A Similarity Measurement Method for Diffuse Lung Disease CT Slice Image Retrieval

Mengjiao Wang; Fujitsu R&D Center CO.,LTD.; Beijing, China
Rujie Liu; Fujitsu R&D Center CO.,LTD.; Beijing, China

Abstract

Diffuse lung disease (DLD) is difficult to diagnose due to the ambiguity of disease patterns, which motivates the development of image retrieval method to facilitate the physicians in diagnosis by retrieving the similar cases from database. In this paper, we propose a similarity measurement method for diffuse lung disease computed tomography (CT) slice image retrieval. In our method, the DLD patterns and the spatial distribution of the diseased area are both integrated to compute the similarity between query and database image. For this purpose, the powerful GoogLeNet is adopted and fine-tuned to locate the diseased area and classify it into different DLD patterns. Moreover, the spatial distribution of the diseased area is calculated based on the distance to the body center. Our method is verified on 324 CT slice images obtained from 53 subjects. The correct ratio among the top-5 retrieved images achieved 86.2%. Based on this performance, we can draw the conclusion that this method has high potential to improve the efficiency for diagnosis of diffuse lung disease in clinical use.

Keywords — Similarity Measurement, Diffuse Lung Disease, CT Slice Image, Image Retrieval, GoogLeNet

1. Introduction

The high ambiguity between diffuse lung disease patterns causes difficulty for diagnosis. Fortunately, the image retrieval procedure can assist the physician for diagnosis and improve the efficiency and accuracy. The key step of image retrieval system is to calculate the similarity between query and database images. The similarity calculation is normally based on the information of disease type and disease area location, which can be obtained by a pre-trained classifier. Afterwards, the top ranked similar database images are retrieved to assist the physician for diagnosis.

For the pre-trained classifier, previous literature has proposed methods using local discrete cosine transform following with random forest [1], local binary pattern with KNN [2], and intensity textons with support vector machine [3], etc. In the recent years, image classification accuracy has been dramatically improved by deep convolutional neural network (DCNN). To train a DCNN model, a large amount of annotated data is required. However, in medical application, the annotated image data is usually very rare, which can easily cause over-fitting problem during training process. Therefore, in our application, we used a transfer learning strategy by fine tuning the pre-trained DCNN model to take advantage of the feature representing power of existing convolutional neural network and to avoid over-fitting problem.

For similarity calculation, the common methods are usually based on the overlapping area ratio between the diseased areas in query and database images. This method doesn't consider the spatial distribution of diseased area, which indicates the important pathogenesis information of diffuse lung disease. One example is shown as Fig. 1. In Fig. 1, the red block indicates the diseased area. In this figure, we can see that the overlapping area ratio between

query and two database (DB) images are roughly the same. However, the spatial distributions of diseased area in DB images are quite different. In Fig. 1, the diseased area in query and DB image I are all located in peripheral lung region, while the area in DB image II spread across the lung region. In fact, based on spatial distribution information, DB image I should have higher ranking order than DB image II. This will be explained later.

Fig. 1 Example for Spatial Distribution

In order to generate more reasonable retrieval result, we proposed a similarity measurement that integrates the spatial distribution information. The spatial distribution can be represented by the distance of diseased area to the body center, which can be calculated as follows: firstly, in order to represent the relative location in lung region, the lung region mask is extracted from the slice image; secondly, the extracted lung region is divided into three sub-regions (central area, peripheral area and middle area) based on the distance to the body center to represent the spatial distribution; thirdly, the similarity is calculated using sub-region weighted bi-directional distance for the subsequent ranking and retrieval.

This paper mainly consists of three parts: firstly, the detailed methodology of our proposed similarity measurement method was introduced; then, the results were evaluated to prove the effectiveness of our method; finally, the performance was concluded in the last part.

2. METHODOLOGY

The methodology consists of three parts: A. discriminant classifier design; B. spatial distribution calculation, and C. similarity measurement.

2.1 Discriminant Classifier Design

In our proposed method, we used a deep convolutional neural network (DCNN) as our discriminant classifier. To apply the DCNN in diffuse lung disease images, three major techniques are feasible: (1) Training the DCNN from scratch; (2) using "off-the-shelf" DCNN models; (3) fine-tuning the pre-trained DCNN model on natural image using the diffuse lung disease images. Due to the small image number of our diffuse lung disease image database, method (1) and (2) will easily over-fit. Therefore, we used the method (3) in our method, i.e. fine-tuning the pre-trained DCNN model.

Here, we used GoogLeNet[5] model for fine-tuning. GoogLeNet is the state-of-art model, which achieved better result on the famous ImageNet Large-Scale Visual Recognition Challenge. Another reason why we choose GoogLeNet is that it

offered the pre-trained model parameters, which makes it lot easier to fine-tune the model.

For fine-tuning procedure, all CNN layers except the last one were fine-tuned at a learning rate 10 times smaller than the default learning rate, since the parameters are very close to the optimal value. The last fully-connected layer was random initialized and freshly trained, in order to accommodate the new object categories in our diffuse lung disease applications. The CNN network was retrained end-to-end starting with the pre-trained model parameters.

The intensity of original pixel value is signed 16-bit integer with range roughly between -2000 ~ 4000. However, the GoogLeNet can only take unsigned 8-bit integer with range of 0~255. To map the original pixel value into the range of 0~255, we added a window to the original value. The window center and window width are -600 and 1500 respectively.

During the classifier training process, we divided the slice images into small image patches using mesh grids to make it easier for practical use. The dimension of each image patch is 16*16. The whole retraining procedure was performed on the annotated image patches, which contains 5 types: Consolidation (CON), GGO, HoneyComb (HON), Emphysema (EMP), and normal(NOR). Afterwards, the fine-tuned DCNN classifier model was applied on slice images to locate the diseased image patches and output the disease type information.

2.2 Spatial distribution calculation

The spatial distribution calculation consists of two major steps: lung region extraction, and lung sub-region dividing based on the distance to body center. The spatial distribution can be represented by the location of diseased area in different sub-regions. The two steps will be demonstrated in details in the following parts.

2.2.1 Lung Region Extraction

For diffuse lung disease CT images, the lung region can be separated as normal region and abnormal region. Each type of region can be extracted separately, and combined afterwards as the final lung region mask.

For the normal lung region, in computed tomography (CT) slice images, the normal lung tissue have intensity within -950 ~ -310. Therefore, a threshold method was used here to extract the normal region.

For the abnormal lung region, it can be categorized as two types: diffuse abnormal area (i.e. the abnormal area without obvious boundary, as shown in a.1 in Fig. 2), and solid abnormal area (i.e. the abnormal area with obvious boundary, as shown in the b.1 in Fig. 2).

For the diffuse abnormal area, we can see from a.1 in Fig. 2 that it usually contains rich texture information (frequently spatial change of pixel intensity) comparing to other lung region. Therefore, the diffuse abnormal area can be extracted using this texture information. The gray level co-occurrence matrix (GLCM) is a well-established tool to represent the spatial distribution of pixel intensity by storing the co-occurrence frequencies of the pairs of gray levels. Afterwards, the Haralick feature[4] is usually applied to extract different information from the GLCM matrix. Here, we found that among 28 Haralick features, sum of entropy can emphasize the small changes of pixel intensity in local region, which can highlight the diffuse abnormal region. Therefore, a threshold procedure can be applied on the sum of entropy image to extract the diffuse abnormal region.

For the solid abnormal area, it mainly appears at the peripheral boundary of the lung contour, which will cause 'caves' in the convex lung contour. Therefore a method using convex hull was applied to extract the solid abnormal area. Firstly, the convex

hull of normal region was calculated. Secondly, the solid abnormal area was obtained by subtracting the normal area from the convex hull area, and eliminating the largest isolated part which corresponds to the area that is surrounded by the two lung region.

Finally, after the normal area, diffuse and solid abnormal area were obtained, these parts can be combined as the final lung region.
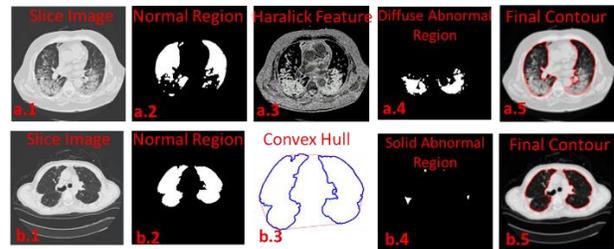
Fig. 2 Illustration for Lung Region Extraction

2.2.2 Lung Sub-region Dividing

The basic idea of lung sub-region dividing is to segment the sub-regions of lung area according to the distance to the body center. The detection of body center was as follows: Firstly, the spine region was extracted by detecting the largest isolated region in bony structure. The central column of spine region will be the column index of body center. Secondly, the bony structure located in the region that is above spine region and bounded by the leftmost and rightmost column of spine was detected. Finally, the middle row of spine's upper bound and detected bony structure's lower bound will be the row index of body center.

After the body center was located, we can divide the lung region into three sub-regions based on the distance to body center. The three sub-regions include: central area (red area in Fig. 3), peripheral area (green area in Fig. 3), and middle area (yellow area in Fig. 3). The central area is the intersection area of lung region and the disk area around body center. The peripheral area can be defined as follows: (1) create the line connecting lung region centroid and body center; (2) find the line which is parallel to the created line and is tangent to the lung region contour; (3) extract the part of lung contour that lies between the two tangent lines, and move the contour along the normal vector for a certain distance to determine the peripheral area. The middle area can be obtained by subtracting the previous two sub-regions from the whole lung region obtained by the first step.
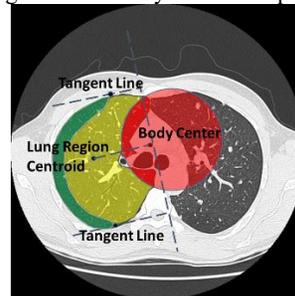
Fig. 3 Illustration of lung sub-regions dividing

2.3 Similarity Metric Design

The principle of similarity metric should satisfies: (1) the same disease type images have the highest similarity; (2) points within each sub-region are close, and points between two sub-

regions are far apart. Only in this way, the disease type information and spatial distribution information can be integrated into the similarity metric.

In this paper, we proposed a similarity metric called sub-region weighted bi-directional distance. The illustration was shown in Fig. 4.
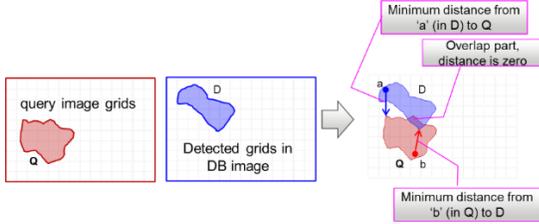
Fig. 4 Illustration for similarity measurement

The similarity metric can be formulated as Eq. 1. In Eq. 1, the smaller v is, the higher the similarity is. Here, Q and DB presents the query and database respectively, and $j_{DB} = \underset{j \in DB}{\operatorname{argmin}} \|P_i - P_j\|$,

$i_Q = \underset{i \in Q}{\operatorname{argmin}} \|P_j - P_i\|$.

To calculate similarity metric, first, the minimum distance between query and database image patches were computed. If the two patches are with same disease type, the distance will be the actual calculated value. Otherwise, it will be set as a fixed large value $MaxDis$ in Eq. 1. Then, the sub-region information was integrated by introducing weight w. If the pair of image patches are located in same sub-region, $w = 1$; if they are located in adjacent sub-regions, $w = 2$; and so on.

Using this method, we can make sure that the images with same disease type and same spatial distribution will have smaller value, which represents higher similarity. Let's revisit the case in Fig. 1. Using our proposed similarity metric, the DB_I will have higher similarity than DB_II. The result is more reasonable than using the overlapped area ratio.

$$v = \sum_{i=1}^{N_Q} w_Q^i \min_{j \in DB} \|P_i - P_j\| + \sum_{j=1}^{N_{DB}} w_{DB}^j \min_{i \in Q} \|P_j - P_i\| \quad (1)$$

$$w_Q^i = \begin{cases} \|P_i - P_{j_{DB}}\| \cdot w & i, j_{DB} \text{ with same disease type} \\ MaxDis \cdot w & i, j_{DB} \text{ with different disease type} \end{cases}$$

$$w_D^j = \begin{cases} \|P_j - P_{i_Q}\| \cdot w & i_Q, j \text{ with same disease type} \\ MaxDis \cdot w & i_Q, j \text{ with different disease type} \end{cases}$$

3. RESULT

3.1 Discriminant Classifier Evaluation

In our paper, we used 39010 annotated image patches to evaluate the discriminant classifier. The image patches are separated into training and testing sets in patient-wise manner, i.e. the training and testing sets are from different patients. The reason is that in practical use, the query image is very unlikely obtained from same patient with database image. The number for each type is shown in Table I.

The accuracy for fine-tuned GoogLeNet is shown in Table II. The accuracy for each disease type was evaluated. Here, we use the average accuracy of our classifier, which is 86.9%. The comparison with reference [6], [7], and [8] was shown in Table III. From this table, we can see this result is better than the previously

reported accuracy. In our opinion, this accuracy is sufficient for practical use.

Table I: Number of training and testing image patches for patient-wise separation

Disease Type	CON	GGO	HON	EMP	NOR
Training Number	3349	3088	2864	3583	14776
Testing Number	884	3437	786	2131	4112

Table II: Accuracy for each disease type, 'AVG' means average

Disease Type	CON	GGO	HON	EMP	NOR	AVG
Accuracy	0.934	0.773	0.910	0.828	0.902	0.869

Table III: Accuracy comparison with previously reported method

Method	Accuracy
AlexNet + SVM[8]	76.7%
GoogLeNet Fine Tuning[6]	81.7%
Home-brewed DNN model[7]	85.6%
Our Method	86.9%

3.2 Spatial Distribution Calculation Evaluation

The spatial distribution calculation consists of two steps: lung region extraction and lung sub-region dividing. The evaluation will be performed on these two steps respectively.

3.2.1 Lung Area Extraction Evaluation

To evaluate the lung area extraction method, we used 1963 thoracic slice image to test the performance of method. The proposed lung area extraction was applied for each image. The extracted region was evaluated by visual observation and categorized into four quality types, including 'Good', 'Small part missing', 'Large part missing', and 'bad'. The criteria for defining the quality type are: Good: the lung region is perfectly extracted; Small Part Missing: most of the lung region is extracted, with small insignificant part missing; Large Part Missing: large part of lung region is missing when applying extraction procedure; Bad: the lung region extraction generates bad result, e.g. the lung region wasn't recognized at all, or the non-lung area was extracted as lung region.

For the purpose of image retrieval, the small part missing result will not significantly change the extracted feature [9]. Therefore, these types of 'Good' and 'Small part missing' result are both of sufficient quality. With this criteria, the sufficient lung area extraction results were 1878 out of 1963, with up to 95.7% of the total number.

3.2.2 Sub-region Dividing Evaluation

For the lung sub-region dividing, the idea of objective evaluation still applies. We employed the lung sub-region dividing method for 324 images. The dividing result were categorized into three types: 'Good', 'Acceptable', and 'bad'. Since the result of lung sub-region dividing mainly depends on the recognition of body center and the result of lung area extraction, the criteria for defining these three types' results are made based these two conditions: Good: the body center and lung region are both correctly detected, the three sub-regions are divided correctly as well; Acceptable: the body center is slightly mis-located, or the lung region extraction result is slightly different with the ground truth. But the sub-region dividing is still reasonable; Bad: the body

center or/and the lung region is totally mis-detected, which generates wrong sub-region dividing result.

The slightly imperfect result doesn't cause significant change of the result. Therefore, the first two types, i.e. 'Good' and 'Acceptable', are both sufficient for our use. The sufficient lung sub-region dividing rate were 293 out of 324, with up to 90.4% of total number.

3.3 Similarity Ranking Evaluation

To test the similarity ranking result, we selected 384 images for the evaluation. From these 384 images, 100 images were selected as query image. In each experiment, we selected one single query image and make the rest 383 images as database images. Therefore, 100 experiments were conducted in total. The top-5 ranked database images in each experiment were used to classified as 'correct' or 'wrong'. The criterion to define 'correct' and 'wrong' is that the top ranked database images should have the same dominant disease type and same spatial distribution of abnormal block with the query image. If the top ranked database image meets this criterion, it will be consider as 'correct', otherwise, it will be considered as 'wrong'. The number of the correct case in the similarity ranking experiment is 431 out of 500 results, up to the rate of 86.2%. To the best of our knowledge, this is the first time reporting the accuracy for diffuse lung disease image retrieval system. In our opinion, this is sufficient for clinical use.

4. CONCLUSION

In this paper, we proposed a similarity measurement method for diffuse lung disease image retrieval. The disease type and the spatial distribution of disease area are integrated to calculate the similarity between query image and database images. In our method, we applied the fine-tuned GoogLeNet classifier to detect the disease type and disease area from CT slice image. Then the bi-directional weighted distance was calculated as similarity metric based on the spatial distribution. The correct ratio among the top-5 retrieved images achieved 86.2%. The result shows our method has highly potential for clinical use to improve the efficiency for diagnosis of diffuse lung disease.

REFERENCES

[1] M. Anthimopoulos, S. Christodoulidis, a Christe, and S. Mougiakakou, "Classification of interstitial lung disease patterns using local DCT features and random forest," *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*., pp. 6040–6043, 2014.

[2] Sørensen, Lauge, Saher B. Shaker, and Marleen De Bruijne. "Quantitative analysis of pulmonary emphysema using local binary patterns." *IEEE Transactions on Medical Imaging* 29, no. 2: 559-569, 2010.

[3] Gangeh, Mehrdad J., Lauge Sørensen, Saher B. Shaker, Mohamed S. Kamel, Marleen De Bruijne, and Marco Loog. "A texton-based approach for the classification of lung parenchyma in CT images." *In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 595-602. Springer Berlin Heidelberg, 2010.

[4] Haralick, R. M. "Statistical and structural approaches to texture". *Proceedings of the IEEE*, 67(5), 786–804, 1979.

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. "Going deeper with convolutions _ GoogLeNet". *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, 1–9, 2015.

[6] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., … Summers, R. M. "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning". *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298, 2016.

[7] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S. "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network". *IEEE Transactions on Medical Imaging*, 35(5), 1207–1216, 2016.

[8] Arik, S., Huang, T., Lai, W. K., Liu, Q. A "Transfer Learning Method with Deep Convolutional Neural Network for Diffuse Lung Disease Classification". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, 9489, 199–207, 2015.

[9] Heuberger, J., Geissbuhler, A., Müller, H. "Lung CT segmentation for image retrieval using the Insight Toolkit (ITK) ". *Medical Imaging and Telemedicine (MIT 2005)*, 5, 2005

Author Biography

Mengjiao Wang received his PhD degree from Tsinghua University (2014). He joined Fujitsu R&D Center Co., LTD. in 2016. His work has focused on the deep learning in medical image analysis and face recognition.

Rujie Liu received his PhD degree from Beijing Jiaotong University (2001). Since then he joined Fujitsu R&D Center Co., LTD. as a researcher. His work has focused on the machine learning and deep learning in image processing.