Thoracic Abnormality Detection
with Data Adaptive Structure Estimation

Yang Song[1], Weidong Cai[1], Yun Zhou[2], and Dagan Feng[1]

[1] Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Australia
[2] The Russell H. Morgan Department of Radiology and Radiological Science,
Johns Hopkins University School of Medicine

Abstract. Automatic detection of lung tumors and abnormal lymph nodes are useful in assisting lung cancer staging. This paper presents a novel detection method, by first identifying all abnormalities, then differentiating between lung tumors and abnormal lymph nodes based on their degree of overlap with the lung field and mediastinum. Regression-based appearance model and graph-based structure labeling are designed to estimate the actual lung field and mediastinum from the pathology-affected thoracic images adaptively. The proposed method is simple, effective and generalizable, and can be potentially applicable to other medical imaging domains as well. Promising results are demonstrated based on our evaluations on clinical PET-CT data sets from lung cancer patients.

1 Introduction

Lung cancer is currently the leading cause of cancer deaths; and staging plays a critical role in defining the prognosis and the best treatment approaches. Imaging-based staging with positron emission tomography – computed tomography (PET-CT) is now widely accepted as the best non-invasive technique.

Since the existence of primary lung tumors and disease spread in regional lymph nodes are the most important factors for classifying the stage of lung cancer, our aim of this study is to develop a computerized method to detect the lung tumors and abnormal lymph nodes from PET-CT thoracic images automatically. PET highlights abnormal areas with high uptake values (Fig. 1b), but it is difficult to identify the type of the abnormality from PET without well-depicted anatomical structures. While such information can be viewed from the integrated CT (Fig. 1a), it is still quite challenging to differentiate lung tumors and abnormal lymph nodes, especially for complex cases with lung tumors invading into the mediastinum or lymph nodes abutting the lung field.

The prior works mainly focus on detecting either lung tumors [1,2] or lymph nodes [3,4] only. They avoid handling the influence from the other type of abnormality by assuming its non-existence [2,4] or with user-defined region of interest [1,3]. For simultaneous detection of both lung tumors and abnormal lymph nodes, a multi-level inference method has recently been proposed [5]. While the proposed local-, spatial- and object-level features are demonstrated effective for

the detection, the feature design appears to be based on empirical study, and hence might be limited to the available scenarios in the data sets and difficult to generalize to a larger variation of cases.

In this work, we propose a new and intuitive idea to the detection problem – after attempting to detect all abnormalities, if we can identify the actual lung field (tumors inclusive), then we can differentiate lung tumors and abnormal lymph nodes based on the degree of overlap between the detected abnormality and the lung field. The main problem is thus how to estimate the pathology-affected lung field. Limited studies exist in this area, and are mostly based on statistical shape models [6,7], with time-consuming registration [6] or complex landmark detections [7]. Since our problem does not require a very precise lung segmentation, but only a fair estimation of the overlap, we design a simpler yet effective atlas-based approach. Our design can be considered similar to [8], which unlike local-level computation [9], obtains brain segmentation mask by minimizing the weighted difference for the whole image with a regression-based approach. However, its direct derivation of segmentation from multiple weak segmenters might impose a stringent requirement on the weight learning, which would be difficult to optimize in our problem domain due to the large variety of thoracic patterns caused by abnormalities. This thus motivates us to opt for an indirect approach, with intermediate multi-atlas modeling of the feature space and a further classification for final labeling.

Our main contributions of this work are five-fold: (i) we approach the detection problem with a more intuitive and generalizable method, based on estimation of the actual lung field and mediastinum; (ii) the estimation is adaptive to each image, by weighted approximation of appearance model and then structure labeling; (iii) we design a regression approach for the appearance model, with enhanced local weights, supervised labeling information, and sparse regularization; (iv) we construct a customized conditional random field (CRF) [10] for globally-optimal structure labeling, encoding global and pairwise contrast information; and (v) simple features are used for structure estimation and abnormality classification, to keep the method adaptable for other imaging domains.

2   Proposed Method

2.1   Initial Abnormality Detection

The PET-CT thoracic images are first preprocessed to remove the background and soft tissues outside of the lung and mediastinum with morphological operations. All images are then aligned based on the carina of tracheae, and rescaled to the same size [4]. Next, the abnormalities are detected by classification of lung field (L), mediastinum (M) or abnormalities (O) (Fig. 1c), based on PET uptake values and CT densities. This classification method is the same as the *local-level modeling* described in [5], and lung tumors and lymph nodes are not differentiated. The high-uptake myocardium is masked out based on its size, spatial location within the thorax and the shape of the left lung field.

(a)   (b)   (c)   (d)   (e)   (f)
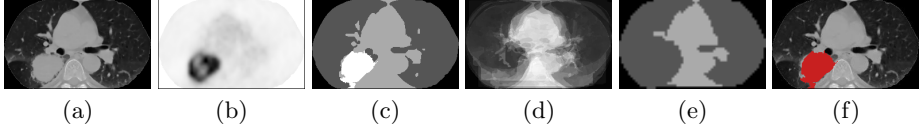
Fig. 1. Method illustration. (a) An axial CT slice (after preprocessing). (b) The co-registered PET slice, where the dark region indicates a lung tumor. (c) Output of the initial abnormality detection, showing the lung field, mediastinum and abnormality with increasing grayscale values. (d) The appearance model generated with regression, approximating the CT intensities if without the lung tumor. (e) Output of the graph-based structure labeling for lung field and medaistinum. (f) The detection output after tumor/lymph node classification, with tumor highlighted in red on CT image.

2.2   Adaptive Structure Estimation

To differentiate between lung tumors and abnormal lymph nodes, a general rule is that lung tumors should be inside the lung field, while lymph nodes are outside. However, as shown in Fig. 1c, due to the lung tumor, only a portion of the right lung field is correctly identified. Such problems are especially common for cases with tumors adjacent to or invading into the mediastinum. Therefore, we need to estimate the actual lung field before the tumor growth (Fig. 1e). Given a 3D PET-CT thoracic volume I, our objective is thus to label each voxel i (excluding the background) to the lung field or mediastinum type. To do this, the thoracic appearance is first modeled from a set of reference images, then the voxels are classified as L/M.

Regression-Based Appearance Model. Although patient-specific conditions introduce variational factors, there is still great similarity between images for the normal structures. It is thus a fair assumption that one image can be approximated by a weighted combination of multiple images. Therefore, at a first stage, we model the CT appearance of the original thoracic structure (Fig. 1d) based on other reference images. PET data is not used here due to its limited capability in depicting the anatomical structures.

We first introduce a basic formulation for the appearance model. Let $y \in \mathbb{R}^{n \times 1}$ be the n-dimensional feature vector (i.e. voxel-wise CT intensities) of I, and $D \in \mathbb{R}^{n \times K}$ be the matrix of K feature vectors from K reference images I_k ($n \gg K$). The difference between y and the weighted combination of D should then be minimized: $\min_x \| y - Dx \|_2^2$, where $x \in \mathbb{R}^{K \times 1}$ is the weight vector; and Dx is the original appearance of I approximated.

With the derived x, each reference image I_k is assigned one weight x_k, and hence all voxels in I_k contribute equally to the approximated appearance. However, due to the non-rigid structure of the thorax and presence of the abnormalities, it is normal that only a portion of I_k is similar to I and the rest should take lower weights. Therefore, we incorporate a voxel-wise similarity-based weight vector for each I_k. For voxel i_k of image I_k, the weight $w_{i,k}$ is computed as:

$$w_{i,k} = \frac{1}{\alpha_i} \exp(-\frac{1}{\beta_i} \parallel i - i_k \parallel_2), \; \beta_i = \sum_{k=1}^{K} \parallel i - i_k \parallel_2 \tag{1}$$

where α_i is to normalize $\sum_k w_{i,k} = 1$. With the weight matrix $W = \{w_{i,k}\} \in \mathbb{R}^{n \times K}$, the regression formulation thus becomes: $\min_x \parallel y - (W \circ D)x \parallel_2^2$.

Furthermore, while the above formulation is sufficient for obtaining a closely matching appearance model, the L/M labeling information is not utilized. Since the final objective is to achieve accurate structure labeling, it is natural to integrate the supervised information to enhance the discriminative power:

$$\min_x \parallel y - (W \circ D)x \parallel_2^2 + \parallel h - (W \circ A)x \parallel_2^2$$
$$= \min_x \parallel \begin{pmatrix} y \\ h \end{pmatrix} - \begin{pmatrix} W \circ D \\ W \circ A \end{pmatrix} x \parallel_2^2 \qquad = \min_x \parallel f - \Omega x \parallel_2^2 \tag{2}$$

where $h \in \{1, 2, 1.5\}^{n \times 1}$ is the label vector of I from the initial detection outputs (1=L, 2=M, and 1.5=O), and $A \in \{1, 2\}^{n \times K}$ for the reference images from the ground truth. The value 1.5 is chosen to have equal distance between O/L and between O/M, to assign no preference for matching such areas with L or M . Both h and A are normalized to the same range as y and D, and the approximated appearance model is then $(W \circ D)x$ and the labeling $(W \circ A)x$.

Finally, to avoid overfitting, we choose not to have all reference images contributing to the appearance approximation, with a sparse regularization:

$$\min_x \parallel f - \Omega x \parallel_2^2, \; s.t. \parallel x \parallel_0 \leq C \tag{3}$$

where C is the constant number of reference images we limit to (set to 5 in this study). The OMP algorithm [11] is then used to solve x.

Implementation details. Due to small correlations between voxels of large distances, and to improve computational efficiency, we divide I into multiple sections, each with three slices, and y is then derived for each section. To construct D, the annotated tumor voxels are replaced with the average intensity of the lung field labeled at the initial detection step.

Graph-Based Structure Labeling. Next, based on the appearance model (Fig. 1d), we would like to classify the lung fields and mediastinum (Fig. 1e). A straightforward idea is to use the approximated labeling $(W \circ A)x$ as the classification output. However, such labelings are sometimes erroneous especially for the boundary areas, as shown in Fig. 2c. Therefore, we design a further graph-based classification step for the structure labeling (Fig. 2d).

We first define a notation for the appearance model: $G = \{g_i\} = (W \circ D)x$, where g_i is the approximated intensity for voxel i. The problem is thus to derive a label set $V = \{v_i \in \{L, M\}\}$, to classify each voxel to category L or M.

Based on the example, we can tell that the mislabeled part in Fig. 2c does appear lighter in G (Fig. 2b), but still darker than the real mediastinum. It thus motivates us to encode contrast information for the labeling. To do this, from G, we first calculate the mean values (m) and the graylevel histograms (d, range 1 to 256)

(a)   (b)   (c)   (d)   (e)   (f)

Fig. 2. Illustration of structure labeling. (a) An axial CT slice. (b) The appearance model. (c) The approximated labeling; compared with (d), a part of the right lung field is misclassified as mediastinum. (d) The structure labeling output. (e) The graylevel histograms of (b) for lung field and mediastinum. (f) The distribution of spatial distances between voxel pairs with nonzeros $s_{i,j}$ in (b).

of the lung field and mediastinum (labeled during the initial abnormality detection). As shown in Fig. 2e, a quite clear separation can be observed between the intensity distributions of L and M; and the probability density of g_i relative to d_L and d_M can be a good indicator of its structure category. A 5-dimensional feature vector q_i is thus computed for each voxel i: (i) g_i; (ii) g_i/m_L; (iii) g_i/m_M; (iv) $Pr[g_i \leq d_L \leq 256]$; and (v) $Pr[1 \leq d_M \leq g_i]$.

In addition to q_i, which incorporates the global-level information m and d, contrast information can also be described in a pairwise fashion. Specifically, for two voxels i and j, if g_i and g_j are similar and they are spatially close, they would likely take the same label. Hence we define the difference $s_{i,j}$ between i and j based on their intensity $\mid g_i - g_j \mid$ and spatial $\parallel i - j \parallel_2$ distances:

$$s_{i,j} = \log(\parallel i - j \parallel_2 + 1) \times \log(\mid g_i - g_j \mid + 1) \qquad (4)$$

A lower $s_{i,j}$ would imply a higher probability of $v_i = v_j$.

We then design a CRF construct to integrate both q_i and $s_{i,j}$ to label G, with the following energy function:

$$E(V|G) = \sum_i \phi(v_i) + \sum_{i,j} \psi(v_i, v_j) \qquad (5)$$

Here $\phi(v_i)$ represents the cost of i taking the label v_i, computed as $1 - p(v_i|q_i)$; and $p(.)$ is the probability estimate from a binary linear-kernel support vector machine (SVM) classifier based on q_i. The pairwise term $\psi(v_i, v_j)$ penalizes the labeling difference between i and j, with a cost value of $\exp(-0.5\gamma^{-1}s_{i,j})\mathbf{1}(v_i \neq v_j)$, where γ is the normalization factor as the average of all $s_{i,j}$ in G.

The pairwise term connects longer distance (beyond neighboring) voxels to encourage consistent labelings for similar voxels. And to ensure a sparse graph, we introduce a constant threshold tr, so that $s_{i,j} = 0$, if $\mid g_i - g_j \mid > tr$ with $tr = 3$; and Fig. 2f indicates that most pairwise terms are formed from non-neighboring voxels. The labeling set V is then derived by minimizing $E(V|G)$ using graph cut [12].

Implementation Details. Since the L/M labeling during the initial abnormality detection is quite accurate for the normal areas of the thorax, we only need to reclassify the detected abnormalities and their surrounding areas. Therefore, the graph-based labeling is conducted for the bounding box volume enclosing the

detected abnormality and with an extended contour (of constant width of 20 voxels) of the bounding box to cover the surrounding areas (denoted as B). For memory efficiency, the image I is rescaled to 1/4 (not too small to affect detection of small lymph nodes) of the size (in xy dimension) and divided into multiple sections (three slices per section), for minimizing $E(V|G)$.

2.3   Feature Extraction and Classification

Based on the estimated thoracic structure V (Fig. 1e), we then classify the detected abnormalities (O) into tumors (T) or abnormal lymph nodes (N) (Fig. 1f). A simple 4-dimensional feature vector is designed: (i) size of O; (ii) size of overlap between O and lung field labeled in V; (iii) size of overlap between O and mediastinum labeled in V; and (iv) size of overlap between O and the convex hull of lung field detected during initial abnormality detection. Features (ii)–(iv) are also normalized by the size of O. A binary linear-kernel SVM is then trained to classify O to T or N. To enhance the error tolerance, the classification is performed on a section basis as well, and the final T/N label is produced based on a weighted averaging of the probability estimates from each section. The weights are computed as $\exp(-d/\eta)$, where d is the distance between the section and center of O, and η is the maximum distance possible for O.

3   Experimental Results

Data Sets. The experiment is performed on 50 sets of 3D PET-CT thoracic images from patients with non-small cell lung cancer (NSCLC), provided by the Royal Prince Alfred Hospital, Sydney. A total of 54 lung tumors and 35 abnormal lymph nodes are annotated as the ground truth. For each data set, the contour of lung field is also roughly delineated. Five images representing the typical cases are selected manually as the training set for both structure labeling and classification between tumors and lymph nodes. The data sets are then randomly divided into five sets; and within each set, each image is used as the testing image, with the other nine as the reference images.

Initial Detection. The initial abnormality detection results in a total of 4 false negatives (2 tumors and 2 lymph nodes), and 5 false positives. This is equivalent to a recall of 95.5% and precision of 94.4% for all abnormalities. These measurements are very similar to the detection rates reported in [5].

Table 1. The labeling accuracy comparing various components of our method. R-* are variations of the regression-based appearance model, and L-* are variations of the graph-based structure labeling. Refer to the text for details.

	R-basic	R-weight	R-label	R-sparse	L-approx	L-global	L-neigh	L-long
Acc (%)	83.9	87.9	89.2	89.7	86.8	88.2	87.4	89.7

Structure Estimation. The usefulness of each component in the structure estimation can be seen from Table 1. The accuracy is computed as *# voxels with correct labeling/ size of the bounding box volume B*. First, with the proposed graph-based structure labeling, we evaluate the appearance model with: basic regression, including voxel-wise weights, labeling information, and sparse regularization, to confirm the benefits over the basic regression model. Next, with the fixed regression-based appearance model, we then evaluate the structure labeling by: using the regression-approximated labeling, classification on global contrast features, and CRF with standard neighboring pairwise terms or long-distance pairwise terms. The results suggest the advantages of the global and pairwise contrast information; and that the standard pairwise terms actually cause lower performance than the non-structured classification. Note that R-sparse and L-long both represent our proposed method.

Table 2. The detection recall and precision

	Tumor (Proposed)	Node (Proposed)	Tumor [5]	Node [5]
Recall (%)	90.7	88.6	84.4	77.8
Precision (%)	89.1	88.6	83.8	76.9

Final Detection. Among the detected abnormalities, three tumors and two lymph nodes are misclassified as the other type. The mislabelings are mainly due to the close resemblance between tumors and lymph nodes at the hilum; and one lymph node is mistaken as tumor due to it connecting into an adjacent tumor volume. Together with the five false positive detections, four of which classified as tumors and one as lymph node, the overall detection recall and precision are shown in Table 2. The results show significant improvement over [5], especially for the abnormal lymph nodes; and it suggests the effectiveness of our approach for differentiating the two abnormalities, by mainly analyzing the degree of overlap between the detected abnormality and the estimated lung structures. Fig. 3 shows three examples with tumors near to the mediastinum and lymph nodes attaching to the lung field, to demonstrate the capability of our proposed method in handling such cases.
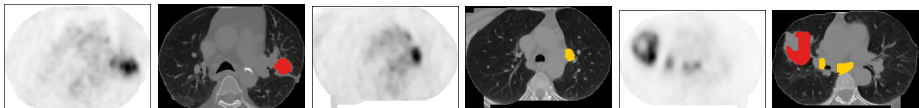
Fig. 3. Three example detection results, with each showing a PET axial slice and the detected tumor or abnormal lymph node highlighted on the CT slice (red for tumors and orange for lymph nodes)

4   Conclusions

We proposed a new detection method for lung tumors and abnormal lymph nodes from PET-CT thoracic images. The actual lung field and mediastinum are estimated with a regression-based appearance model and graph-based structure labeling, and the detected abnormalities are then classified based on their degree of overlap with the estimated structures. We have also shown improved detection performance compared to the existing method. The proposed method would assist the physicians in the image interpretation process and potentially also provide a second opinion for staging.

References

1. 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, pp. 97–100 (2009)
2. 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, 65–76 (2011)
3. Feulner, J., Zhou, S.K., Huber, M., Hornegger, J., Comaniciu, D.: Lymph Nodes Detection in 3-D Chest CT Using a Spatial Prior Probability. In: CVPR, pp. 2926–2932 (2010)
4. Feuerstein, M., Glocker, B., Kitasaka, T., Nakamura, Y., Iwano, S., Mori, K.: Mediastinal Atlas Creation from 3-D Chest Computed Tomography Images: Application to Automated Detection and Station Mapping of Lymph Nodes. Med. Image Anal. 16(1), 63–74 (2011)
5. Song, Y., Cai, W., Eberl, S., Fulham, M.J., 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, pp. 191–198. Springer, Heidelberg (2011)
6. Sluimer, I., Prokop, M., van Ginneken, B.: Toward Automated Segmentation of the Pathological Lung in CT. IEEE Trans. Med. Imag. 24(8), 1025–1038 (2005)
7. Sofka, M., Wetzl, J., Birkbeck, N., Zhang, J., Kohlberger, T., Kaftan, J., Declerck, J., Zhou, S.K.: Multi-stage Learning for Robust Lung Segmentation in Challenging CT Volumes. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 667–674. Springer, Heidelberg (2011)
8. Chen, T., Vemuri, B.C., Rangarajan, A., Eisenschenk, S.J.: Mixture of Segmenters with Discriminative Spatial Regularization and Sparse Weight Selection. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 595–602. Springer, Heidelberg (2011)
9. Rousseau, F., Habas, P.A., Studholme, C.: Human Brain Labeling Using Image Similarities. In: CVPR, pp. 1081–1088 (2011)
10. Wu, D., Lu, L., Bi, J., Shinagawa, Y., Boyer, K., Krishnan, A., Salganicoff, M.: Stratified Learning of Local Anatomical Context for Lung Nodules in CT Images. In: CVPR, pp. 2791–2798 (2010)
11. Tropp, J.: Greed Is Good: Algorithmic Results for Sparse Approximation. IEEE Trans. Inform. Theory 50, 2231–2242 (2004)
12. Kolmogorov, V., Zabih, R.: What Energy Functions Can Be Minimized via Graph Cuts? IEEE Trans. Pattern Anal. Machine Intell. 26(2), 147–159 (2004)