Multi-scale Residual Network with Two Channels of Raw CT Image and Its Differential Excitation Component for Emphysema Classification

Liying Peng[1], Lanfen Lin[1(✉)], Hongjie Hu[2], Huali Li[2],
Qingqing Chen[2], Dan Wang[2], Xian-Hua Han[3], Yutaro Iwamoto[3],
and Yen-Wei Chen[3]

[1] College of Computer Science and Technology, Zhejiang University,
Hangzhou, China
llf@zju.edu.cn
[2] Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou, China
[3] College of Information Science and Engineering, Ritsumeikan University,
Kusatsu, Japan

Abstract. Automated tissue classification is an essential step for quantitative analysis and treatment of emphysema. Although many studies have been conducted in this area, there still remain two major challenges. First, different emphysematous tissue appears in different scales, which we call "inter-class variations". Second, the intensities of CT images acquired from different patients, scanners or scanning protocols may vary, which we call "intra-class variations". In this paper, we present a novel multi-scale residual network with two channels of raw CT image and its differential excitation component. We incorporate multi-scale information into our networks to address the challenge of inter-class variations. In addition to the conventional raw CT image, we use its differential excitation component as a pair of inputs to handle intra-class variations. Experimental results show that our approach has superior performance over the state-of-the-art methods, achieving a classification accuracy of 93.74% on our original emphysema database.

Keywords: Emphysema classification · Multi-scale
Differential excitation component

1 Introduction

Emphysema is a major component of chronic obstructive pulmonary disease (COPD), which is emerging as a worldwide health problem. Generally, as shown in Fig. 1, emphysema can be classified into three subtypes: centrilobular emphysema (CLE) that generally appears as scattered small low attenuation areas; paraseptal emphysema (PSE) which is shown as low attenuation areas aligned in a row along a visceral pleura [1]; and panlobular emphysema (PLE) that usually manifests as a wide range low attenuation region with fewer and smaller lung vessels [1]. They have different pathophysiological significance [2]. Hence, classification and quantification of emphysema are important.
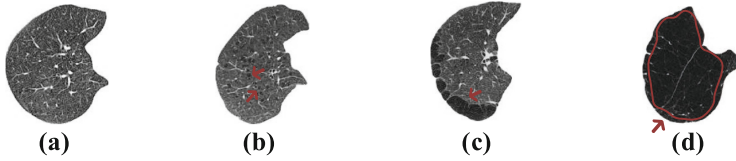
Fig. 1. (a) Normal tissue (NT). (b) CLE. (c) PSE. (d) PLE.

Much research has been conducted to classify the lung tissue of different emphysema subtypes. One common way is based on the local intensity distribution, such as kernel density estimation (KDE) [3]. Another class of approaches describes the morphology of emphysema using texture analysis techniques [1, 4–6]. In the last years, some attempts have revealed the potential of deep learning techniques on lung disease classification, but it has been applied in only two studies [7, 8] for emphysema classification. The networks in these two studies are very preliminary, using two or three convolutional layers, so they are not able to capture the high-level features. Since the classification of emphysema mainly depends on features of texture and intensity, there still remain two major challenges. (1) "inter-class variations": as can be seen in Fig. 1, different emphysematous tissue appears in different scales. Since existing methods ignore the scales of different emphysema which are useful clues for diagnosing emphysema, it is highly desirable to develop new models that can take full advantage of the information from multiple scales. (2) "intra-class variations": in clinical practice, the intensities of CT images acquired from different patients, scanners or scanning protocols may vary [9]. The variation in CT images will affect the classification accuracy of emphysema, so it is necessary to design models which are robust to such variability. In addition, existing methods for emphysema classification are limited to extracting low-level features or mid-level features, which have limited abilities to distinguish different patterns.

In this paper, we focus on the supervised classification of emphysema. We propose a novel deep learning method using the multi-scale residual network (MS-ResNet) [16] with two channels of the raw CT image and its differential excitation component. In contrast to previous works, our proposed method discovers high-level features that can better characterize the emphysema lesions. We incorporate multi-scale information into our networks to address the challenge of inter-class variations. Moreover, to handle intra-class variations, we first transform the raw image data into the differential excitation domain of human perception based on weber's law, which is robust to intensity variability. Then we use the raw CT images and the transformed images as different channels of the inputs of networks. The experiments show that our method can achieve higher classification accuracy than the state-of-the-art methods. Based on the classification results, we calculate the area percentage of each class (CLE%, PLE%, PSE%, respectively). Then, we show the relationship between the quantitative results (area percentages) and the forced expiratory volume in one second dividing with a predicted value ($FEV_{1\%}$), which is the primary indicator of pulmonary function tests (PFTs).

2   Methods

In this section, we first describe how to transform the raw CT image into the differential excitation domain. Subsequently, we present our multi-scale residual network with two channels of the raw CT image and its differential excitation component. An overview of the proposed method is shown in Fig. 2.
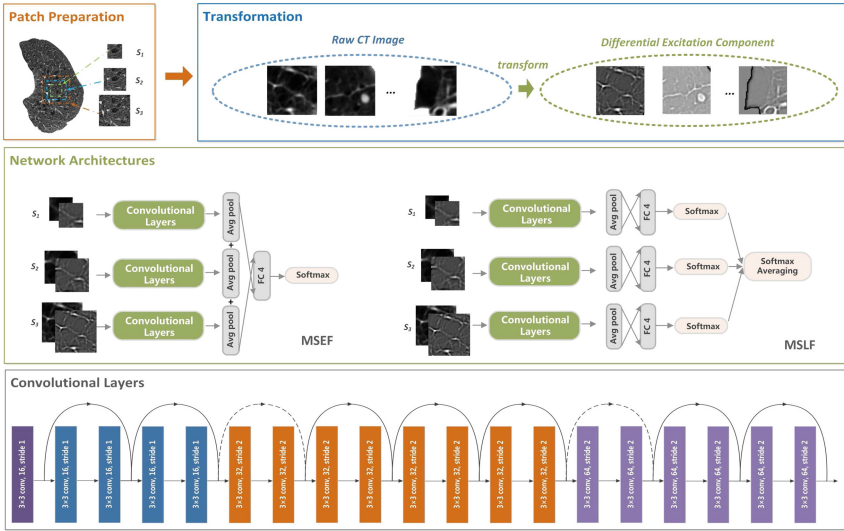

Fig. 2. Overview of the proposed approach

2.1   Differential Excitation Component

Ernst Heinrich Weber, an experimental psychologist in the 19th century, observed that the ratio of the perceived change in stimulus to the initial stimulus is a constant [10], which is well-known as Weber's law and can be defined as $\Delta I/I = \alpha$, where ΔI denotes the perceived change in stimulus, I denotes the initial stimulus, and α is referred to as the Weber fraction for detecting changes in stimulus.

Inspired by Weber's law, which shows that human perception of a pattern depends not only on the absolute intensity of the stimulus but also on the relative variance of the stimulus, we transform the raw image into the differential excitation domain of human perception which is robust to intensity variability [10]. In order to do so, we first compute the difference between a focused pixel and its neighbors, which can be formulated as

$$\Delta I_c = \sum_{i=0}^{p-1} \left(\Delta I_c^i\right) = \sum_{i=0}^{p-1} \left(I_c^i - I_c\right) \tag{1}$$

where I_c is the intensity at position x_c, $I_c^i (i = 0, 1, \ldots, p - 1)$ is intensity of the ith neighbor of c, and p is the number of neighbors. The differential excitation component of the focused pixel c is defined as

$$E_c = \arctan[\frac{\Delta I_c}{I_c + \lambda}] = \arctan[\sum_{i=0}^{p-1} \frac{(I_c^i - I_c)}{I_c + \lambda}] \tag{2}$$

where λ is a constant which avoids the situation in which there is zero intensity. λ is set to one in our experiments.

2.2   MS-ResNet with Raw and Excitation Channels

MS-ResNet. Due to the inter-class variations of emphysema, one target category tends to be identified on a certain scale and the most suitable scales for different target classes may vary. That is, we cannot find the best scale for all cases. Thus, it is essential to incorporate information from different scales into our deep neural networks [16].

For a baseline, we build a 20-layer ResNet [11], which has been shown to achieve the excellent performance on image classification. For the sake of adapting it to our problem (small inputs and only 4 classes), we remove the pooling layer and modify the configuration for some layers. Figure 2 (bottom) presents the details of our ResNet. As shown in Fig. 2 (top), for each annotated pixel, we can extract patches with different scales from its neighborhood. The label assigned to each patch is the same as label of the central pixel. Note that, in this paper, different scales mean various sizes of inputs. Figure 2 (middle) presents two ways for fusing information from different scales: multi-scale early fusion (MSEF) and multi-scale late fusion (MSLF). For the MSEF, we employ the independent convolutional layers for each scale. The outputs of average pooling layers are combined and fed into a 4-way shared fully connected layer with softmax to compute a cross entropy classification loss. For the MSLF, we train three separate networks, each focusing on a certain scale. During the fusion step, we first sum up the values of probability vectors yielded by different networks, and then compute the average of them.

Fused Representation of Raw Image and its Differential Excitation Component. As mentioned in Introduction part, there exists the challenge of intra-class variations for emphysema classification. As shown in Fig. 2 in order to reduce the impact of intensity variability, we first transform the raw image data into the differential excitation domain of human perception, which is robust to intensity variability. Then we use the raw CT images and their differential excitation components as different channels of the inputs of networks.

3   Experimental Results

3.1   Materials

Our dataset contains 101 HRCT volumes. The first part of our dataset includes 91 HRCT volumes annotated manually by two experienced radiologists and checked by one experienced chest radiologist. Four types of patterns were annotated: CLE, PLE, PSE, and non-emphysema (NE) which corresponds to tissue without emphysema. This part of dataset is used for evaluation of classification accuracy shown in Sect. 3.2. Since the first part of dataset does not include complete pulmonary function evaluations, we collected additional 10 HRCT volumes from patients who have a complete pulmonary function evaluation for a quantitative analysis of emphysema shown in Sect. 3.3. All data came from two hospitals and were acquired using seven types of CT machines with a slice collimation of 1 mm–2 mm, a matrix of 512×512 pixels, and an in-plane resolution of 0.62 mm–0.71 mm.

3.2   Evaluation of Classification Accuracy

Experimental Setup. Our classification experiments are conducted on 91 annotated subjects (the first part of dataset): 59 subjects (about 720,000 patches) for training, 14 subjects (about 140,000 patches) for validation, and 18 subjects (about 160,000 patches) for testing. A 20-layer ResNet is chosen as the baseline in this work (we found 8-layer, 32-layer, 44-layer, and 56-layer ResNet decrease the performance, compared to 20-layer ResNet, on our data). We have done extensive experiments for selecting patch sizes and the experimental results show that the most suitable scales (patch sizes) for different target categories are different: for non-emphysema tissue, the inputs of 27×27 generate the best result; for CLE, the best scale is 41×41; for PLE and PSE, the highest classification accuracy is obtained with inputs of size 61×61. Therefore, patches of sizes 27×27, 41×41, and 61×61 are selected as inputs of the multi-scale neural networks.

Single Scale versus Multiple Scales. In this section, to investigate the effect of fusing multi-scale information on the classification accuracy, we use only raw images as inputs of networks. As shown in Table 1, both MSEF model and MSLF model outperform the single-scale models (27×27, 41×41, and 61×61). To test the statistical significance of the classification accuracy differences between single-scale models and multi-scale models, we calculated the classification accuracy of each patient, and then employed t-test. The results of analysis confirmed the statistically significant (p-value < 0.05) superior performance of the multi-scale models against all single-scale models. Fusion of multi-scale information leads to higher accuracy, so we can conclude that the multi-scale methods are beneficial compared to the single scale setting.

Single Channel versus Multiple Channels. This part compares the classification accuracy between the single-channel models (use only raw images as inputs) and the multi-channel models (use raw CT images and their differential excitation components

Table 1. The comparison between the single-scale models and the multi-scale models.

	27×27	41×41	61×61	MSEF	MSLF
NE	93.19%	91.77%	86.04%	**94.05%**	91.98%
CLE	86.85%	88.87%	86.50%	**91.17%**	89.02%
PLE	83.61%	92.18%	**95.06%**	89.48%	93.78%
PSE	87.35%	89.52%	95.52%	**95.89%**	92.36%
Avg.	87.77%	90.58%	90.81%	**92.68%**	91.80%

as different channels of inputs). As shown in Table 2, for both single-scale setting and multi-scale setting, the multi-channel models offer superior performance to the single-channel models (p-value < 0.05).

Table 2. The comparison between the single-channel models and the multi-channel models.

	27×27	41×41	61×61	MSEF	MSLF
Single-channel	87.77%	90.58%	90.81%	92.68%	91.80%
Multi-channel	**89.39%**	**91.47%**	**91.84%**	**93.74%**	**92.90%**

Comparison to the State-of-the-Art Methods. In this section, our approaches are compared to other state-of-the-art methods. The comparison between our methods and the machine learning (ML) methods for emphysema classification is provided in the first five rows. The results prove the superior performance of our methods that significantly outperform the rest by 14% to 20%. The rest of Table 3 shows a comparison to other deep learning methods. Since existing deep learning methods for emphysema classification [7, 8] are very primary using only two or three convolutional layers, we also compare our approaches with other CNNs for interstitial lung disease (ILD) classification [12, 14]. The results show that our approaches have superior performance over other deep learning methods.

3.3   Emphysema Quantification

In this section, based on the classification results, we quantify the whole lung area of 10 subjects (the second part of dataset with complete pulmonary function evaluations) by calculating the area percentage of each class (CLE%, PLE%, PSE%, respectively), and show the relationship between the quantitative results (area percentages) and the forced expiratory volume in one second dividing with a predicted value ($FEV_{1\%}$), which is the primary indicator of pulmonary function tests (PFTs). Some visual results of full lung classification are shown in Fig. 3. It can be seen that, auto-annotations (or classification results) of proposed method are similar to annotations of radiologists (manual annotations). The relationship between the quantitative results (area percentages) and $FEV_{1\%}$ of 10 subjects are shown in Table 4. According to [15], $FEV_{1\%}$ is an effective indicator that indicates both functional and symptomatic impairment of COPD. Symptoms arise in individuals in relation to a relative loss of FEV_1. More specifically,

Table 3. The comparison of classification accuracy (Acc.) to the state-of-the-art approaches.

ML methods	Acc.
LBPINT [1]	78.67%
Texton-based [4]	79.06%
KDE [3]	76.67%
Sparse representation [5]	72.96%
JWRIULTP [6]	79.31%
DL methods	Acc.
Karabulut [7]	65.51%
Pei [8]	72.34%
AlexNet-TL [12]	81.79%
GoogLeNet-TL [12]	85.75%
Wang [13]	73.62%
Anthimopoulos [14]	85.00%
Proposed method	**93.74%**

$FEV_{1\%}$ can reflect the severity of airflow obstruction in the lungs. The lower value of FEV1% means the more severe the airflow obstruction in the lungs. Our results show that a larger CLE% (or PLE%) corresponds to a lower $FEV_{1\%}$ (the more severe the airflow obstruction in the lungs). From our experiments, we found there is no relationship between PSE% and $FEV_{1\%}$. According to the literature [1], PSE is often not associated with significant symptoms or physiological impairments, which is in close agreement with our experimental results.
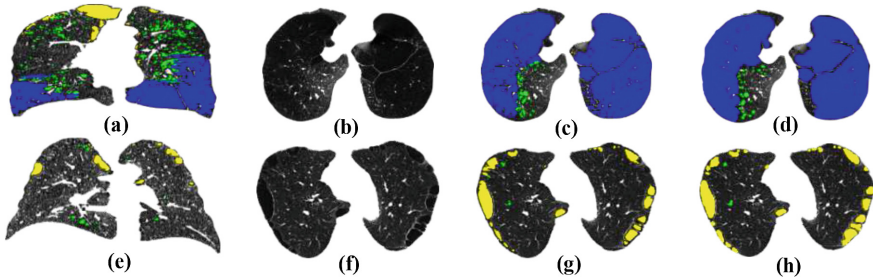
Fig. 3. Examples of the classification results. Each row represents a subject. (a), (e) Classification results in coronal view. (b), (f) Typical original HRCT slices from subjects of (a), (e), respectively. (c), (g) Auto-annotated mask of our proposed method. (d), (h) Manual annotated mask of radiologists. Green mask: CLE lesions. Blue mask: PLE lesions. Yellow mask: PSE lesions.

Table 4. Relationship between quantitative results and $FEV_{1\%}$.

	CLE%	PLE%	PSE%	FEV_1 %
Case #1	13.120%	11.311%	0.002%	57.000%
Case #2	1.464%	0.002%	0.004%	77.200%
Case #3	16.878%	4.962%	0.000%	36.000%
Case #4	1.138%	1.437%	15.546%	76.100%
Case #5	10.207%	38.158%	0.161%	20.500%
Case #6	0.124%	0.000%	0.001%	81.000%
Case #7	0.057%	0.003%	0.000%	86.000%
Case #8	0.000%	42.368%	0.000%	25.000%
Case #9	13.244%	5.860%	1.769%	53.000%
Case #10	4.467%	0.622%	0.812%	69.000%

4   Conclusions

In this paper, we proposed a novel deep learning approach for emphysema classification, using the multi-scale ResNet with two channels of raw CT image and its differential excitation component. Our proposed approach achieved a classification accuracy of 93.74%, which is superior to the state-of-the-art methods.

References

1. Sørensen, L., et al.: Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Trans. Med. Imaging **29**(2), 559–569 (2010)
2. Lynch, D.A., et al.: CT-definable subtypes of chronic obstructive pulmonary disease: a statement of the fleischner society. Radiology **277**(1), 192–205 (2015)
3. Mendoza, C.S., et al.: Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions. In: ISBI 2012, pp. 474–477 (2012)
4. Gangeh, M.J., Sørensen, L., Shaker, S.B., Kamel, M.S., de Bruijne, M., Loog, M.: A texton-based approach for the classification of lung parenchyma in CT images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 595–602. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15711-0_74
5. Jie, Y., et al.: Texton and sparse representation based texture classification of lung parenchyma in CT images. In: EMBC 2016, pp. 1276–1279 (2016)
6. Liying, P., et al.: Joint weber-based rotation invariant uniform local ternary pattern for classification of pulmonary emphysema in CT images. In: ICIP 2017, pp. 2050–2054 (2017)
7. Karabulut, E.M., et al.: Emphysema discrimination from raw HRCT images by convolutional neural networks. In: ELECO 2015, pp. 705–708 (2015)

8. Pei, X.: Emphysema classification using convolutional neural networks. In: Liu, H., Kubota, N., Zhu, X., Dillmann, R., Zhou, D. (eds.) ICIRA 2015. LNCS (LNAI), vol. 9244, pp. 455–461. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22879-2_42
9. Cheplygina, V., et al.: Transfer learning for multi-center classification of chronic obstructive pulmonary disease. JBHI (2018)
10. Xianhua, H., et al.: Integration of spatial and orientation contexts in local ternary patterns for HEp-2 cell classification. Pattern Recognit. Lett. **82**, 23–27 (2016)
11. Kaiming, H., et al.: Deep residual learning for image recognition. In: CVPR 2016, pp. 770–778 (2016)
12. Shin, H.C., et al.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016)
13. Qian, W., et al.: Multiscale rotation-invariant convolutional neural networks for lung texture classification. JBHI **22**, 184–195 (2017)
14. Anthimopoulos, M., et al.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans. Med. Imaging **35**(5), 1207–1216 (2016)
15. Jakeways, N., et al.: Relationship between FEV1 reduction and respiratory symptoms in the general population. Eur. Respir. J. **21**(4), 658–663 (2003)
16. Liying, P., et al.: Classification of pulmonary emphysema in CT images based on multi-scale deep convolutional neural networks. In: ICIP (2018, in Press)