

# Chapter 7

## Interstitial Lung Diseases via Deep Convolutional Neural Networks: Segmentation Label Propagation, Unordered Pooling and Cross-Dataset Learning

Mingchen Gao, Ziyue Xu and Daniel J. Mollura

**Abstract** Holistically detecting interstitial lung disease (ILD) patterns from CT images is challenging yet clinically important. Unfortunately, most existing solutions rely on manually provided regions of interest, limiting their clinical usefulness. We focus on two challenges currently existing in two publicly available datasets. First of all, missed labeling of regions of interest is a common issue in existing medical image datasets due to the labor-intensive nature of the annotation task which requires high levels of clinical proficiency. Second, no work has yet focused on predicting more than one ILD from the same CT slice, despite the frequency of such occurrences. To address these limitations, we propose three algorithms based on deep convolutional neural networks (CNNs). The differences between the two main publicly available datasets are discussed as well.

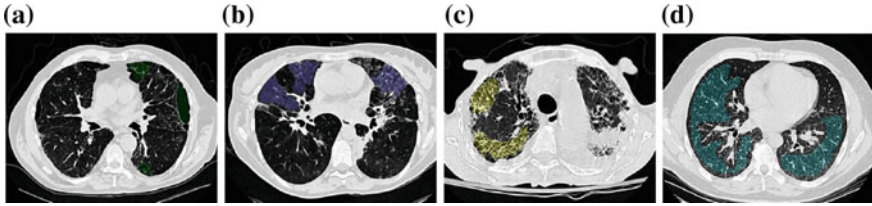
### 7.1 Introduction

Interstitial lung disease (ILD) refers to a group of more than 150 chronic lung diseases that causes progressive scarring of lung tissues and eventually impairs breathing. The gold standard imaging modality for diagnosing ILD patterns is high-resolution computed tomography (HRCT) [1, 2]. Figures 7.1 and 7.2 depict examples of the most typical ILD patterns.

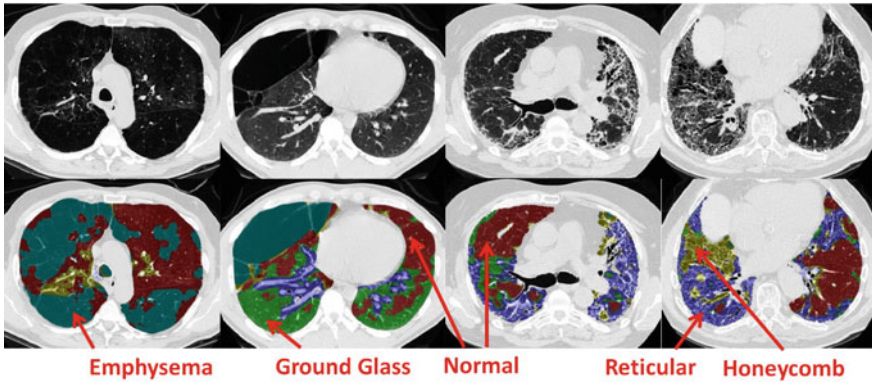
Automatically detecting ILD patterns from HRCT images would help the diagnosis and treatment of this morbidity. The majority of previous work on ILD detection is limited to patch-level classification, which classifies small patches from manu-

---

M. Gao (✉) · Z. Xu · D.J. Mollura  
Department of Radiology and Imaging Sciences,  
National Institutes of Health (NIH), Bethesda, MD 20892, USA  
e-mail: mingchen.gao@nih.gov



**Fig. 7.1** Visual aspects of the most common lung tissue patterns in HRCT axial slices in UHG dataset. Infected regions are annotated with different colors in the publicly available dataset [1]. **a** Emphysema (EM). **b** Ground Glass (GG). **c** Fibrosis (FB). **d** Micronodules (MN)



**Fig. 7.2** Examples of ILD patterns. Every voxel in the lung region is labeled as healthy or one of the four ILDs: ground glass, reticular, honeycomb or emphysema. The *first row* is the lung CT images. The *second row* is their corresponding labelings

ally generated regions of interest (ROIs) into one of the ILDs. Approaches include restricted Boltzmann machines [3], convolutional neural networks (CNNs) [4], local binary patterns [5, 6] and multiple instance learning [7]. An exception to the patch-based approach is the recent work of Gao et al. [8], which investigated a clinically more realistic scenario for ILD classification, assigning a *single* ILD label to any holistic two-dimensional axial CT slice without any pre-processing or segmentation. Although holistic detection is more clinically desirable, the underlying problem is much harder without knowing the ILD locations and regions a priori. The difficulties lie on several aspects, which include the tremendous amount of variation in disease appearance, location, and configuration and also the expense required to obtain delicate pixel-level ILD annotations of large datasets for training.

We would like to tackle these challenges from several aspects. There are two main publicly available datasets for CT imaging based ILD classification [1, 2]. The first one is from University Hospital of Geneva (UHG), with limited annotations [1]. As

shown in Fig. 7.1, we find that only less than 15% of the lung region in the pixel coverage measure is labeled, which significantly restricts the number of available training image pixels, patches or data. Assigning semantic labels to each pixel of a CT image is tedious, time-consuming and error-prone, or simply is not affordable and feasible for a large amount of patients. Supervised learning, the most common technique for integrating domain knowledge, usually needs the manual annotation from expensive medical experts to assign a label to every pixel. This hinders the learning scalability both in the amount of training data and in the number of classes. On the other hand, we have witnessed the success of many applications in computer vision and medical imaging analysis when a large-scale well-annotated dataset is available [9].

Therefore automated image annotation or labeling methods are needed to assist doctors during the labeling process. In an ideal framework, computerized algorithms would complete most of the tedious tasks, and doctors would merely validate and fine-tune the results, if necessary. We propose a segmentation propagation algorithm that combines the cues from the initial or partial manual annotations, deep convolutional neural networks (CNN) based single pixel classification and formulate into a constrained dense fully connected conditional random field (CRF) framework [10]. Our main technical novelties are the **constrained unary** (*manually labeled pixels are hard-enforced with their original ILD image labels; pixels outside of lung are considered as hard-encoded background; unlabeled lung pixels are the key subjects to be assigned ILD labels using our method*) and **pairwise terms** (*message passing is only allowed for any pair of lung image pixels*) and their efficient implementation in [11]. The proposed method is applicable to other problems as a generic semi-supervised image segmentation solution. This work is partially inspired by interactive graph-cut image segmentation [12] and automatic population of pixelwise object-background segmentation from manual annotations on ImageNet database [13].

Another challenge we would like to solve is detecting multiple ILDs simultaneously without the locations which has not been addressed by previous studies [3, 4, 8, 14], including that of Gao et al. [8], which all treat ILD detection as a single-label classification problem. When analyzing the Lung Tissue Research Consortium (LTRC) dataset [2], the most comprehensive lung disease image database with detailed annotated segmentation masks, we found that there are significant amounts of CT slices associated with two or more ILD labels. For this reason, and partially inspired by the recent natural image classification work [15], we model the problem as multi-label regression and solve it using a CNN [16]. We note that multi-label regression has also been used outside of ILD contexts for heart chamber volume estimation [17, 18]. However, this prior work used hand-crafted features and random-forest-based regression, whereas we employ learned CNN-based features, which have enjoyed dramatic success in recent years over hand-crafted variants [9]. Thus, unlike prior ILD detection work [3–6, 8], our goal is to detect multiple ILDs on holistic CT slices simultaneously, providing a more clinically useful tool.

While CNNs are powerful tools, their feature learning strategy is not invariant to the spatial locations of objects or textures within a scene. This order-sensitive feature encoding, reflecting the spatial layout of the local image descriptors, is effective in

object and scene recognition. However, it may not be beneficial or even be counter-productive for texture classification. The spatial encoding of order-sensitive image descriptors can be discarded via unordered feature encoders such as bag of visual words (BoVW), Fisher vectors (FV) [19], or aggregated by order-sensitive spatial pyramid matching (SPM). Given the above considerations, we enhance our CNN-regression approach using spatial-invariant encodings of feature activations for multi-label multi-class ILD detection.

## 7.2 Methods

Our algorithms are of two respects. The first one would extend the limited labels in the UHG dataset to every pixel in the lung region. Specifically, we explore the possible ways to **propagate** the ILD labels from the limited manually drawn regions to the whole lung slice as a **per-pixel multi-class image segmentation and labeling**. The fully connected conditional random field builds the pairwise potentials densely on all pairs of pixels in the image. The CRF optimization is conducted as message passing that can naturally handle multi-class labeling. The CRF unary energies are learned from CNN-based image patch labeling. Ground truth labels by radiologists are also integrated into the CRF as hard constraints. The proposed algorithm is evaluated on a publicly available dataset [1] and the segmentation/labeling results are validation by an expert radiologist.

The second method focuses on predicting multiple labels simultaneously on the same slice and is tested on the LTRC dataset. We propose two variations of multi-label deep convolutional neural network regression (MLCNN-R) models to address the aforementioned challenges. First, an end-to-end CNN network is trained for multi-label image regression. The loss functions are minimized to estimate the actual pixel numbers occupied per ILD class or the binary [0,1] occurring status. Second, the convolutional activation feature maps at different network depths are spatially aggregated and encoded through the FV [19] method. This encoding removes the spatial configurations of the convolutional activations and turns them into location-invariant representations. This type of CNN is also referred as FV-CNN. The unordered features are then trained using a multivariate linear regressor (Mvregress function in MATLAB) to regress the numbers of ILD pixels or binary labels. Our proposed algorithm is demonstrated using the LTRC ILD dataset [2], composed of 533 patients. Our experiments use fivefold cross-validation (CV) to detect the most common ILD classes of ground glass, reticular, honeycomb and emphysema. Experimental results demonstrate the success of our approach in tackling the challenging problem of multi-label multi-class ILD classification.

### 7.2.1 Segmentation Label Propagation

We formulate the segmentation problem as a maximum a posteriori (MAP) inference in a CRF defined over pixels. To take into account of long-range image interactions, an efficient fully connected CRF method is adapted [11].

The CRF representation captures the conditional distribution of the class labeling  $X$  given an image  $I$ . Consider a random field  $X$  defined over a set of variables  $\{X_1, \dots, X_N\}$ , with  $X_i \in X$  being associated with every pixel  $i \in V$  and taking a value from the label set  $L = \{l_1, \dots, l_K\}$  of label categories. The labeling of  $X$  from images is obtained with a maximum a posterior (MAP) estimation of the following conditional log-likelihood:

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j), \quad (7.1)$$

where  $i$  and  $j$  range from 1 to  $N$ .  $\psi_u(x_i)$ , the unary potential, is computed independently by the convolutional neural network classifier for each pixel/patch. The pairwise potentials in our model have the form

$$\begin{aligned} \psi_p(x_i, x_j) &= u(x_i, x_j) \sum_{m=1}^K k(f_i, f_j) \\ &= u(x_i, x_j) \sum_{m=1}^K \omega^{(m)} k^{(m)}(f_i, f_j). \end{aligned} \quad (7.2)$$

Each  $k^{(m)}$  is a Gaussian kernel

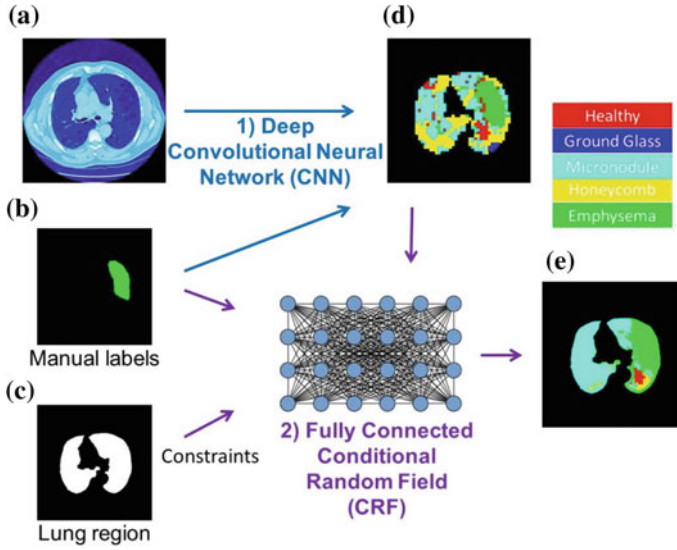
$$k^{(m)}(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i - f_j)^T \Lambda^{(m)}(f_i - f_j)\right), \quad (7.3)$$

where the vectors  $f_i$  and  $f_j$  are feature vectors for pixels  $i$  and  $j$  in an arbitrary feature space;  $u$  is a label compatibility function; and  $\omega^{(m)}$  are linear combination weights.

In our implementation, we use two-kernel potentials, defined in terms of the CT attenuation vectors  $I_i$  and  $I_j$  (introduced in [8]) and positions  $p_i$  and  $p_j$ :

$$\begin{aligned} k(f_i, f_j) &= \omega^{(1)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right) \\ &\quad + \omega^{(2)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right). \end{aligned} \quad (7.4)$$

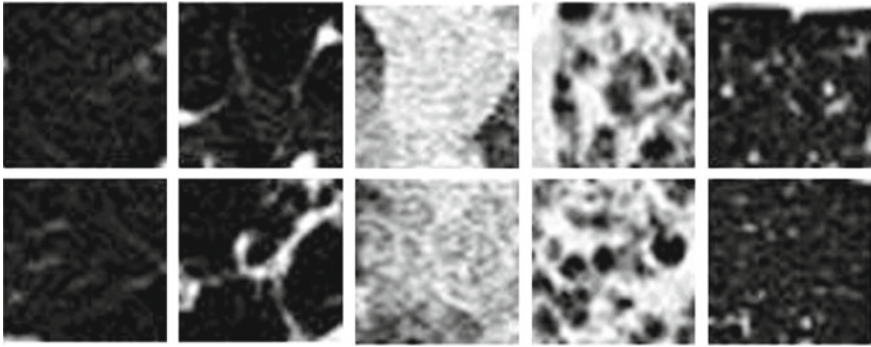
The first term presents the appearance kernel, which represents the affinities of nearby pixels with similar CT attenuation patterns. The second term presents the smoothness



**Fig. 7.3** Major intermediate results. **a** Three channels of different HU windows illustrated as RGB values. **b** Annotated ROI. **c** Annotated lung mask. **d** CNN classifier at a spatial interval of 10 pixels. **e** Final result integrating image features, unary prediction and hard constraints

kernel, which removes small isolated regions. The parameters  $\theta_\alpha$ ,  $\theta_\beta$  and  $\theta_\gamma$  are used to control the degree of nearness and similarity. The inference of fully connected conditional random field is efficiently approximated by an iterative message passing algorithm. Each iteration performs a message passing, a compatibility transform and a local update. The message passing can be performed using Gaussian filtering in feature space. The complexity of the algorithm reduces from quadratic to linear in the number of variables  $N$  and sublinear in the number of edges in the model (Fig. 7.3).

**Unary classifier using Convolutional Neural Network:** At present, there is a vast amount of relevant work on computerized ILD pattern classification. The majority focuses on image patch based classification using hand-crafted [5, 6, 20] or CNN learned features [8]. We use the CNN-based CRF unary classifier because of its state-of-the-art performance: classification accuracy of 87.9% reported in [8]. To facilitate comparison, five common ILD patterns are studied in this work: healthy, emphysema, ground glass, fibrosis and micronodules (Fig. 7.4). Image patches of size  $32 \times 32$  pixels within the ROI annotations of these five classes, are extracted to train a deep CNN classifier. The well known CNN AlexNet model [9] trained on ImageNet is used to fine-tune on our image patch dataset.  $32 \times 32$  pixel images patches are rescaled to  $224 \times 224$  and three channels of different HU windows [8] are generated to accommodate the CNN model.



**Fig. 7.4** Examples of  $32 \times 32$  patches for each ILD category. From *left to right* columns: healthy, emphysema, ground glass, fibrosis and micronodules

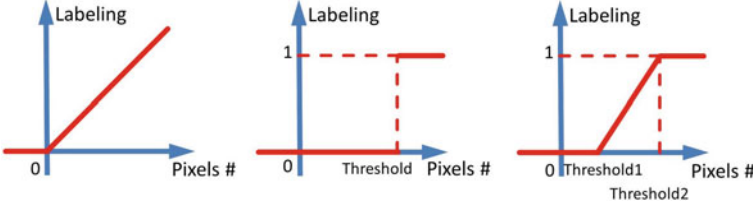
**Hard Constraints:** The image labels given by radiologists from the dataset [1] are considered as ground truth and ought to be strictly enforced. During each CRF message passing iteration, the hard constrained image regions are hard-reset to be consistent with their ground truth labels. In such cases, there is only message passing out of the hard constrained regions towards unlabeled lung image pixels. On the other hand, we assume that the ILD label map should only be inferred within the lung field. The lung field CRF ILD labeling is conditionally independent of image pixel patterns outside the lung mask. In implementation of Eq. 7.4, the parameters  $\theta_\alpha$ ,  $\theta_\beta$  and  $\theta_\gamma$  are set to be a small constant (0.001) for any pixel pairs linking lung and non-lung spatial indexes  $(p_i, p_j)$  so the associated  $k(f_i, f_j)$  has a numerically vanishing value, which is equivalent to no message passing.

Specifically, we explore the possible ways to **propagate** the ILD labels from the limited manually drawn regions to the whole lung slice as a **per-pixel multi-class image segmentation and labeling**. The fully connected conditional random field builds the pairwise potentials densely on all pairs of pixels in the image. The CRF optimization is conducted as message passing that can naturally handle multi-class labeling. The CRF unary energies are learned from CNN based image patch labeling. Ground truth labels by radiologists are also integrated into the CRF as hard constraints. The proposed algorithm is evaluated on a publicly available dataset [1] and the segmentation/labeling results are validation by an expert radiologist.

### 7.2.2 Multi-label ILD Regression

Our algorithm contains two major components: (1) we present a squared  $L2$  loss function based multi-label deep CNN regression method to estimate either the observable ILD areas (in the numbers of pixels), or the binary  $[0,1]$  status of “non-appearing” or “appearing”. This regression-based approach allows our algorithm to naturally





**Fig. 7.5** Three functions for mapping the number of pixels to the regression label

preserve the co-occurrence property of ILDs in CT imaging. (2) CNN activation vectors are extracted from convolutional layers at different depths of the network and integrated using a Fisher vector feature encoding scheme in a spatially unordered manner, allowing us to achieve a location-invariant deep texture description. ILD classes are then discriminated using multivariate linear regression.

**CNN Architecture:** Deep CNN regression is used to calculate the presence or the area of spatial occupancy for IDL in the image, where multiple pathology patterns can co-exist. The squared  $L_2$  loss function is adopted for regression [15] instead of the more widely used softmax or logistic-regression loss for CNN-based classification [4, 8, 9]. There are multiple ways to model the regression labels for each image. One straightforward scheme is to count the total number of pixels annotated per disease to represent its severity, e.g., Fig. 7.5 left. We can also use a step function to represent the presence or absence of the disease, as shown in Fig. 7.5 middle, where the stage threshold  $T$  may be defined using clinical knowledge. For any ILD in an image, if its pixel number is larger than  $T$ , the label is set to be 1; otherwise as 0. A more sophisticated model would have a piecewise linear transform function, mapping the pixel numbers towards the range of  $[0,1]$  (Fig. 7.5 right). We test all approaches in our experiments.

Suppose that there are  $N$  images and  $c$  types of ILD patterns to be detected or classified, the label vector of the  $i^{th}$  image is represented as a  $c$ -length multivariate vector  $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{ic}]$ . An all-zero labeling vector indicates that the slice is healthy or has no targeted ILD found based on the ground truth annotation. The  $L_2$  cost function to be minimized is defined as

$$L(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \sum_{i=1}^N \sum_{k=1}^c (y_{ik} - \hat{y}_{ik})^2, \quad (7.5)$$

There are several successful CNN structures from previous work, such as AlexNet [9] and VGGNet [21]. We employ a variation of AlexNet, called **CNN-F** [22], for a trade-off between efficiency and performance based on the amount of available annotated image data. **CNN-F** contains five convolutional layers, followed by two fully connected (FC) layers. We set the last layer to the squared  $L_2$  loss function. Four classes of ILDs are investigated in our experiments: ground glass, reticular, honeycomb and emphysema (other classes have too few examples in the LTRC



database [2]). The length of  $\mathbf{y}_i$  is  $c = 4$  to represent these four ILD classes. Based on our experience, random initialization of the CNN parameters worked better than ImageNet pre-trained models. Model parameters were optimized using stochastic gradient descent.

**Unordered Pooling Regression via Fisher Vector Encoding:** In addition to CNN-based regression, we also test a spatially invariant encoding of CNN feature activations. We treat the output of each  $k$ -th convolutional layer as a 3D descriptor field  $\mathbf{X}_k \in \mathbb{R}^{W_k \times H_k \times D_k}$ , where  $W_k$  and  $H_k$  are the width and height of the field and  $D_k$  is the number of feature channels. Therefore, the whole deep feature activation map is represented by  $W_k \times H_k$  feature vectors and each feature vector is of dimension  $D_k$ .

We then invoke FV encoding [19] to remove the spatial configurations of total  $W_k \times H_k$  vectors per activation map. Following [19], each descriptor  $x_i \in \mathbf{X}_k$  is soft-quantized using a Gaussian mixture model. The first- and second-order differences  $(u_{i,m}^T, v_{i,m}^T)$  between any descriptor  $x_i$  and each of the Gaussian cluster mean vectors  $\{\mu_m\}$ ,  $m = 1, 2, \dots, M$  are accumulated in a  $2MD_k$ -dimensional image representation:

$$\mathbf{f}_i^{FV} = [u_{i,1}^T, v_{i,1}^T, \dots, u_{i,M}^T, v_{i,M}^T]^T. \quad (7.6)$$

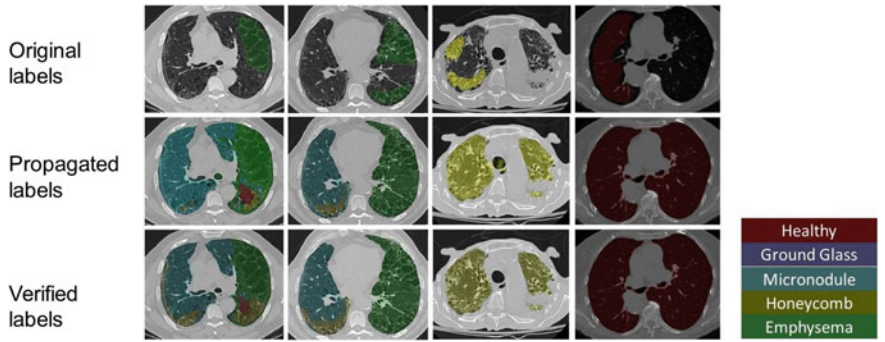
The resulting FV feature encoding results in very high  $2MD_k$  (e.g.,  $M = 32$  and  $D_k = 256$ ) dimensionality for deep features of  $\mathbf{X}_k$ . For computational and memory efficiency, we adopt principal component analysis (PCA) to reduce the  $\mathbf{f}_i^{FV}$  features to a lower-dimensional parameter space. Based on the ground truth label vectors  $\mathbf{y}_i$ , multivariate linear regression is used to predict the presence or non-presence of ILDs using the low-dimensional image features  $PCA(\mathbf{f}_i^{FV})$ .

## 7.3 Experiments and Discussion

The proposed algorithms are testing on two datasets, UHG and LTRC, respectively. The training folds and testing fold are split at the patient level to prevent overfitting (i.e., no CT slices from the same patient are used for both training and validation). CNN training was performed in MATLAB using MatConvNet [23] and was run on a PC with an Nvidia Tesla K40 GPU.

### 7.3.1 Segmentation Label Propagation

UHG dataset [1] is used for training and validation under twofold cross-validation for the segmentation propagation problem. ROIs of total 17 different lung patterns and lung masks are also provided along with the dataset. Figure 7.6 shows the annotation provided by the dataset, the labeling obtained from our algorithm and the ground



**Fig. 7.6** Comparison between the annotation provided by the UHG dataset, our labeling results, and the final annotation verified by experienced radiologists

**Table 7.1** Confusion matrix, precision, recall and F-score of ILD pattern labeling

Ground truth	Prediction				
	NM	EM	GG	FB	MN
NM	<b>0.9792</b>	0.0067	0.0029	0.0020	0.0092
EM	0.2147	<b>0.7389</b>	0	0.0170	0.0294
GG	0	0	<b>1.0000</b>	0	0
FB	0.0414	0.0271	0.0118	<b>0.8046</b>	0.1151
MN	0.0007	0.0013	0.0174	0.0058	<b>0.9748</b>
Precision	0.9500	0.9320	0.8175	0.9060	0.9666
Recall	0.9792	0.7389	1	0.8046	0.9748
F-scail	0.9644	0.8243	0.8996	0.8523	0.9707

truth validated by radiologists. In our implementation, we use the lung mask provided from the dataset. Please note that trachea is included in the lung mask provided from the dataset. This misleads our algorithm to give a prediction in the trachea region. A recent rough lung segmentation method [24] can be used to automate this process.

Quantitative evaluation is given in Table 7.1 with the total accuracy reaching 92.8%. More importantly, the amount of auto-annotated pixels is 7.8 times greater than the amount of provided annotation [1]. Thus the labeled training dataset [1] is significantly enlarged via segmentation label propagation. This data expansion is a critical contribution of this paper. The CRF solver is implemented in C++. The most time consuming part is the unary classification of densely sampled image patches. To speed up testing, a relatively coarse prediction map of image patches is sufficient. This map can be bi-linearly interpolated and later refined by the CRF pairwise constraints. In our implementation, we predict the labels of image patches at a spatial interval of 10 pixels. Parameters  $\theta_\alpha$ ,  $\theta_\beta$  and  $\theta_\gamma$  are set to be 80, 13, and 3 through a small calibration dataset within training. We set  $\omega^{(1)} = \omega^{(2)} = 1$ , which is found to work well in practice.

### 7.3.2 Multi-label ILD Regression

LTRC dataset [2] enjoys complete ILD labeling at the CT slice level [10]. We use the LTRC dataset to evaluate the algorithm detecting multiple labels simultaneously. Every pixel in the CT lung region is labeled as healthy or one of the four tissue types: ground glass, reticular, honeycomb or emphysema. Only 2D axial slices are investigated here, without taking successive slices into consideration. Many CT scans for ILD study have large inter-slice distances (for example 10 mm in [1]) between axial slices, making direct 3D volumetric analysis implausible. The original resolution of the 2D axial slices is  $512 \times 512$  pixels. All images are resized to the uniform size of  $214 \times 214$  pixels.

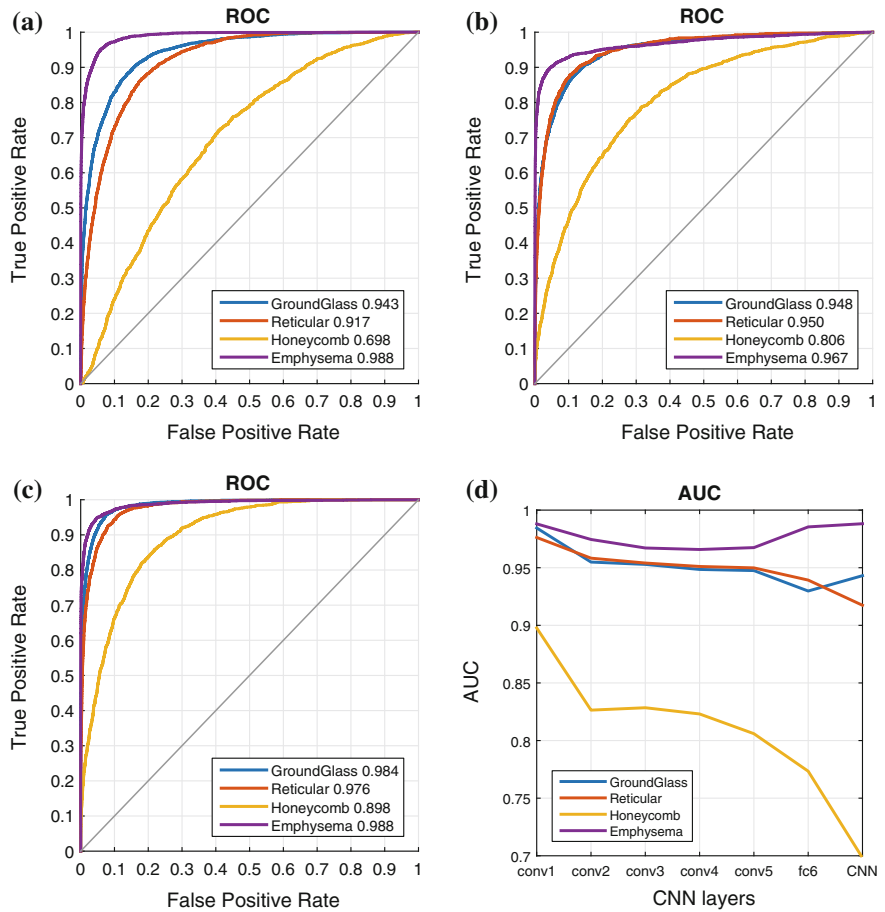
To conduct holistic slice based ILD classification [8], we first convert the pixelwise labeling into slice-level labels. There are 18883 slices in total for training and testing. Without loss of generality, if we set  $T = 6000$  pixels as the threshold to differentiate the presence or absence of ILDs, there are 3368, 1606, 1247 and 2639 positive slices for each disease, respectively. In total there are 11677 healthy CT images, 5675 images with one disease, 1410 images with two diseases, 119 images with three diseases, and 2 images with four diseases. We treat the continuous values after regression (in two types of pixel numbers or binary status) as “classification confidence scores”. We evaluate our method by comparing against ground truth ILD labels obtained from our chosen threshold.

Each ILD pattern is evaluated separately by thresholding the “classification confidence scores” from our regression models to make the binary presence or absence decisions. Classification receiver operating characteristic (ROC) curves can be generated in this manner. We experimented with Fig. 7.5’s three labeling converting functions. Regression using the ILD occupied pixel numbers or the binary status labels produced similar quantitative ILD classification results. However, the piecewise linear transformation did not perform well.

When constructing the FV-encoded features,  $f_i^{FV}$ , the local convolutional image descriptors are pooled into 32 Gaussian components, producing dimensionalities as high as 16K dimensions [19]. We further reduce the FV features to 512 dimensions using PCA. Performance was empirically found to be insensitive to the number of Gaussian kernels and the dimensions after PCA.

All quantitative experiments are performed under fivefold cross-validation. The training folds and testing fold are split at the patient level to prevent overfitting (i.e., no CT slices from the same patient are used for both training and validation). CNN training was performed in MATLAB using MatConvNet [23] and was run on a PC with an Nvidia Tesla K40 GPU. The training for one fold takes hours. The testing could be accomplished in seconds per image.

We show the ROC results directly regressed to the numbers of ILD pixels in Fig. 7.7. The area under the curve (AUC) values are marked in the plots. In Fig. 7.7d, AUC scores are compared among configurations using FV encoding on deep image features pooled from different CNN convolutional layers. Using activations based on the first fully connected layer (fc6) are also evaluated. Corresponding quantitative



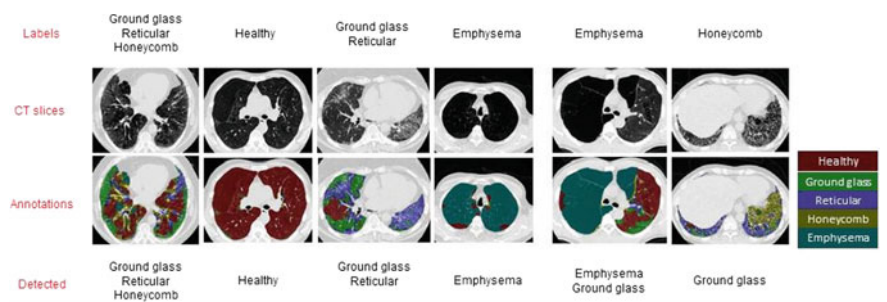
**Fig. 7.7** ILD detection results shown in ROC curves. Both CNN and FV-CNN regression are used to regress to the numbers of pixels. **a** Detection results of CNN regression. **b, c** Detection results of FV-CNN via the unordered feature pooling using conv5 and conv1 layer, respectively. **d** AUC versus FV pooling at different convolutional layers

results are shown in Table 7.2. Both deep regression models achieve high AUC values for all four major ILD patterns. FV unordered pooling operating on the first CNN convolutional layer **conv1** produces the overall best quantitative results, especially for Honeycomb. Despite residing in the first layer, the filters and activations on **conv1** are still part of a deep network since they are learned through back-propagation. Based on these results, this finding indicates that using FV encoding with deeply learned **conv1** filter activations is an effective approach to ILD classification.

Figure 7.8 presents some examples of successful and misclassified results. First four cases of examples are well successfully detected all types of ILD patterns. In the second to last, although it is marked as misclassified (compared to the ground truth

**Table 7.2** Quantitative results comparing the AUC between different layers. Both CNN and multi-variant linear regression regress to pixel numbers

Disease	Area under the curve (AUC)						
	conv1	conv2	conv3	conv4	conv5	fc6	CNN
Ground glass	<b>0.984</b>	0.955	0.953	0.948	0.948	0.930	0.943
Reticular	<b>0.976</b>	0.958	0.954	0.951	0.950	0.939	0.917
Honeycomb	<b>0.898</b>	0.826	0.828	0.823	0.806	0.773	0.698
Emphysema	<b>0.988</b>	0.975	0.967	0.966	0.967	0.985	0.988



**Fig. 7.8** Examples of correctly detected and misclassified ILD slices

binary labels with  $T = 6000$  pixels), our method finds and classifies emphysema and ground glass correctly that do occupy some image regions. These qualitative results visually confirm the high performance demonstrated by our quantitative experiments.

## 7.4 Conclusion

In this work, we present several solutions related to ILD pattern detections. The first segmentation label propagation method efficiently populates the labels from the annotated regions to the whole CT image slices. High segmentation/labeling accuracy are achieved. The amount of labeled training data in [1] is significantly expanded and will be publicly shared upon publication<sup>1</sup>.

We also present a new ILD pattern detection algorithm using multi-label CNN regression combined with unordered pooling of the resulting features. In contrast to previous methods, our method can perform multi-label multi-class ILD detection. Moreover, this is performed without the manual ROI inputs needed by much of the state-of-the-art [3–5]. We validate on a publicly available dataset of 533 patients using five-fold CV, achieving high AUC scores of 0.982, 0.972, 0.893 and 0.993 for Ground-Glass, Reticular, Honeycomb and Emphysema, respectively. Future work includes

<sup>1</sup><http://www.research.rutgers.edu/minggao>.

performing cross-dataset learning and incorporating weakly supervised approaches to obtain more labeled training data. Nonetheless, as the first demonstration of effective multi-class ILD classification, this work represents an important contribution toward clinically effective CAD solutions.

## References

1. Depeursinge A, Vargas A, Platon A, Geissbuhler A, Poletti P-A, Müller H (2012) Building a reference multimedia database for interstitial lung diseases. *CMIG* 36(3):227–238
2. Holmes III D, Bartholmai B, Karwoski R, Zavaletta V, Robb R (2006) The lung tissue research consortium: an extensive open database containing histological, clinical, and radiological data to study chronic lung disease. *Insight J*
3. van Tulder G, de Bruijne M (2016) Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines. *TMI*
4. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 35(5):1207–1216
5. Song Y, Cai W, Huang H, Zhou Y, Feng D, Wang Y, Fulham M, Chen M (2015) Large margin local estimate with applications to medical image classification. *TMI* 34(6):1362–1377
6. Song Y, Cai W, Zhou Y, Feng DD (2013) Feature-based image patch approximation for lung tissue classification. *TMI* 32(4):797–808
7. Hofmanninger J, Langs G (2015) Mapping visual features to semantic profiles for retrieval in medical imaging. In: *CVPR*, pp 457–465
8. Gao M, Bagci U, Lu L, Wu A, Buty M, Shin H-C, Roth H, Papadakis GZ, Depeursinge A, Summers RM, et al (2016) Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis* 1–6
9. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: *NIPS*, pp 1097–1105
10. Gao M, Xu Z, Lu L, Nogues I, Summers R, Mollura D (2016) Segmentation label propagation using deep convolutional neural networks and dense conditional random field. In: *IEEE international symposium on biomedical imaging*
11. Krähenbühl P, Koltun V (2011) Efficient inference in fully connected CRFs with Gaussian edge potentials. In: *NIPS*, pp 109–117
12. Boykov YY, Jolly M-P (2001) Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: *Eighth IEEE international conference on proceedings*, vol 1. IEEE, pp 105–112
13. Guillaumin M, Küttel D, Ferrari V (2014) Imagenet auto-annotation with segmentation propagation. *IJCV* 110(3):328–348
14. Gong Y, Wang L, Guo R, Lazebnik S (2014) Multi-scale orderless pooling of deep convolutional activation features. In: *ECCV 2014*. Springer, pp 392–407
15. Wei Y, Xia W, Huang J, Ni B, Dong J, Zhao Y, Yan S (2014) CNN: single-label to multi-label. *arXiv preprint arXiv:1406.5726*
16. Gao M, Xu Z, Lu L, Harrison AP, Summers RM, Mollura DJ (2016) Multi-label deep regression and unordered pooling for holistic interstitial lung disease pattern detection. *Machine learning in medical imaging*
17. Zhen X, Islam A, Bhaduri M, Chan I, Li S (2015) Direct and simultaneous four-chamber volume estimation by multi-output regression. In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp 669–676
18. Zhen X, Wang Z, Islam A, Bhaduri M, Chan I, Li S (2014) Direct estimation of cardiac bi-ventricular volumes with regression forests. In: *MICCAI*, pp 586–593

19. Perronnin F, Sánchez J, Mensink T (2010) Improving the fisher kernel for large-scale image classification. In: ECCV, pp 143–156
20. Depeursinge A, Van de Ville D, Platon A, Geissbuhler A, Poletti P-A, Muller H (2012) Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames. *IEEE Trans Inf Technol Biomed* 16(4):665–675
21. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
22. Chatfield K, Simonyan K, Vedaldi A, Zisserman A (2014) Return of the devil in the details: delving deep into convolutional nets. arXiv preprint [arXiv:1405.3531](https://arxiv.org/abs/1405.3531)
23. Vedaldi A, Lenc K (2015) MatConvNet: convolutional neural networks for MATLAB. In: *Proceedings of the 23rd annual ACM conference on multimedia conference*. ACM, pp 689–692
24. Mansoor A, Bagci U, Xu Z, Foster B, Olivier KN, Elinoff JM, Suffredini AF, Udupa JK, Mollura DJ (2014) A generic approach to pathological lung segmentation