Discriminative Pathological Context Detection in Thoracic Images based on Multi-level Inference

Yang Song[1], Weidong Cai[1], Stefan Eberl[1,2], Michael J Fulham[1,2,3],
Dagan Feng[1,4]

[1]Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Australia
[2]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia
[3]Sydney Medical School, University of Sydney, Australia
[4]Center for Multimedia Signal Processing (CMSP), Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hong Kong

Abstract. Positron emission tomography - computed tomography (PET-CT) is now accepted as the best imaging technique to accurately stage lung cancer. The consistent and accurate interpretation of PET-CT images, however, is not a trivial task. We propose a discriminative, multi-level learning and inference method to automatically detect the pathological contexts in the thoracic PET-CT images, i.e. the primary tumor and its spatial relationships within the lung and mediastinum, and disease in regional lymph nodes. The detection results can also be used as features to retrieve similar images with previous diagnosis from an imaging database as a reference set to aid physicians in PET-CT scan interpretation. Our evaluation with clinical data from lung cancer patients suggests our approach is highly accurate.

1 Introduction

Lung cancer is among the most common malignancies in the Western world, and accurate staging is critical for the selection of the most appropriate therapy, be it surgery, chemotherapy, radiotherapy or combined therapies. The size and extent of the primary tumor and the status of mediastinal lymph nodes are critical for staging the thorax; automated methods to achieve this goal can shorten the time a physician needs to read an image.

PET-CT is now accepted as the best imaging technique to accurately stage the most common form of primary lung cancer, non-small cell lung cancer (NSCLC). PET-CT scanners produce co-registered anatomical (CT) and functional (PET) patient information from a single scanning session. The PET tracer ^{18}*F-fluoro-deoxy-glucose* (FDG) is the most commonly used tracer for clinical PET-CT diagnosis, and tumors typically take up more FDG than surrounding normal structures.

Our aim is to develop a method to automatically detect the primary tumor, the spatial relationships of the tumor within the lung and to the mediastinum, and the location of disease in lymph nodes. The objective is not to perform a precise segmentation, but to provide an inference of the pathological context and function as a robust localization system to assist the reading physician. The detection output can also serve as input to a content-based image retrieval (CBIR) system to retrieve similar imaging cases to help interpretation.

Related work. The majority of existing work focuses on segmentation on CT images using various classification techniques [1–3]. Our method is partially motivated by these approaches. However, they do not support concurrent detection of tumors and abnormal lymph nodes, and do not consider the complexity caused by two pathological types within one image. Recent work by Wojak et.al. introduced a tumor and lymph node segmentation method on PET-CT images using energy minimization [4]. However, the work does not address the differentiation between tumors and lymph nodes, and the spatial context of the tumors.

The work most similar to ours was reported by Wu et.al. [5], for detecting lung nodules and the connectivity with vessel, fissure and lung wall, and did not aim for perfect segmentations. However, it differs from our approach in several aspects: (1) it works on CT subvolumes with the nodule appearing at the center, while our method works on raw PET-CT images of the entire thorax; (2) our method detects abnormal lymph nodes and differentiates them from the primary tumors; and (3) we are interested in the higher-level spatial relationships, i.e. the connectivity between tumors and the chest wall and mediastinum.

Our work has also been provoked by the idea of multi-class object detection proposed for general computer vision problems [6–8]. Different from these methods, we design three levels of features to exploit the specific characteristics of PET-CT thoracic images, and a different multi-level discriminative model for more effective inference of the pathological context.

2 Method

2.1 Discriminative Structure Localization

At the first stage we detected four types of structures – the lung fields (L), mediastinum (M), tumor (T) and disease in lymph nodes (N) – from the thoracic images. We formulated the detection as a multi-level, multi-class (L, M, T or N) object localization problem. For an image **I**, the classification score with labeling **Y** (the label matrix of **I**) is defined as:

$$S(\mathbf{I}, \mathbf{Y}) = \sum_{l} \alpha_{y_l} \cdot f_l + \sum_{s} \beta_{y_s} \cdot f_s + \sum_{o} \gamma_{y_o} \cdot f_o \qquad (1)$$

where f_l, f_s and f_o are the three levels of feature vectors (local, spatial, and object levels) of **I**; α, β and γ are the respective feature weights; y_l, y_s and y_o are the class labels at each level, representing the four classes; and l, s and o are the indices of the regions formed at each level in the transaxial slices. The goal was to find the labeling $\mathbf{Y}$ that maximized the score S for image $\mathbf{I}$.

Our approach was region-based for effective modeling of the higher-level features, and we designed a cascaded learning approach for the classification. The higher-level spatial and object features were important for differentiating the four types of structures, especially for between T and N, and between T and M, as described in more details in the following sections. We also employed a two-phase design by exploring first the 2D features at the local and spatial levels, then the 3D features at the object level; this was to optimize the classification for each image slice first before considering the inter-slice relationships.

Local-level Modeling. Each image slice was first clustered into a number of regions of various sizes and shapes using the mean-shift algorithm [9]. The regions were generated separately for PET and CT slices, and then merged into one set for each slice pair. Each region R_l was then represented by the local feature f_l: the mean CT density; and the mean standardized uptake values (SUV), which was computed by normalizing the mean SUV of R_l based on an adaptive threshold [10].

At the local level, f_l could not differentiate between T and N, because both had high CT densities and high SUV values. So, we limited y_l to take three values, L, M or T/N, to focus on differentiating the pathological tissues from lung fields and mediastinum.

Spatial-level Modeling. Besides an inability to distinguish T and N, another major problem with local-level modeling was that areas surround the tumor were often misclassified as M, which could subsequently cause T to be confused as N. To better classify the surrounding area, we observed that the spatial information played an important part, e.g. its proximity with T and L and distance from M, and the differences between its average CT density and SUV and those of the other regions. Similar spatial features could also help to improve the labeling of some misclassified regions in the mediastinum.

The spatial-level features were thus computed as the following feature vector f_s for region R_s in 11 dimensions: (Dim. 1-3) the average spatial distance from region R_s to other regions R_i of type k ($k \in \{L, M, T/N\}$); (Dim. 4) the size of R_s; (Dim. 5-7) the difference between the mean CT density of R_s and the average CT densities of all regions of type k; (Dim. 8-10) the difference between the mean SUV of R_s and the average SUV of all regions of type k; and (Dim. 11) the local-level labeling at R_s.

The regions R_s at this level were different from the local-level ones. We first performed another mean-shift clustering for areas around the detected abnormal regions, to discover finer-scale details. For regions not connected with the abnormal areas, and with high confidence of being L or M (based on the classification score), we also merged the connected regions of the same type into one region. And similarly to y_l, y_s could be either L, M or T/N.

Object-level Modeling. So far, T and N were still treated as one type, and the transaxial slices were processed separately. Based on the classification results of the previous level, by merging connected regions with the same label into one region, a slice was then represented as a relatively small number of regions, roughly corresponding to the anatomical structures, but with some discontinuous segments. The goal was thus to differentiate tumors from abnormal lymph nodes and smooth the labeling, and we observed that the object-level information was the main distinctive factor. For example, T should be within L and possibly invading into M while N should be within M; hence, the distance between T and L regions should be small and the size of L surrounding T should be large, while N should have similar properties relating to M.

At this level, we thus explored the intra- and inter-slice object-level features. For each merged region R_o, a 32-dimensional feature vector f_o was computed: (Dim. 1-15) the minimum distance from R_o to the type k areas in the d direction (above, below, left, right, and the z direction); (Dim. 16-30) the average size of type k in the d direction relative to R_o, normalized by the dimension of R_o; (Dim. 31) the size of R_o; and (Dim. 32) the spatial-level labeling at R_o. Unlike y_l and y_s, the labeling y_o should then take four possible values: L, M, T or N.

Cascaded Learning and Inference. To create the discriminative classifier, we performed piecewise learning for the feature weights α, β and γ (Eq. (1)). We first trained a one-versus-all multi-class support vector machine (SVM) for the local-level model, then another multi-class SVM for the spatial level, and lastly a third one for the object level. At each stage, the training focused on the features of that level only, with classification results of the previous level as the input for feature computation.

Although we could rewrite the score function into structural-SVM type [6], we chose to do SVM-based piecewise learning mainly because: (1) a feature vector combining all three levels generated based on the training data would not capture the cascaded nature of higher-level features dependent on the lower levels, thus would not achieve the optimal performance; and (2) our features were designed to be independent between regions at the same level, so optimization for the entire image collectively was not necessary.

A three-level inference based on mean-shift clustering with the three learned multi-class SVMs was then performed. The final labeling was chosen as the class type with the highest combined margin from three levels. The classification could be done per region using SVM, without considering inter-dependencies between regions, because the spatial relationships were derived based on the labeling of the previous level, not within the same level.

2.2 Pathological Context Description

We described the pathological context for the detected tumor (T) and abnormal lymph nodes (N) in three aspects: (1) texture features: the mean, standard deviation, skewness and kurtosis of the Gabor filtered T and N areas for both

CT and PET; (2) shape features: the volume, eccentricity, extent and solidity of T and N; and (3) spatial features: the distance to the chest wall and mediastinum for tumor, and distance to two lung fields for lymph nodes, normalized by the size of the tumor or lymph node itself. The distances were computed in four directions per slice, and averaged across all slices weighted by the detection score S. So, slices with more obvious T or N regions would contribute more to the spatial feature.

Besides extracting the feature vectors of the detected T/N areas, we also extended the context description with an image retrieval component, to retrieve a set of images with similar pathological patterns for a given query image. The retrieved images, which were stored in the database with diagnosis information, could be used to aid image interpretation. Given the query image I and the image J, the distance was defined as:

$$D_{I,J} = \omega \cdot (|v_I - v_J|/(v_I + v_J)) = \omega \cdot v_{I,J} \tag{2}$$

where v was the feature vector of the image (concatenation of the texture, shape and spatial features of T and N), and ω was the feature weights. A training set was constructed of Q triplets: $\langle I, J, K \rangle$, where I was similar to J, and dissimilar to K. It was thus expected to satisfy $D_{I,K} > D_{I,J}$, and the weight vector ω was computed based on the large-margin optimization method [11]:

$$argmin_{\omega,\xi \geq 0} \frac{1}{2}\|\omega\|^2 + C\sum_q \xi_q, \quad s.t. \forall q : \omega \cdot (v_{I,K} - v_{I,J}) \geq 1 - \xi_q \tag{3}$$

The training data $\langle I, J, K \rangle$ captured the search preference, e.g. based on tumor characteristics only, or including lymph nodes. By changing the training data, the derived weights ω would vary and result in different retrievals.

2.3 Materials and Preprocessing

In this study, a total of 1279 transaxial PET-CT image slice pairs were selected from 50 patients with NSCLC. The images were acquired using a Siemens TrueV 64-slice PET-CT scanner (Siemens, Hoffman Estates, IL) at the Royal Prince Alfred Hospital, Sydney. All 50 cases contained primary lung tumors, and 23 of them contained abnormal lymph nodes. The locations of tumors and disease in regional lymph nodes were annotated manually, and for each patient study, the other 49 patient studies were marked similar or dissimilar, as the ground truths. A fully-automatic preprocessing was performed on each CT slice to remove the patient bed and soft tissues outside of the lung and mediastinum, based on simple thresholding, connected component analysis and filling operations. The resulting mask was then mapped to the co-registered PET slice.

3 Results

The structure localization performance for the 50 patient studies is summarized in Table 1a. Based on visual inspections, a volume (case-level) that was classified accurately with its boundary matching closely to the ground truth was considered correct. The multi-level model was trained on slice pairs randomly selected from 10 imaging studies. To further evaluate the localization performance at the slice level, Table 1b shows the measurements for all 1279 slices. We also compared our method with two other approaches (Table 1c and 1d): a four-class SVM for voxel-level classification; and a four-class SVM for region-level classification after mean-shift clustering (identical to our local-level modeling, except training for four classes). Both approaches were trained using the same set of data as our local-level modeling. Our multi-level modeling showed clear advantages, especially in differentiating tumor and abnormal lymph nodes. As a component of our model, Table 1d illustrated the benefit of region-based processing compared to Table 1c. Some visual results are shown in Figure 1.

Table 1. The pairwise confusion matrix of the four region classes tested on 50 patient studies. (a) Our method - image/case level results. (b) Our method - finer slice-level results. (c) Gabor+SVM - image/case level results. (d) Gabor+Mean-shift+SVM - image/case level results.

Ground	Prediction (%)			
Truth	L	M	T	N
Lung lobe	100	0	0	0
Mediastinum	0	94.3	3.8	1.9
Tumor	0	1.6	84.4	14.1
Lymph node	0	3.7	18.5	77.8

(a)

Ground	Prediction (%)			
Truth	L	M	T	N
Lung lobe	99.2	0.8	0	0
Mediastinum	0	97.1	2.1	0.8
Tumor	1.7	6.1	87.8	4.3
Lymph node	0	7.5	12.3	80.2

(b)

Ground	Prediction (%)			
Truth	L	M	T	N
Lung lobe	100	0	0	0
Mediastinum	13.3	60.2	21.7	4.8
Tumor	6.8	10.2	42.4	40.7
Lymph node	5.6	11.1	27.8	55.6

(c)

Ground	Prediction (%)			
Truth	L	M	T	N
Lung lobe	100	0	0	0
Mediastinum	0	83.3	5.0	11.7
Tumor	0	16.5	43.7	39.8
Lymph node	0	6.5	35.5	58.1

(d)

The sensitivity and specificity of tumor/lymph node localization relative to the lung and mediastinum are listed in Table 2a. In testing, the distances between the tumor and the chest wall and mediastinum/hilum, and between the abnormal lymph nodes and the left and right lung lobes, were assessed to determine the sensitivity and specificity. The remaining errors were mainly caused by misclassifications between tumors near the mediastinum and the abnormal lymph nodes. Our method resulted in higher sensitivity and specificity in deriving the spatial relationships, compared to using only local-level features (Table 2b), because of the highly effective structure localization.

Finally, we evaluated the retrieval performance by using each imaging study as the query to retrieve the most similar cases, and the average precision and recall were computed. We compared our method with techniques based on weighted histogram and bag-of-SIFT [12] features for global and local feature extraction; and both approaches were trained in the same way as our method for similarity measure. As shown in Table 3, our method achieved much higher precision and recall. The results showed that our method could extract the salient (pathological) features more effectively than the general techniques; and suggest that the detected context could be used in a CBIR system.

Table 2. The sensitivity (SE) and specificity (SP) of the tumor and lymph node localization relative to the lung lobes and mediastinum tested on 50 cases. (a) Our method and, (b) Gabor+Mean-shift+SVM.

	Tumor		Lymph node	
	Wall	Hilum	Left	Right
SE (%)	100	97.2	92.9	88.9
SP (%)	98.0	84.8	89.4	91.5

(a)

	Tumor		Lymph node	
	Wall	Hilum	Left	Right
SE (%)	83.3	82.9	93.3	87.5
SP (%)	98.0	77.8	65.0	69.2

(b)

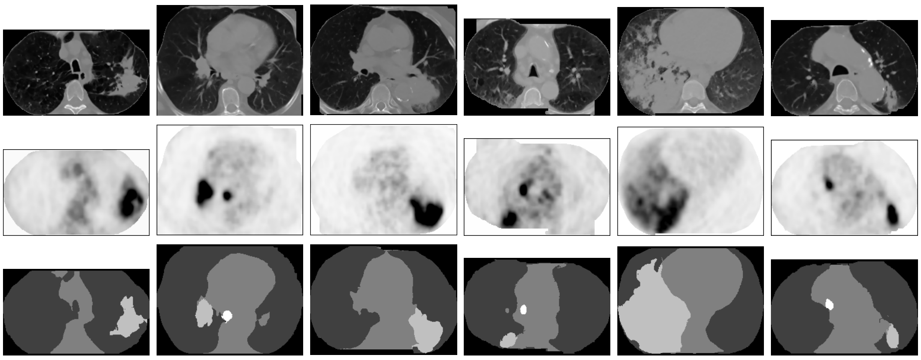

Fig. 1. Six examples of structure localization, showing one transaxial slice pair per case. The top row is the CT image slice (after preprocessing); the middle row is the co-registered PET slice; and the bottom row shows the localization results, with 5 different gray scale values (black to white) indicating background, L, M, T and N.

4 Conclusions

We proposed a new method to automatically detect the primary tumor and disease in lymph nodes, and the spatial relationships with the lung and mediastinum on PET-CT thoracic images. By exploring a comprehensive set of features at the local, spatial and object levels, the discriminative classification achieves an accurate localization of the various structures in the thorax. The work is an initial step towards a computer aided system for PET-CT imaging

Table 3. The precision-recall measure of the retrieval results of the top one, three or five most similar matches on 50 cases. Our method is compared with the histogram (HIST) and bag-of-SIFT [12] features (BoSF) based approaches.

	Precision (%)			Recall (%)		
	Ours	HIST	BoSF	Ours	HIST	BoSF
Top-1	84.0	46.0	32.0	14.1	7.5	6.0
Top-3	70.7	34.7	29.3	31.4	14.1	12.7
Top-5	63.2	28.8	25.6	44.3	19.7	17.9

diagnosis for lung cancer staging. The extracted pathological contexts also show high precision when used to retrieve the most similar images.

References

1. Tao, Y., Lu, L., Dewan, M., Chen, A.Y., Corso, J., Xuan, J., Salganicoff, M., Krishnan, A.: Multi-level ground glass nodule detection and segmentation in CT lung images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 724-731. Springer, Heidelberg (2009)
2. Kakar, M., Olsen, D.R.: Automatic segmentation and recognition of lungs and lesions from CT scans of thorax. Computerized Medical Imaging and Graphics 33(1), 72-82 (2009)
3. Feulner, J., Zhou, S.K., Huber, M., Hornegger, J., Comaniciu, D., Cavallaro, A.: Lymph nodes detection in 3-D chest CT using a spatial prior probability. In: CVPR, pp. 2926-2932 (2010)
4. Wojak, J., Angelini, E.D., Bloch, I.: Joint variational segmentation of CT-PET data for tumoral lesions. In: ISBI, pp. 217-220 (2010)
5. 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)
6. Desai, C., Ramanan, D., Fowlkes, C.: Discriminative models for multi-class object layout. In: ICCV, pp. 229-236 (2009)
7. Galleguillos, C., McFee, B., Belongie, S., Lanckriet, G.: Multi-class object localization by combining local contextual interactions. In: CVPR, pp. 113-120 (2010)
8. Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, LNCS, vol. 3951, pp. 1-15. Springer, Heidelberg (2006)
9. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603-619 (2002)
10. 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(Supplement 1), 211 (2011)
11. Frome, A., Singer, Y., Sha, F., Malik, J.: Learning globally-consistent local distance functions for shape-based image retrieval and classification. In: ICCV, pp. 1-8 (2007)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91-110 (2004)