect that its training step is lot more time-consuming.

5. Concluding remarks

In this initial study, we present a new methodology called AAR-DQ for disease quantification on PET/CT images, keeping in mind the primary tasks for which diagnostic imaging is employed in the clinical management of cancer patients. AAR-DQ extended the previously-developed AAR technology (Udupa et al., 2014) to disease quantification by stopping at the object recognition step and performing disease quantification directly from object location information, permitting the methodology to skip the rather challenging and somewhat ill-defined step of explicit object delineation. Our long-term goal is to adapt AAR-DQ for body-wide automatic disease quantification on PET/CT images building on the generality of AAR in body-wide object recognition/delineation. AAR-DQ adopts a fuzzy strategy throughout – for object modeling, object recognition, disease mapping, and disease quantification. This al-

lows handling of disease gradation without the need to perform explicit partial volume correction and committed binary classification. Three methods were presented for disease quantification, two at the object (e.g., organ) level – DQ-AO and DQ-MO, and one at the lesion level – DQ-ML. Commercially available clinical software was used to generate reference disease measurements at the lesion level with the assistance of a manually guided method of quantification as needed. By using these reference quantities, evaluations were carried out using both NEMA phantoms and clinical FDG-PET/CT image data sets in patients with metastatic cancer.

A challenge that we encountered in the development of AAR-DQ is the establishment of a reliable method of deriving reference disease quantities at the lesion, organ, and even body region levels. This is a serious hurdle to automatic disease quantification. As mentioned above, in case of ill-defined, small, or multifocal lesions, commercially available software often fails to properly delineate the lesions. Although the use of phantoms offers the possibility for ground truth assessments, they do not tender the same challenges as those encountered in real patient PET/CT images. Another obstacle is the lack of normative PET data sets. Whatever is the approach for quantification, knowledge of normative distribution of SUV in organs becomes crucial to establish disease quantity to express deviation from normality.

One shortcoming of this investigation is the small number of data sets used for evaluation. However, as notable from Table 7, our study is not an outlier in this regard. One of the barriers we experienced in employing larger data sets was the effort needed in establishing ground truth disease quantity irrespective of conspicuous or non-conspicuous lesions. AAR-DQ has no inherent limitations in being applied to other organs throughout the body on a substantially larger number of data sets. We are in the process of gathering more normative data sets and generating ground truth disease quantities for demonstrating body-wide application of AAR-DQ on a much larger cohort.

Another potential limitation of AAR-DQ ensues from possible inaccuracies in object recognition. When pathology is extensive, the AAR recognition algorithm may position the model in such a manner that the disease quantity $Q_X(O)$ may show substantial errors. At the lesion level, this may lead to an increase in false positives and false negatives. We are actively pursuing extensions of AAR recognition algorithms for handling such situations. Note that in DQ-MO with organ-level manual recognition, the accuracy of recognition, even in the case of extensive pathology, is not an issue. Even confounding organs such as the kidneys which exhibit high radiotracer content due to FDG excretion can be handled via the specification of an ROI to exclude such objects. We can have a composite ROI consisting of additive and subtractive ROIs, the main goal being accurate and efficient interactive means of offering recognition help. Once we accurately localize objects, the disease quantification procedure can be smoothly followed.

In this paper, we did not explore fully automatic lesion level DQ where recognition not just at organ level but also at lesion level is automatic (we may refer to such a strategy by DQ-AL). Once an organ O is recognized in I^{CT} and the disease map $d_O(x)$ is computed, we have a fuzzy membership image given by $d_O(I_S(v))$ corresponding to SUV image I_S (cf Eq. (4)). The fuzzy connectedness machinery (Udupa et al., 1996) can then be applied to this disease membership image to label all fuzzy components in the image automatically by using the homogeneity of disease membership as the affinity function. Subsequently, the disease quantity Q_X can be computed for each separate fuzzy component. The background non-lesion region will also be labeled as a separate fuzzy component by this approach. We will investigate such a DQ-AL approach in the future for outputting $Q_X(O)$ at the lesion level automatically. We will also consider more sophisticated functional forms for $d_O(x)$ which may allow for more accurate disease mapping. Since radiotracers other than FDG are also being actively studied as molecular probes for imaging of disease conditions, generalization of AAR-DQ to PET images acquired following administration of non-FDG radiotracers is also a future goal of ours.

Conflict of interest

There is no any conflict of interest or industry support of the project.

Main contributions

1. A general single framework for disease quantification via PET/CT images that is independent of objects and body regions and that can operate seamlessly at the body region level, organ level, and lesion level.
2. Disease quantification performed without explicit delineation of body regions, organs, or lesions following immediately after organs are roughly localized (recognized) via the previously reported AAR approach.
3. Disease quantification on PET images without explicitly accounting for or correcting for partial volume effects and disease heterogeneity but by using fuzzy principles for object models, object localization, and disease mapping that considers SUV distributions within both normal organs and lesions.
4. A comprehensive evaluation based on phantom as well as patient data sets and analysis of results at object and lesion level and on conspicuous and more challenging non-conspicuous lesions.

Acknowledgments

The authors acknowledge philanthropic support from James and Frances Maguire, Margarita Louis-Dreyfus, and Sharon Berman for the Lymphoma Program at the Abramson Cancer Center of the University of Pennsylvania.

References

Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. Neuroimage 45, 333–341.

Bagci, U., Udupa, J.K., Mendhiratta, N., Foster, B., Xu, Z., Yao, J., Chen, X., Mollura, D.J., 2013. Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. Med. Image Anal. 17, 929–945.

Cui, H., Wang, X., Lin, W., Zhou, J., Eberl, S., Feng, D., Fulham, M., 2016. Primary lung tumor segmentation from PET-CT volumes with spatial-topological constraint. Int. J. Comput. Assist. Radiol. Surg. 11 (1), 19–29 PubMed PMID: 26133651. doi:10.1007/s11548-015-1231-0.

Doot, R.K., Scheuermann, J.S., Christian, P.E., Karp, J.S., Kinahan, P.E., 2010. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. Med. Phys. 37 (11), 6035–6046 PubMed PMID: 21158315; PMCID: PMC2988835. doi:10.1118/1.3499298.

Doyle, W., 1962. Operations useful for similarity-invariant pattern recognition. J. ACM 9, 259–267.

Fagundes, T.C., Mafra, A., Silva, R.G., Castro, A.C.G., Silva, L.C., Aguiar, P.T., Silva, J.A., E, P.J., Machado, A.M., Mamede, M., 2018. Individualized threshold for tumor segmentation in 18F-FDG PET/CT imaging: the key for response evaluation of neoadjuvant chemoradiation therapy in patients with rectal cancer? Rev. Assoc. Med. Bras. 64 (2), 119–129.

Falcao, A., Udupa, J.K., Samarasekera, S., Sharma, S., Hirsch, B.E., Lotufo, R., 1998. User-steered image segmentation paradigms: live wire and live lane. Graph. Models Image Process. 60 (4), 233–260.

Ford, E.C., Kinahan, P.E., Hanlon, L., Alessio, A., Rajendran, J., Schwartz, D.L., Phillips, M., 2006. Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. Med. Phys. 33 (11), 4280–4288 PubMed PMID: 17153406. doi:10.1118/1.2361076.

Geworski, L., Karwarth, C., Fitz, E., Plotkin, M., Knoop, B., 2010. [Quality control in PET/CT systems: experiences and requirements]. Z. Med. Phys. 20 (1), 46–50.

Goceri, E., Gürcan, M., Dicle, O., 2014. Fully automated liver segmentation from spir image series. Int. J. Comput. Biol. Med. 53 (1), 265–278. doi:10.1016/j.compbiomed.2014.08.009.

Grevera, G., Udupa, J., Odhner, D., Zhuge, Y., Souza, A., Iwanaga, T., Mishra, S., 2007. CAVASS: a computer-assisted visualization and analysis software system. J. Digit Imaging 20 (Suppl 1), 101–118 PubMed PMID: 17786517; PMCID: PMC2039829. doi:10.1007/s10278-007-9060-5.

Hillner, B.E., Siegel, B.A., Liu, D., et al., 2008. Impact of positron emission tomography/computed tomography and positron emission tomography (PET) alone on expected management of patients with cancer: initial results from the National Oncologic PET Registry. J. Clin. Oncol. 26 (13), 2155–2161.
Hofheinz, F., Poetzsch, C., van den Hoff, J., 2007. Quantitative 3D ROI delineation in PET: algorithm and validation. J Nucl Med 48 (Supplement 2), 407 P.
Hofheinz, F., Dittrich, S., Potzsch, C., Hoff, J., 2010. Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. Phys. Med. Biol. 55 (4), 1099–1113.
Hu, P., Wu, F., Peng, J., Bao, Y., Chen, F., Kong, D., 2017. Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. Int. J. Comput. Assist. Radiol. Surg. 12 (3), 399–411 PubMed PMID: 27885540. doi:10.1007/s11548-016-1501-5.
Ju, W., Xiang, D., Zhang, B., Wang, L., Kopriva, I., Chen, X., 2015. Random walk and graph cut for co- segmentation of lung tumor on PET-CT images. IEEE Trans. Image Process. 24 (12), 5854–5867 PubMed PMID: 26462198. doi:10.1109/TIP.2015.2488902.
Kashyap, S., Zhang, H., Rao, K., Sonka, M., 2018. Learning-based cost functions for 3-D and 4-D multi-surface multi-object segmentation of knee MRI: data from the osteoarthritis initiative. IEEE Trans. Med. Imaging 37 (5), 1103–1113 PubMed PMID: 29727274; PMCID: PMC5995124. doi:10.1109/TMI.2017.2781541.
Kinahan, P.E., Fletcher, J.W., 2010. PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy. Semin. Ultrasound CT MR 31 (6), 496–505.
Kohlmann, P., Strehlow, J., Jobst, B., Krass, S., Kuhnigk, J.M., Anjorin, A., Sedlaczek, O., Ley, S., Kauczor, H.U., Wielputz, M.O., 2015. Automatic lung segmentation method for MRI-based lung perfusion studies of patients with chronic obstructive pulmonary disease. Int. J. Comput. Assist. Radiol. Surg. 10, 403–417.
Kwee, T.C., Torigian, D.A., Alavi, A., 2013a. Oncological applications of positron emission tomography for evaluation of the thorax. J. Thorac. Imaging 28 (1), 11–24.
Kwee, T.C., Torigian, D.A., Alavi, A., 2013b. Overview of positron emission tomography, hybrid positron emission tomography instrumentation, and positron emission tomography quantification. J. Thorac. Imaging 28 (1), 4–10.
Lee, J., Kim, S.H., Styner, M., 2016. Multi-object model-based multi-atlas segmentation for rodent brains using dense discrete correspondences. Proc. SPIE Int. Soc. Opt. Eng. 9784 PubMed PMID: 27065200; PMCID: PMC4825178. doi:10.1117/12.2217709.
Mansoor, A., Bagci, U., Xu, Z., Foster, B., Olivier, K.N., Elinoff, J.M., Suffredini, A.F., Udupa, J.K., Mollura, D.J., 2014. A generic approach to pathological lung segmentation. IEEE Trans. Med. Imaging 33, 2293–2310.
Mena, E., Sheikhbahaei, S., Taghipour, M., Jha, A.K., Vicente, E., Xiao, J., Subramaniam, R.M., 2017. 18F-FDG PET/CT metabolic tumor volume and intratumoral heterogeneity in pancreatic adenocarcinomas: impact of dual-time point and segmentation methods. Clin. Nucl. Med. 42 (1), e16–e21 PubMed PMID: 27819858; PMCID: PMC5360463. doi:10.1097/RLU.0000000000001446.
Namias, R., D'Amato, J.P., Del Fresno, M., Venere, M., Pirro, N., Bellemare, M.E., 2016. Multi-object segmentation framework using deformable models for medical imaging analysis. Med. Biol. Eng. Comput. 54 (8), 1181–1192 PubMed PMID: 26392182. doi:10.1007/s11517-015-1387-3.
Narasimhan, R., Fornango, J.P., 1963. Some further experiments in the parallel processing of pictures. IEEE Trans. Electron. Comput. EC -13, 748–750.
Okada, T., Linguraru, M.G., Hori, M., Summers, R.M., Tomiyama, N., Sato, Y., 2015. Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors. Med. Image Anal. 26 (1), 1–18.
Oliveira, B., Queiros, S., Morais, P., Torres, H.R., Gomes-Fonseca, J., Fonseca, J.C., Vilaca, J.L., 2018. A novel multi-atlas strategy with dense deformation field reconstruction for abdominal and thoracic multi-organ segmentation from computed tomography. Med. Image Anal. 45, 108–120 PubMed PMID: 29432979. doi:10.1016/j.media.2018.02.001.
Piert, M., Shankar, P.R., Montgomery, J., Kunju, L.P., Rogers, V., Siddiqui, J., Rajendiran, T., Hearn, J., George, A., Shao, X., Davenport, M.S., 2018. Accuracy of tumor segmentation from multi-parametric prostate MRI and (18)F-choline PET/CT for focal prostate cancer therapy applications. EJNMMI Res. 8 (1), 23 PubMed PMID: 29589155; PMCID: PMC5869349. doi:10.1186/s13550-018-0377-5.
Powell, M.J.D., 2006. The NEWUOA software for unconstrained minimization without derivatives. In: Di Pillo, G., Roma, M. (Eds.), Large-Scale Nonlinear Optimization. Springer, pp. 255–297.
Priyadarsini, S., Selvathi, D., 2012. Survey on segmentation of liver from epaCT images,. IEEE Int. Conf. Adv. Commun. Control Comput. Technol. 1 (1), 234–238. doi:10.1109/ICACCCT.2012.6320777.
Taghanaki, S.A., Duggan, N., Ma, H., Hou, X., Celler, A., Benard, F., Hamarneh, G., 2018. Segmentation-free direct tumor volume and metabolic activity estimation from PET scans. Comput. Med. Imaging Graph. 63, 52–66.
Tan, S., Li, L., Choi, W., Kang, M.K., D'Souza, W.D., Lu, W., 2017. Adaptive region-growing with maximum curvature strategy for tumor segmentation in (18)F-FDG PET. Phys. Med. Biol. 62 (13), 5383–5402 PubMed PMID: 28604372; PMCID: PMC5497763. doi:10.1088/1361-6560/aa6e20.
Tong, T., Wolz, R., Wang, Z., Gao, Q., Misawa, K., Fujiwara, M., Mori, K., Hajnal, J.V., Rueckert, D., 2015. Discriminative dictionary learning for abdominal multi-organ segmentation. Med. Image Anal. 23 (1), 92–104 PubMed PMID: 25988490. doi:10.1016/j.media.2015.04.015.
Tong, Y.B., Udupa, J.K., Odhner, D., Schuster, S.J., Torigian, D.A., 2017. Disease quantification on PET/CT images without object delineation. Proc. SPIE Int. Soc. Opt. Eng. 10137 101370V1-6.
Torigian, D.A., Lopez, R.F., Alapati, S., Bodapati, G., Hofheinz, F., van den Hoff, J., Saboury, B., Alavi, A., 2011. 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), 8–14.
Udupa, J.K., Odhner, D., Zao, L., Tong, Y., Matsumoto, M.M.S., Ciesielski, K.C., Falcao, A.X., Vaideeswaran, P., Ciesielski, V., Saboury, B., Mohammadianrasanani, S., Sin, S., Arens, R., Torigian, D.A., 2014. Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images. Med. Image Anal. 18, 752–771.
Udupa, J.K., Samarasekera, S., 1996. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation, CVGIP. Graph. Model. Image Process. 58 (3), 246–261.
Wang, H., Udupa, J.K., Odhner, D., Tong, Y., Zhao, L., Torigian, D.A., 2016. Automatic anatomy recognition in whole-body PET/CT images. Med. Phys. 43, 613–629. doi:10.1118/1.4939127.
Withofs, N., Bernard, C., Van der Rest, C., Martinive, P., Hatt, M., Jodogne, S., Visvikis, D., Lee, J.A., Coucke, P.A., Hustinx, R., 2014. FDG PET/CT for rectal carcinoma radiotherapy treatment planning: comparison of functional volume delineation algorithms and clinical challenges. J. Appl. Clin. Med. Phys. 15 (5), 4696 PubMed PMID: 25207560; PMCID: PMC5711099. doi:10.1120/jacmp.v15i5.4696.
Xu, W., Yu, S., Ma, Y., Liu, C., Xin, J., 2017. Effect of different segmentation algorithms on metabolic tumor volume measured on 18F-FDG PET/CT of cervical primary squamous cell carcinoma. Nucl. Med. Commun. 38 (3), 259–265 PubMed PMID: 28118260; PMCID: PMC5318156. doi:10.1097/MNM.0000000000000641.
Zhou, J., Rajapakse, J.C., 2005. Segmentation of subcortical brain structures using fuzzy templates. NeuroImage 28, 915–924.
Zhuang, X., Bai, W., Song, J., Zhan, S., Qian, X., Shi, W., Lian, Y., Rueckert, D., 2015. Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection. Med. Phys. 42, 3822–3833.
Zhuang, X., Shen, J., 2016. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. 31, 77–87.
Ziegler, S., Jakoby, B.W., Braun, H., Paulus, D.H., Quick, H.H., 2015. NEMA image quality phantom measurements and attenuation correction in integrated PET/MR hybrid imaging. EJNMMI Phys. 2 (1), 18 PubMed PMID: 26501819; PMCID: PMC4542864. doi:10.1186/s40658-015-0122-3.