Lesion Detection and Characterization with Context Driven Approximation in Thoracic FDG PET-CT Images of NSCLC Studies

Yang Song, *Student Member, IEEE,* Weidong Cai, *Member, IEEE,* Heng Huang, Xiaogang Wang, *Member, IEEE,* Yun Zhou, Michael J Fulham, and David Dagan Feng, *Fellow, IEEE*

***Abstract*—We present a lesion detection and characterization method for ^{18}F-fluorodeoxyglucose positron emission tomography – computed tomography (FDG PET-CT) images of the thorax in the evaluation of patients with primary non-small cell lung cancer (NSCLC) with regional nodal disease. Lesion detection can be difficult due to low contrast between lesions and normal anatomical structures. Lesion characterization is also challenging due to similar spatial characteristics between the lung tumors and abnormal lymph nodes. To tackle these problems, we propose a context driven approximation (CDA) method. There are two main components of our method. First, a sparse representation technique with region-level contexts was designed for lesion detection. To discriminate low-contrast data with sparse representation, we propose a reference consistency constraint and a spatial consistent constraint. Second, a multi-atlas technique with image-level contexts was designed to represent the spatial characteristics for lesion characterization. To accommodate inter-subject variation in a multi-atlas model, we propose an appearance constraint and a similarity constraint. The CDA method is effective with a simple feature set, and does not require parametric modeling of feature space separation. The experiments on a clinical FDG PET-CT dataset show promising performance improvement over the state-of-the-art.**

***Index Terms*—Detection, characterization, approximation, sparse representation, multi-atlas model**

This work was supported in part by ARC grants. H. Huang was partially supported by US NSF IIS-1117965, 1302675, 1344152. *Asterisk indicates corresponding author.*

Y. Song* and W. Cai are with Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, NSW 2006, Australia (e-mail: yson1723@uni.sydney.edu.au).

H. Huang is with Department of Computer Science and Engineering, University of Texas, Arlington, TX 76019, USA.

X. Wang is with Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong.

Y. Zhou is with the Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA.

M. J. Fulham is with the Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, NSW 2050, Australia, and Sydney Medical School, University of Sydney, NSW 2006, Australia.

D. D. Feng is with BMIT Research Group, School of Information Technologies, University of Sydney, NSW 2006, Australia, and with the Center for Multimedia Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, and also with Med-X Research Institute, Shanghai Jiaotong University, Shanghai 200030, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

I. Introduction

Lung cancer is the leading cause of cancer death in many countries and NSCLC accounts for about 80% of all cases [1], [2]. The 5-year survival rate and the treatment approach vary according to stage. Accurate staging is essential to choose the type of treatment and helps determine prognosis. NSCLC staging takes into account the location and extent of the primary tumor, spread to regional lymph nodes and to sites beyond the thorax. Within the thorax, the size and spatial extent of the primary lung tumor and the degree of involvement of regional lymph nodes are critical factors.

FDG PET-CT is the most accurate imaging modality for lung cancer staging [3]. While FDG PET was a valuable non-invasive imaging technique to detect functional rather than anatomical data, its main limitation was the lack of spatial resolution. The combination of PET and CT in one device helps overcome this limitation. Manual interpretation of a PET-CT study is time-consuming due to the large volume of data and requires extensive experience and so can suffer from inter-observer differences. An automated lesion detection methodology would be valuable. In this study, we focus on detecting sites of abnormal pathology or ‘lesions’, in the lung parenchyma, which is the primary lung tumor, and in the pulmonary hilar regions and in the mediastinum, which are involved lymph nodes.

In thoracic PET-CT studies, the main challenges to automated lesion detection are low contrast between normal anatomical structures and lesions, and inter-subject variations in tumoural FDG uptake. On CT, lesions usually appear similar to the soft tissues, which is problematic when lesions are located adjacent to the chest wall or mediastinal structures. For PET, although lesions are typically more FDG-avid than surrounding structures, some lesions have only mild FDG uptake; and there can be elevated FDG uptake in normal structures. Such low contrast implies that detecting lesions based on the within-subject information can be complicated. Furthermore, adding supervised information from other subjects might not be very useful as well, due to the inter-subject variation. Different subjects can display different ranges of FDG uptake, in both the normal anatomical structures and lesions. The separation criteria would thus vary between subjects.

Another challenge is to characterize the lesion that is detected, i.e. is it a primary lung tumor or an abnormal lymph node? The main distinguishing feature between the two types of lesions is spatial information. Primary lung tumors are usually located in the lung parenchyma and lymph nodes at the pulmonary hilar regions and in the mediastinum. However, lung tumors can invade the mediastinum and lymph nodes can be adjacent to the lung parenchyma. In both examples, the two lesion types have similar spatial characteristics and it is difficult to differentiate between them. The key problem is thus how to effectively extract and represent the spatial information for computerized processing.

A. Related Work

Most lesion detection methods on PET-CT images are based on thresholding. The standard uptake value (SUV), which is a semi-quantitative measure of FDG uptake, is widely used to determine the threshold. Traditionally fixed SUV values have been used as the threshold [4], [5]. More recently, adaptive threshold values have been proposed to better accommodate the subject- or region-level characteristics, such as the contrast-based threshold [6]–[8], and the iterative threshold [9], [10]. However, the derived thresholds might not represent a suitable level of spatial scope, and this can affect the detection performance. Apart from the unsupervised SUV-based algorithms, other studies incorporate more complex features and apply classification techniques to detect lesions [11]–[13]. To avoid the inter-subject variation, within-subject information are used as references for patch-wise labeling [14]. However, without structural labeling, this approach could result in over- or under-detected lesion volumes.

Some studies have reported on the detection of only one lesion type – the primary lung tumor or lymph nodes – with the assumption that there is only single lesion type in the image [8], [15], [16]. Without such an assumption, the lung fields can be firstly segmented, and then lung tumors are detected within the segmented lung fields [9], [10], [17]. However, in cases where the lung tumors invade the mediastinum, the segmentation of the lung fields is often unreliable. Another approach is to include false positive reduction to remove lesions that are not likely to be lung tumors, based on tumor-specific features [18]–[20]. These features, however, might not be able to discriminative situations where abnormal lymph nodes are similar to the lung tumor.

On lesion characterization in the thorax, an initial study proposed three levels of features and cascaded classification [21]. The feature set was then improved and a three-stage discriminative model with structural volume-level classification was designed [22]. In further refinements, the feature set was simplified with data-adaptive structure estimation based on image-level [23] and patch-level [14] information. The simpler feature set implies that the refined approaches are more generalizable to unseen data. However, the image-level labeling [23] involves multiple complex models and the patch-level labeling [14] provides less accurate structure estimation.

Learning-based approaches have the advantage that lesion detection or characterization can be guided by prior knowledge inferred from training data. Classification techniques are usually used in such approaches. The commonly used classifiers include the support vector machine (SVM) [13], [16], [20], [22], [23], artificial neural network (ANN) [12], and linear discriminant analysis (LDA) [13]. These classifiers normally work well if there is good separation in the feature space between different classes. However, it would be difficult to design a discriminative feature set, due to the low contrast between lesions and normal structures, and the complex spatial characteristics of lesions. In other words, there could be large similarities between different classes and considerable differences within the same class, causing difficulties in creating a clear feature space separation. The large inter-subject variation also implies that a parametric model learned from the training data might not be generalizable for the testing data.

Non-parametric models, including multi-atlas and sparse representation methods, have thus recently been proposed. These models can be considered similar to the k-nearest neighbor (kNN), which is based on local affinities between the testing and training data. Consequently, the classifier is constructed adaptively to the testing data and the feature set need not be globally discriminative among all training data.

With the multi-atlas method, reference data are referred to as atlases. Majority voting or weighted combination of multiple atlases transfers to the labeling of a testing image. Most commonly the weights are determined based on predefined formulas, such as local similarity between atlases and the test image [24]–[27]. Optimization-based algorithms have been proposed to combine the multiple atlases in a more data-adaptive manner. For example, the weights can be learned, based on reconstruction using the atlases [23], [28]. Sparse regularization has also been incorporated to limit the number of atlases involved in labeling [23], [29]–[31].

Sparse representation has been successfully applied to solve classification problems with applications in face recognition [32], and recently in the medical imaging domain [14], [33]–[37]. Briefly, a reference set is constructed to represent each class, and a reconstruction difference is computed for the test data based on each reference set. The class corresponding to the lowest difference is then the class label of the test data. This method can be considered similar to multi-atlas with sparsity constraints, with each reference set containing multiple atlases of the same class.

In multi-atlas and sparse representation approaches, improvement over the basic sparse regularization has mainly focused on spatial constraints, such as group sparsity [30], [38]. Another technique is to incorporate dictionary learning in place of the raw reference data [29], [33]–[35]. However, in these methods, the optimization is usually to improve the reconstruction, which might not correspond to better classification. Combining dictionary learning with classification has also been demonstrated effective [36], [39]. However, dictionary learning complicates the method design, and its necessity over using raw reference data is usually not considered. A very recent method is to modify the reference data based on logical reasoning, with the intention that such a modification would improve classification. Some examples include graph-guided fusion [40], patch-adaptive sparse approximation [37], and similarity guided labeling [14]. While good performance has been reported [14], [37], [40], there is still scope for improvement by incorporating spatial relationships.

B. Our Contribution

In this work, we propose a lesion detection and characterization method for thoracic FDG PET-CT images in patients with NSCLC. For lesion detection, image patches that clearly represent the lung fields, mediastinum or the lung lesion are first detected with thresholding operations. The remaining patches are then labeled as lesion or mediastinum based on their approximation of the detected lesion and mediastinum patches. For lesion characterization, the original lung fields and mediastinum inclusive of the lesions are first estimated by approximation from other images. A detected lesion is then labeled as lung tumor or abnormal lymph nodes based on its spatial relationship with the estimated anatomical structures.

We refer to our method as context driven approximation (CDA). For lesion detection, an approximation method based on sparse representation was designed to label image patches as a lesion or mediastinum. Region-level contexts extracted from the test image were used as reference data in the approximation. Compared to the existing sparse representation techniques, we designed reference and spatial consistency constraints for a more effective discrimination of low-contrast data. For lesion characterization, an approximation method based on multi-atlas was designed to estimate the original lung fields and mediastinum. Image-level contexts obtained from the reference images were used as atlases in the approximation. We improved the multi-atlas model with appearance and similarity constraints, to handle the inter-subject variations caused by lesions and normal anatomical differences.

For this work, when compared to our prior work [21], [22], we designed a non-parametric approximation method with a much simpler feature set. The approximation method adapts the reference data to the test image with constraint modeling and reconstruction-based optimization, and does not rely on feature space separation by classifiers. This suggests that simple image features could lead to good labeling, without crafting complex features based on the available dataset. Our proposed CDA method is thus less coupled to the current dataset and thus should work well on unseen data.

Preliminary data from this work were reported in abstract form [14], [23]. Our lesion detection method is based on the sparse representation technique [14]. In this work, we designed a spatial consistency constraint in a graphical model to improve the detection performance. Our lesion characterization method is based on the multi-atlas approach [23]. We have improved the appearance constraint for better structure estimation and lower method complexity without the additional structure delineation step. We also designed a simple rule-based algorithm to label lung tumors and abnormal lymph nodes, replacing the more complex SVM classification [23] and the more heuristic post-processing [14].

II. Lesion Detection

Lesions typically exhibit high CT densities similar to the mediastinum and higher FDG uptake than the normal anatomical structures. To locate such areas, contrast information between a region of interest (ROI) and the other regions in the 3D test image is important. In this section, we describe our lesion detection method based on approximation with region-level contexts.

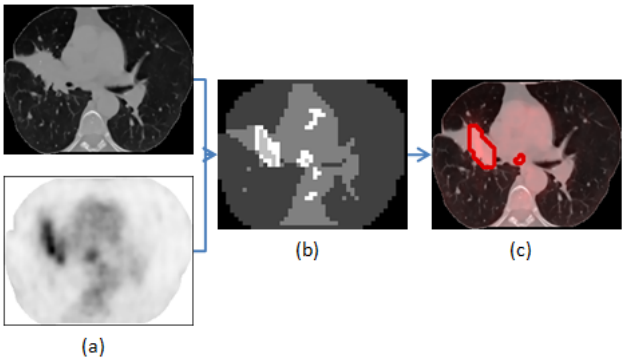


Fig. 1. Illustration of lesion detection. (a) PET-CT axial slice pair with CT on top and PET below showing increased FDG uptake in the right lung. (b) The four-class labeling output, with LF depicted as dark gray, MS as medium gray, LS as light gray and UN as white. (c) The lesion detection output, with LS contour shown as red on the fused PET-CT slice.

Briefly, regions that are obviously representative of lung fields (LF), mediastinum (MS) or lesion (LS) are labeled first (Fig. 1b). The remaining regions (UN) are further labeled as MS or LS based on their approximation of the detected MS and LS regions (Fig. 1c). A connected component of the LS regions is then a lesion object. Our detection approach does not require prior learning and is adaptive to each 3D test image.

Our design motivation can be explained by: (1) While most areas on a FDG PET-CT image can be easily identified as LF, MS or LS, there are some regions that show indistinct features and can be easily misidentified. For example, a non-lesion mediastinal region can have elevated FDG uptake, and a lesion can have relatively low FDG uptake. We hypothesize that these UN regions can be labeled based on their degree of approximation of the identified MS and LS regions. (2) The degree of approximation is computed only based on the MS/LS regions within the same image as the UN region, without involving the other images in the database. This is to reduce the effect of inter-subject variation, considering that the mediastinal regions and lesions in different subjects often display different levels of FDG uptake.

In the following subsections, we describe our methodology of the approximation-based labeling, which is relatively independent of PET-CT imaging and the specific lesion detection problem. Then, a detailed description of the lesion detection approach is presented.

A. Approximation with Region-level Contexts

Let $f_i \in \mathbb{R}^{H\times 1}$ represent the H-dimensional feature vector of an image patch p_i. Suppose a set of image patches $\{p_i\}$ is given with known labels $\mathcal{L}(p_i) = l \in \{1, ..., L\}$, where L denotes the number of distinct labels. A reference dictionary is defined as: $D_l = \{f_i : \mathcal{L}(p_i) = l\} \in \mathbb{R}^{H\times N_l}$, where N_l denotes the number of image patches of label l, and $f_i \in D_l$ is a reference vector representing a reference patch p_i. L reference dictionaries $\{D_l : l = 1, ..., L\}$ are then constructed. The problem is to determine the label of a test image patch p_x based on $\{D_l\}$.

A basic sparse representation approach can be applied to the labeling problem, by finding the reference dictionary D_l that produces the smallest approximation difference for p_x. Specifically, based on each reference dictionary D_l, an approximation of f_x is derived as f_x^l:

$$\begin{aligned} v_l = &\ \underset{v_l}{\operatorname{argmin}} \|f_x - D_l v_l\|_2^2 \quad s.t. \|v_l\|_0 \leq C \\ f_x^l = &\ D_l v_l \end{aligned} \tag{1}$$

where $v_l \in \mathbb{R}^{N_l \times 1}$ is a sparse coefficient/weight vector with C nonzero elements, and C is a constant. Then, the reference dictionary producing the smallest approximation difference between f_x^l and f_x leads to the label of p_x:

$$\mathcal{L}(p_x) = \underset{l}{\operatorname{argmin}} \|f_x - f_x^l\|_2 \tag{2}$$

The potential problems with this basic sparse representation approach are: (1) the best approximation f_x^l from a certain reference dictionary D_l might not correspond to the correct label for p_x, without limiting the possible values in the weight vector v_l; and (2) by labeling image patches individually, inconsistent labels might occur in a local region comprising multiple visually similar patches. Therefore, we improve the sparse representation method with a reference consistency constraint and a spatial consistency constraint.

1) Reference Consistency Constraint: To reduce the possibility that a good approximation could be obtained from a mismatched reference dictionary, our suggestion is to restrict the freedom of approximation. Specifically, we hypothesize that logically, (1) reference patches that are visually similar should contribute similarly to the approximation; and (2) reference patches that are more visually similar to the test patch p_x should contribute more to the approximation. We formulate these hypotheses as the additional reference consistency constraint in the sparse representation. Such a constraint affects the optimization for v_l, and we expect the resultant approximation f_x^l would lead to better classification.

Formally, the following construct is defined:

$$\begin{aligned} v_l = &\ \underset{v_l}{\operatorname{argmin}} \|f_x - D_l v_l\|_2^2 + \Theta(v_l) + \Phi(v_l) \\ &\ s.t.\ \|v_l\|_0 \leq C \\ \Theta(v_l) = &\ \sum_{i=1}^{N_l-1} \sum_{j=i+1}^{N_l} \exp\{-d(f_i, f_j)\} |v_l(i) - v_l(j)| \\ \Phi(v_l) = &\ \sum_{i=1}^{N_l} d(f_x, f_i) v_l(i) \end{aligned} \tag{3}$$

where i and j are indices to the reference vectors in D_l, and $v_l(i)$ denotes the ith element in v_l corresponding to the reference patch p_i. The variable $d(f_1, f_2)$ represents the Euclidean distance between the two feature vectors f_1 and f_2, and is normalized to $[0, 1]$.

The term $\Theta(v_l)$ promotes to assign similar weights in v_l to visually similar reference patches. Given two reference patches p_i and p_j, if they appear similar, i.e. higher $\exp\{-d(f_i, f_j)\}$, then the corresponding elements in v_l should preferably be similar, i.e. smaller $|v_l(i) - v_l(j)|$. The term $\Phi(v_l)$ encourages to assign smaller weights in v_l to the reference patches that are visually different from the test patch p_x. Consider a reference patch p_i, if it is different from p_x with large $d(f_x, f_i)$, then its corresponding element $v_l(i)$ is expected to be small.

To make Eq. (3) easier to solve, we first construct a similarity matrix $U_l \in \mathbb{R}^{0.5N_l(N_l-1) \times N_l}$, with each element defined as [40]:

$$U_l((i,j),k) = \begin{cases} \exp\{-d(f_i, f_j)\} & \text{if } k = i \\ -\exp\{-d(f_i, f_j)\} & \text{if } k = j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where (i, j) and k are the row and column indices of U_l. Each row of U_l corresponds to a pair of reference patches p_i and p_j, with $i = 1, ..., N_l - 1$ and $j = i + 1, ..., N_l$. For the column index, $k = 1, ..., N_l$. The $\Theta(x_l)$ term can then be rewritten as: $\Theta(x_l) = \|U_l v_l\|_1$. Note that if the size of the reference dictionary N_l is large, the dimension of U_l would be large and affect the computational efficiency. Therefore, we devise a simple strategy to restrict the dictionary size, by selecting only the top N_l reference patches that are the most visually similar to the test patch p_x, from all available reference patches of label l.

We also construct a distance vector $V_l \in \mathbb{R}^{1 \times N_l}$ with each element defined as:

$$V_l(i) = d(f_x, f_i) \tag{5}$$

where $i = 1, ..., N_l$. The $\Phi(x_l)$ term can thus be rewritten as: $\Phi(x_l) = \|V_l v_l\|_1$.

Then, we relax the L1 norms with L2 norms so that v_l can be easily computed using the orthogonal matching pursuit (OMP) [41]. The final formulation is thus defined as:

$$\begin{aligned} v_l = &\ \underset{v_l}{\operatorname{argmin}} \|f_x - D_l v_l\|_2^2 + \|U_l v_l\|_2^2 + \|V_l v_l\|_2^2 \\ = &\ \underset{v_l}{\operatorname{argmin}} \| \begin{pmatrix} f_x \\ 0^{0.5N_l(N_l-1) \times 1} \\ 0 \end{pmatrix} - \begin{pmatrix} D_l \\ U_l \\ V_l \end{pmatrix} v_l \|_2^2 \\ = &\ \underset{v_l}{\operatorname{argmin}} \|F_x - \Omega_l v_l\|_2^2 \quad s.t. \quad \|v_l\|_0 \leq C \end{aligned} \tag{6}$$

where F_x denotes the new feature vector for p_x, and Ω_l is the new reference dictionary of label l. Note that Ω_l is adaptive to the test patch p_x with the constructs of U_l and V_l. The approximation of F_x is derived as:

$$F_x^l = \Omega_l v_l \tag{7}$$

2) Spatial Consistency Constraint: To encourage spatially consistent labels in a local region, our idea is to encode the spatial consistency preference into the labeling phase. In particular, rather than determining the label of a test patch p_x based on the approximation difference only, as in Eq. (2), we propose that the label of p_x should also be affected by the surrounding patches, which could contain other test patches or patches with known labels. If p_x is visually similar to the surrounding patches, similar labels should be assigned among them. We formulate this as the additional spatial consistency constraint in a graphical model.

Assume an image contains multiple isolated local regions. A local region R is defined as a connected component comprising a number of image patches $\{p_x\}$ with unknown labels, and R is surrounded by image patches $\{p_i\}$ with known labels. The surrounding patches $\{p_i\}$ are just one layer and immediately adjacent to $\{p_x\}$. We define a graph $\mathcal{G}$ for R with $(M_u + M_n)$ nodes. Each node $p_m \in \mathcal{G}$ represents an image patch, M_u and M_n denote the number of image patches inside and surrounding R. Edges are linked between the nodes p_m and $p_{m'}$ of neighboring patches. The following energy function is then defined:

$$E(\mathcal{L}|\mathcal{G}) = \sum_m \varphi_m(l_m) + \alpha \sum_{m,m'} \psi_{m,m'}(l_m, l_{m'}) \qquad (8)$$

where $l_m \in \{1, ..., L\}$ denotes the possible label of node p_m, $\varphi(\cdot)$ and $\psi(\cdot)$ are the unary and pairwise costs, and α is a constant parameter balancing between the two costs. Minimizing this energy function with graph cut [42] derives the label set $\mathcal{L}$ for the nodes in $\mathcal{G}$. Each image patch $p_m \in R$ is thus assigned the label $\mathcal{L}(p_m)$.

The unary cost $\varphi(\cdot)$ is defined differently for the nodes with or without known labels. For the nodes with unknown labels, i.e. test patches in R, the cost is computed based on the approximation difference:

$$\varphi_m(l_m) = \beta^{-1} \| F_m - F_m^{l_m} \|_2 \qquad (9)$$

where F_m and $F_m^{l_m}$ follow the definitions in Eq. (6) and (7), and β is the maximum of $\|F_m - F_m^l\|_2, \forall(m, l)$ in R so that the maximum cost value is 1. A lower cost $\varphi_m(l_m)$ means p_m is better approximated with D_{l_m}, and higher probability of p_m labeled with l_m. For the nodes with known labels $\mathcal{L}(p_m)$, i.e. patches surrounding R, the cost is computed as:

$$\varphi_m(l_m) = \begin{cases} 0 & \text{if } l_m = \mathcal{L}(p_m) \\ 1 & \text{otherwise} \end{cases} \qquad (10)$$

Such a cost function effectively determines that p_m would only take $\mathcal{L}(p_m)$ to achieve minimum energy.

The pairwise cost $\psi(\cdot)$ penalizes the difference in labels of the neighboring patches p_m and $p_{m'}$. The cost function is defined based on the feature distance with a Pott's model:

$$\psi_{m,m'}(l_m, l_{m'}) = \exp(-\frac{\| f_m - f_{m'} \|^2}{2\gamma}) \mathbf{1}(l_m \neq l_{m'}) \qquad (11)$$

where γ is the normalization factor as the average of $\| f_m - f_{m'} \|^2, \forall(m, m')$ in $\mathcal{G}$. If p_m and $p_{m'}$ are visually similar, the cost of assigning different labels would be high. This thus encourages the spatial smoothness of labeling.

B. Patch-based Detection

We designed a patch-based labeling method for the lesion detection. Given a 3D thoracic FDG PET-CT image I, the image is divided into $5 \times 5 \times 3$ voxel patches $\{p_i\}$. The patches are non-overlapping in xy-dimension but overlap in z-dimension with one-voxel spacing. We choose a small patch size, to avoid smoothing of image features on small lesions. The objective is to label the patches into LF, MS or LS categories: $\mathcal{L}(p_i) \in \{LF, MS, LS\}$, and the LS patches would be the lesions detected.

In the first step, we perform a four-class labeling: $\mathcal{L}(p_i) \in \{LF, MS, LS, UN\}$, where UN represents the patches of unsure (i.e. MS or LS) category (Fig. 2a). Specifically, Otsu thresholding [43] is used to separate the LF patches from the rest based on the average CT densities of patches, since the lung fields exhibit much lower CT densities compared to the other three categories. The PET image is then used to differentiate MS, LS and UN. Consider the difference of LS from MS is mainly due to the high FDG uptake. However, it is difficult to identify the actual tissue type with slightly increased FDG uptake. Therefore, we choose to label the patches that are obviously representative of MS or LS, and mark the remaining patches $\{p_x\}$ as UN for further processing. To do this, the FDG uptake is converted to SUV based on the injected dose and patient weight, and an image-level SUV threshold $t(I)$ is derived in a similar way to our previous work [6]. The only difference is that the threshold $t(I)$ is computed as an average for the 3D image I, rather than the slice-level computation as in the prior work [6]. A non-LF patch p_i is then labeled based on its average SUV s_i: (1) MS, if $s_i < t(I)$; (2) LS, if $s_i > 1.5t(I)$; and (3) UN, otherwise. The parameter 1.5 is chosen based on our empirical study on an initial subset (about 10%) of data.

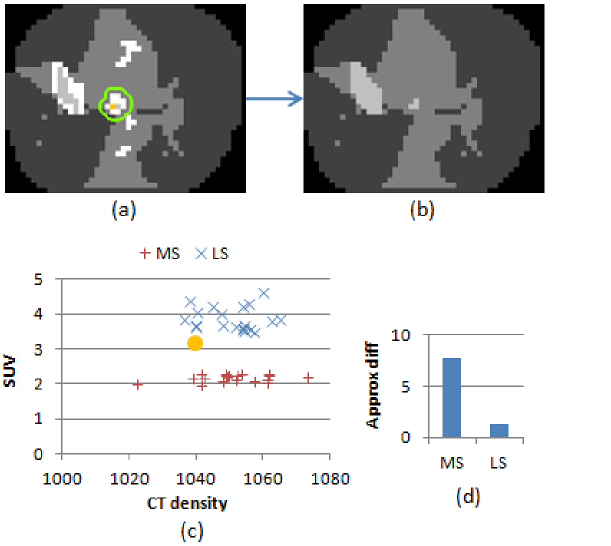


Fig. 2. Illustration of approximation-based labeling for lesion detection. (a) The four-class labeling output, with varying grayscales depicting the four classes of labels; the green circle indicates a local region R, and the yellow dot denotes a test patch $p_x \in R$ of the LS type. (b) The approximation-based labeling output, with light gray depicting LS. (c) Scatter plot describing the mean CT density and SUV of the two reference dictionaries, D_{MS} and D_{LS}, constructed for the test patch p_x, and the yellow disc indicates the feature value of p_x. (d) The approximation differences of p_x, showing smaller difference for LS than MS.

In the second step, the UN patches $\{p_x\} \subset \{p_i\}$ are further labeled as MS or LS (Fig. 2b). To do this, each image patch p_i is represented by a four-dimensional feature vector f_i: its mean and standard deviation of CT densities, and mean and standard deviation of SUV. Next, for a UN patch p_x, two reference dictionaries are constructed based on the labeled MS and LS patches in image I: D_{MS} and D_{LS}. Two feature approximations F_x^{MS} and F_x^{LS} are thus derived using Eq. (7). Then, $\{p_x\}$ are clustered into regions of connected components, and for each region R, a graph $\mathcal{G}$ is constructed. If LF patches exist in the surrounding area of R, these patches are omitted, so $\mathcal{G}$ presents a two-class (MS/LS) problem. The graph-based labeling Eq. (8) is then conducted to obtain the patch-wise labels: $\mathcal{L}(p_x) \in \{MS, LS\}$. The regions of connected LS patches are thus the lesions detected from image I.

Fig. 2c and 2d show examples of the reference dictionaries and the approximation differences derived for a UN patch p_x of the LS type. An example of R is visualized in Fig. 2a. In this study, the parameters N_l and C are set to 20 and 10. Small numbers are used to reduce the computational complexity. The parameter α is set to 0.2. These settings are chosen based on our empirical study on an initial subset (about 10%) of data.

III. Lesion Characterization

The detected lesions in thoracic FDG PET-CT images comprise two types: primary lung tumor (LT) and abnormal lymph nodes (LN). There can also be false positive detection that is actually high FDG uptake in the myocardium (MC), which is normal and non-pathological. In this section, we describe our method for lesion characterization, i.e. differentiation of LT and LN, and filtering of MC, based on approximation with image-level contexts.

The main distinguishing feature among LT, LN and MC is the spatial characteristics. Generally, LT lies inside the lung fields, LNs are in the hilar or mediastinal regions, and MC is a large area in the mediastinum near the left lung field. Therefore, our underlying algorithm identifies the lung fields and mediastinum to extract the spatial characteristics. The LF/MS regions labeled during lesion detection exclude the lesions, hence this makes them unsuitable to determine if LT is inside the lung fields or LNs are inside the mediastinum. In addition, in some cases LT could also invade the mediastinum, with part of the tumor residing outside the lung fields. These considerations suggest that we need to estimate the actual lung fields and the mediastinum from the image, i.e. to reconstruct normal thoracic structures as if the subject was normal.

Briefly, to characterize the lesions, the actual lung fields and mediastinum are firstly estimated (Fig. 3b). Simple spatial features are then computed for the detected lesions, and a lesion is categorized to LT, LN or MC (Fig. 3c). We designed an approximation approach based on multi-atlas for the estimation of the lung fields and mediastinum. The motivation for designing a multi-atlas model, rather than a sparse representation as for lesion detection, is that the approximation is better performed at image level, so that spatial relationships between the overall lung fields and mediastinum are modeled. A reference vector would thus represent an image, and contain mixture of labels (lung fields and mediastinum). This is different from sparse representation that one reference dictionary represents a single label. Additional constraints are also formulated for effective labeling with the image-level contexts.

In the following subsections, we first describe our approximation-based labeling method for estimating the lung fields and mediastinum. Then, a detailed description of the lesion characterization approach is presented.

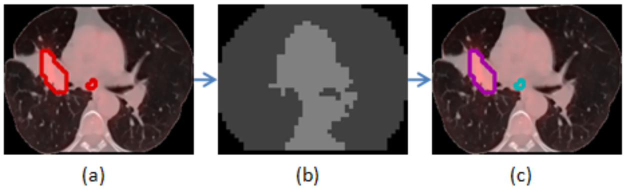


Fig. 3. Illustration of lesion localization. (a) The lesion detection output. (b) The estimation output of lung fields (dark gray) and mediastinum (medium gray). (c) The lesion characterization output, with contours of LT and LN shown as purple and blue on the fused PET-CT slice.

A. Approximation with Image-level Contexts

Our hypothesis is that, an image can be well approximated by a weighted combination of multiple other images. This is based on the observation that there is great similarity in the normal anatomical appearances between images, even though patient-specific conditions introduce variation. Therefore, given a test image I, the approach is to estimate the lung fields and mediastinum based on reference images. The estimation is equivalent to relabeling the patches in I as LF or MS.

To do this, for each axial slice I_s in the test image I, a feature vector $g_x \in \mathbb{R}^{Q \times 1}$ is computed by concatenating the patch-wise labels derived during lesion detection. In g_x, each element $g_x(p_i)$ represents the label (LF, MS or LS) of patch $p_i \in I_s$, with numeric value 1, 2 or 1.5. The value 1.5 means that the LS patches can be relabeled as LF or MS with equal probabilities. Q denotes the number of patches in I_s.

Assume K reference images $\{J_k : k = 1, ..., K\}$ are given, each with annotated lung fields and mediastinum. A reference dictionary $T \in \mathbb{R}^{Q \times K}$ is created: $T = \{g_k : k = 1, ..., K\}$, where g_k denotes the reference label vector from $J_{k,s}$ (the sth slice of J_k). The elements in T are of value 1 or 2, representing LF and MS.

A multi-atlas model with sparse regularization is then formulated to derive the LF/MS labels for I_s:

$$\begin{aligned} w = & \ \underset{w}{\operatorname{argmin}} \|g_x - Tw\|_2^2 \quad s.t. \|w\|_0 \leq C \\ g_x^* = & \ Tw \end{aligned} \tag{12}$$

where $w \in \mathbb{R}^{K \times 1}$ is a weight vector with C nonzero elements, and C is a constant. The vector g_x^* contains real numbers that are approximations of LF/MS labels. To derive the discrete LF/MS labels, a patch $p_i \in I_s$ is relabeled as: (1) LF, if $g_x^*(p_i) < 1.5$; and (2) MS, otherwise. With such patch-wise labels, the lung fields and mediastinum in the test image I are thus estimated.

In this approach, we made two main design choices. First, the feature vector g_x describes an entire slice I_s, rather than individual image patches. The concatenation of patch-wise labels implicitly represents the spatial arrangement of the image patches. The regions that are already identified as LF or MS during lesion detection are thus effectively incorporated as spatial contextual priors to constrain the approximation of the LS patches. With obvious inter-subject variations at small scales, such structural labeling is important for a good estimation of the lung fields and the mediastinum. Second, the approximation is performed per image slice I_s, rather than in 3D. It is observed that data from adjacent slices add little benefit due to large inter-slice spacing, and using the structural information within an axial slice is sufficient and helps to reduce the computational cost.

To further enhance the labeling performance, we designed two additional constraints – appearance and similarity constraints – for the approximation algorithm.

1) Appearance Constraint: One limitation with the multi-atlas model in Eq. (12) is that, the feature vector g_x contains the label information only. Such a single dimension of data can be considered as a quantized representation of the thoracic appearance and is thus less descriptive. It also restricts the approximation target in accommodating the appearance variations between subjects. We thus include the patch-wise average CT density as a second feature vector. PET data are not used given the low spatial resolution and relatively large inter-subject variations compared to the CT data.

Specifically, for I_s, a feature vector $h_x \in \mathbb{R}^{Q\times 1}$ is computed as the concatenation of average CT densities from all patches in I_s. A reference dictionary $A \in \mathbb{R}^{Q\times K}$ is also constructed from $\{J_{k,s}\}$ as: $A = \{h_k : k = 1, ..., K\}$. The multi-atlas model is then reformulated as:

$$\begin{aligned} w = \; & \underset{w}{\text{argmin}} \, \|g_x - Tw\|_2^2 + \|h_x - Aw\|_2^2 \\ & s.t. \|w\|_0 \le C \\ = \; & \underset{w}{\text{argmin}} \, \| \begin{pmatrix} g_x \\ h_x \end{pmatrix} - \begin{pmatrix} T \\ A \end{pmatrix} w\|_2^2 \\ & s.t. \|w\|_0 \le C \\ g_x^* = \; & Tw \end{aligned} \tag{13}$$

Note that the LS patches in I_s need some special handling while creating h_x. In particular, the CT density of a LS patch would be high and similar to MS. If the patch actually represents a lung tumor, it should have been part of the lung fields with low CT density. Using the high density in h_x thus causes an unsuitable approximation target. On the other hand, the high CT density would be suitable for patches representing abnormal lymph nodes. Therefore, to establish a more accurate approximation target, the CT density of a LS patch p_i is redefined as:

$$h_x(p_i) = \lambda c_1 + (1 - \lambda) c_2 \tag{14}$$

where c_1 and c_2 are the average CT densities of the labeled LF and MS patches in I_s. λ is computed as the proportion between the sizes of the LS region containing p_i and a quarter of the MS region detected in I_s, to make $h_x(p_i)$ lower with a larger LS region (i.e. higher probability of being a lung tumor).

2) Similarity Constraint: Another issue with the formulation Eq. (12) is that, each reference image $J_{k,s}$ corresponds to a single weight element w_k in w. This means that all patches in $J_{k,s}$ would contribute equally to the approximation. However, due to the non-rigid structure of the thorax and presence of abnormalities, it is normal that only a portion of $J_{k,s}$ is similar to I_s. The less-similar patches should then carry lower weights towards the approximation. We thus include the patch-wise similarity information between the test image and the reference images to allocate different weights to different patches.

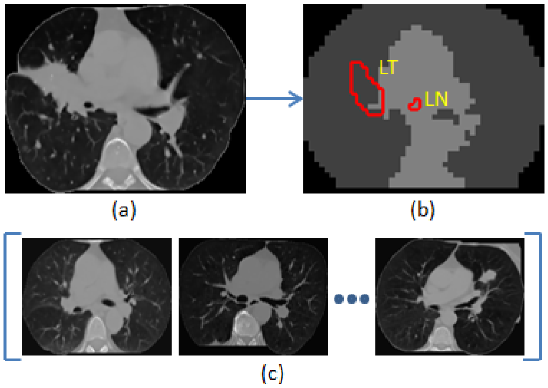


Fig. 4. Illustration of approximation-based labeling for lesion characterization. (a) A CT slice. (b) The estimated lung fields and mediastinum; the red contours indicate the detected lesions, to show that LT and LN can be correctly identified based on their overlap with the estimated lung fields and mediastinum. (c) The reference images for the estimation.

Specifically, for a patch $p_i \in I_s$ and a patch $p_i \in J_{k,s}$ at corresponding spatial locations, a similarity factor $y_{i,k}$ is computed based on the feature distance:

$$\begin{aligned} y_{i,k} = \; & \frac{1}{\eta_i} \exp(-\frac{1}{\theta_i} |h_x(p_i) - h_k(p_i)|) \\ \theta_i = \; & \sum_{k=1}^{K} |h_x(p_i) - h_k(p_i)| \end{aligned} \tag{15}$$

where η_i is to normalize $\sum_k y_{i,k} = 1$. A similarity matrix $Y \in \mathbb{R}^{Q\times K}$ is then created by concatenating all pairwise similarity factors between the patches in I_s and $\{J_{k,s}\}$: $Y = \{y_{i,k} : i = 1, ..., Q; k = 1, ..., K\}$. The multi-atlas model is thus redefined by incorporating the similarity matrix:

$$\begin{aligned} w = \; & \underset{w}{\text{argmin}} \, \| \begin{pmatrix} g_x \\ h_x \end{pmatrix} - \begin{pmatrix} Y \circ T \\ Y \circ A \end{pmatrix} w\|_2^2 \\ & s.t. \|w\|_0 \le C \\ g_x^* = \; & (Y \circ T) w \end{aligned} \tag{16}$$

Based on the approximated g_x^*, the LF/MS labels are then determined for the LS patches, and the lung fields and mediastinum are thus estimated for the test image.

B. Object-based Characterization

In this work, to estimate the lung fields and mediastinum in a test image I, all reference images are firstly linearly rescaled to the same size as the test image I. $K = 10$ images (excluding the test image) are randomly selected from the database as the reference images. The parameter C is set to 5, as half of the reference size. The number of reference images is kept small for performance efficiency. $K = 10$ is also found sufficient, based on our empirical study. Fig. 4 shows an example of the reference images and the estimation output.

Once the lung fields and mediastinum are estimated, a detected lesion is characterized as LT or LN as follows. First, a 3D region consisting of connected LS patches is extracted from the test image I as a lesion object O. A set of spatial features is computed for O: (1) z_1 – size of overlap between O

and the estimated lung fields; (2) z_2 – size of overlap between O and the estimated mediastinum; (3) z_3 – the spatial location of O in the thorax, represented by the z coordinate of the first axial slice of O and the x and y coordinates of the geometric centroid of O; and (4) z_4 – the average cross-section size of O in axial dimension. Note that if $z_4 < \tau$, O is dilated by $\lfloor\sqrt{\tau}\rfloor$ voxels. The parameter τ is set to 250, based on the average z_4 of lymph nodes. The dilation is implemented since the small size of O implies a small margin of error in estimating the lung fields and mediastinum. The inclusion of the surrounding area helps to incorporate more spatial information.

Then, a simple rule-based procedure is used to label O: (1) O is labeled as MS (i.e. MC), if $z_1 < z_2$ and z_3 represents an expected location for myocardium; (2) if $z_4 \geq \tau$, O is labeled as LT if $z_2/z_1 < 1$ and LN otherwise; and (3) if $z_4 < \tau$, O is labeled as LT if $z_2/z_1 < 0.5$ and LN otherwise. Rule (2) is based on the spatial characteristics that O should be LT if it mainly overlaps with the lung fields. Rule (3) introduces a stricter threshold 0.5 for LT, so that a small O is more likely to be labeled as LN.

IV. Dataset and Evaluation Metrics

The dataset comprised 85 sets of 3D thoracic FDG PET-CT image from NSCLC patients (50 men, 35 women; mean age, 68.1 years; age range, 33–86 years). The images were acquired using a Siemens Biograph Truepoint 64-slice PET-CT scanner at the Royal Prince Alfred Hospital, Sydney. The scanner has four rings of detector blocks with a z-axis field of view (FOV) of 21.6 cm. Approximately 400 MBq of ^{18}F-FDG was injected intravenously; the uptake period was 60 minutes. The acquisition time of PET was 1.5 to 4 min per bed position, depending on the patient's weight. PET images were reconstructed using the 3D ordered-subset expectation maximization (3D-OSEM) method [44] with 21 subsets and 3 iterations and point spread function (PSF) based resolution recovery (Siemens HD reconstruction). A Gaussian post reconstruction filter with full width at half maximum of 4 mm was applied. CT-derived attenuation correction, random counts correction, ^{18}F decay correction, and Siemens proprietary scatter correction were incorporated in the reconstruction. The reconstructed matrix size of each transaxial CT slice was 512×512 voxels with a voxel size of $0.98 \times 0.98 \times 3$ mm^3. For PET images, the matrix size was 168×168 with a voxel size of $4.07 \times 4.07 \times 5$ mm^3. The number of PET-CT slices produced at the thorax in a single scanning session was in the range of 65 to 97. During the preprocessing, the PET images were linearly interpolated to the same voxel size as the CT images. Upsampling of PET images was chosen over downsampling of CT images, to avoid losing information in the patch-based feature representation. The background and soft tissues outside the lung and mediastinum were removed automatically [22].

The dataset and the associated ground truth were the same as those we used in our previous study [22]. A total of 93 lung tumors and 65 abnormal lymph nodes were annotated. A summary of the lesion characteristics is listed in Table I. To prepare the ground truth, for each FDG PET-CT image, a senior expert provided a brief description of the lesions that were detected manually. This senior expert has read over 9000 PET-CT lung cancer studies. We then translated the description into 3D masks indicating the image regions of the various lesions. The regions were roughly marked without emphasis on boundary delineation, since precise segmentation was not the goal of this study.

Based on the annotation, we further created reference images of lung fields and mediastinum, for lesion characterization. Specially, in the ground truth, only LT and LN regions were annotated. To obtain the lung fields and mediastinum, the non-annotated regions were firstly labeled as LF or MS using Otsu thresholding based on CT densities. Then the annotated LT regions were merged into the lung field, by changing the voxel labels to LF, and replacing the CT densities with the average CT density of LF regions. The annotated LN regions were merged into the mediastinum, by changing the voxel labels to MS. A reference image thus comprised image patches of two labels: LF and MS.

To evaluate lesion detection, three different outcomes for detection were examined: (1) true positive (TP): an annotated lesion was correctly identified and the volume of the detected lesion overlapped the ground truth annotation by at least 50%; (2) false positive (FP): an extra lesion was detected; and (3) false negative (FN): an annotated lesion was not detected, or the overlap between the detected volume and the annotated volume was smaller than 50%. The overlap formula and the 50% threshold were applied following the PASCAL standard for evaluating object detection [45]. Since our objective was lesion detection rather than segmentation, the ground truth annotation of lesions was not required to delineate the lesion volumes precisely. Misalignment between the detected and annotated volumes was thus expected, and the 50% threshold was used to account for such misalignment. Recall (ratio of TP to TP+FN), precision (ratio of TP to TP+FP), and F-score (harmonic mean of precision and recall) were then computed as the performance metrics for object detection.

For lesion characterization, recall, precision and F-score were also evaluated. For each lesion type (LT or LN), TP, FP and FN were defined. Take LT as an example: (1) TP: a lesion was correctly classified as LT; (2) FP: a lesion was misclassified as LT; and (3) FN: a LT lesion was classified as another type (LN or MS). In addition, receiver operating characteristics (ROC) curves were analyzed. The ROC curve was a plot of true positive rates (TPR) versus false positive rates (FPR), based on the classification probability used in rules (2) and (3) for object-based characterization. The area under the curve (AUC) was then computed to quantify the characterization performance.

In our experiments, we evaluated the various components of the method design, especially focusing on analyzing the effects of added constraints. We also compared the performance with our previous work [22], which was the latest work in lesion detection and characterization with comprehensive performance evaluation. The results reported in [22] were obtained using the same dataset.

TABLE I
SUMMARY OF THE DATASET. (A) PRIMARY LUNG TUMORS. (B) ABNORMAL LYMPH NODES.

Type	# volumes
Well-within lung fields	21
Adjacent to pleural	26
Adjacent to mediastinum	27
Invading mediastinum	19

(a)

Type	# volumes
Well-within mediastinum	27
Adjacent to left lung field	19
Adjacent to right lung field	19

(b)

TABLE II
RESULTS OF LESION DETECTION, COMPARED TO THE PREVIOUS WORK [22].

Method	TP	FP	FN	Recall	Precision	F-score
CDA	157	12	1	0.994	0.929	0.960
[22]	155	20	3	0.981	0.886	0.931

TABLE III
RESULTS OF LESION DETECTION AT PATCH-LEVEL, COMPARING VARIOUS COMPONENTS OF THE PROPOSED METHOD.

Method	Recall-patch	Precision-patch
CDA	0.841	0.867
$t(I)$	0.991	0.512
$1.5t(I)$	0.586	0.921
SR	0.725	0.690
SR+R	0.823	0.773
SR+S	0.693	0.839

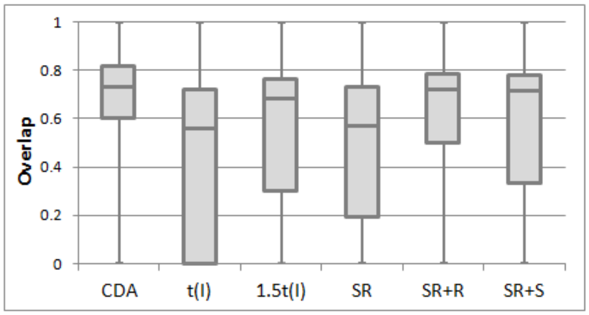


Fig. 6. Box plots of overlap between the detected lesions and ground truth, comparing various components of the proposed method.

V. Results and Discussions

A. Lesion Detection

As shown in Table II, our proposed lesion detection method (CDA) achieved higher performance than the previous work [22]. The CDA method detected almost all lesions, except one FN in abnormal lymph nodes. Some instances of elevated FDG uptake in the mediastinum were detected as FP, which included nine cases that were high-uptake in the myocardium (MC). The same nine MC cases were also detected in the prior study [22]. These FPs were expected at this stage, since MC exhibited high FDG uptake and were not separable from true lesions with the intensity-based image features. Among the non-MC FPs, the CDA method reduced the number of FPs to about 27% compared to the number detected in the work [22].

Fig. 5 shows the overlap in SUV between the mediastinum (MS) and the lesions (LS) in the dataset. The mean and standard deviation of the lower (5%) and upper (95%) range of SUV in MS and LS were computed based on the ground truth annotation. The 5% and 95% values were used rather than the minimum and maximum, to accommodate the imprecise delineation in the ground truth. The overlap between the 95% MS and 5% LS implied that a thresholding technique would have difficulty with patch labeling in this range. The figure also shows that the derived threshold $t(I)$ was normally below the 95% MS whereas the $1.5t(I)$ was normally above the 5% LS. The range between $t(I)$ and $1.5t(I)$ was thus the target for further labeling with the proposed approximation approach.

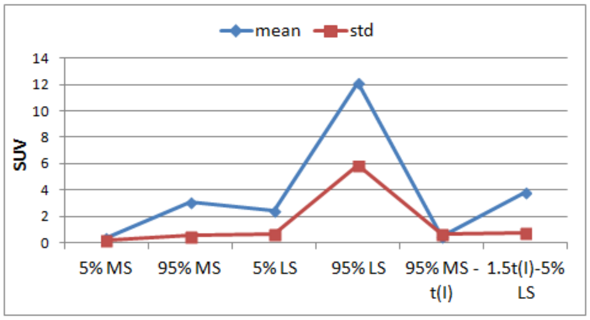


Fig. 5. Statistics of SUV in the dataset.

The labeling performance of lesions was measured at patch-level to evaluate the effectiveness of various components in the proposed method. As shown in Table III, the CDA method achieved the best combination of recall and precision. If the detection was performed using thresholding only, i.e. $t(I)$ or $1.5t(I)$, the recall and precision would be very unbalanced. The basic sparse representation (SR) produced relatively balanced recall and precision, but both measures were not high. The benefits of incorporating the reference consistency constraint (SR+R) and the spatial consistency constraint (SR+S) on top of SR are evident from the table. SR+R improved recall and precision, whereas SR+S was mainly helpful for precision. Specifically, since the surrounding patches were of MS type, the preference of spatial consistency would tend to label the region of interest to MS. Especially if the unary costs in Eq. (8) were derived using SR, the graphical model would not be discriminative enough to obtain accurate labels. SR+R was effective in deriving approximation differences that were consistent with patch labels. Integrating SR+R with SR+S (i.e. CDA) thus helped to enhance the discriminative power of the model, and led to better detection.

Object-level labeling performance was also measured for the comparison. The volume overlap between the detected lesion and the annotation was used as the measurement metric. With 50% used as the detection criterion, the number of detections with more than 50% overlap was important. As

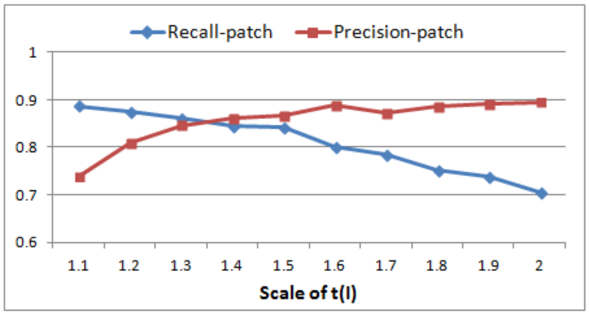


Fig. 7. Results of patch-level lesion detection with CDA, with different settings of the upper threshold.

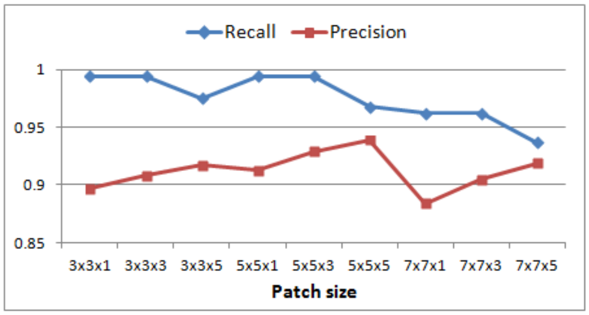


Fig. 8. Results of lesion detection comparing various patch sizes.

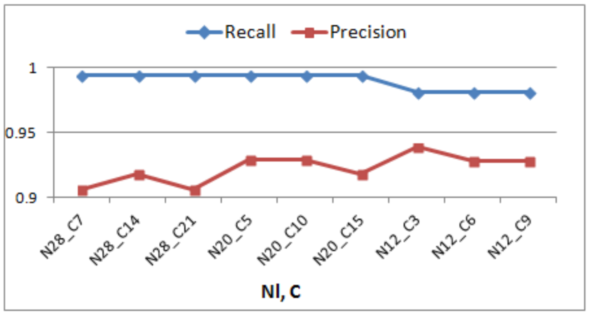


Fig. 9. Results of lesion detection comparing various settings of parameters N_l and C. The notation N20_C10 means $N_l = 20$ and $C = 10$.

shown in Fig. 6, the overlap was mostly above 60% using our CDA method. The problem with $t(I)$ was mainly FPs, while under-estimation of volumes and FNs affected the results of $1.5t(I)$. SR+R largely improved the approximation-based patch labeling, and reduced both FPs and FNs. More detections above the 50% mark were thus obtained. Integrating SR+S with SR+R then helped to further refine the patch labeling and achieved the best detection performance as in CDA.

The effect of the scale parameter, 1.5 in the threshold $1.5t(I)$, on lesion detection is shown in Fig. 7. With the scale varying between 1.1 and 2.0, the best balance between recall and precision was observed at 1.4 and 1.5. Higher scales meant that the reference dictionary of LS D_{LS} would represent higher CT density and SUV. This would cause more LS patches with relatively low CT density or SUV to be labeled as MS and fewer MS patches to be labeled as LS. This thus led to lower recall but higher precision at large scales. The reverse similarly explained for small scales.

Fig. 8 shows the effect of patch size on detection recall and precision. With a certain size in xy-dimension, a larger span in z-dimension resulted in lower recall and higher precision, e.g. $3 \times 3 \times 5$ compared to $3 \times 3 \times 3$. It was mainly because with a larger z size and hence a larger patch, the mean SUV usually became lower; this reduced the number of FPs detected in the mediastinum but also introduced several more FNs in abnormal lymph nodes. The same explanation also applied to the difference between xy sizes of 3×3 and 5×5. With xy size of 7×7, the precision levels were lower than the smaller xy sizes, e.g. $7 \times 7 \times 3$ compared to $5 \times 5 \times 3$. It was because when the xy span became too large, the reference patches representing LS would exhibit relatively low SUV; this effectively lowered the threshold between MS and LS and thus caused some FPs in the mediastinum.

In regard to the subdivision of patches, besides creating patches that were non-overlapping in xy-dimension and one-voxel spacing in z-dimension, we tested two other schemes. First, with half-overlapping in xy-dimension, the detection performance remained unchanged. However, the processing complexity was increased due to the larger amount of patches. Second, with non-overlapping in z-dimension, precision was improved to 0.940 with fewer FPs. However, the lesion volume that was labeled often became distorted because of the coarse subdivision of patches, and the criterion of volume overlap had to be lowered to 30% to maintain the recall level.

As shown in Fig. 9, the detection performance was not sensitive to the parameters N_l and C. There were fewer FPs with a small N_l and fewer FNs with a large N_l. With a certain N_l, different C settings did not affect the recall but resulted in different precision levels. The varying performance occurred in cases with lesions exhibiting relatively low FDG uptake. We only tested small values of N_l, for computational efficiency.

The parameter α produced the best results at 0.2 and 0.3. At larger values (0.4 to 1), small lesions with low FDG uptake would be missed, due to spatial smoothing of labels with the surrounding mediastinum. This could reduce the recall to 0.975. A smaller $\alpha = 0.1$ would reduce the precision to 0.897, due to insufficient spatial information causing FPs at small regions of elevated FDG uptake.

Examples of lesion detection are shown in Fig. 13. While the initial four-class labeling produced some FPs, they were removed with the approximation-based labeling. The first and third examples contain abnormal lymph nodes with relatively low FDG uptake. These lesions were effectively detected and isolated from the mediastinum. The examples also show that with the approximation-based labeling, the volumes of the detected lesions were better delineated. This was important to determine TP of detection, and also important for spatial feature representation at the lesion characterization stage.

With a Matlab implementation on a PC with a 2.66-GHz dual core CPU, lesion detection needed on average 62s per 3D PET-CT image. About 96% of the time was spent on the approximation-based labeling. Compared to the basic sparse representation, CDA took about 50s more time. The extra time

TABLE IV
RESULTS OF LESION CHARACTERIZATION, COMPARED TO THE PREVIOUS WORK [22].(A) PRIMARY LUNG TUMOR. (B) ABNORMAL LYMPH NODES.

Method	TP	FP	FN	Recall	Precision	F-score
CDA	91	6	2	0.979	0.938	0.958
[22]	91	12	2	0.979	0.884	0.924

(a)

Method	TP	FP	FN	Recall	Precision	F-score
CDA	59	5	6	0.908	0.922	0.915
[22]	56	7	9	0.862	0.889	0.875

(b)

TABLE V
THE CAUSES OF INCORRECT CHARACTERIZATION.

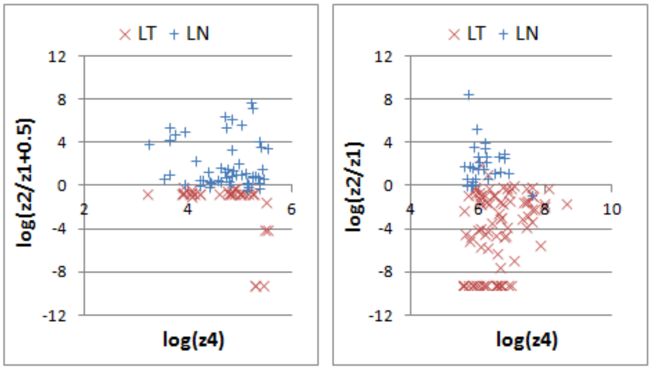
	LT		LN	
	FP	FN	FP	FN
Mislabel with MS	1	0	3	1
Mislabel LT/LN	5	2	2	5

was due mainly to the computation of the similarity matrix and larger dictionary size in the reference consistency constraint.

Note, that we assumed lesions would exhibit distinctive CT densities and FDG uptake from the normal anatomical structures. If the visualization of lesions was largely affected by motion artifact, e.g. normal FDG uptake was shown at the lesion site, an initial correction procedure would be necessary prior to applying the CDA method. Such a procedure was not within the aims of this study.

B. Lesion Characterization

Table IV shows the results of lesion characterization. The causes of FP and FN are listed in Table V. During lesion detection, 12 FPs were initially detected, including 9 MC cases. After lesion characterization, 8 of the MC cases were correctly filtered and relabeled as MS, and 1 was misidentified as LT. The 3 non-MC FPs were in the mediastinum and were thus labeled as LN based on the spatial characteristics. Together with 2 LT cases that were mislabeled as LN, there were 5 FP LN cases. These 2 LT cases appeared near the hilar region, hence were difficult to differentiate from LN. In addition, 5 LN cases were mislabeled as LT. These were mainly cases with LN appearing to attach to the left lung field. The estimation of the left lung field tended to over-estimate slightly and render these LN cases to overlap more with the lung fields. Note that among all the LN cases in the dataset, 19 were adjacent to the left lung field; hence, a majority of such cases were correctly labeled.



(a) (b)

Fig. 10. Scatter plots of proportion between z_1 and z_2 for lesion characterization. (a) Lesions with $z_4 < \tau$. (b) Lesions with $z_4 \geq \tau$.

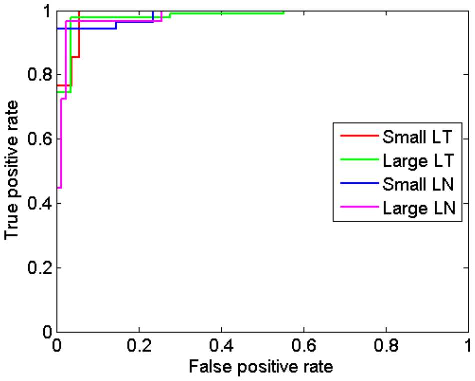


Fig. 11. ROC curves of lesion characterization.

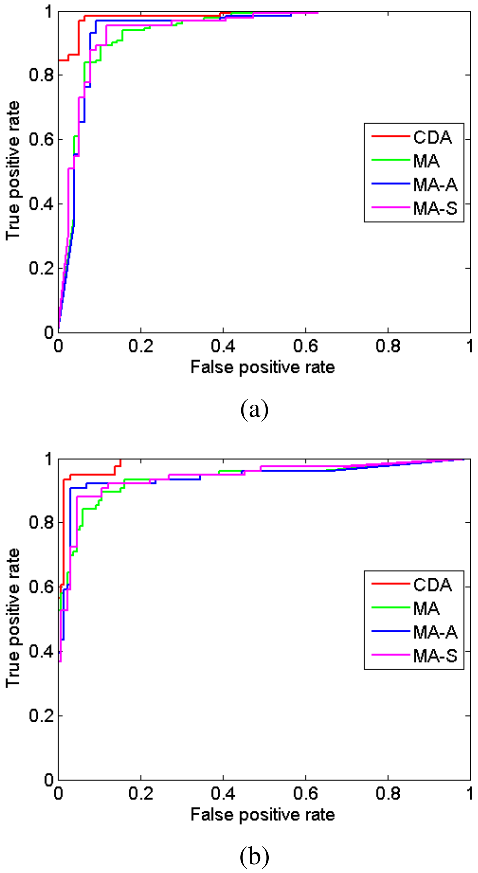
Performance comparison with the work [22] is also shown in Table IV. Compared to this previous work, the CDA method achieved an improved performance, except for the same recall of LT. The improvement in precision in detecting LT mainly related to the lesion detection stage, in which fewer FPs were detected and subsequently all 3 FPs were characterized as LN. The improvement in precision in detecting LN was attributed mainly to better differentiation between LT and LN, such that only 2 LT cases were mislabeled as LN. The recall of LN also showed marked improvement, with fewer FNs resulting from lesion detection and fewer LN cases mislabeled as LT during lesion characterization.

Fig. 10 shows the evaluation of the rule-based classification between LT and LN. Recall that the classification criterion was based on z_2/z_1 and different thresholds (1 or 0.5) were applied depending on the size of the detected lesion (z_4). The evaluation was thus performed separately for small ($z_4 < \tau$) and large ($z_4 \geq \tau$) lesions, as shown in Fig. 10a and 10b. Log scales were used in the plots to accommodate the large value ranges of z_4 and z_2/z_1. In both plots, value 0 on the y-axis indicates the separation line between LT and LN. Fig. 10a shows more points clustered around the separation line. This suggested that a lower threshold (e.g. 0.5) was necessary for accurate classification of small lesions. Overall, good feature space separation can be seen in both cases. This helped to validate the design of the rule-based classification and the thresholds defined.

The characterization performance with varying thresholds of z_2/z_1, i.e. not fixed at 1 or 0.5, was evaluated with ROC analysis. The curves were generated by using z_1/z_2 as the input for LT, and z_2/z_1 as the input for LN. Larger inputs indicated higher probabilities. As shown in Fig. 11, the classification could achieve high TPR with low FPR for all types of lesions. The AUC values were above 0.98 for all

TABLE VI
RESULTS OF LESION CHARACTERIZATION, COMPARING VARIOUS COMPONENTS OF THE PROPOSED METHOD.

	LT		LN	
Method	Recall	Precision	Recall	Precision
CDA	0.979	0.938	0.908	0.922
MA	0.893	0.865	0.800	0.800
MA+A	0.957	0.918	0.877	0.891
MA+S	0.936	0.897	0.846	0.859



(a)

(b)

Fig. 12. ROC curves of lesion characterization, comparing various components of the proposed method. (a) Primary lung tumor. (b) Abnormal lymph nodes.

four cases. For small LT, 100% TPR could be easily reached with about 6% FPR, since all small LT cases produced very high z_1/z_2. It was more difficult to achieve 100% TPR for large LTs, because certain cases exhibited very similar spatial characteristics to LN.

Table VI shows the results of lesion characterization with various components of CDA. Compared to the basic multi-atlas model (MA), CDA had improved performance. The inclusion of the appearance (MA+A) and similarity constraints (MA+S) improved the estimation of the lung fields and mediastinum and led to higher recall and precision of LT and LN. With MA, the lung fields tended to over-estimate when there was LN adjacent to the lung fields, and under-estimate in cases where the LT was invading the mediastinum. Such problems with the estimated lung fields would cause mislabeling between LT and LN. MA+A was particularly effective in handling the estimation in these conditions. This was attributed to the customized feature vector computed for the LS regions in Eq. (14), which effectively added a preference of labels. While MA+S introduced similar but smaller improvement over MA, integrating MA+S with MA+A was helpful in further refining the estimation and achieving high performance of lesion characterization as in CDA.

The comparison between the various components of the proposed method was further evaluated with ROC analysis. For LT, the input to ROC analysis was computed as: z_1/z_2 if $z_4 \geq \tau$, and $0.5z_1/z_2$ if $z_4 < \tau$. For LN, the input to ROC analysis was computed as: z_2/z_1 if $z_4 \geq \tau$, and $2z_2/z_1$ if $z_4 < \tau$. The separation point was aligned to 1 for both small and large lesions, and larger values indicated higher probabilities. As shown in Fig. 12, the proposed CDA method achieved the best performance. Compared to MA, MA+A and MA+S, the CDA method reached higher TPR with lower FPR. The AUC was 0.987, which was higher than the compared approaches by 0.04 to 0.06.

Besides the slice-based formulation for structure estimation, we also tested the inclusion of adjacent slices. They were included by concatenating the slice-level features into one feature vector, and the labeling method remained the same. We found that the additional data actually degraded the characterization performance. With the immediately adjacent slices, 3 more LNs and 2 more MCs were mislabeled as LTs. If the next two adjacent slices were included, another 2 more LNs were mislabeled as LTs. These results demonstrated that a better estimation of the lung fields and mediastinum was obtained with slice-level information. With large inter-slice spacing and inter-subject variation, the adjacent slices introduced more variations in anatomical structures between the approximation target and reference images. The approximation would then be biased towards accommodating these additional variations and often result in less accurate estimation for the test slice.

Examples of lesion characterization are shown in Fig. 13. The examples mainly show LT cases that invade the mediastinum. These lesions would appear outside the lung fields based on the lesion detection output. The estimated lung fields and mediastinum achieved good correspondence with the actual anatomical structures. The different aspect ratios of the images were well accommodated by the estimation algorithm. The third example shows LN abutting the lung fields. This lesion could be easily mislabeled as LT if the lung fields were over-estimated at the lesion area. The MC case shown in the third example was also correctly relabeled as MS.

Lesion characterization needed on average 4s per 3D PET-CT image. About 1s was spent on estimation of lung fields and mediastinum. The formulation of the appearance and similarity constraints was low cost with negligible impact on the computational efficiency.

For a 3D thoracic PET-CT image, the time taken for lesion detection and characterization was on average 71s, including preprocessing. For manual interpretation, an experienced image specialist may require between 3–5 minutes; for a less experienced reader this may be extended to 7–10 minutes; this mainly relates to navigating, assimilating and visualizing the

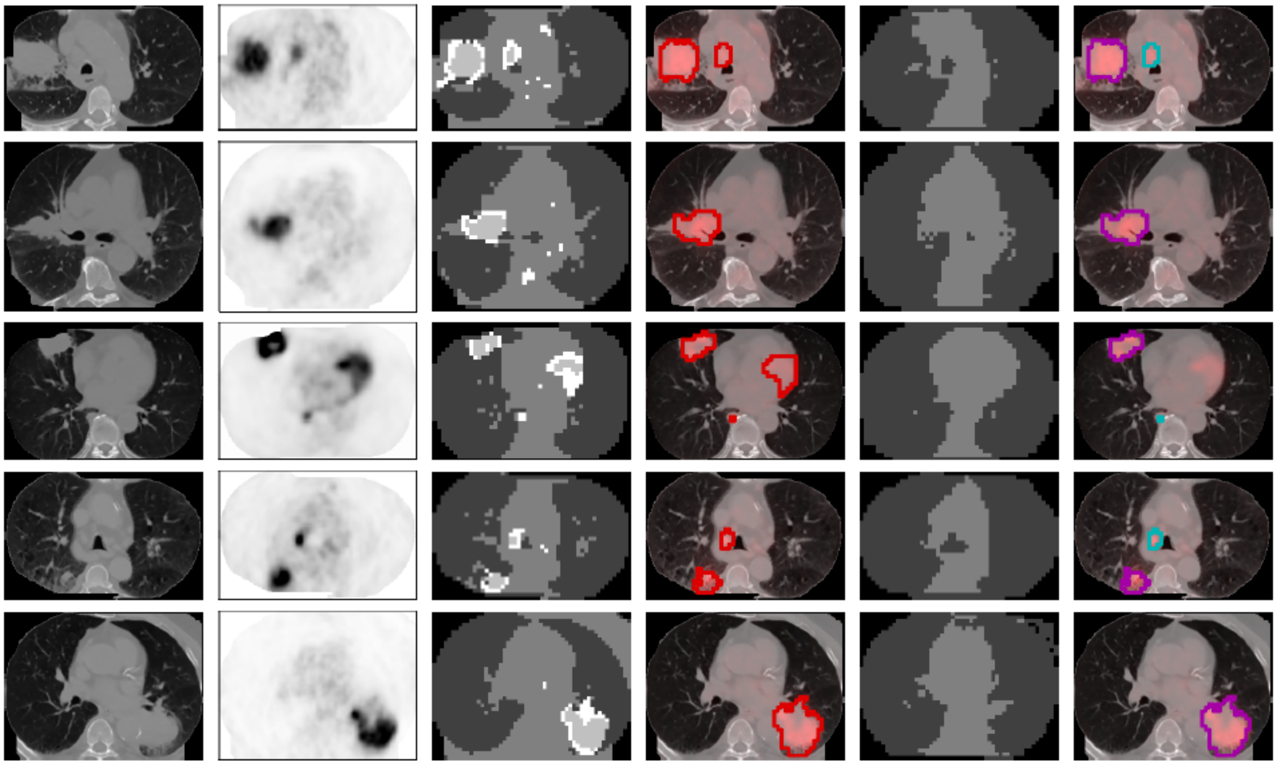


Fig. 13. Example results of lesion detection and characterization from five different NSCLC studies. From left to right columns: a CT slice; the corresponding PET slice; the initial four-class labeling output; the lesion detection output; the estimation output of lung fields and mediastinum; and the lesion characterization output. The color coding follows Fig. 1 and 3.

large volume of image data. While lesions with relatively low FDG uptake could be easily missed with manual interpretation, the proposed method effectively detected such lesions and obtained low FN rate as shown in Table II. On the other hand, manual interpretation typically incorporates more knowledge about the anatomy and subject conditions to filter FPs and differentiate LT and LN. While the proposed method produced some mislabelings as shown in Table V, skilled, accurate expert reading can require years of experience and a less experienced reader can make similar mistakes. In addition, accuracy of manual interpretation can be affected by other distractions in the reading environment. Therefore, we suggest that our proposed method could be a useful adjunct to the clinical workflow.

C. Discussion on Extensibility

We suggest that our CDA methodology could also be extended to other problem domains such as in the evaluation of head and neck cancers. To detect tumors and involvement of regional lymph nodes, the underlying algorithm remains the same: first identify image patches that are obviously representative of normal structures or lesions, and then differentiate the remaining patches based on approximation. In the head and neck region the initial labeling needs to adapt to the different visual characteristics, by labeling three classes – normal tissues (excluding the bones), lesions and unsure areas – with similar thresholding-based techniques. Labeling of the unsure patches is then conducted based on approximation of the labeled normal tissues and lesions, using the proposed sparse representation model.

The detected lesions would contain three types: tumors, abnormal lymph nodes, and high FDG uptake in normal tissues such as tonsils. To characterize them, the proposed approximation-based approach can be applied by first estimating the anatomical structures without lesions. The estimated image would comprise normal tissues, lymph node regions, and areas normally showing high FDG uptake. The proposed multi-atlas model would be applied, by redefining the label vector to represent the three types of structures. The feature vector used in the appearance constraint would be updated with texture features to better describe the three structures, since they exhibit similar CT densities. The detected lesions are then differentiated based on their spatial overlap with the estimated structures.

Overall, we suggest that the proposed sparse representation and multi-atlas models can be generally applicable for identifying lesions. Problem-specific changes would mainly be related to the label and feature vectors.

VI. Conclusion

We present an automated method to detect and characterize the primary lung tumor and disease in regional lymph nodes in thoracic FDG PET-CT images from NSCLC studies. We propose a context driven approximation method to distinguish between lesions and soft tissues, and between lung tumors and abnormal lymph nodes. New sparse representation and multi-atlas models were designed with additional constraints to improve the labeling performance. Patch-based lesion detection and object-based lesion characterization were designed based on approximation with region- and image-level contexts. We evaluated our method on a clinical dataset and showed that our method outperformed the state-of-the-art approaches.

References

[1] World Health Organization, "Cancer, fact sheet no. 297," 2013, http://www.who.int/mediacentre/factsheets/fs297/en/.

[2] S. Edge, D. Byrd, C. Compton, A. Fritz, F. Greene, and A. Trotti (Eds.), *AJCC cancer staging handbook, 7th ed.* Springer, 2010.

[3] W. Wever, S. Stroobants, J. Coolen, and J. Verschakelen, "Integrated PET/CT in the staging of nonsmall cell lung cancer: technical aspects and clinical integration," *Eur. Respir. J.*, vol. 33, pp. 201–212, 2009.

[4] Y. Hashimoto, T. Tsujikawa, C. Kondo, M. Maki, M. Momose, A. Nagai, T. Ohnuki, T. Nishikawa, and K. Kusakabe, "Accuracy of PET for diagnosis of solid pulmonary lesions with 18F-FDG uptake below the standardized uptake value of 2.5," *J. Nucl. Med.*, vol. 47, pp. 426–431, 2006.

[5] H. Guan, T. Kubota, X. Huang, X. S. Zhou, and M. Turk, "Automatic hot spot detection and segmentation in whole body FDG-PET images," *in Proc. ICIP*, pp. 85–88, 2006.

[6] Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng, "Automatic detection of lung tumor and abnormal regional lymph nodes in PET-CT images," *J. Nucl. Med.*, vol. 52, no. Supplement 1, p. 211, 2011.

[7] Y. Song, W. Cai, and D. D. Feng, "Global context inference for adaptive abnormality detection in PET-CT images," *in Proc. ISBI*, pp. 482–485, 2012.

[8] H. Cui, X. Wang, and D. Feng, "Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis," *in Proc. EMBC*, pp. 5384–5387, 2012.

[9] H. Ying, F. Zhou, A. F. Shields, O. Muzik, D. Wu, and E. I. Heath, "A novel computerized approach to enhancing lung tumor detection in whole-body PET images," *in Proc. EMBC*, pp. 1589–1592, 2004.

[10] N. Zsoter, P. Bandi, G. Szabo, Z. Toth, R. A. Bundschuh, J. Dinges, and L. Papp, "PET-CT based automated lung nodule detection," *in Proc. EMBC*, pp. 4974–4977, 2012.

[11] G. Saradhi, G. Gopalakrishnan, A. Roy, R. Mullick, R. Manjeshwar, K. Thielemans, and U. Patil, "A framework for automated tumor detection in thoracic FDG PET images using texture-based features," *in Proc. ISBI*, pp. 97–100, 2009.

[12] M. S. Sharif, M. Abbod, A. Amira, and H. Zaidi, "Artificial neural network-based system for PET volume segmentation," *Int. J. Biomed. Imag.*, vol. 2010, pp. 1–11, 2010.

[13] C. Lartizien, S. Marache-Francisco, and R. Prost, "Automatic detection of lung and liver lesions in 3-D positron emission tomography images: a pilot study," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 1, pp. 102–112, 2012.

[14] Y. Song, W. Cai, H. Huang, X. Wang, S. Eberl, M. Fulham, and D. Feng, "Similarity guided feature labeling for lesion detection," *in MICCAI, LNCS*, vol. 8149, pp. 284–291, 2013.

[15] D. Hellwig, T. P. Graeter, D. Ukena, A. Groeschel, G. W. Sybrecht, H. J. Schaefers, and C. M. Kirsch, "18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense," *J. Nucl. Med.*, vol. 48, pp. 1761–1766, 2007.

[16] C. Lartizien, M. Rogez, A. Susset, F. Giammarile, E. Niaf, and F. Ricard, "Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information," *in Proc. ISBI*, pp. 118–121, 2012.

[17] I. Jafar, H. Ying, A. Shields, and O. Muzik, "Computerized detection of lung tumors in PET/CT images," *in Proc. EMBC*, pp. 2320–2323, 2006.

[18] Y. Cui, B. Zhao, T. Akhurst, J. Yan, and L. Schwartz, "CT-guided automated detection of lung tumors on PET images," *in SPIE Med. Imaging*, vol. 6915, p. 69152N, 2008.

[19] C. Ballangan, X. Wang, S. Eberl, M. Fulham, and D. Feng, "Automated detection and delineation of lung tumors in PET-CT volumes using a lung atlas and iterative mean-SUV threshold," *in SPIE Med. Imaging*, vol. 7259, p. 72593F, 2009.

[20] J. Gubbi, A. Kanakatte, T. Kron, D. Binns, B. Srinivasan, N. Mani, and M. Palaniswami, "Automatic tumour volume delineation in respiratory-gated PET images," *J. Med. Imag. Radia. Oncol.*, vol. 55, pp. 65–76, 2011.

[21] Y. Song, W. Cai, S. Eberl, M. Fulham, and D. Feng, "Discriminative pathological context detection in thoracic images based on multi-level inference," *in MICCAI, LNCS*, vol. 6893, pp. 185–192, 2011.

[22] Y. Song, W. Cai, J. Kim, and D. D. Feng, "A multistage discriminative model for tumor and lymph node detection in thoracic images," *IEEE Trans. Med. Imag.*, vol. 31, no. 5, pp. 1061–1075, 2012.

[23] Y. Song, W. Cai, Y. Zhou, and D. Feng, "Thoracic abnormality detection with data adaptive structure estimation," *in MICCAI, LNCS*, vol. 7510, pp. 74–81, 2012.

[24] F. Rousseau, P. A. Habas, and C. Studholme, "Human brain labeling using image similarities," *in Proc. CVPR*, pp. 1081–1088, 2011.

[25] S. Parisot, H. Duffau, S. Chemouny, and N. Paragios, "Graph-based detection, segmentation & characterization of brain tumors," *in Proc. CVPR*, pp. 988–995, 2012.

[26] P. A. Yushkevich, H. Wang, J. Pluta, and B. B. Avants, "From label fusion to correspondence fusion: a new approach to unbiased groupwise registration," *in Proc CVPR*, pp. 956–963, 2012.

[27] R. Wolz, C. Chu, K. Misawa, K. Mori, and D. Rueckert, "Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases," *in MICCAI, LNCS*, vol. 7510, pp. 10–17, 2012.

[28] T. Chen, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, "Mixture of segmenters with discriminative spatial regularization and sparse weight selection," *in MICCAI, LNCS*, vol. 6893, pp. 595–602, 2011.

[29] S. Zhang, Y. Zhan, Y. Zhou, M. Uzunbas, and D. N. Metaxas, "Shape prior modeling using sparse representation and online dictionary learning," *in MICCAI, LNCS*, vol. 7512, pp. 435–442, 2012.

[30] F. Shi, L. Wang, G. Wu, Y. Zhang, M. Liu, J. H. Gilmore, W. Lin, and D. Shen, "Atlas construction via dictionary learning and group sparsity," *in MICCAI, LNCS*, vol. 7510, pp. 247–255, 2012.

[31] S. Liao, Y. Gao, J. Lian, and D. Shen, "Sparse patch-based label propagation for accurate prostate localization in CT images," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 419–434, 2013.

[32] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[33] M. Liu, L. Lu, X. Ye, S. Yu, and M. Salganicoff, "Sparse classification for computer aided diagnosis using learned dictionaries," *in MICCAI, LNCS*, vol. 6893, pp. 41–48, 2011.

[34] X. Huang, D. P. Dione, C. B. Compas, X. Papademetris, B. A. Lin, A. J. Sinusas, and J. S. Duncan, "A dynamical appearance model based on multiscale sparse representation: segmentation of the left ventricle from 4D echocardiography," *in MICCAI, LNCS*, vol. 7512, pp. 58–65, 2012.

[35] Y. Gao, S. Liao, and D. Shen, "Prostate segmentation by sparse representation based classification," *in MICCAI, LNCS*, vol. 7512, pp. 451–458, 2012.

[36] T. Tong, R.Wolz, P. Coupe, J. V. Hajnal, D. Rueckert, and ANDI, "Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.

[37] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 797–808, 2013.

[38] J. Huang, C. Chen, and L. Axel, "Fast multi-contrast MRI reconstruction," *in MICCAI, LNCS*, vol. 7510, pp. 281–288, 2012.

[39] Y. Song, W. Cai, H. Huang, Y. Wang, and D. D. Feng, "Object localization in medical images based on graphical model with contrast and interest-region terms," *in Proc. CVPR Workshop*, pp. 1–7, 2012.

[40] Y. Han, F. Wu, J. Shao, Q. Tian, and Y. Zhuang, "Graph-guided sparse reconstruction for region tagging," *in Proc. CVPR*, pp. 2981–2988, 2012.

[41] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, 2004.

[42] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, 2004.

[43] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Sys. Man. Cyber*, vol. 9, no. 1, pp. 62–66, 1979.

[44] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.

[45] M. Everingham, L. Gool, C. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.