PULMONARY TEXTURES CLASSIFICATION USING A DEEP NEURAL NETWORK WITH APPEARANCE AND GEOMETRY CUES

Rui Xu[⋆†] *Zhen Cong*[⋆†] *Xinchen Ye*[*⋆†] *Yasushi Hirano*[‡] *Shoji Kido*[‡]

[⋆]DUT-RU International School of Information Science & Engineering, Dalian University of Technology, China
[†]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, China
[‡]Graduate School of Science and Technology for Innovation, Yamaguchi University, Japan

ABSTRACT

Classification of pulmonary textures on CT images is essential for the development of a computer-aided diagnosis system of diffuse lung diseases. In this paper, we propose a novel method to classify pulmonary textures by using a deep neural network, which can make full use of appearance and geometry cues of textures via a dual-branch architecture. The proposed method has been evaluated by a dataset that includes seven kinds of typical pulmonary textures. Experimental results show that our method outperforms the state-of-the-art methods including feature engineering based method and convolutional neural network based method.

***Index Terms*—** residual network, pulmonary texture, Hessian matrix, CAD, CT

1. INTRODUCTION

Diffuse lung diseases exhibit several kinds of opacities that are widely distributed inside lungs on computed tomography (CT) images [1]. Due to various variations of lung diseases, the opacities show complex textures that are difficult to be distinguished, even for some experienced radiologists. A computer-aided diagnosis (CAD) system is required to help radiologists automatically and precisely diagnose diffuse lung diseases on CT images. An essential technique to establish such a CAD system is to automatically classify which kinds of pulmonary textures local 2D or 3D region-of-interests (ROIs) inside lungs belong to [2]. Fig. 1 shows seven categories of typical pulmonary textures on CT images. Since these textures are very complicated due to variations of both global appearance and local structures with specific geometry, it is not an easy task to achieve classification results with high accuracy.

Traditional methods to classify pulmonary textures usually extract powerful features followed by the training of a discriminative classifier [2]. These methods usually adopt existing classifiers including artificial neural network [3], support vector machine [4][5], K-nearest neighbor classifier [6][7], naive baysian classifier [8] and random forest [9]. Besides, they concentrate on designing excellent features to describe lung tissues. In earlier time, researchers make an effort to design handcrafted features, such as six features derived from basic image processing techniques [3] and statistical measures to quantify spacial variations of intensity [8][10][7]. Sørensen et al. [6] combine local binary pattern and intensity histogram to classify emphysema textures. Later, learning based methods are introduced in feature engineering to yield more powerful feature representations. For example, the bag-of-features and sparse representation based schemes are utilized to classify typical textures for diffuse lung dieseases [4][5][11].

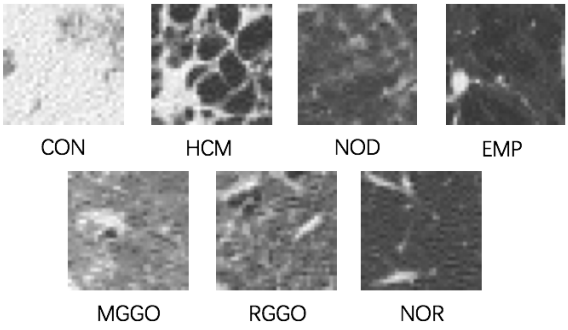

Fig. 1. Examples of pulmonary textures on CT images. They are consolidation (CON), honeycombing (HCM), nodular opacity (NOD), emphysema (EMP), multi-focal ground-glass opacity (M-GGO), reticular ground-glass opacity (R-GGO) and normal pulmonary tissues (NOR).

Recently, Deep networks have had a big impact in many fields of computer vision. It unifies feature engineering and classifier training into an end-to-end framework, and has achieved superior performance [12][13][14]. Deep learning has also been utilized in the classification of pulmonary textures. For example, restricted Boltzmann machines are employed to classify pulmonary opacity and detect airway [15], which show better performance than standard filter banks. Anthimopoulos et al. [16] have designed a convolutional

This work was supported by National Natural Science Foundation of China (NSFC) under Grant 61772106 and Grant 61702078, and by the Fundamental Research Funds for the Central Universities. *Corresponding author. E-mail: yexch@dlut.edu.cn

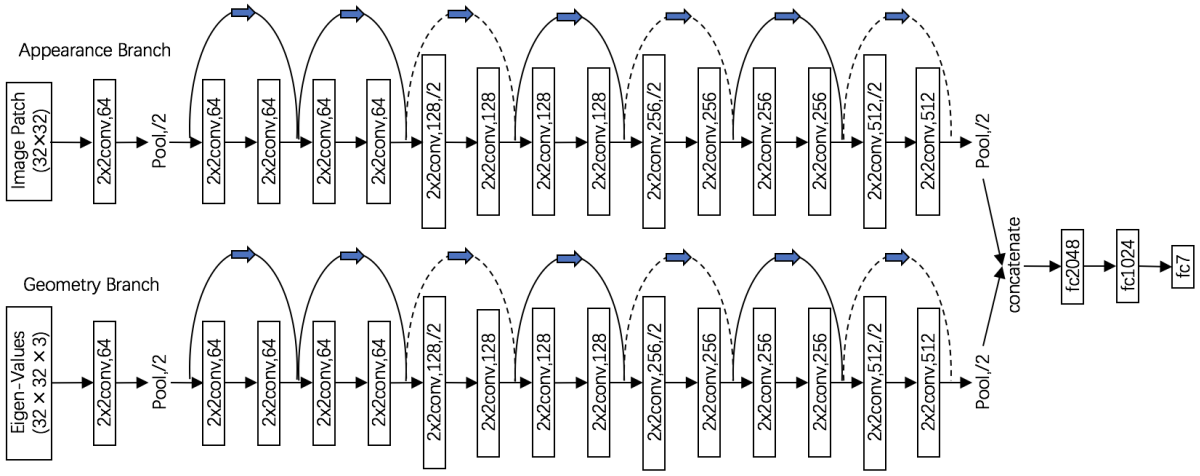

Fig. 2. The architecture of the deep neural network with dual-branch.

neural network (CNN) to classify seven types of pulmonary textures. Then, transfer learning is applied in their later work [17], where classification performance is slightly improved by pre-training the CNN-based network on publicly available texture databases followed by fine-tuning on lung tissue data.

Although these attempts utilizing deep learning have achieved better results than the methods based on handcrafted features, the classification accuracy is still not satisfied for the application in a practical CAD system. We notice that these methods do not have a very deep architecture, which is essential to exploit underlying information from data and demonstrated to be effective [18]. For example, the CNN-based network proposed in [16] only contains eight layers including five convolutional layers and three fully connected layers. This shallow network does not have enough representation capacities to fully reveal the potential information embedded in tiny data structures and leads to unsatisfied classification results. Inspired from the recent work of residual network [19] that designs a very deep network with skip connections on each residual block, we use the similar network architecture to deal with the pulmonary textures classification problem.

Besides, some categories of pulmonary textures are similar in appearance. For example, MGGO looks quite identical with RGGO because both of them contain ground-glass opacity, which increases the difficulty of classification with the appearance attribute only. As we observe, although these similar textures cannot easily be distinguished from their appearance, they are characterized by structures with specific geometry information, e.g., RGGO exhibits structures with reticular shape that is not presented in MGGO. Therefore, geometry information is another important attribute that is complementary to appearance, and should be utilized together with appearance to develop a deep neural network. So far, we have not found any works, which consider to extract the geometry information via deep networks for the pulmonary textures classification.

In this paper, we propose a novel method to classify pulmonary textures by using a deep neural network. Our network has a dual-branch architecture separately extracting appearance and geometry cues via each deep residual network and fusing both information for texture classification. We evaluate the proposed method on a dataset, which contains seven types of pulmonary textures. Experimental results show the efficiency of the carefully designed deep network.

2. PROPOSED METHOD

2.1. Network Architecture

As we notice that pulmonary textures contain both appearance and geometry cues, we design a deep neural network to utilize both of them for classification. As shown in Fig. 2, our network architecture contains two branches, i.e., appearance branch and geometry branch, each of which is composed by fifteen convolutional layers coupled with skip connections and two max-pooling layers. Then three fully-connected layers are used to concatenate the two branches together for the classification. The appearance branch extracts information of intensity variation from CT image-patches directly, while the geometry branch extracts valid shape information of tiny structures from the eigen-values derived from Hessian matrices of original CT image-patches. The two branches have almost the same configurations but different inputs. Details about the calculation of the eigen-values are shown in the section 2.2.

Specifically, the inputs of the appearance branch are CT image-patches with the size of 32×32, while the counterparts of the geometry branch are the 3D tensor fields with the same size of corresponding image-patches. Each branch is constructed with simple block structures including convolutional layer, max-pooling layer and skip connections similar to the deep residual network [19]. Note that, the number of

convolutional kernels is gradually increased with a factor of 2, while the feature map size is halved when convolutional kernels are doubled to save computing costs.

2.2. Eigen-Values Derived from Hessian Matrices

The inputs of the geometry branch are derived from the Hessian matrices of image intensity function at each pixel of CT image-patches. Hessian matrix is embedded with rich geometry information of local intensity structure, since its eigen-values correspond to the second order derivatives of intensity along specific directions. Due to this character, Hessian matrix analysis are widely used to extract intensity structures with specific shapes from medical images [20], such as vessels or blobs. In this paper, we calculate eigen-values from Hessian matrices at each pixel to extract geometry cues embedded in pulmonary textures.

Given $I(x, y, z)$ is the CT intensity at a pixel on CT image-patches, we can calculate the Hessian matrix by using the definition illustrated in Eq. 1.

$$\mathbf{H}(x,y,z) = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial x \partial z} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} & \frac{\partial^2 I}{\partial y \partial z} \\ \frac{\partial^2 I}{\partial z \partial x} & \frac{\partial^2 I}{\partial z \partial y} & \frac{\partial^2 I}{\partial z^2} \end{pmatrix} \quad (1)$$

The 3×3 Hessian matrices are calculated on every pixel of CT image-patches. By decomposing these Hessian matrices, we can calculate three eigen-values at each pixel. These eigen-values are arranged to be cubes with dimension of $32 \times 32 \times 3$, which are embedded with geometry cues of textures. These cubes are fed into the geometry branch of the deep network to extract the underlying geometry information.

2.3. Implementation

The proposed deep network is implemented by using the Tensorflow framework [21]. We utilize two kinds of active functions in the network. Specifically, rectified linear units (ReLU) are used for all convolutional layers and leaky rectified linear units (Leaky-ReLU) are applied for fully-connected layers. The parameter α in Leaky-ReLU is 0.9 and the kernel size of max-pooling layers is 2×2. The network is trained by using the cost function of cross entropy which is minimized by stochastic gradient descent.

3. EXPERIMENTS

3.1. Data and Experimental Protocol

Our database includes 217 CT images scanned from different patients in Osaka University hospital. All CT images are captured from a GE Discovery CT750 HD CT scanner with the tube voltage of 120kVp and the current of 213mAs. The image data on the axial direction is reconstructed by a 512×512 matrix with the slice thickness of 0.67mm. There are 30 CT images captured from different patients with mild or no diffuse lung diseases. Since no pulmonary opacities exhibit on these CT images, they are used to extract 2D image-patches for the texture of normal lung tissues. The rest 187 CT images are captured from different patients with severe diffuse lung diseases and they are used to extract patches for different textures of diffuse lung diseases.

Image-patches are extracted from CT images according to the following procedure. An experienced radiologist firstly selects three axial plains where there are typical pulmonary textures on each CT image. Together with the other two experienced radiologists, they respectively delineate typical regions containing pulmonary textures on previously selected CT plains by using a graphical user interface based drawing tool. The delineated regions from different radiologists are processed by a logical operation of AND to extract the common-agreed regions, which are sampled by raster-scanning with a stride of 8 to extract potential image-patches with the size of 32×32. Only image-patches whose centers are inside the common-agreed regions are reserved for experiments. Finally, we totally extract 72,348 CT image-patches.

All methods are evaluated by the following training-validation-testing protocol. We train different methods on a training set, and turn their hyper parameters on a separate validation set. After all hyper parameters are optimized, we evaluate the performance on a testing set. In experiments,

Table 1. Recognition accuracy of deep neural network with different configurations. The proposed dual-branch network is denoted with the prefix of DB-ResNet-18, and the single branch network is denoted as ResNet-18.

Models	Kernel	Pooling	Dropout	Accuracy
DB-ResNet-18-a	4x4	Max	0.8	0.8733
DB-ResNet-18-b	3x3	Max	0.8	0.9022
DB-ResNet-18-c	2x2	Max	0.8	**0.9367**
DB-ResNet-18-d	2x2	- -	0.5	0.9136
ResNet-18	2x2	Max	0.8	0.9247

Table 2. Comparison of performances of different methods, which are the LeNet [12], bag-of-features based method [4], the CNN-based method (CNN-8) [16], residual network (ResNet-18) and the proposed dual-branch deep network (DB-ResNet-18).

Methods	Accuracy	F_{avg}
LeNet	0.7897	0.7921
Bag-of-Features	0.9014	0.9003
CNN-8	0.9025	0.9039
ResNet-18	0.9247	0.9247
DB-ResNet-18 (proposed)	**0.9367**	**0.9365**

54,392 patches are used for the training and validation, and the rest 17,956 patches for testing. CT images that are used to extracted patches for testing set are not mixed with the ones for training/validation sets. We adopt two measures to evaluate the classification performance for different methods. One is the recognition accuracy, which is defined as the ratio of the number of correctly classified samples and the total number of testing samples. The other is the average F-value over different classes whose definition are given by Eq. 2.

$$F_{avg} = \frac{2}{7}\sum_{c=1}^{7}\frac{r_c \times p_c}{r_c + p_c} \tag{2}$$

where r_c is the recall of the class c, which is defined as the ratio of the number of samples correctly classified as c and the number of samples in class c, and p_c is the precision of class c, which is defined as the ratio of the number of samples correctly classified as c and the number of samples classified as class c.

3.2. Results and Discussions

We train several models of the proposed network with different configurations, which are listed in Table 1. Kernel size is evaluated from 2×2 to 4×4 for the convolutional layers (DB-ResNet-18-a, b, c). Experimental results show that the kernel size alters the performance largely and a smaller convolutional kernel leads to better classification results. A deep network has a huge number of parameters and it requires lots of data for training. However, data are usually limited in practice. A smaller kernel in convolution can largely reduce the parameters, and makes the training converge better. Similar results have also been reported in [16], where the kernel size of 2×2 makes the CNN-based method achieve the best result. Besides, we also evaluate the necessity of the max-pooling layers for the proposed method. We train a deep network without the two max-pooling layers (DB-ResNet-18-d) and find that the performance is decreased. This could be because that the max-pooling layers reduce dimensions for image-patches and eigen-values, and leads to less parameters in the deep network.

In order to verify the efficiency of the dual-branch architecture extracting features from both appearance and geometry cues, we train a residual network that has a single-branch to only exploit appearance information (ResNet-18 in Table 1). It has the same depth as the dual-branch network to ensure a fair comparison. Experimental results show that the dual-branch architecture can achieve superior results. Especially, the classification of both MGGO and RGGO can be improved by using the dual-branch network, illustrated by the confusion matrices given in (c) and (d) of Table 3. Therefore, it is necessary to design a deep network with dual-branch architecture to fully exploit underlying information of both appearance and geometry.

Table 3. Confusion matrices of different methods, together with recalls (r_c) and precisions (p_c) for each kind of textures. Ground-truth labels are given in each row and predicted labels are given in each column.

	CON	MGGO	HCM	RGGO	EMP	NOD	NOR	r_c
CON	1812	3	1	55	0	3	3	0.965
MGGO	140	2244	2	144	5	10	12	0.878
HCM	0	3	2594	133	0	0	0	0.950
RGGO	112	332	92	1804	0	15	0	0.766
EMP	0	0	12	0	2402	20	16	0.980
NOD	1	23	0	2	0	2200	567	0.788
NOR	0	4	0	2	0	58	3130	0.980
p_c	0.877	0.860	0.960	0.843	0.998	0.954	0.840	

(a) Bag-of-Features

	CON	MGGO	HCM	RGGO	EMP	NOD	NOR	r_c
CON	1876	0	0	1	0	0	0	0.999
MGGO	16	2188	10	214	13	103	13	0.856
HCM	0	1	2674	55	0	0	0	0.979
RGGO	6	463	111	1775	0	0	0	0.754
EMP	0	17	48	0	2198	21	166	0.897
NOD	0	117	0	0	5	2471	200	0.885
NOR	0	65	0	0	21	84	3024	0.947
p_c	0.988	0.767	0.941	0.868	0.983	0.922	0.889	

(b) CNN-8

	CON	MGGO	HCM	RGGO	EMP	NOD	NOR	r_c
CON	1877	0	0	0	0	0	0	1.000
MGGO	22	2006	10	336	34	133	16	0.785
HCM	0	8	2634	88	0	0	0	0.965
RGGO	25	190	97	2043	0	0	0	0.868
EMP	0	0	1	0	2404	17	28	0.981
NOD	0	15	0	0	18	2509	251	0.898
NOR	0	9	0	0	8	47	3130	0.980
p_c	0.976	0.900	0.961	0.828	0.976	0.927	0.914	

(c) ResNet-18

	CON	MGGO	HCM	RGGO	EMP	NOD	NOR	r_c
CON	1868	0	0	0	3	6	0	0.995
MGGO	23	2069	2	324	56	68	15	0.809
HCM	0	0	2662	68	0	0	0	0.975
RGGO	9	128	81	2137	0	0	0	0.907
EMP	1	2	38	0	2367	8	34	0.966
NOD	0	36	0	0	7	2560	190	0.917
NOR	0	16	0	0	10	12	3156	0.988
p_c	0.983	0.919	0.957	0.845	0.969	0.965	0.930	

(d) DB-ResNet-18

The proposed network is compared with four methods including two state-of-the-art methods [4][16]. Table 2 summarizes the comparison results, and the detailed confusion matrices are given in Table 3. LeNet [12] is the first convolutional network applied for image recognition task, however it achieves the worst result. Since it cannot be comparable with the others, we omit its confuse matrix in Table 3 to save the space. The bag-of-features based method was proposed in our previous research [4], where K-Means clustering was utilized to extract discriminate features followed by the classification based on a SVM classifier. The performance of the bag-of-features based method is approximate to the CNN-based method [16], where the network depth is only eight. Note that, the parameters of the CNN-based method have been carefully turned and the performance reported in this paper is higher than the original research [16]. This result illustrates that a shallow network cannot reveal the capacity of deep learning and a deeper network is required in pulmonary texture classification. Through the detailed comparison, it can be seen that the proposed deep network with dual-branch architecture is efficient to classify seven kinds of pulmonary textures and outperforms all competitors. Therefore, we conclude that the proposed deep network has achieved the state-of-the-art performance in the classification of pulmonary textures.

4. CONCLUSION

We proposed a novel method to classify pulmonary textures by utilizing a deep network with a dual-branch architecture, which favors exploiting underlying information of both appearance and geometry on pulmonary textures. Experimental results showed that the proposed method outperformed four methods, including two state-of-the-art methods [4][16]. This method can be applied for the segmentation of pulmonary opacities for the CAD system of diffuse lung diseases. We will research it in future.

5. REFERENCES

[1] W. M. Thurlbeck, R. R. Miller, and et. al., *Diffuse Diseases of the Lung: A Team Approac*, PA Decker, 1991.

[2] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 385–405, 2006.

[3] Y. Uchiyama, S. Katsuragawa, and et. al., "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Medical Physics*, vol. 30, no. 9, pp. 2440–2454, 2003.

[4] R. Xu, Y. Hirano, R. Tachibana, and S. Kido, "Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach," in *International Conference on Medical Image Computing & Computer-assisted Intervention (MICCAI)*, 2011, p. 183.

[5] W. Zhao, R. Xu, and et. al., "Classification of diffuse lung diseases patterns by a sparse representation based method on hrct images," in *Proc. 35th Int. Conf. IEEE Eng. Med. Biol. Soc*, 2013, pp. 5457–5460.

[6] L. Sørensen, S. B. Shaker, and M. de Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 559–569, 2010.

[7] P. D. Korfiatis, A. N. Karnhaliou, and et. al., "Texture-based identification and characterization of interstitial pneumonia patterns in lung multidetector ct," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 675–680, 2010.

[8] Y. Xu, M. Sonka, and et. al., "Mdct-based 3-d texture classification of emphysema and early smoking related lung pathologies," *IEEE Transactions on Medical Imaging*, vol. 13, no. 8, pp. 969–978, 2006.

[9] M. Anthimopoulos, S. Christodoulidis, and et. al., "Classification of interstitial lung disease patterns using local dct features and random forest classification of interstitial lung disease patterns using local dct features and random forest," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 6040–6043.

[10] I. Sluimer, M. Prokop, L. Hartmann, and B. van Ginneken, "Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution ct of the lung," *Medical Physics*, vol. 33, no. 7, pp. 2610–2620, 2006.

[11] Y. Song, W. Cai, and et. al., "Feature-based image patch approximation for lung tissue classification," *IEEE Transactions on Medical Imaging*, vol. 32, no. 4, pp. 797–808, 2013.

[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[14] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221, 2017.

[15] Tulder G Van and Bruijne M De, "Combining generative and discriminative representation learning for lung ct analysis with convolutional restricted boltzmann machines," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1262–1272, 2016.

[16] M. Anthimopoulos, S. Christodoulidis, and et. al., "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[17] S Christodoulidis, M Anthimopoulos, and et. al., "Multi-source transfer learning with convolutional neural networks for lung pattern analysis," *IEEE Journal of Biomedical & Health Informatics*, vol. 21, no. 1, pp. 76–84, 2017.

[18] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[19] Kaiming He, Xiangyu Zhang, and et. al., "Deep residual learning for image recognition," in *Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[20] Yoshinobu Sato, Carl Fredrik Westin, and et. al., "Tissue classification based on 3d local intensity structures for volume rendering.," *Visualization & Computer Graphics IEEE Transactions on*, vol. 6, no. 2, pp. 160–180, 2000.

[21] Martn Abadi, Ashish Agarwal, and et. al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv:1603.04467*, 2016.