Latent Discriminative Modeling for Lesion Detection in PET-CT Images

Yang Song[1], Weidong Cai[1], Fan Zhang[1], Heng Huang[2], Yun Zhou[3], Dagan Feng[1,4], Michael J. Fulham[5,6]

[1]BMIT Research Group, School of IT, University of Sydney, Australia
[2]Computer Science and Engineering, University of Texas at Arlington, USA
[3]Johns Hopkins University School of Medicine, USA
[4]Med-X Research Institute, Shanghai Jiaotong University, China
[5]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Australia
[6]Sydney Medical School, University of Sydney, Australia

Abstract. ^{18}*F-fluorodeoxyglucose* positron emission tomography – computed tomography (FDG PET-CT) is now the best imaging technique to accurately stage non-small cell lung cancer (NSCLC). While an automated lesion detection system would help to reduce the time consumption and inter-observer differences in manual image interpretation, the accuracy of automated lesion detection is often limited by the inter-subject variation in normal and tumoural FDG uptake. In this paper, we propose a latent discriminative model to detect the lesions from thoracic images. With this model design, slice-level thresholds of FDG uptake are derived via latent variables during discriminative learning. We thus leveraged the benefit of supervised learning to address the inter-subject variation while explicitly modeling the functional characteristics. We evaluated our proposed method on a clinical dataset and obtained good detection performance.

1 Introduction

Lung cancer is the leading cause of cancer death in many countries. FDG PET-CT is now the best imaging technique to accurately stage non-small cell lung cancer (NSCLC), which is the main primary lung malignancy. Accurate detection of lesions is an essential step towards the staging and treatment planning of patients, but is a time-consuming process in the clinical work flow. The inter-observer differences can also introduce mistakes or inconsistencies in manual detection of lesions.

An automated lesion detection system helps to alleviate these problems with manual interpretation, but the accuracy of detection might be unsatisfactory. Thresholding of the standard uptake value (SUV) in PET has been a common way of lesion detection in clinical research, since lesions typically appear as focal regions of increased FDG uptake, i.e. higher SUVs, in PET images. A simple method to determine the threshold is to use an SUV of value 2.5 or 40% of the maximum SUV [1–5]. Such thresholding techniques have been widely used in

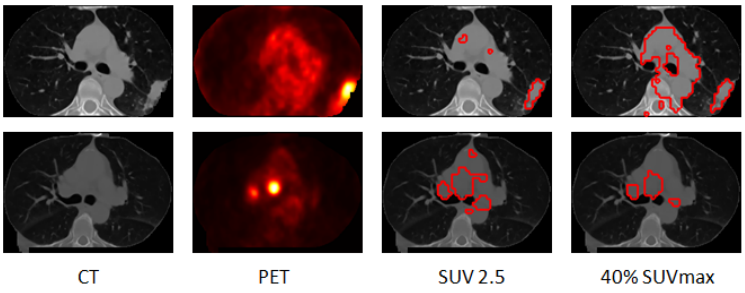

Fig. 1. Two examples showing the thresholding outputs using the standard SUV threshold of 2.5 or 40% of the maximum SUV. Axial slices (CT and PET) after preprocessing (see Section 2.3) are shown. The red contours drawn on the CT images indicate the detected lesions using the respective thresholding techniques. Both techniques result in some false positives in both cases.

clinical studies due to their simplicity. However, with individual heterogeneity in the degree of FDG uptake, these techniques often result in excessive false positive or negative detection of lesions. Fig. 1 shows two examples when the standard SUV thresholds are not effective. The first example illustrates that 40% of maximum SUV identifies many false positives when there is low contrast between the lesion and background. The problem with the second example is caused by the slightly elevated FDG uptake in normal tissue.

Some studies suggest to use subject-adaptive threshold based on background SUV from manually annotated regions in the mediastinum or liver [6, 7]. Following these findings, methods with automatic background estimation have reported performance advantage in lesion detection [8, 9]. However, there could be pathologies in the mediastinum or liver, and the automatic estimation of background regions could mistakenly include these areas and affect the resultant SUV threshold. One way to address this issue is to further process the thresholding output based on the similarity information between image patches within a subject [10, 11]. The method design, however, is relatively complex.

Another type of methods includes more features to identify the lesions. For example, spatial features or information from CT have been used to refine the thresholding results [12–18]. However, the rule-based refinement is generally heuristic and might not generalize well to unseen data. Other studies have incorporated supervised learning with mainly texture feature descriptors for lesion detection [19–25]. Although these approaches are more data adaptive compared to using predefined rules, the effectiveness of these methods can be limited by the discriminative power of the feature descriptors and the ability of classifiers to cope with the feature space overlap between normal tissue and lesion.

For this study we designed a latent discriminative model to detect lesions in thoracic FDG PET-CT images from NSCLC patients. Our method aims to compute subject-adaptive SUV threshold for distinguishing the normal tissue

and lesion. Different from the existing approaches, we model the slice-level SUV thresholds as latent variables in a discriminative construct, so that the thresholds are linked with the classification objective via feature representation, and are optimized at the slice-level with supervised learning. Subject-level threshold is then obtained by averaging the slice-level thresholds.

Our method design is inspired by the latent support vector machine (LSVM) in the deformable parts model [26], which has been widely used in general computer vision [27, 28]. With LSVM, important but unlabeled data (e.g. object part locations), are treated as latent variables and optimized during an iterative learning process. In our study, we consider that a subject-adaptive SUV threshold is essential for correctly detecting lesions. It would be sub-optimal to obtain the threshold based on some heuristic rules or learn a monolithic discriminative classifier (e.g. SVM) from a training population. We suggest that the latent discriminative model provides an efficient way to associate the adaptive threshold computation with the classification objective.

2 Methods

In this work, we model the SUV threshold as a latent variable that affects the feature descriptors of image regions, which are used to classify normal tissue and lesion in a supervised learning construct. To better present our method, we first give some background of the existing work [11] then provide a detailed description of our model.

2.1 Background

It is proposed in the work [11] that the thoracic image (after preprocessing to remove the background) can be divided into four areas: (i) the normal lung fields; (ii) the area clearly representing the normal mediastinum; (iii) the abnormal area with high FDG uptake; and (iv) the area with slightly elevated FDG uptake and potentially containing a mixture of normal and abnormal tissues. Area_(i) is obtained by Otsu thresholding of the CT densities. The remaining image patches are partitioned into the latter three areas based on an automatically computed SUV threshold SUV_{th}: image patches with average SUV $< \mathrm{SUV}_{th}$ are categorized as Area_(ii); image patches with average SUV $> 1.5\mathrm{SUV}_{th}$ as Area_(iii); and the rest as Area_(iv). With an improved sparse representation algorithm incorporating the reference and spatial consistency constraints [11], the image patches in Area_(iv) are further classified as normal or abnormal patches; and the abnormal patches together with Area_(iii) become the detected lesions. SUV_{th} is subject specific and is computed as:

$$\mathrm{SUV}_{th} = 0.15\,\mathrm{SUV}_{max} + \mathrm{SUV}_{med}, \tag{1}$$

where SUV_{max} is the maximum SUV of the subject, SUV_{med} is the average SUV in the mediastinum and this mediastinum is a rough estimation based on the

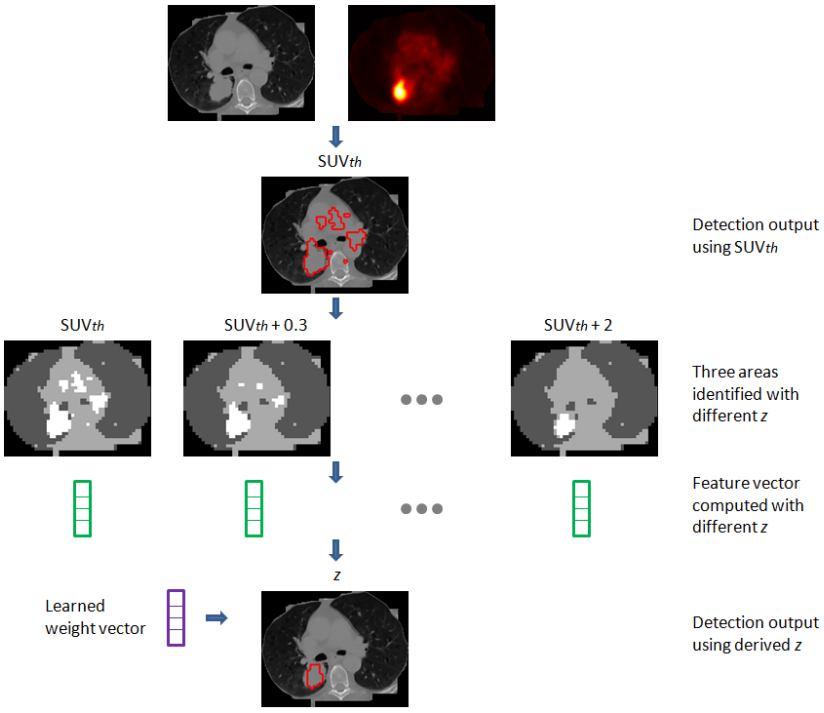

Fig. 2. Illustration of method flow. First, an initial threshold SUV_{th} is computed. Next, a feature vector is computed based on the three areas identified with the threshold. By varying the threshold from SUV_{th} to $\text{SUV}_{th} + 2$, a set of feature vectors is obtained. Then, with the learned weight vector ω, the threshold producing the highest score S is selected as the optimal threshold. Such slice-level thresholds are averaged to obtain the subject-level threshold, and lesion objects are finally detected.

CT density, SUV and location in the thorax. While this method [11] produces highly accurate lesion detection results, the design of the sparse representation algorithm is quite complex with relatively high computational cost and does not involve discriminative learning.

2.2 Latent Discriminative Model

In this work, we propose a latent discriminative model to embed slice-level SUV thresholds as latent variables in an LSVM classifier, so that optimal thresholds are iteratively derived during an SVM optimization for classification between normal tissue and lesion. Compared to the work [11], our model is expected to be more efficient and provide more effective lesion detection by incorporating discriminative learning based on supervised information. The method flow is illustrated in Fig. 2.

We first use the SUV_{th} from Eq. (1) to perform an initial region-of-interest (ROI) detection. Image patches of 5×5 voxels with average SUVs $> \text{SUV}_{th}$ are identified, and a set of connected patches forms an ROI. These ROIs typically include many false positives and the size of a true positive can be much larger than the actual size of a lesion.

A feature descriptor for an ROI r is then computed to represent the likeliness of r being a lesion. To do this, assume a slice-level SUV threshold z is given for the axial slice containing r. Based on this z, the slice is partitioned into three areas: (a) the normal lung fields by Otsu thresholding of CT densities; (b) the area with average SUV $< z$ excluding the lung fields; and (c) the area with average SUV $\geq z$. A four-dimensional feature vector $\phi(r, z)$ is then computed as $\{u_r/u_2, u_r/u_3, u_2/u_1, u_3/u_1\}$, where u_1, u_2, and u_3 are the average SUVs of the three areas, u_r is the average SUV of r, and the threshold z is implicitly utilized in the feature vector via u_2 and u_3. The first two dimensions describe the SUV contrasts between the ROI r and areas (b) and (c), and are helpful in deciding whether r is a lesion. The latter two dimensions represent the overall characteristics of the slice, and are useful as reference contrasts to complement the first two dimensions.

If the given threshold z is optimal, we expect the highest score from the following function:

$$S = \sum_j |\omega \cdot \phi(r_j, z) - b|, \tag{2}$$

where $\{r_j\}$ denotes the set of ROIs in the slice, ω is a four-dimensional weight vector, and b is the bias term. In other words, a better threshold is expected to generate a larger positive or negative value with $\omega \cdot \phi(r_j, z) - b$ for lesion or normal ROI; and the total absolute score S from all ROIs in the slice is larger as well. Therefore, the optimal slice-level threshold z can be obtained by searching over a range of possible thresholds from $z^0 = \text{SUV}_{th}$ to $z^0 + 2$:

$$z = \underset{k}{\text{argmax}} \sum_j |\omega \cdot \phi(r_j, k) - b|, \; k = z^0, z^0 + 0.1, ..., z^0 + 2. \tag{3}$$

The above feature vector computation and threshold derivation are conducted for each slice in the thoracic image. The derived slice-level thresholds of a subject are then averaged to generate the subject-level threshold, which helps to characterize the overall FDG uptake distribution of a subject. Subsequently, the image patches with average SUV higher than this subject-level threshold are labeled as lesion regions; and these regions are then connected in 3D to obtain the lesion objects.

To learn the model parameters ω and b, we formulate an LSVM model with the threshold z as the latent variable:

$$\begin{aligned} &\underset{\omega,\xi,b}{\text{argmin}} \; \frac{1}{2}\|\omega\|^2 + C\sum_i \xi_i \\ &s.t. \; y_i(\omega \cdot \phi(r_i, z_i) - b) \geq 1 - \xi_i, \; \xi_i \geq 0, \; \forall i \end{aligned} \tag{4}$$

where $\{r_i\}$ denotes a set of ROIs as the training data, with label $y_i \in \{-1, 1\}$ indicating normal tissue or lesion. The parameters ξ_i and C follow the definitions in a standard SVM. The key idea is the inclusion of latent variable z_i, which denotes the optimal slice-level SUV threshold to classify r_i. It affects the feature vector $\phi(r_i, z_i)$ of r_i, but its value is unknown for the training and testing sets. An iterative approach is applied to solve Eq. (4):

Step 1: Initialize $\omega = \mathbf{1}$ and $b = 0$.
Step 2: With ω and b fixed, infer z_i using Eq. (3).
Step 3: With z_i fixed, learn ω and b using linear SVM.
Steps 2 and 3 are repeated until convergence.

2.3 Experiment

In this study, we used 95 sets of FDG PET-CT images from NSCLC patients as the test dataset. A total of 184 lesions were annotated based on the clinical notes from a senior radiologist. A separate dataset of 5 subjects with 12 lesions was used as the training set and not included for testing. The training set exhibited the typical lesion characteristics with various degrees of contrast between the lesion and normal tissues. The images were first preprocessed automatically to remove the background and soft tissues outside the lung fields and mediastinum, based on simple morphological operations including Otsu thresholding and connected component analysis. The myocardium area was localized based on the location in the thorax and excluded from the lesion detection and evaluation.

We used recall, precision and F-score as the object-level evaluation metrics. A detected lesion object with at least 50% overlap with the annotated ground truth was marked as true positive. The Dice coefficient, which computed the degree of overlap between the detected lesion and ground truth at the voxel-level, was also measured. For performance comparison, we applied the following techniques to detect the lesions: (i) the sparse representation algorithm [11], which we introduced in Section 2.1; (ii) thresholding using SUV_{th}; (iii) thresholding using SUV of 2.5; (iv) thresholding using 40% of the maximum SUV; (v) classification of initially detected ROIs using a linear-kernel SVM, for which our four-dimensional feature vector was computed based on SUV_{th} without the iterative searching of an optimal threshold.

3 Results

Fig. 3 shows the object-level detection performance. Compared to the sparse representation approach [11] and SVM technique, our method achieved higher recall, precision and F-score. A main advantage of our method over [11] was the lower false negatives. The lesions with relatively low FDG uptake were especially difficult to detect, and our method was more effective in handling some cases. An example is shown in Fig. 4. There were still some false positives detected in the mediastinum, which typically had irregular shapes and could potentially be filtered based on morphological analysis. The SVM technique failed to detect

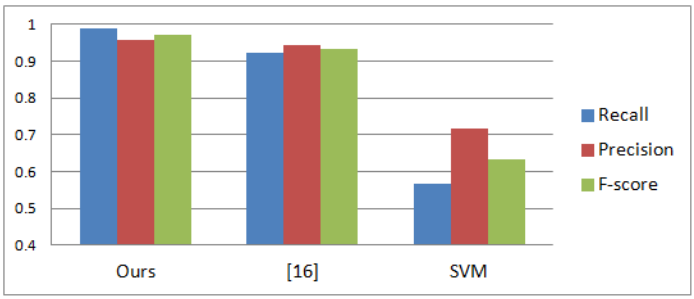

Fig. 3. Lesion detection results using the various techniques.

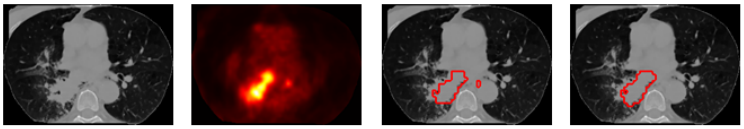

Fig. 4. Example results showing that our method was more effective in detecting lesions with relatively low FDG uptake when compared to [11].

many lesions, even for lesions showing relatively high contrast from the background. It could only detect lesions with average SUVs that were high in a global (i.e. dataset-wise) scope. For subjects with relatively high FDG uptake in normal tissues, more false positives were produced as well. These findings indicate the limitation of a monolithic classifier such as SVM in accommodating the large variations in FDG uptake between subjects.

The results of Dice coefficient are shown in Fig. 5. The improvement of our method over [11] was mainly because the latter approach often produced a larger volume for a detected lesion while our method could better approximate the actual object boundaries. Sometimes if two lesions are spatially close, the sparse representation approach [11] could result in a detected region enclosing both lesions (example in Fig. 6). With a simple 3D connected component analysis, the two lesions would be linked as one lesion object, resulting in a false negative detection at the object-level for the small lesion. The low Dice value of the SVM technique was mainly caused by the large number of false negatives. The SUV thresholding techniques (SUV_{th}, SUV 2.5 and 40% maximum SUV) generated many false positives, causing the lower Dice values. Fig. 7a shows the statistics of the derived subject-level SUV thresholds, i.e. maximum, minimum and three quantiles, using our method, SUV_{th} and 40% maximum SUV. Fig. 7b shows the statistics of the threshold differences between our method and the three thresholding techniques. It can be seen that our method generally produced higher thresholds, hence largely reducing the amount of false positives. Note that we

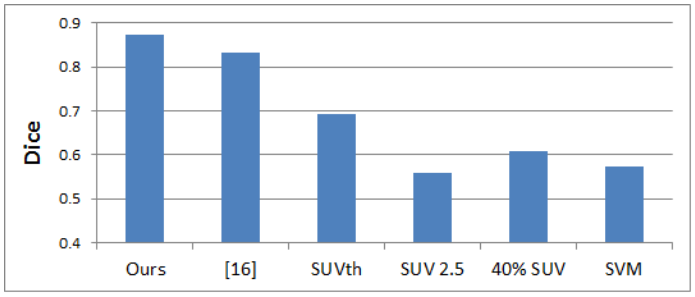

Fig. 5. Dice results of the various detection techniques.

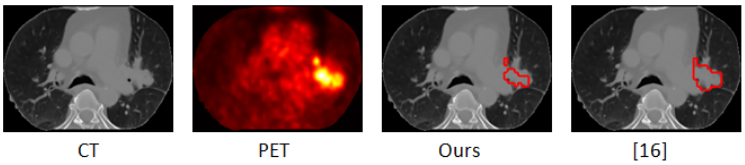

Fig. 6. Example results showing that our proposed method obtained better approximation of the lesion boundaries when compared to [11].

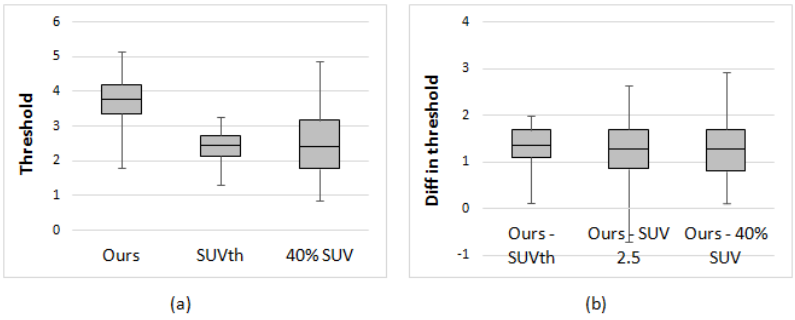

Fig. 7. (a) Box plots of the derived subject-level SUV thresholds using the various methods. (b) Box plots of the differences in SUV thresholds between our method and the three thresholding techniques (SUV_{th}, SUV 2.5 and 40% maximum SUV).

did not report the object-level recall, precision and F-score measures for the three thresholding techniques, because there were excessive amounts of false positives and their sizes were often very small or large. We suggest that the object-level measures were not fair indicators to quantify such detection results. In addition, we would like to mention that our method was not intended as a segmentation approach and the ground truth delineations were not optimized for voxel-wise accuracy, hence the Dice results were not used to show the segmentation accuracy but the extent of false positive or negative voxels detected.

Fig. 8 shows the Dice coefficients when different numbers of iterations were used to train the LSVM model for learning the model parameters. At iteration 0, the weight vector ω was set to $\mathbf{1}$ and the bias term b as 0, and the slice-level SUV thresholds were searched using these initial settings. At subsequent iterations, the learned model parameters were used to search for the optimal thresholds. It can be seen that the Dice value improved considerably during the first three iterations, then stayed relatively flat afterwards. These results indicate that the training process was quite efficient requiring only a small number of iterations.

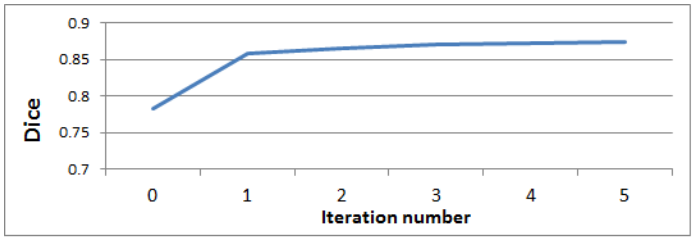

Fig. 8. Dice results corresponding to the various numbers of iterations used for LSVM training.

Our method was implemented in Matlab on a standard PC. On average for each image/subject, the time for lesion detection was 1.8 s. This was more efficient when compared to the sparse representation approach [11], which required on average 62 s per image.

4 Conclusions

We present an adaptive method for detecting lesions from thoracic FDG PET-CT images in this paper. We designed a latent discriminative model, in which the slice-level SUV thresholds are incorporated as the latent variables to compute the feature vectors. Based on the model parameters learned from this discriminative model, subject-adaptive thresholds are derived by maximizing a score function, and then used to detect the lesions. We conducted the experiment on a clinical dataset of NSCLC patients, and demonstrated improved performance over the existing approaches.

References

1. Hashimoto, Y., Tsujikawa, T., Kondo, C., Maki, M., Momose, M., Nagai, A., Ohnuki, T., Nishikawa, T., Kusakabe, K.: Accuracy of PET for diagnosis of solid pulmonary lesions with 18F-FDG uptake below the standardized uptake value of 2.5. J. Nucl. Med. **47** (2006) 426–431
2. Guan, H., Kubota, T., Huang, X., Zhou, X.S., Turk, M.: Automatic hot spot detection and segmentation in whole body FDG-PET images. In: ICIP (2006) 85–88
3. Hellwig, D., Graeter, T.P., Ukena, D., Groeschel, A., Sybrecht, G.W., Schaefers, H.J., Kirsch, C.M.: 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense. J. Nucl. Med. **48** (2007) 1761–1766
4. Bagci, U., Yao, J., Miller-Jaster, K., Chen, X., Mollura, D.J.: Predicting future morphological changes of lesions from radiotracer uptake in 18F-FDG-PET images. PLoS ONE **8**(2) (2013) e57105
5. Foster, B., Bagci, U., Mansoor, A., Xu, Z., Mollura, D.J.: A review on segmentation of positron emission tomography images. Computers in Biology and Medicine **50** (2013) 76–96
6. Nestle, U., Kremp, S., Schaefer-Schuler, A., Sebastian-Welsch, C., Hellwig, D., Rube, C., Kirsch, C.M.: Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. J. Nucl. Med. **46**(8) (2005) 1342–1348
7. Wahl, R.L., Jacene, H., Kasamon, Y., Lodge, M.A.: From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J. Nucl. Med. **50**(Suppl 1) (2009) 122S–150S
8. Song, Y., Cai, W., Eberl, S., Fulham, M.J., Feng, D.: Automatic detection of lung tumor and abnormal regional lymph nodes in PET-CT images. J. Nucl. Med. **52**(Suppl 1) (2011) 211
9. Song, Y., Cai, W., Feng, D.: Global context inference for adaptive abnormality detection in PET-CT images. In: ISBI (2012) 482–485
10. Song, Y., Cai, W., Huang, H., Wang, X., Eberl, S., Fulham, M., Feng, D.: Similarity guided feature labeling for lesion detection. In: Mori, K. et al. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, (2013) 284–291, Springer, Heidelberg
11. Song, Y., Cai, W., Huang, H., Zhou, Y., Wang, Y., Feng, D.: Lesion detection and characterization with context driven approximation in thoracic FDG PET-CT images of NSCLC studies. IEEE Trans. Med. Imag. **33**(2) (2014) 408–421
12. Ying, H., Zhou, F., Shields, A.F., Muzik, O., Wu, D., Heath, E.I.: A novel computerized approach to enhancing lung tumor detection in whole-body PET images. In: EMBC (2004) 1589–1592
13. Jafar, I., Ying, H., Shields, A., Muzik, O.: Computerized detection of lung tumors in PET/CT images. In: EMBC (2006) 2320–2323
14. Cui, Y., Zhao, B., Akhurst, T., Yan, J., Schwartz, L.: CT-guided automated detection of lung tumors on PET images. In: SPIE Med. Imaging **6915** (2008) 69152N
15. Renisch, S., Opfer, R., Wiemker, R.: Towards automatic determination of total tumor burden from PET images. In: SPIE Med. Imaging **7624** (2010) 76241T
16. Zsoter, N., Bandi, P., Szabo, G., Toth, Z., Bundschuh, R.A., Dinges, J., Papp, L.: PET-CT based automated lung nodule detection. In: EMBC (2012) 4974–4977
17. Teramoto, A., Adachi, H., Tsujimoto, M., Fuita, H., Takahashi, K., Yamamuro, O., Tamaki, T., Nishio, M., Kobayashi, T.: Automated detection of lung tumors in PET/CT images using active contour filter. In: SPIE Med. Imaging **9414** (2015) 94142V

18. Song, Y., Cai, W., Zhou, Y., Feng, D., Fulham, M.: Lesion detection with extremal regions in thoracic PET-CT images. J. Nucl. Med. **56**(Suppl 3) (2015) 1783
19. Saradhi, G., Gopalakrishnan, G., Roy, A., Mullick, R., Manjeshwar, R., Thielemans, K., Patil, U.: A framework for automated tumor detection in thoracic FDG PET images using texture-based features. In: ISBI (2009) 97–100
20. Sharif, M.S., Abbod, M., Amira, A., Zaidi, H.: Artificial neural network-based system for PET volume segmentation. Int. J. Biomed. Imag. **2010** (2010) 1–11
21. Song, Y., Cai, W., Eberl, S., Fulham, M., Feng, D.: Thoracic image case retrieval with spatial and contextual information. In: ISBI (2011) 1885–1888
22. Gubbi, J., Kanakatte, A., Kron, T., Binns, D., Srinivasan, B., Mani, N., Palaniswami, M.: Automatic tumour volume delineation in respiratory-gated PET images. J. Med. Imag. Radia. Oncol. **55** (2011) 65–76
23. Song, Y., Cai, W., Eberl, S., Fulham, M., Feng, D.: Discriminative pathological context detection in thoracic images based on multi-level inference. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, (2011) 191–198, Springer, Heidelberg
24. Lartizien, C., Marache-Francisco, S., Prost, R.: Automatic detection of lung and liver lesions in 3-D positron emission tomography images: a pilot study. IEEE Trans. Nucl. Sci. **59**(1) (2012) 102–112
25. Song, Y., Cai, W., Kim, J., Feng, D.D.: A multistage discriminative model for tumor and lymph node detection in thoracic images. IEEE Trans. Med. Imag. **31**(5) (2012) 1061–1075
26. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9) (2010) 1627–1645
27. Raptis, M., Kokkinos, I., Soatto, S.: Discovering discriminative action parts from mid-level video representations. In: CVPR (2012) 1242–1249
28. Zhou, G., Lan, T., Yang, W., Mori, G.: Learning class-to-image distance with object matchings. In: CVPR (2013) 795–802