A RANKING-BASED LUNG NODULE IMAGE CLASSIFICATION METHOD USING UNLABELED IMAGE KNOWLEDGE

Fan Zhang[1], *Yang Song*[1], *Weidong Cai*[1], *Yun Zhou*[2], *Michael Fulham*[3,4], *Stefan Eberl*[3], *Shimin Shan*[5], *Dagan Feng*[1]

[1]BMIT Research Group, School of IT, University of Sydney, Australia
[2]The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, USA
[3]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Australia
[4]Sydney Medical School, University of Sydney, Australia
[5]School of Software, Dalian University of Technology, China

ABSTRACT

In this paper, we propose a novel semi-supervised classification method for four types of lung nodules, *i.e.*, *well-circumscribed*, *vascularized*, *juxta-pleural* and *pleural-tail*, in low dose computed tomography (LDCT) scans. The proposed method focuses on classifier design by incorporating the knowledge extracted from both training and testing datasets, and contains two stages: (1) bipartite graph construction, which presents the direct similar relationship between labeled and unlabeled images, (2) ranking score calculation, which computes the possibility of unlabeled images for each of the given four types. Our proposed method is evaluated on a publicly available dataset and clearly demonstrates its promising classification performance.

Index Terms— Lung nodule, classification, bipartite graph, ranking score.

1. INTRODUCTION

Automatic differentiation of potentially malignant lung nodules from the benign ones is highly desirable for early detection of lung cancers. Normally, the nodules that are near surrounding anatomical structures are more likely to be benign than those solitarily located in the lung [1]. Therefore, lung nodules are divided into four different types according to their relative locations [2]: well-circumscribed nodule (W) located centrally in the lung without any connection to other structures; vascularized nodule (V) is also central in the lung, but closely attached to the neighbouring vessels; juxta-pleural nodule (J) has a large portion connected to the pleura; and pleural-tail nodule (P) is near the pleural surface connected by a thin tail. Sample images are shown in Fig.1, with the nodule encircled in red.

Image-based computerized diagnosis in clinical settings has focused on establishing the relationship quantitatively among various lung nodules based on their feature descriptions, and thus can automatically classify them into the proper categories [3]. Manual interpretation of the lung nodule images by the radiologists, which is highly dependent on their levels of expertise, and mental and physical status [4], can be error-prone. Therefore, automatic classification is highly desirable, providing a "second opinion" for the decision making. Hence, the aim of this work is to classify the lung nodules into the above four types automatically.

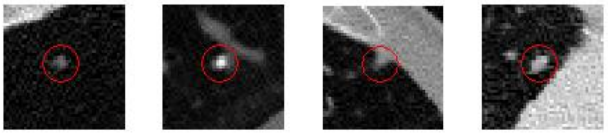

Fig. 1. Sample images from the four categories with well-circumscribed, vascularized, juxta-pleural and pleural-tail cases from left to right respectively.

Image classification in medical or general imaging normally contains two major stages: feature extraction and classification. Considering that the overlapping spatial structure of the lung nodule image content can lead to large intra-type and small inter-type variations [5, 6], even with the comprehensive feature design, the current classifiers still show their deficiencies on classifying the nodule images precisely. Usually, feature-based image classification is based on supervised learning approaches [7, 8, 9], such as SVM and *k*-nearest neighbour (*k*-NN) which have been used in lung nodule classification [3, 10]. However, such methods, which predict the unlabeled images merely based on the information from labeled items, can be error prone due to large intra-variation feature spaces resulting from the complex structure of the nodules. In particular, new unlabeled images sometimes struggle to form a sufficient

similarity with the labeled ones in the same category; however, they might be associated with them through other unlabeled images indirectly. For example, suppose there are three images *A*, *B* and *C* in the same category with *A* already labeled and the other two not. Labeling *C* directly based on *A* would fail if *C* is not similar to *A*, but would work if *C* is highly related to *B* which is similar to *A*. Hence, better performance can potentially be obtained by extracting such indirect associations from labeled and unlabeled images.

In light of above, we designed a new classification method in a semi-supervised way exploring the information from unlabeled images. The proposed method contains two stages: first, a bipartite graph is constructed between the training and testing datasets (with labeled and unlabeled images respectively) for each type of nodules based on the feature similarities; second, a ranking method is designed to calculate the probability of unlabeled images given certain type based on the bipartite graph. The designed ranking method works on two rules in an iterative way. The two rules define the ranking score of unlabeled image, which indicates its possibility belonging to this type, and the contribution score of labeled image, which indicates its contribution degree for the decisions of unlabeled images.

2. METHODS

2.1. Dataset

In this study, we use the publicly available Early Lung Cancer Action Program (ELCAP) [11] database for experiments. The ELCAP database contains 50 sets of low-dose CT lung scans with 379 unduplicated lung nodules annotated at the centroid. The nodules are further classified into well-circumscribed (W-15%), vascularized (V-16%), juxta-pleural (J-30%) and pleural-tail (P-39%) by an expert reader of the images. In the experiments, various percentages (10% to 90%) of training datasets were randomly selected from for each of the four categories, and the remaining were used for testing. The selection of various sizes of training datasets is to make sure that the proposed method can work even in adverse circumstances.

2.2. Bipartite graph construction

At the first stage, we aim to present the direct similar relationship between training and testing datasets. Denote the whole training dataset as $Dtr = \{Dtr(tp) : tp = \{w, v, j, p\}\}$ and the testing dataset as *Dte*. For each sub-training dataset $Dtr(tp)$, a bipartite graph $BG_{Dtr(tp)\text{-}Dte}(I_i, I_j)$ is built according to the feature similarities with the images in the testing dataset, where image $I_i \in Dtr(tp)$ and image $I_j \in Dte$, and $BG(I_i, I_j)=1$ if I_j belongs to top *N* highest similar neighbors of I_i; otherwise, $BG(I_i, I_j)=0$. The overall description is shown in Fig.2.

Specifically, feature vector is first extracted to describe the image numerically, and further to calculate the similarity between different images. For our classifier, without over-emphasizing on the feature design, we simply adopt the scale-invariant feature transform (SIFT) algorithm followed by SVM probability estimates given the four types of nodules, which has been proven effective to capture the similarity relationship for lung nodule classification [5]. The feature extraction step forms a 4-length feature vector for each image *I* with each element indicating the probability given certain type, and cosine value between two vectors is computed for each pair of images across the whole dataset to obtain the similarity network. Then, for each image I_i in $Dtr(tp)$, the first *N* images in testing dataset *Dte* with the highest similarities are selected, making $BG_{Dtr(tp)\text{-}Dte}(I_i, I_j)=1$. From our experiments, selecting the first 23% of the most similar testing images proved to generate the best performance (introduced in Section 3).

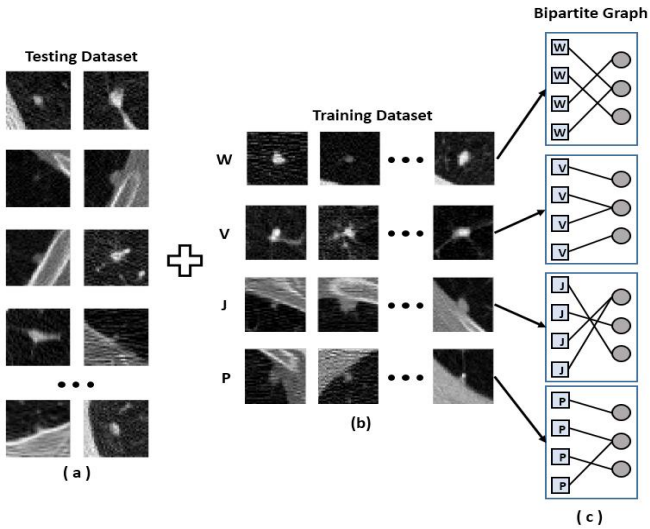

Fig. 2. Bipartite graph generation: (a) the testing dataset, (b) the training dataset, which contains four sub-datasets representing type W, V, J and P respectively, (c) the obtained bipartite graph for each type. The line connecting training image I_i (rectangle) and testing image I_j (circle) indicates $BG(I_i, I_j)=1$.

2.3. Ranking score calculation

The second stage of our method is to calculate the ranking score for each image in testing dataset given certain sub-training dataset $Dtr(tp)$ using the relationship obtained from the bipartite graph. The ranking score of testing image I_j is regarded as the possibility of I_j belonging the type *tp*. In addition, the contribution score of training image $I_i \in Dtr(tp)$ is also computed to indicate the contribution of I_i to the decision of testing images belonging to type *tp*. Instead of ranking the testing images merely based on the similarities with training items, knowledge existed in the testing dataset is also incorporated.

Specifically, for $I_j \in Dte$ and $I_i \in Dtr(tp)$, the ranking score $S_{COtp}(I_j)$ and contribution score $Con_{tp}(I_i)$ are computed based on the following two rules, given the type tp:

(1) $Sco_{tp}(I_j)$ would be higher if I_j is similar to lots of training images with high Con;
(2) $Con_{tp}(I_i)$ would be higher if I_i is similar to lots of testing images with high Sco.

The key of classifier design is to identify the most related items for the unlabeled image. Although the bipartite graph construction has already provided a relatively coherent collection by eliminating some irrelevant items, the extracted items are still not reliable to determine the category of unlabeled image. In particular, the classification conducted directly based on these extracted similar items would result in mislabeling because part of them might come from other categories due to the small inter-type variations. The designed two rules work twofold to reduce the influence of these labeled items: rule (1) is to determine potential category of unlabeled images based on the most related identified labeled images, and rule (2) is to further re-rank these labeled images based on the preferences of unlabeled images. Hence, their influence in terms of the contribution score would be reduced through the above two rules due to the fact that they are not highly similar with other images that belong to same category with the unlabeled image.

The similarity relationship between training and testing images is represented by the obtained bipartite graph; thus, the above two rules can be formally expressed as the following two operations, working in a circular manner:

$$\text{Operation 1: } Sco_{tp}(I_j) = \sum_{I_i \text{ is connected to } I_j} Con_{tp}(I_i) \quad (1)$$

$$\text{Operation 2: } Con_{tp}(I_i) = \sum_{I_j \text{ is connected to } I_i} Sco_{tp}(I_j) \quad (2)$$

Inputs: bipartite graph $BG_{Dtr(tp)\text{-}Dte}$, number of iteration T
Outputs: the ranking score of all testing images S_{COtp}
Steps:
1. Initializing Con_0 with 1 for all training images.
2. for t = 1:T
   a. Compute Sco'_t using operation 1 with Con_{t-1};
   b. Compute Con'_t using operation 2 with Sco'_t;
   c. Obtain Sco_t by normalizing Sco'_t;
   d. Obtain Con_t by normalizing Con'_t;
   end for
Return: Sco_T

Fig. 3. The procedure of the proposed ranking method. The *Con* and *Sco* would be convergent if T is large enough; however, T = 20 proved to be efficient to obtain a sufficiently accurate result in our experiment.

We make use of the bipartite relationship through an iterative algorithm that maintains and updates numerical weights for images from both two datasets according to Eq. (1) and Eq. (2). The procedure of the algorithm is shown in Fig.3. The proposed algorithm can associate the testing images together so that the knowledge from them can be further utilized. For testing image I_j, $Sco(I_j)$ is calculated according to the other testing images through the bridge of the training images. Formally, $Sco_t(I_j)$ is computed from $Con_{t-1}(I_i)$ which was derived from Sco_{t-1} in the last iteration. Therefore, the knowledge from the testing images, represented by the temporary ranking scores Sco_t ($t<T$), is incorporated with the knowledge from training images, represented by the direct similar relationship in the bipartite graph, for the final decision Sco_T. The final label of image I_j is determined by selecting the type $tp \in \{w, v, j, p\}$ which makes $Sco_{tp}(I_j)$ maximum.

3. RESULTS

We evaluate the performance of the proposed method by computing the overall classification rate, *i.e.*, the ratio of correctly classified images out of the whole dataset. Throughout the experiments, parameter T, the number of iteration, was set at 20, which sufficiently converged both *Sco* and *Con*.

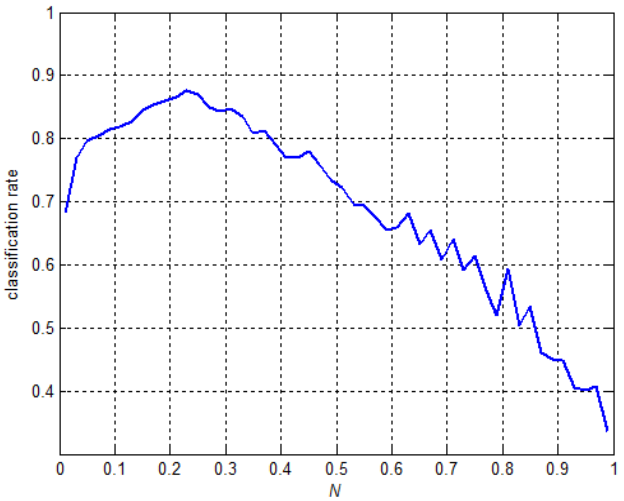

Fig. 4. Effect of parameter N. The peak value is obtained when N is around 23%. The classification rates are relatively high when the most similar testing images are incorporated (the first 23%); however, it keeps going down significantly when further relatively dissimilar testing images are incorporated (the remaining after 23%).

Firstly, the influence of how many images are selected for constructing the bipartite graph is discussed, which is determined by parameter N. The average classification rate for each N, in terms of the ratio of selected similar testing items out of all testing images, across all training dataset percentages (20% - 90%) is shown in Fig. 4. We observe that incorporating too few or too many testing images into the bipartite graph results in lower performance, which arrives at the peak when N is around 23%. For our proposed method, the ranking score is highly dependent upon how many similar testing images are incorporated for constructing the bipartite graph; thus, too few testing images

would lead to the loss of important information and too many would involve noise, both of which would reduce the performance. Therefore, throughout the whole experiments, we fixed *N* at the 23%.

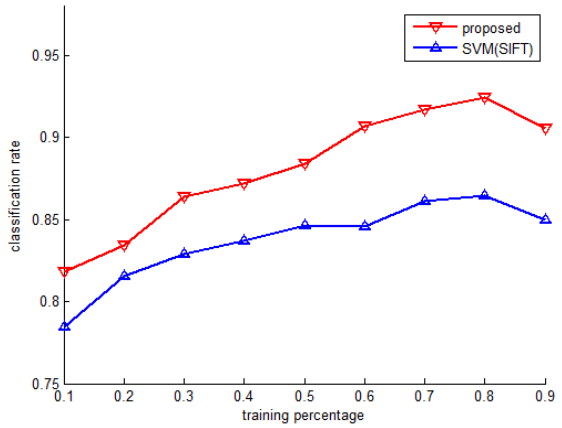

Fig. 5. The comparison between the proposed classifier and the SVM classifier based on the same SIFT descriptor.

In addition, the improvement from the proposed classifier is shown in Fig.5, comparing to the SVM classifier. Based on the same feature descriptor, *i.e.*, SIFT, the proposed approach outperforms SVM for all training datasets percentages. This illustrates that our proposed method's improved performance in finding similar items for a range of circumstances. Furthermore, it can be observed that greater enhancements are derived when more images are incorporated for training, indicating that the proposed method could make the more accurate classification with more information.

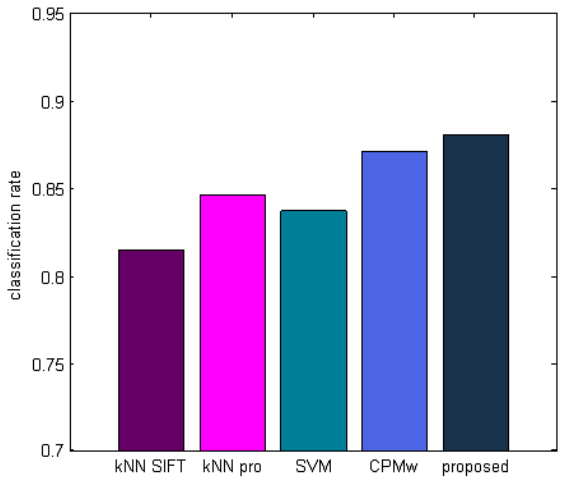

Fig. 6. The overall classification rates comparing various methods.

The evaluation for the proposed method regarding to other classifiers is shown in Fig. 6. The comparisons were conducted among the following approaches: *k*-NN based on raw SIFT descriptor, *k*-NN based on SVM probability estimates, SVM based on raw SIFT descriptor, the weighed clique percolation method (CPMw) upon SVM probability estimates [5] and the proposed method. The value shown in Fig.6 is the average classification rate for each method across all training datasets (10% - 90%). It is apparent that our proposed method obtains the best performance, and an average 6% more of images are correctly classified comparing to the other three approaches. Therefore, it suggests that incorporating knowledge from the unlabeled images can further improve the performance of lung nodule image classification.

4. CONCLUSIONS

In this paper, we propose a novel method for lung nodule image classification into four types, *i.e.*, well-circumscribed, juxta-vascular, juxta-pleural and pleural-tail. The bipartite graph is firstly constructed for a certain sub-training dataset according to the similarities with the testing images. Then, a new ranking algorithm is designed to calculate the possibility of testing image given certain type of nodule by incorporating the knowledge from the testing dataset. Our method was evaluated using ELCAP database with a larger number of correctly classified images, which shows its promising performance for lung nodule image classification.

5. REFERENCES

[1] A. A. Farag, "A Variational Approach for Small-Size Lung Nodule Segmentation," in *ISBI*, pp. 81-84, 2013.

[2] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, et al., "3-D segmentation algorithm of small lung nodules in spiral CT images," *IEEE Trans. Information Technology in Biomedicine*, vol. 12, pp. 7-19, 2008.

[3] Y. Song, W. Cai, Y. Wang, and D. Feng, "Location classification of lung nodules with optimized graph construction," in *ISBI*, 2012, pp. 1439-1442.

[4] D. Unay and A. Ekin, "Dementia diagnosis using similar and dissimilar retrieval items," in *ISBI*, 2011, pp. 1889-1892.

[5] F. Zhang, W. Cai, et al., "Overlapping Node Discovery for Improving Classification of Lung Nodules," in *EMBC*, 2013, pp. 5461-5464.

[6] F. Zhang, et al, "Context Curves for Classification of Lung Nodule Images", in *DICTA*, 2013, pp. 185-191.

[7] Y. Song, W. Cai, et al., "Feature-based Image Patch Approximation for Lung Tissue Classification," *IEEE Trans. Medical Imaging*, vol.32, no.4, pp.797-808, 2013.

[8] Y. Song, et al, "Discriminative Pathological Context Detection in Thoracic Images based on Multi-level Inference", in *MICCAI*, 2011, pp191-198..

[9] Y. Song, et al, "A Multi-Stage Discriminative Model for Tumor and Lymph Node Detection in Thoracic Images", I*EEE Trans. Medical Imaging*, Vol.31, No.5, pp1061-1075, 2012.

[10] A. Farag, S. Elhabian, J. Graham, et al., "Toward precise pulmonary nodule descriptors for nodule type classification," in *MICCAI*, 2010, pp. 626-633.

[11] "ELCAP public lung image database, " http://www.via.cornell.edu/databases/lungdb.html.