Similarity Guided Feature Labeling for Lesion Detection

Yang Song[1], Weidong Cai[1], Heng Huang[2], Xiaogang Wang[3], Stefan Eberl[4], Michael Fulham[4,5], and Dagan Feng[1]

[1]BMIT Research Group, School of IT, University of Sydney, Australia
[2]Computer Science and Engineering, University of Texas at Arlington, USA
[3]Department of Electronic Engineering, Chinese University of Hong Kong, China
[4]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Australia
[5]Sydney Medical School, University of Sydney, Australia

Abstract. The performance of automatic lesion detection is often affected by the intra- and inter-subject feature variations of lesions and normal anatomical structures. In this work, we propose a similarity-guided sparse representation method for image patch labeling, with three aspects of similarity information modeling, to reduce the chance that the best reconstruction of a feature vector does not provide the correct classification. Based on this classification model, we then design a new approach for detecting lesions in positron emission tomography – computed tomography (PET-CT) images. The approach works well with simple image features, and the proposed sparse representation model is effectively applied for both detection of all lesions and characterization of lung tumors and abnormal lymph nodes. The experiments show promising performance improvement over the state-of-the-art.

1 Introduction

Automatic lesion detection is highly desirable for computed aided diagnosis. The detection system can be used in early screening or to provide second opinions for decision making. While it is conceptually simple that lesions are just regions with features distinctive from the normal anatomical structures, the detection performance is often hindered by large intra- and inter-subject variations of visual patterns. Such variations are common for both normal anatomical structures and lesions within the same subject or across different subjects.

Lesion detection is usually based on customized feature extraction and classification [6,8]. These classifiers are mainly based on parametric models and work well if there is good feature separation between lesions and normal structures. Complex and domain-specific feature design might be necessary, but could become ineffective for unseen data. Non-parametric classifications, such as multi-atlas and sparse representation methods, have also been recently proposed [4,5,11,3]. The basic principle of both types of approaches can be considered as weighted combination of reference images. While the weights for multi-atlas are normally computed using predetermined formula, the weights in sparse representation are derived by minimizing the reconstruction error.

K. Mori et al. (Eds.): MICCAI 2013, Part I, LNCS 8149, pp. 284–291, 2013.

A potential issue with sparse representation is that, since it is aimed at minimizing the reconstruction errors, it does not necessarily lead to good classification. Various improvements have thus been proposed to incorporate extra constraints into the formulation, such as discriminative labeling [2], group and locality information [10,12], and similarity relationships between references [1]. To better address problem of lesion detection, we design a new similarity-guided sparse representation method for image patch classification. Based on the basic sparse representation, we model the between-reference similarity, similarities between the testing patch and references, and similarities between the testing patch and its neighborhood. The design is motivated by the propositions that 1) to achieve labeling-consistent reconstruction, similar references should get similar weights, and references that are more similar to the testing patch should have higher weights; and 2) neighboring patches should get similar labels if they exhibit similar visual features.

The proposed classification model is common to different application domains. As a case study, in this work, we design a new three-stage approach based on the proposed similarity-guided sparse representation method for lesion detection on FDG PET-CT images of the thorax. The objectives are: 1) to detect different types of lesions; and 2) to characterize a lesion that is detected as a lung tumor or an abnormal lymph node. Compared to the lesion detection method [8], the proposed approach relies on much simpler feature design and uses a single classification model for detection and characterization.

2 Similarity-Guided Sparse Representation

Suppose an image I contains N_I non-overlapping patches, and given that some patches exhibiting typical anatomical features are already labeled (Section 3.1), the objective is to label the remaining patches. Denote the feature vector of an image patch p_i as f_i, with $f_i \in \mathbb{R}^{H\times 1}$. A reference dictionary D_l of class l can be constructed by concatenating the feature vectors of Q_l labeled patches of class l into a matrix: $D_l \in \mathbb{R}^{H\times Q_l}$. To determine the labeling of a testing patch p with feature f, a sparse representation approach can be used, by first deriving the reconstructed feature vectors $\{f'_l\}$ for all classes:

$$x_l = \underset{x_l}{\operatorname{argmin}} \|f - D_l x_l\|_2^2 \;\; s.t. \|x_l\|_0 \leq C; \quad f'_l = D_l x_l \tag{1}$$

where $x_l \in \mathbb{R}^{Q_l\times 1}$ is a weight vector. Then the patch p is labeled as the class with the lowest reconstruction difference: $L(p) = \operatorname{argmin}_l \|f - f'_l\|_2$. The classification performance, however, is often found unsatisfactory, since the linear combination is actually optimized for reconstruction but not classification. To improve the classification performance using sparse representation, we propose a similarity-guided design, which is detailed below.

2.1 Pairwise Reference Similarity

It is natural to expect that visually similar references in a dictionary would preferably contribute similarly to the reconstruction. We thus design a modified sparse reconstruction to obtain similar weights in x_l for similar references:

$$\begin{aligned} x_l &= \text{argmin}_{x_l} \|f - D_l x_l\|_2^2 + \Theta(x_l) \quad s.t. \ \|x_l\|_0 \leq C \\ \Theta(x_l) &= \textstyle\sum_{(a,b),a<b} s(q_a, q_b)|x_l(a) - x_l(b)| \end{aligned} \tag{2}$$

where q_a and q_b denote the feature vectors of two reference patches a and b, $s(q_a, q_b)$ measures the similarity between a and b, and $x_l(a)$ and $x_l(b)$ denote the corresponding weight elements in the vector x_l. The addition of the $\Theta(x_l)$ term helps to encourage similar weights $x_l(a)$ and $x_l(b)$ if q_a and q_b are similar.

To represent the similarity between references, a pairwise distance $d(q_a, q_b)$ between a pair of references is computed: $d(q_a, q_b) = \|q_a - q_b\|_2$. Then based on the normalized distance $\overline{d}(q_a, q_b) \in [0, 1]$, a degree of similarity is derived: $s(q_a, q_b) = \exp\{-\overline{d}(q_a, q_b)\}$ Next, to make Eq. (2) easier to solve, we construct a similarity matrix $U_l \in \mathbb{R}^{0.5Q_l(Q_l-1)\times Q_l}$, with each element defined as [1]:

$$U_l((a,b),k) = \begin{cases} s(q_a, q_b) & \text{if } k = a \\ -s(q_a, q_b) & \text{if } k = b \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where (a, b) and k are the row and column indexes of U_l. Each row of U_l corresponds to a pair of references a and b. Then the $\Theta(x_l)$ term can be rewritten as: $\Theta(x_l) = \|U_l x_l\|_1$. We relax it with L2 norm $\|U_l x_l\|_2^2$ so that Eq. (2) would be easily solvable using orthogonal matching pursuit (OMP) [9].

2.2 Patch Reference Similarity

It is also expected that references that are more similar to the testing patch should be assigned with higher weights, so that it becomes more likely to obtain a good reconstruction only for the correct dictionary. To incorporate the similarity preference between the testing patch and the references, the sparse reconstruction is further modified as:

$$\begin{aligned} x_l &= \text{argmin}_{x_l} \|f - D_l x_l\|_2^2 + \|U_l x_l\|_2^2 + \Phi(x_l) \quad s.t. \ \|x_l\|_0 \leq C \\ \Phi(x_l) &= \textstyle\sum_c \overline{d}(f, q_c) x_l(c) \end{aligned} \tag{4}$$

where c indexes the reference patches and q_c denotes its feature vector, $x_l(c)$ is the corresponding weight element in the vector x_l, and $\overline{d}(f, q_c)$ measures the normalized distance between patch p and the reference c. Here minimization of the $\Phi(x_l)$ term would lead to a smaller weight $x_l(c)$ if $\overline{d}(f, q_c)$ is larger.

Similarly, by defining a distance vector $V_l \in \mathbb{R}^{1\times Q_l}$ with each element of $\overline{d}(f, q_c)$, and relaxing with L2 norm, $\Phi(x_l)$ can be rewritten as: $\Phi(x_l) = \|V_l x_l\|_2^2$.

And the overall sparse reconstruction is thus now defined as:

$$\begin{aligned} x_l &= \text{argmin}_{x_l} \|f - D_l x_l\|_2^2 + \|U_l x_l\|_2^2 + \|V_l x_l\|_2^2 \\ &= \| \begin{pmatrix} f \\ 0^{0.5Q_l(Q_l-1)\times 1} \\ 0 \end{pmatrix} - \begin{pmatrix} D_l \\ U_l \\ V_l \end{pmatrix} x_l \|_2^2 \quad = \|\mathbf{f} - \Omega_l x_l\|_2^2 \quad s.t. \ \|x_l\|_0 \leq C \end{aligned} \tag{5}$$

The OMP algorithm is applied to solve x_l efficiently, and the reconstructed vector is thus: $\mathbf{f}'_l = \Omega_l x_l$, and the labeling is $L(p) = \text{argmin}_l \|\mathbf{f} - \mathbf{f}'_l\|_2$.

2.3 Neighborhood Similarity

Considering that label of a testing patch would be similar to its neighboring patches (if they are visually similar), the collective information of the neighborhood is thus also important. To refine the label based on the neighborhood information surrounding p, a different labeling scheme is designed:

$$L(p) = \underset{l}{\text{argmin}} \|\mathbf{f} - \mathbf{f}'_l\|_2 + \sum_j s(f,g)\gamma(j,l) \tag{6}$$

where j indexes a neighboring patch of p and g denotes its feature vector, $s(f,g)$ measures the degree of similarity between p and j, and $\gamma(j,l)$ is the cost of j labeled as class l:

$$\gamma(j,l) = \left\{ \begin{array}{ll} 0 & \text{if } L(j) = l \\ \|\mathbf{f} - \mathbf{f}'_l\|_2 & \text{if } L(j) \neq l \\ \|\mathbf{g} - \mathbf{g}'_l\|_2 & \text{if } L(j) \text{ unknown} \end{array} \right\} \tag{7}$$

If the label $L(j)$ of the neighboring patch j is already known after the initial labeling (Section 3.1), $\gamma(j,l)$ is then determined based on the reconstruction difference of p. Otherwise, $\gamma(j,l)$ is equal to its own reconstruction difference. In this way, spatial smoothness is encouraged, and the contribution from j is higher if it is more similar to p.

3 Lesion Detection

3.1 Initial Tissue Labeling

In PET-CT images, the lung field normally exhibits low CT densities. Lesions are usually prominent on PET because they display increased FDG uptake. However, some lesions can exhibit relatively low uptake, and non-lesion high uptake regions can also occur in the mediastinum. Such cases might cause incorrect classification between lesions and mediastinal regions. Therefore, in the first step, we would like to label areas that are obviously representative of the lung field (LF), mediastinum (MS) or lesion (LS). The MS and LS areas then serve as references to further classify the unlabeled (UN) areas (Section 3.2).

To do this, we first divide the image into non-overlapping 5×5 voxel patches. An image patch p_i can be confidently categorized as LF if its average CT density is less than a domain-knowledge based value (e.g. 800). Then for non-LF patches, labeling is derived based on its average FDG uptake v: $L(p_i)$ =MS if $v < \alpha(Z)$, or $L(p_i)$ =LS if $v > 2\alpha(Z)$, with $\alpha(Z)$ an image set specific threshold [7]. The remaining patches with $v \in [\alpha(Z), 2\alpha(Z)]$ are thus the UN ones (Fig. 1a).

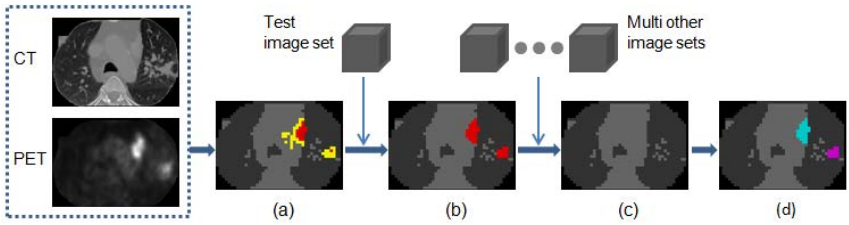


Fig. 1. Method illustration. (a) The initial tissue labeling output, with LF depicted as dark gray, MS as light gray, LS as red and UN patches as yellow. (b) The lesion detection output, derived based on reference dictionaries constructed from the test image set. (c) The labeling output for LF and MS patches, derived based on reference dictionaries constructed from multiple other image sets. (d) The lesion characterization output, with tumor shown as purple and abnormal lymph nodes as blue.

3.2 Intra-image Lesion Detection

In the second step, we further classify the UN patches as LS or MS, so that all true lesions would be detected (Fig. 1b). To do this, first, for each p_i, a 4-dimensional feature vector is computed: its mean and standard deviation of the CT densities, and mean and standard deviation of the FDG uptake. Next, two reference dictionaries $D_{LS}(Z)$ and $D_{MS}(Z)$ are constructed, by concatenating the feature vectors of the labeled LS or MS patches. Note that instead of using the entire database, such patches are gathered from the 3D image set containing p_i only, to avoid inter-image variations. Then, for a UN patch p_i, its labeling $L(p_i) \in \{\text{LS,MS}\}$ is determined using the similarity-guided sparse representation Eq. (6). The logic here is that, if p_i is more similar to the LS patches, it is more likely that p_i is also LS; and similarly for the MS case.

3.3 Inter-image Lesion Characterization

In the last step, the detected lesion is characterized as a tumor or an abnormal lymph node (Fig. 1d). Similar to [8], we estimate what normal anatomical structure (LF or MS) could be present at the lesion location if the subject had been healthy. Then, if this anatomical region is originally LF (or MS), the lesion would be a tumor (or abnormal lymph node). Different from [8], here we use the proposed similarity-guide sparse representation method.

The objective is to relabel each LS patch p_i as LF or MS (Fig. 1c). We consider that at similar spatial locations, collective information from multiple image sets would estimate well the original anatomical structure of p_i. Therefore, two reference dictionaries $D_{LF}(p_i)$ and $D_{MS}(p_i)$ are constructed for p_i, using labeled LF and MS patches at similar locations as p_i but from images excluding the test subject. A patch is considered spatially similar to p_i if the distance between them is less than 10% of the thorax size. In order to reduce the dictionary size for computational efficiency, only 1/5 of the database is used for dictionary construction. Then, with the two reference dictionaries, $L(p_i)$ is derived as LF or MS, using Eq. (6). Note that the x and y coordinates of the patch center are also included in the feature vector for location-based estimation. The patch-wise labels are finally combined by majority voting to classify a lesion object as tumor or abnormal lymph node.

Based on this approach, a small number of detected lymph nodes, however, would actually be tumors (those affecting mostly the mediastinum rather than lung fields) or myocardium (large bright area in the mediastinum). Therefore, for the detected abnormal lymph nodes that are very large, those in the upper left area of the medastinum are filtered as myocardium, and others are marked as tumors. The size criteria are determined based on two-fold cross validation.

4 Experimental Results

Our dataset comprises 50 sets of 3D thoracic FDG PET-CT images from subjects with non-small cell lung cancer, provided by the Royal Prince Alfred Hospital, Sydney. An expert reader of the images annotated 54 lung tumors and 35 abnormal lymph nodes. During preprocessing, the background and soft tissue areas outside of the lung and mediastinum are removed automatically with Otsu thresholding and connected component analysis. The 3D image sets are also aligned in the z-direction based on the location of the carina in the central part of the thorax, to obtain the spatially-similar reference patches (Section 3.3).

Table 1. Patch-level labeling performance. (a) Initial tissue labeling. (b) Lesion detection. (c) LF/MS labeling for detected lesions.

Ground truth	Labeling		
	MS	LS	UN
MS	0.962	0	0.038
LS	0.024	0.405	0.571

(a)

Ground truth	Labeling	
	MS	LS
MS	0.991	0.009
LS	0.011	0.989

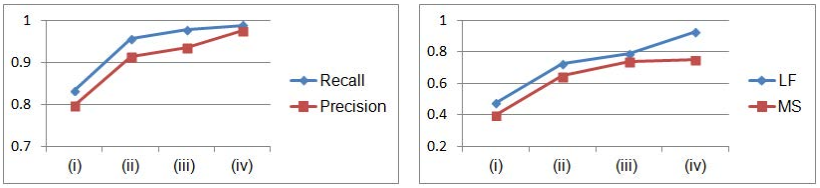
(b)

Ground truth	Labeling	
	LF	MS
LF	0.927	0.073
MS	0.249	0.751

(c)

As shown in Table 1a, after initial tissue labeling, most of the patches that are labeled as MS or LS are indeed MS or LS types. Then, with the intra-image lesion detection, the UN patches are further categorized, and most of the MS and LS patches are now correctly labeled (Table 1b). Mislabeled patches are mainly in areas with relatively high uptake in the mediastinum or where there is lower

uptake than is expected for tumors. Finally, with the inter-image lesion characterization, the detected lesions are categorized as tumors or abnormal lymph nodes. The performance of labeling LS patches as LF or MS is listed in Table 1c. In cases with lymph nodes adjacent to the lung fields, about 1/4 of patches are labeled as LF rather than the expected MS. However, this is usually not a problem for characterizing abnormal lymph nodes, since the majority of patches are correctly labeled. The performance comparison using different constructs of the sparse representation is shown in Fig. 2.



(a) (b)

Fig. 2. (a) Patch labeling performance for LS (Section 3.2). (b) True positive rates of labeling LF and MS (Section 3.3). Comparing (iv) proposed similarity-guided sparse representation with: (i) basic sparse representation, (ii) basic plus pairwise reference similarity, and (iii) basic plus pairwise reference and patch reference similarities.

The performance of object-level detection is listed in Table 2. A lesion object is counted as true positive if at least 60% of its volume is labeled correctly. Some false positive lesions are detected in the mediastinum where there is elevated uptake, and are thus mostly characterized as lymph nodes. This affects the precision of detecting abnormal lymph nodes. The lymph nodes, especially those at the hilum, are sometimes difficult to differentiate from lung tumors and this then affects the recall of lymph nodes and precision of detecting tumors. Our results are overall better than the state-of-the-art [8]. On a standard PC with a Matlab implementation, the detection method takes on average 50s per 3D PET-CT image. Compared to [8], this method is more efficient without requiring additional structure delineation.

Table 2. The object-level detection performance for tumors and abnormal lymph nodes, compared with state-of-the-art [8].

	Tumor	Node	Tumor [8]	Node [8]
Recall (%)	96.3	91.4	90.7	88.6
Precision (%)	92.9	86.5	89.1	88.6

5 Conclusion

In this work, we present a new similarity-guided sparse representation model for patch-wise feature classification, incorporating the pairwise reference similarity, patch reference similarity and neighborhood similarity. Based on this model, we then design a new three-stage approach to detect lung tumors and abnormal lymph nodes from thoracic FDG PET-CT images. Our method tackles the challenges caused by intra- and inter-subject variations effectively, and achieves promising performance improvement. In future work, we will investigate if more comprehensive feature design helps to improve the detection performance.

References

1. Han, Y., Wu, F., Shao, J., Tian, Q., Zhuang, Y.: Graph-guided sparse reconstruction for region tagging. In: CVPR, pp. 2981–2988 (2012)
2. Jiang, Z., Lin, Z., Davis, L.S.: Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: CVPR, pp. 1697–1704 (2011)
3. Liao, S., Gao, Y., Shen, D.: Sparse patch based prostate segmentation in CT images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 385–392. Springer, Heidelberg (2012)
4. Liu, M., Lu, L., Ye, X., Yu, S., Salganicoff, M.: Sparse classification for computer aided diagnosis using learned dictionaries. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 41–48. Springer, Heidelberg (2011)
5. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Graph-based detection, segmentation & characterization of brain tumors. In: CVPR, pp. 988–995 (2012)
6. van Ravesteijin, V.F., van Wijk, C., Vos, F.M., Truyen, R., Peters, J.F., Stoker, J., van Vliet, L.J.: Computer-aided detection of polyps in CT colonography using logistic regression. IEEE Trans. Med. Imag. 29(1), 120–131 (2010)
7. 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(supp. 1), 211 (2011)
8. Song, Y., Cai, W., Zhou, Y., Feng, D.: Thoracic abnormality detection with data adaptive structure estimation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 74–81. Springer, Heidelberg (2012)
9. Tropp, J.: Greed is good: algorithmic results for sparse approximation. IEEE Trans. Inf. Theory 50, 2231–2242 (2004)
10. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR, pp. 3360–3367 (2010)
11. Wolz, R., Chu, C., Misawa, K., Mori, K., Rueckert, D.: Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 10–17. Springer, Heidelberg (2012)
12. Xu, D., Huang, Y., Zeng, Z., Xu, X.: Human gait recognition using patch distribution feature and locality-constrained group sparse representation. IEEE Trans. Image Process. 21(1), 316–326 (2012)