Lung Image Patch Classification with Automatic Feature Learning

Qing Li, *Student Member, IEEE,* Weidong Cai, *Member, IEEE,* and David Dagan Feng, *Fellow, IEEE*

***Abstract*— Image patch classification is an important task in many different medical imaging applications. The classification performance is usually highly dependent on the effectiveness of image feature vectors. While many feature descriptors have been proposed over the past years, they can be quite complicated and domain-specific. Automatic feature learning from image data has thus emerged as a different trend recently, to capture the intrinsic image features without manual feature design. In this paper, we propose to create multi-scale feature extractors based on an unsupervised learning algorithm; and obtain the image feature vectors by convolving the feature extractors with the image patches. The auto-generated image features are data-adaptive and highly descriptive. A simple classification scheme is then used to classify the image patches. The proposed method is generic in nature and can be applied to different imaging domains. For evaluation, we perform image patch classification to differentiate various lung tissue patterns commonly seen in interstitial lung disease (ILD), and demonstrate promising results.**

I. Introduction

Image pattern classification has a wide range of applications in medical imaging, such as analysis of various lung diseases [1], [2], [3], [4], [5], [6], [7], [8], [9]. Commonly the classification task comprises of two main components: 1) image feature extraction and 2) feature classification. Both components have been extensively studied over the past years with increasingly good performance. Popular feature extraction techniques include the more traditional ones such as grayscale or color distributions, gray-level co-occurrence matrix, Gabor filters and wavelets; and the more recent feature descriptors such as the scale-invariant feature transform (SIFT) [10], local binary patterns (LBP) [11], and histogram of oriented gradients (HOG) [12].

Different from the manually crafted feature descriptors, automatic feature learning has been recently proposed to solve some challenging image classification problems [13], [14]. They demonstrate good generalizability that could be applied to different problems without major changes. Both supervised and unsupervised machine learning algorithms could automatically adapt to the training dataset, and extract useful features for classification. Restricted Boltzmann Machine (RBM) [15], [16], an unsupervised learning algorithm, is used widely to perform automatic feature learning from various types of input. Unsupervised learning algorithms can learn from unlabelled input, which is a great advantage over supervised learning algorithms, as accurately labelled training data is usually more expensive to prepare than unlabelled data.

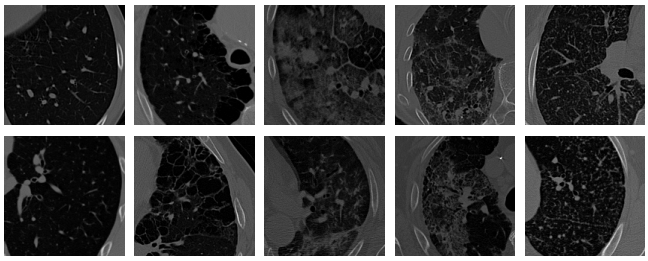


Fig. 1. Two example images (part of original CT slices) are shown for each tissue category. (a) Normal. (b) Emphysema. (c) Ground glass. (d) Fibrosis. (e) Micronodules.

Our aim of this study is to classify different lung tissue patterns depicting ILDs. ILD represents a group of more than 150 different disorders of the lung parenchyma [17]. They cause scarring of the lung tissues and patients would develop breathing difficulties. Identifying the specific type of ILD is important for treatment, and the computed tomography (CT) imaging is usually employed during the process. However, due to individual differences between patients, even the same type of ILD often displays different image patterns; and hence interpreting the images is quite a challenging task for physicians. A computed-aided system for automatic classification of the tissue patterns is thus very helpful for the clinical practice, and has been studied recently focusing on customized feature design [3], [7].

The main difficulty in accurate classification of lung tissue patterns is the high intra-class feature variation and low inter-class feature distinction. For example, as shown in Fig. 1, within each tissue category, there are considerable feature variations. And for different categories, such as the ground glass and fibrosis, the images often look rather similar. Therefore, it is quite challenging to design a representative and discriminative feature set to effectively group fairly different images into the same tissue category and differentiate between similar images of different categories.

In this work, we propose a new image patch classification method based on fully automatic feature learning. Rather than defining a set of features manually, we design a unsupervised learning approach to construct multi-scale feature extractors, and such extractors are convolved with the image patches to obtain feature vectors. Compared to using existing popular feature descriptors [2], the features generated with the proposed method are more data-adaptive with automatic

*This work was supported in part by ARC grants.

Q. Li, W. Cai and D. D. Feng are with the Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia.

exploration of the intrinsic image characteristics. Also different from customized problem-specific features [7], it avoids the empirical nature of feature design, and can be easily adapted to any other imaging domains. The extracted patch-wise features are then classified using a standard supervised classifier to derive the tissue categories of the image patches.

II. Methods

First, to capture image features of different scales, feature extractors of different sizes are learned using unsupervised algorithm (Section II-A). Next, each image patch is convoluted with the learned feature extractors to obtain a group of feature maps; and sum of all pixel values in feature maps of one image patch is used as the feature vector (Section II-B). The patch-wise feature vector is then classified into one of the five tissue categories (Section II-C).

A. Extractor Learning

RBM is a stochastic and generative neural network that is capable of capturing and reproducing the statistical structure of a given dataset [18]. The RBM method and its variants have been proven effective in learning filters, i.e. feature extractors, from natural images for classification and generation tasks. In this article, we use Gaussian RBM (GRBM) [13], [16] to learn a model that could represent patches from lung images. GRBM is chosen, because it is capable to model real value input like natural images.

GRBM consists two layers of neuron nodes, the visible layer and the hidden layer. The nodes in visible layer and hidden layer are connected by undirected weight matrix. Each node also has a bias value associated. An energy function of the network is defined as:

$$E(v,h) = \sum_{i=1}^{n_v} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v}\sum_{j=1}^{n_h} W_{ij} h_j \frac{v_i}{\sigma_i} \tag{1}$$

where v_i and h_j are values of visible and hidden neurons respectively. a_i and b_j are biases corresponding to visible and hidden neurons. W_{ij} is the weight matrix connecting visible and hidden neurons, and σ_i is the standard deviation of the visible neuron v_i

The probability of a visible state given a hidden state is as:

$$p(v,h) = \frac{1}{Z}\exp^{-E(v,h)} \tag{2}$$

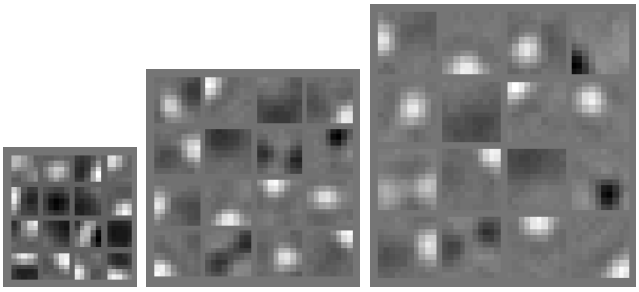
where Z is the partition function that accumulates over all possible visible and hidden states:

$$Z = \sum_{v,h} \exp^{-E(v,h)} \tag{3}$$

The probability of the network assigned to a visible state is given by summing over all possible hidden states:

$$p(v) = \frac{1}{Z}\sum_{h} \exp^{-E(v,h)} \tag{4}$$

The goal of the training process is to increase the probability of the visible state $p(v)$ by adjusting the weights and biases. In other words, the network learns the pattern of the dataset based on training samples. The contrastive divergence algorithm [19], [20] is used to derive a set of weight and bias configurations so that the probabilities assigned to the training images are raised. And the derived weight matrices of the GRBM networks are thus the feature extractors.



(a) (b) (c)

Fig. 2. Feature extractors learned from various input size. (a)16 filers from 5×5 patch. (b)16 filers from 8×8 patch. (c)16 filers from 10×10 patch.

The size of lung image patch to be classified is 32×32 pixels. However unlike human face or hand written characters, there is no obvious structure on 32×32 patches of lung image. Therefore, instead of training the networks using 32×32 patches as inputs, we use smaller patches to capture features of the finer structures within the 32×32 patches. And instead of using single-scale patches, it is intuitive to consider a multi-scale approach to represent image features at different resolutions. Three networks are thus trained using random sampled input patches of size 5×5, 8×8 and 10×10 respectively. This generates a multi-scale network capturing features of different sizes. The training patch sizes are chosen based on experimental results. All networks are configured with 16 hidden neurons. We choose such a small size of hidden neuron number to avoid over-fitting problem during GRBM training and the subsequent training for classifiers. The feature extractors learned from the contrastive divergence training process can be visualized in Fig. 2.

B. Convolutional Feature Extraction

With the feature extractors generated using the above unsupervised learning algorithm, an image patch is then transformed into a feature vector in the following way.

First, let X represent an image patch with P pixels x_p: $X = \{x_p : p = 1, ..., P\}$, and $F^{s,n}$ represent a feature extractor of scale s and index of hidden neuron n. A group of $S \times N$ feature maps $\{M^{s,n}(X) : s = 1, .., S, n = 1, ..., N\}$ are then created for image patch X by convolving it with the set of feature extractors $\{F^{s,n} : s = 1, .., S, n = 1, ..., N\}$:

$$M^{s,n}(X) = X * F^{s,n} \tag{5}$$

Then, a feature element $g^{s,n}(X)$ is computed by summing all values in one feature map:

$$g^{s,n}(X) = \sum_{k} M_k^{s,n}(X), \forall M_k^{s,n}(X) \in M^{s,n}(X) \tag{6}$$

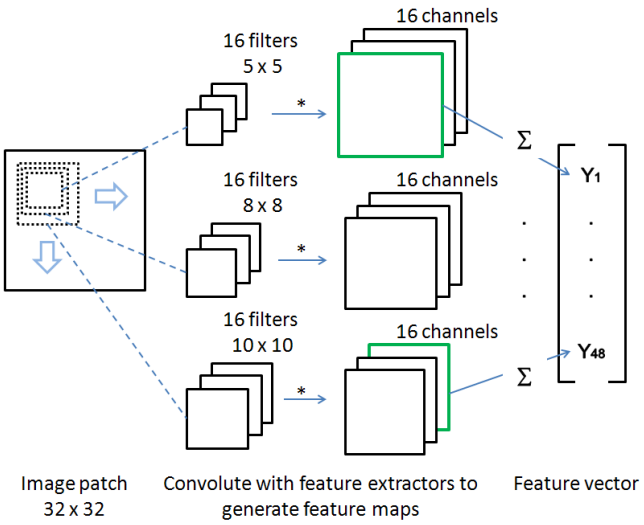


Fig. 3. Illustration of the proposed feature vector computation. Three groups of feature extractors of different sizes are convolved with the 32×32 input patch, to produce 48 feature maps. Sum of pixel values in each feature map is used as a feature element. The concatenation of all feature elements as a 48-dimensional feature vector is used for SVM classification.

where $M_k^{s,n}(X)$ denotes a value in the feature map $M^{s,n}(X)$. The concatenation of all feature elements $g^{s,n}(X)$ is thus the feature vector of image patch X:

$$G(X) = \{g^{1,1}(X), ..., g^{S,N}(X)\} \tag{7}$$

and the feature vector $G(X)$ is of $S \times N$ dimensions.

In this study, since we have a three-scale GRBM network each with 16 hidden neurons, i.e. $S = 3$ and $N = 16$, the feature dimension is thus $3 \times 16 = 48$. This is independent of the image patch size, which is set to 32×32 following the convention used in [7]. The feature dimension is also considered quite small, compared to the popular hand-crafted feature descriptors such as SIFT and LBP. The process of feature extraction based on learned feature extractors is visualized in Fig. 3.

C. Classification

The image patch X is finally classified into one of the five tissue categories based on its feature vector $G(X)$. A linear-kernel support vector machine (SVM) is applied for effective classification using the LIBSVM package [21]. The training is conducted following the leave-one-subject-out scheme. Therefore, no patches from the same subject are used during training to avoid over-fitting for the testing patches.

III. Experimental Results

A. Datasets

The publicly available database of ILD cases [22] is used in this study. The database contains 113 sets of high-resolution CT (HRCT) images. Each slice contains 512×512 pixels. A set of 2062 2D regions of interest (ROIs) are also provided, each with its tissue type annotated. Among the total 16 abnormal tissue patterns, four of them are commonly seen in ILD patients, and have been mostly studied in the past [22]: emphysema (E), ground glass (G), fibrosis (F) and micronodules (M). Together with the normal tissue (N), the method is thus expected to classify five tissue types.

TABLE I
Summary of the dataset used.

Tissue type	# Images	# Patches
Normal (N)	14	4348
Emphysema (E)	9	1047
Ground glass (G)	29	1953
Fibrosis (F)	37	2591
Micronodule (M)	18	6281

The images are divided into half-overlapping image patches of 32×32 pixels. An image patch with less than 75% of its pixels falling inside of an annotated ROI is filtered out from the experiment; and the patch-wise classification performance is evaluated for all the valid patches [7]. The experiment is thus performed on 92 HRCT image sets with a total of 16220 image patches. The numbers of images and patches of each tissue type are summarized in Table I. Note that one image set might contain multiple tissue types, hence the sum of images is larger than the actual number of images used for experiment.

B. Classification Results

The classification results using the proposed method are shown in Table II. Good classification accuracy of 84% is obtained on micronodules. The results also show that there are relatively large degree of confusions between normal and emphysema types, and between ground glass and fibrosis types. Such confusions are likely due to high level of similarity between the two pairs of tissue patterns, as can be seen from Fig. 1.

The parameter settings in this study, mainly the variable patch sizes used in the multi-scale GRBM network, are then evaluated. It is found that by using only a single-scale network, the learned feature vector is most discriminative with 5×5 input patches for extractor learning. By incorporating a multi-scale design, the classification performance further improves by about 3% recall and 4% precision. The results suggest that using features learnt and extracted from different scales helps to improve classification performance. It is also worth mentioning that while such parameters are empirical to the problem domain, the proposed feature learning method is general to other imaging applications and the parameters would be easily tuned.

We then compare the proposed learned feature vector with several popular feature descriptors, including (i) grayscale histogram with 64 bins (Hist); (ii) LBP with 3 resolutions and rotation invariance; (iii) HOG with 9 orientations and 4 cells; (iv) SIFT with keypoint placing at the center of the patch; and (v) concatenation of these four descriptors (H+L+H+S). For each descriptor (i)–(iv), the required parameters are chosen empirically to achieve the best feature discriminative power, and polynomial-kernel SVM is found most suitable. As

TABLE II

THE CONFUSION MATRIX OF TISSUE PATTERN CLASSIFICATION.

Ground Truth	Prediction				
	N	E	G	F	M
N	0.76	0.15	0.03	0.03	0.04
E	0.26	0.67	0	0.07	0.00
G	0.04	0.02	0.70	0.12	0.11
F	0.05	0.03	0.10	0.74	0.08
M	0.01	0.01	0.05	0.10	0.84

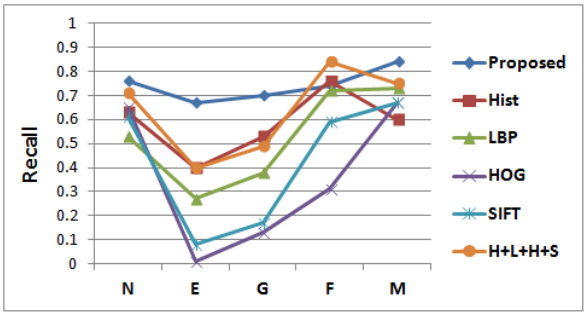


Fig. 4. The classification recall comparing various feature descriptors.

shown in Fig. 4 and 5, the proposed feature vector achieves higher classification recall and precision than the compared descriptors. It is especially encouraging that while H+L+H+S involves combination of different and complementary types of feature descriptors, the proposed learning approach delivers higher performance. Such an approach thus avoids the empirical nature of feature design that is commonly employed, yet is able to discover the intrinsic image features useful for classification.

IV. CONCLUSIONS

In this work, we have designed an image patch classification method based on fully automatic feature learning. With a multi-scale GRBM network, feature extractors of different scales are learned from raw image patches. Then the learned feature extractors are convoluted with image patches to produce the feature vectors for SVM classification. The test results on ILD HRCT lung images show that the proposed method was effective in automatic extraction of high quality image features for lung disease classification task; and higher classification performance was observed comparing to various feature descriptors that are currently widely popular in many imaging applications.

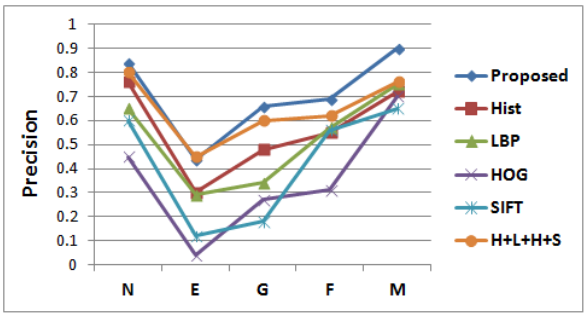


Fig. 5. The classification precision comparing various feature descriptors.

REFERENCES

[1] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman, "MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 464–475, 2006.

[2] L. Sorensen, S. B. Shaker, and M. de Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, 2010.

[3] A. Depeursinge, A. Foncubierta-Rodriguez, D. V. de Ville, and H. Muller, "Lung texture classification using locally-oriented riesz components," *in MICCAI LNCS*, vol. 6893, pp. 231–238, 2011.

[4] Y. Song, W. Cai, and D. D. Feng, "Global context inference for adaptive abnormality detection in PET-CT images," *in Proc. ISBI*, pp. 482–485, 2012.

[5] C. Jacobs, C. I. Sanchez, S. C. Saur, T. Twellmann, P. A. de Jong, and B. van Ginneken, "Computer-aided detection of ground glass nodules in thoracic CT images using shape, intensity and context features," *in MICCAI LNCS*, vol. 6893, pp. 207–214, 2011.

[6] Y. Song, W. Cai, J. Kim, and D. D. Feng, "A multistage discriminative model for tumor and lymph node detection in thoracic images," *IEEE Trans. Med. Imag.*, vol. 31, no. 5, pp. 1061–1075, 2012.

[7] A. Depeursinge, D. V. de Ville, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Muller, "Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 665–675, 2012.

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

[9] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification detection in thoracic images," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 797–808, 2013.

[10] D. G. Lowe, "Distinctive iamge features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE. Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *in Proc. CVPR*, pp. 886–893, 2005.

[13] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Master's thesis, Department of Computer Science, University of Toronto*, 2009.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *in Proc. NIPS*, pp. 1–9, 2012.

[15] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[16] A. Krizhevsky and G. Hinton, "Using very deep autoencoders for content-based image retrieval," *in Proc. ESANN*, pp. 489–494, 2011.

[17] W. R. Webb, N. L. Muller, and D. P. Naidich, *High-resolution CT of the lung*. Lippincott Williams Wilkins, 2008.

[18] N. Le Roux and Y. Bengio, "Representational power of restricted boltzmann machines and deep belief networks," *Neural Computation*, vol. 20, no. 6, pp. 1631–1649, 2008.

[19] G. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[20] G. Hinton, "A practical guide to training restricted boltzmann machines," *Technical Report UTML TR 2010003, University of Toronto*, 2010.

[21] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machine," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[22] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Muller, "Building a reference multimedia database for interstitial lung diseases," *Comput. Med. Imaging Graph.*, vol. 36, no. 3, pp. 227–238, 2012.