GLOBAL CONTEXT INFERENCE FOR ADAPTIVE ABNORMALITY DETECTION IN PET-CT IMAGES

Yang Song[1], *Weidong Cai*[1], *David Dagan Feng*[1,2,3]

[1]Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia
[2]Center for Multimedia Signal Processing (CMSP), Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hong Kong
[3]Med-X Research Institute, Shanghai Jiao Tong University, China

ABSTRACT

PET-CT is now accepted as the best imaging technique for non-invasive staging of lung cancers, and a computer-based abnormality detection is potentially useful to assist the reading physicians in diagnosis. In this paper, we present a new fully-automatic approach to detect abnormalities in the thorax based on global context inference. A max-margin learning-based method is designed to infer the global contexts, which together with local features are then classified to produce the detection results adaptively. The proposed method is evaluated on clinical PET-CT images from NSCLC studies, and high detection precision and recall are demonstrated.

***Index Terms*—** PET-CT, abnormality, global contexts, max-margin, detection

1. INTRODUCTION

Lung cancer is the most common cause of cancer-related death in men and women, and in particular, non-small cell lung cancer (NSCLC) is the most prevalent type of lung cancer, accounting for about 80% of all cases. Positron emission tomography – computed tomography (PET-CT) with ^{18}F-fluoro-deoxy-glucose (FDG) tracer is now accepted as the best imaging technique for non-invasive staging of NSCLC. Abnormalities such as tumors and enlarged lymph nodes in FDG-PET appear as hot spots due to increased FDG uptake, and the co-registered CT image provides the anatomical information helpful for localizing the abnormalities.

Detecting abnormalities is the first step in the image-based staging, and an automatic computer-based detection system can potentially improve the clinical efficiency and provide a second opinion for diagnosis. Although the general idea is to identify regions with higher uptake in PET, in practice, difficulties arise due to high uptake variations [1].

The standard uptake value (SUV), which is a widely accepted normalization technique for reducing inter-subject variations, is often used instead of the raw FDG uptakes for tumor detections, such as the very popular thresholds of SUV-2.5 and SUV 40% or 50% [2], thresholds specific to anatomical regions [3], and texture-based analysis [4]. However, since SUV merely reduces the variation problem, by applying the same detection criteria to all image scans, these approaches largely neglected the inter-subject variations and would still result in high false positives or negatives.

The idea is thus to incorporate case-adaptive information in the detection. As suggested [5], normalizations based on reference tissues could reduce the deviations of the SUV distributions; but the normalization model was mainly aimed at classifying normal organ tissues. A particular interesting idea was proposed to adaptively compute the threshold based on a weighted combination of the maximum and background SUVs for lung tumor delineation [6], which required user marking of the background. Inspired by this approach, we recently designed a similar thresholding method, but with fully automatic background/mediastinum estimation [7]. However, the mediastinum was approximated by taking the central area of the thorax without considering case-specific characteristics. For example, tumors might invade into the mediastinum and cause high uptakes in the approximated area, and hence leading to inaccuracies in the threshold computation.

In this paper, we present a new automatic method for abnormality detection in thoracic PET-CT images with NSCLC, by exploring the global contexts adaptively. The proposed method has the following main contributions: (i) a max-margin learning method is designed for automatic global context inference; (ii) the global contexts are integrated with local features to describe the contrast information adaptively; and (iii) abnormalities are detected by classifying the voxels based on the contrast features. In our study, we limit the region of interest to the thorax, and aim to detect all suspicious areas including lung tumors and abnormal lymph nodes. The characteristics of the thorax are exploited in the detection algorithm; however, the main idea is also generalizable to other anatomical regions.

This work was supported in part by ARC and PolyU grants.

2. METHODS

2.1. Datasets

40 sets of 3D PET-CT thoracic images from NSCLC patient studies are used. The images were acquired using a Siemens TrueV 64-slice PET-CT scanner. The total number of annotated abnormalities is 64. The datasets exhibit a high degree of inter-subject variations, with the maximum SUVs of abnormalities in the range of 2 to 16, while the maximum SUVs of the mediastinum are in the range of 1.5 to 3. A preprocessing is performed on each image slice to remove the soft tissues outside of the lung field and mediastinum using thresholding and connected component analysis.

2.2. Global Context Inference

Due to the overlapping between the uptake ranges of abnormalities and mediastinum, detection based on the absolute SUV values would not be sufficient. The intuition is thus to detect abnormalities based on the contrast information on a global scale, since within one image case, the abnormal areas should always depict higher uptakes than the background.

The global contexts of an image $\mathbf{I}$ are defined as its average SUVs of the lung field V_L and the mediastinum V_M, and the maximum SUV V_R, which are the reference values of the normal and abnormal tissues in the thorax. The methods for the global context inference are described in the following.

2.2.1. Initial Labeling of Structures

We first cluster each image slice using mean-shift [8], to incorporate local spatial information around voxels. A 3D image set $\mathbf{I}$ is then represented as a set of regions: $\mathbf{I} = \{r_i : i = 1, ..., N_r\}$, and we classify the regions into three categories: lung field (L), region-of-interest (R), and others (O), using support vector machine (SVM) [9] with linear kernel and default parameters. The classification is based on the regional local features $f_i = \{c_i, p_i\}$, which are the average CT intensity and SUV of r_i.

The L regions depict the actual lung field (excluding tumors) closely, because of their clear feature distinctions from the other categories. Ideally, the R category represents the abnormal areas; however, with only the local features f_i, the R labeling would result in a number of false positives and negatives. And due to the limitation of local features, the O category is mainly the mediastinum, and possibly with other non-mediastinal regions including lower-uptake abnormalities and borders areas of R regions. Examples of the initial labeling outputs are shown in Fig. 1c.

2.2.2. Derivation of V_L

Since the L regions delineate the normal lung field closely, V_L is simply the average SUV of the L regions in $\mathbf{I}$.

2.2.3. Derivation of V_M

Due to the presence of non-mediastinal regions in the O category, the average SUV of the O regions would be too high to establish good contrast with the abnormal regions. Furthermore, a CT-based separation between lung field and mediastinum would not be reliable either, in cases with tumor invasions into the mediastinum. The problem is thus how to derive V_M, without a precise delineation of the mediastinum.

Here we propose to identify a portion of the image that best represents the mediastinum, which is then used to compute V_M. Specifically, given the 3D image set $\mathbf{I} = \{P_j : j = 1, ..., N_P\}$, which comprises of N_P slices, we aim to detect a bounding box $M_j = \{x_j, y_j\}$ representing the mediastinum from each image slice P_j, where x_j and y_j is the coordinate of the top left corner of M_j.

We fix the bounding box as a narrow rectangle with a preset x-span, and y-span of the width of P_j, as shown in Fig. 1c. Such fixed parameters are chosen to cover a large part of the mediastinum while minimizing the inclusion of other types of voxels; and the fixed size also facilitates a fast localization of the bounding box.

To identify this bounding box, we define a score function $S(M_j)$ as:

$$S(M_j) = \omega^T F(M_j) \tag{1}$$

where $F(M_j)$ is the feature vector of M_j and ω is the associated feature weights. This bounding box M_j^* is to best represent the mediastinum, hence maximizing the score:

$$M_j^* = \underset{M_j}{\operatorname{argmax}}\ S(M_j) \tag{2}$$

Feature vector. Besides the local features, the spatial distribution of the voxels in a bounding box is important. For example, the center area should mostly be the mediastinum, while the top and bottom ends have higher tendency of depicting the lung fields. However, at different z locations in the thorax, and in tumor cases, the spatial distribution would vary. Therefore, we require a feature that can accommodate small variations while describing the spatial distributions effectively. To do this, M_j is divided into multi-scale hierarchical rectangular regions in the structure of spatial pyramid matching (SPM) [10]. And for each rectangle, the mean and standard deviation of its CT intensity and SUV are computed. The ordered concatenation of these features from all multi-scale rectangles is then $F(M_j)$.

Weight learning. Since the detected bounding box is expected to have the highest score, we can use the max-margin learning method to train for the feature weights ω. Assuming the best bounding box for P_j is M_j^*, and given another bounding box M_j^o, the following condition should be met: $S(M_j^*) > S(M_j^o)$. By scaling the feature weights, we can ensure that ideally the following inequalities hold:

$$\omega^T F(M_j^*) \geq \omega^T F(M_j^o) + 1 \tag{3}$$

$$\Rightarrow \omega^T (F(M_j^*) - F(M_j^o)) \geq 1 \tag{4}$$

Deviations from the above constraint is the empirical loss, and the optimization goal is to minimize the total empirical loss of all training samples, with an L_2 regularization term:

$$\begin{array}{rl} \text{argmin}_{\omega,\xi_n \geq 0} & \frac{1}{2} \parallel \omega \parallel^2 + C \sum_n \xi_n \\ s.t. & \forall n: \ \omega^T \Delta_n \geq 1 - \xi_n \end{array} \quad (5)$$

where n indexes one training sample, i.e. a pair of bounding boxes, Δ_n is the feature distance $(F(M_j^*) - F(M_j^o))$, ξ is the slack variable as defined in the standard soft-margin SVM, and C is the trade-off parameter between the regularization term and the empirical loss. We then apply the efficient dual-variable method [11] to solve this optimization for ω.

Bounding box localization. With the trained feature weights ω, the best bounding box M_j^* is then localized by searching over P_j. Since the width and length of the bounding box are fixed, the search space is quite small and M_j^* can be quickly localized. Note that although the detected bounding box best represents the mediastinum, due to its rectangular shape and fixed size, it would still contain some L or R areas. They can however, be easily excluded from the bounding box. Some other O areas would also possibly remain, especially if the lung tumor invades a large portion of the mediastinum; but such noises should be minimized in the detected M_j^* comparing to the other regions in the image.

Computation of V_M. For each slice P_j, we exclude those L and R voxels from M_j^*, and an average SUV $V_{M,j}$ is computed. V_M is then the mean value of $V_{M,j}$ over all slices of the image set $\mathbf{I}$.

2.2.4. Derivation of V_R

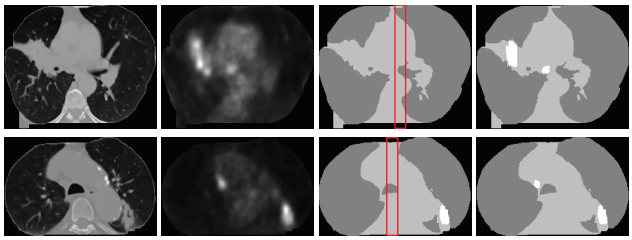
The maximum SUV value of an image is useful because regions with much lower SUVs should belong to the normal tissues. However, since an image can contain multiple tumors or abnormal lymph nodes, and each can exhibit different SUV range, a single value would be too global to adapt to individual abnormality. Therefore, V_R is defined as a set of values $V_R = \{V_{R,j} : j = 1, ..., N_p\}$, with $V_{R,j}$ associated with each slice and computed as the maximum SUV in the three adjacent slices P_{j-1}, P_j and P_{j+1}. Note that for slices without any abnormalities, $V_{R,j}$ would represent normal tissues; this is not a problem though, since the final detection is based on the collection of features, as described below.

2.3. Abnormality Detection

Recall that a 3D image set $\mathbf{I}$ is represented as a set of regions: $\mathbf{I} = \{r_i : i = 1, ..., N_r\}$. For each region, a feature vector g_i is defined as:

$$g_i = \{c_i, p_i, p_i/V_L, p_i/V_M, p_i/V_{R,j}\} \quad (6)$$

where c_i and p_i are the average CT intensity and SUV of r_i, and $r_i \in P_j$. A three-class SVM is then trained to classify the regions into L, R and O categories based on this new feature set. The resultant R regions then represent the detected abnormalities.



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

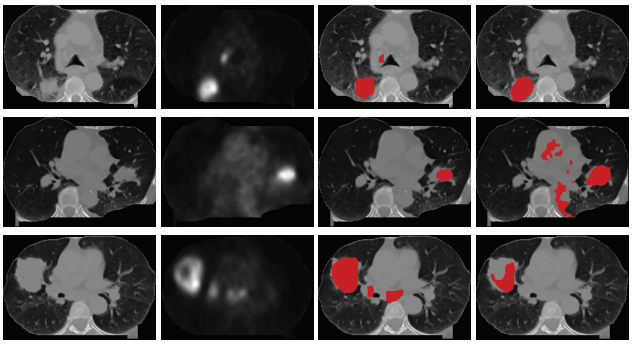
Fig. 1. Example detection outputs using our proposed method, with each row showing one example. (a) Transaxial CT image slices (showing the thorax after preprocessing). (b) Co-registered PET image slices. (c) The initial labeling outputs, with dark gray for category L, light gray for category O and white for category R, and the localized bounding boxes for calculating V_M highlighted in red. (d) The final detection outputs, with abnormalities in white.

3. RESULTS

Fig. 1 shows a couple of examples of the detection process. After preprocessing (Sec 2.1), the CT and PET images are cropped to the lung and mediastinum area, as the 2D slices shown in Fig. 1a and 1b. Each example contains a lung tumor and abnormal lymph nodes, which are easier to see as hot spots on PET images. During the initial labeling (Sec 2.2.1), the L, R and O regions are differentiated; however, some abnormalities are mistaken as O category (Fig. 1c). By incorporating the global contexts, the final labeling (Sec 2.3) manages to identify all abnormalities correctly, as indicated in Fig. 1d. Note that the bounding boxes of the estimated mediastinums (Fig. 1c) tend to minimize the inclusions of actually abnormal regions for computing a more accurate V_M.

Fig. 2 shows several example outputs comparing our proposed method with the the following three approaches: (i) the adaptive SUV threshold [7]; (ii) threshold of SUV 2.5; and (iii) threshold as 50% of the maximum SUV. We did not compare with [3] or [4] because the former focused on organ-specific thresholds for whole-body and the latter omitted the mediastinum from the processing. The proposed method detects all abnormalities correctly in all three cases, while (i) misses the abnormal lymph node due to its relatively low SUV, (ii) includes many false positives with the fixed threshold, and (iii) under-estimates the tumor volume and fails to detect the abnormal lymph nodes.

The quantitative results are summarized in Table 1, and our proposed method demonstrates the highest detection recall, precision and F-score. The previous method [7] has diffi-



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

Fig. 2. Comparisons of detection outputs, with each row showing one example. (a) CT image slices. (b) Co-registered PET image slices. (c) Detection outputs of our proposed method, with abnormalities highlighted in red. (d) Comparison outputs, with top row showing the result from [7], and middle and bottom rows showing the results using SUV 2.5 and 50% maximum SUV as the thresholds respectively.

Table 1. The detection results of various methods.

	Proposed	[7]	SUV-2.5	SUV-50%
Recall (%)	98.4	93.8	95.3	84.4
Precision (%)	92.7	88.2	64.9	81.8
F-score (%)	95.5	90.9	77.2	83.1

culties with small and low SUV abnormalities, which are now better detected with the contrast information based on global contexts, and hence the higher detection recall. Furthermore, because of incorporating the contrast information, we are also able to restrict the initial labeling of R regions to those with high SUVs, and postpone the detection of lower SUV abnormalities to the final stage. Such an approach effectively improves the detection precision of the proposed method. The error patterns shown in Fig. 2 of the other two compared approaches also tend to generalize. With a fixed SUV threshold of 2.5, which is the most widely used method due to its simplicity, many false positives are obtained, while some low uptake abnormalities are still undetected. The threshold at 50% maximum SUV produces reasonable results in most cases; however, the detection performance depends largely on the actual value of the maximum SUV, and changing the proportion to others (e.g. 40%) does not improve the overall performance much.

4. CONCLUSIONS

In this paper, we present a new method for detecting abnormalities from the thoracic PET-CT images automatically based on global context inference. We explain that due to the large inter-subject variations, case-adaptive information is necessary for robust detections, and we propose a classification-based method integrating both regional local features and the global contexts describing the contrast information depicted in an image. We also design a max-margin learning based method to localize a bounding box that best represents the mediastinum to derive one of the main global contexts. Our proposed method is evaluated on a clinical dataset from 40 NSCLC patients with 64 annotated abnormalities, and high detection performance is observed.

5. REFERENCES

[1] W.D. Wever, S. Stroobants, J. Coolen, and J.A. 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.

[2] H. Zaidi and I. E. Naqa, "PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques," *Eur J. Nucl Med Mol Imaging*, vol. 37, no. 11, pp. 2165–2187, 2010.

[3] 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.

[4] G.V. Saradhi, G. Gopalakrishnan, A.S. 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.

[5] V. Potesil, G. Platsch, D. Slosman, and T. Kadir, "Image normalization strategies for organ specific models of functional uptake in FDG-PET/CT," *in Proc. NSS*, pp. 5225–5228, 2008.

[6] U. Nestle, S. Kremp, A. Schaefer-Schuler, C. Sebastian-Welsch, D. Hellwig, C. Rube, and C.M. Kirsch, "Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer," *J. Nucl. Med.*, vol. 46, pp. 1342–1348, 2005.

[7] 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, pp. 211, 2011.

[8] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[9] C.C. Chang and C.J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[10] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," *CVPR*, pp. 2169–2178, 2006.

[11] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," *ICCV*, pp. 1–8, 2007.