Deep learning framework for infiltrative lung disease classification

S. Tarando, C. Fetita, Y-W Kim, H. Cho and P-Y Brillet

1 ARTEMIS Department, TELECOM SudParis, Institut Mines-Telecom, Evry, France
2 CNRS UMR8145 (MAP5), UMR5157 (SAMOVAR)
3 Université Paris13, Paris, France
4 Avicenne Hospital, AP-HP, Bobigny, France
5 Pitié-Salpêtrière Hospital, AP-HP, Paris, France

sebastian.tarando@telcom-sudparis.eu

Abstract
Infiltrative lung diseases enclose a large group of irreversible lung disorders which require regular follow-up with CT imaging. This paper addresses an automated quantitative assessment of different disorders based on lung texture classification. The proposed approach exploits a cascade of convolutional neural networks and a specific preprocessing of input data based on locally connected filtering. The classification targeting the whole lung parenchyma achieves an average of 84% accuracy (75.8% for normal, 90% for emphysema and fibrosis, 81.5% for ground glass).

Keywords
fibrosis, emphysema, ground glass, convolutional neural networks, connected filtering

1 Introduction
Infiltrative lung diseases (ILD) enclose a large group of irreversible lung disorders which require regular follow-up with computed tomography (CT) imaging. A quantitative assessment is mandatory to establish the (regional) disease progression and/or the therapeutic impact. This implies the development of automated computer-aided diagnosis (CAD) tools for pathological lung tissue segmentation, problem addressed as pixel-based texture classification. Traditionally, such classification relies on a two-dimensional analysis of axial CT images by means of handcrafted features [1]. Recently, the use of deep learning techniques, especially Convolutional Neural Networks (CNNs) for visual tasks, has shown great improvements with respect to handcrafted heuristics-based methods. Deep learning methods have the advantage that the features on which relies the classification are learnt directly from the raw image data [2][3]. We have previously demonstrated the effectiveness of CNNs in boosting the performance of an ILD CAD system even with reduced training dataset [4]. In [8], a specific CNN for lung texture classification was developed showing good performance in ILD detection, considered as the state-of-the-art approach. Note however that most of the systems mentioned were evaluated on a ROI-basis whereas the clinical target is to achieve lung tissue categorization for the entire parenchyma. This probably explains the failure to reproduce the results in [8] on our database, for similar training conditions. The main drawbacks observed were an overestimation of the dense pathological regions, partly attributed to a bias introduced by the vascular network, and low precision for fibrosis and ground glass (GDG) discrimination due to an ontological overlap between these classes (Fig. 1).

This paper addresses the previous limitations and develops a CAD system based on a cascade of two CNNs to better discriminate the highest correlated classes (Fig.2). A specific preprocessing using connected filtering is applied to the lung images in order to attenuate the vessel densities while preserving high opacities related to pathologies. This also allows completing the structural knowledge related to the lung mask [6] and airways [7], by the vascular network. The
2 Materials and methods

The preliminary dataset for this study includes a proprietary database of Avicenne Hospital, Bobigny, France consisting of 60 volumetric acquisitions with sparse 2D axial annotations. The CT scanning protocol details are as follows: slice thickness 1.25 mm; modality MDCT; contrast agent none; axial pixel matrix 512x512 or 768x768; mean x-y spacing 0.7 mm (with slightly different pixel spacing).

2.1 Data preprocessing using locally connected filtering and lung masking

Locally connected filters (LCF) exploit grayscale topological connectivity of each point on the support of a function \( f \) in order to filter out \( f \) values weakly connected with a pre-defined subset \( Y \subset\text{supp}\ f \) in the local environment of that point. The topological connectivity is verified by means of a morphological filter, namely the grayscale reconstruction by dilation \( R^d_f(\cdot,Y) \), or the grayscale reconstruction by erosion \( R_f(\cdot,Y) \).

Since the targeted structures here are the blood vessels, exhibiting a positive contrast with respect to their environment, the former operator \( R^d_f(\cdot,Y) \) will be used. The shape and size of the reference subset \( Y \) will determine the type of the researched connectivity. Reminding the definition of \( R^d_f(\cdot,Y) \) as [6]:

\[
R^d_f(\cdot,Y) = \delta_f^{(c)}(g),
\]

where

\[
\delta_f^{(c)}(g) = (g \oplus H) \land f
\]

and

\[
\delta_f^{(s)}(g) = \delta_f^{(1)}(\delta_f^{(s-1)}(g))
\]

with \( H \) denoting the unitary structuring element, \( \oplus \) the morphological dilation operator, \( \land \) the infimum operator and

\[
g(x) = \begin{cases} f(x), & \forall x \in Y \\ -\infty, & \text{elsewhere} \end{cases}
\]

the LCF by dilation is defined as:

\[
\forall x \in \text{supp}\ f, LCF_d^d(x, k) = R^d_f(x, Y) \mid Y \subset \mathcal{N}(k),
\]

where \( \mathcal{N}(k) \) denotes the k-size spatial neighborhood of \( x \),

\[
\mathcal{N}(k) = \{ y \in \text{supp}\ f \mid d(x,y) \leq k \},
\]

with \( d \) a distance function.

LCF thus provides in each point \( x \) of the function support the connectivity strength between \( f(x) \) and the positive-contrast structures in a region \( Y \) of the \( x \) neighborhood. If \( Y \) is chosen as the set of points on the border of \( \mathcal{N}(k) \), eq. 4 becomes

\[
\forall x \in \text{supp}\ f, LCF_d^d(x, k) = R^d_f(x, \mathcal{N}(k)-\mathcal{N}(k-1)).
\]

\( LCF_d^d(\cdot,k) \) reconstructs locally the \( f \) value from a k-distant neighborhood by morphological dilation. Its effect is to attenuate (or suppress) the \( f \) values which are not “linked” with their k-distant neighborhood via a high-intensity path. On contrary, when such connection exists, the structures are preserved via the reconstruction operator with a slight “flattening” of the grayscale levels. \( LCF_d^d \) thus not only presents a denoising property similar to the median filter, but also preserves spatial structures selected by a local connectivity configuration. Fig. 3 illustrates the effect of \( LCF_d^d \) on two neighborhood configurations (defined by a \( d_k \) distance function in eq. 5).

In order to remove vascular structures in the 3D image datasets, we take into account the linear shape of the vessels and their cross-sectional size, and set-up a filtering scheme combining 2D \( LCF_d^d \) applied across planes oriented orthogonal to different directions in space. We chose 9 spatial directions corresponding with the 18-connectivity (excluding symmetric orientations), Fig. 4a. We define the vascular removal LCF of size \( k \) as:

\[
\forall x \in \text{supp}\ f, VLCF_d^d(x, k) = \inf_{d \in C_{18}} \{ LCF_d^d(x, k_d) \},
\]

with \( k_d \) denoting the border subset of the 2D spatial neighborhood of size \( k \) orthogonal to the \( d \) direction, and \( C_{18} \) the set of orientations for the 18-connectivity (Fig. 4b).

Fig. 5b illustrates the 3D filtering effect of VLCF at the level of an axial CT image for \( k=3 \). The vascular structures are detected by adaptive thresholding and directional reconstruction by erosion (Fig. 5d) applied to the difference \( f - VLCF_d^d \) (Fig. 5c). Note that the sheet-like structures are minimally affected by the filter (in the example, the lung fissures).
Fig. 4. Directional filtering: (a) non-symmetric orientations in 18-connectivity, \( d_i \in C_{18} \), (b) example of orthogonal neighborhood \( k_d \) for \( d = d_4 \) in (a).

Fig. 5. Illustration of VLCF on a lung CT image, for \( k = 3 \): (b) filtered image, (c) difference, (d) vascular structures up to size \( k \).

In order to remove vascular structures of different sizes, the VLCF is applied in a multiresolution scheme using 2 levels of decimation. The input at each resolution level is the filtered image from the previous level, enhanced by a multiplication with a small constant (\( \alpha > 1 \)) without exceeding the input (Fig. 6).

In addition, the vascular structures are detected at each level of resolution and combined together prior to the final adaptive thresholding and filtering, which selects the highest confidence vessels. The vascular structure thus extracted is added to the lung mask obtained in [6] (and completed by the airway structures [7]) to reinforce the train-and-test process of the network as discussed later on. Fig. 7 shows the multiresolution filtering result together with the extracted vascular information for the image in Fig. 5a.

As preprocessing steps for the lung image patches, due to the different pixel spacing between cases, we rescale all scans to match 0.4 mm/pixel spacing [8]. Furthermore, the native pixel values for each patch was cropped in the range [-1000, +1000] HU, and mapped to [0,1] interval.

Fig. 6. Flowchart of the multiresolution VLCF filtering. \( \wedge \) and \( \vee \) denote infimum and supremum, while \( \downarrow \) and \( \uparrow \) decimation and interpolation, respectively. \( f_{\text{VLCF}} \) is the filtering result and \( v_f \) vascular regions for all scales.

Fig. 7. Result of vascular removal/detection in Fig. 5a using multiresolution VLCF (notations from Fig. 6): (a) filtered image \( f_{\text{VLCF}} \), (b) vascular information \( v_f \), (c) vascular segmentation, (d) lung and vessels mask.

In contrast to studies in [1][8][9] which analyze the classification performance on a patch-based approach by mixing patients from all database, in this paper we target a patient-based evaluation. This means we perform the classification and analysis on the entire lung field from several axial images. We avoid misclassification of some points near to the lungs edge using a patch filling procedure exploiting the lung masking. In this way, patch values outside the lung mask or inside detected vessels are replaced with neighboring values using a recursive symmetric replication.
2.2 Classification using a cascade of CNNs

A hierarchical classification of each image patch is performed using the architecture in Fig. 2. The network of [8] (T-CNN) is first used to discriminate the less correlated classes, by grouping fibrosis and GDG in one single class. The latter two will be further differentiated using a shallow network architecture with random filters (referred to as R-CNN in the following) inspired from a texture synthesis approach [10][11].

The T-CNN network consists of 5 cascaded layers of convolutional filters to compute image features, with 2x2 kernels and LeakyReLU activations, followed by just one average pooling, with size equal to the size of final feature maps and three dense layers that act as the classifier (Fig. 8). On the contrary of standard CNN architectures adopted in the literature, this T-CNN does not integrate pooling layers operations between the convolutional layers, and also uses the smallest kernel size to help extracting the best texture representation. The T-CNN ends with a final 3-way SoftMax layer for pathological or normal classification. DropOut is used to avoid overfitting [2]. Torch7 framework is used as it efficiently trains the CNN using GPU acceleration [12].

Fig. 8. T-CNN architecture

We used 50 patients for training (from scratch) the T-CNN with the purpose of reaching a numerous and representative set required by the CNN to perform a valid generalization. All these cases are not considered for testing. The rest of 10 different DICOM image series of whole thorax scans are used as test cases. From the training set, non-overlapping patches of 32x32 pixels were extracted. We consider those patches falling 80% inside the annotated ROIs. Horizontal flipping and rotations are applied in order to artificially increase the number of samples and avoid over-fitting the neural network. Thus, the training dataset consisted of 24265 image patches: 8978 normal, 8871 fibrosis+GDG, 6416 emphysema. The validation set consists of 1045 patches (442 normal, 558 fibrosis+GDG, 75 emphysema). Fig. 9 illustrates the associated training curves and confusion matrices.

Similarly, 32x32 pixels patches were extracted from the test dataset as input for the T-CNN, centered to each pixel in the test image to which a class will be assigned with a given probability.

Fig. 9. T-CNN learning curves: loss and accuracy vs. number of epochs (top) and confusion matrices (bottom) for training and validation. The dotted line indicates the early stopping epoch to avoid overfitting the training set.

The second module of the cascade, R-CNN, tries to tackle the subclassification between ground glass and fibrosis by developing a new texture descriptor. Basically, a texture is characterized by the arrangement of local patterns. Gatys et al. [10] demonstrated that CNNs are able to extract not only the content of an image but also the style representation, or the texture. The second one, of our interest, was achieved by computing the Gram matrix (G) between the feature maps given a depth of the CNN. By including the feature correlations of multiple layers, they obtained a stationary, multi-scale representation of the input image, which captures its texture information but not the global spatial arrangement, which is more appropriate at describing textures than objects. In [11] it is shown that even with shallow CNNs with random filter, the style representation could be found for a given input image. These mentioned works use such information to synthesize texture images based on examples. We will use this texture space as a feature descriptor of the pathologies under analysis.

Merging these concepts, we use a Random-Multiscale network with seven different filter sizes fxf with f = 3; 5; 7; 11; 15; 23; 29, and 128 feature maps each (896 feature maps in total) to transform the input data. Filters are obtained from a uniform distribution according to [14]. Notice that no training is needed in contrast with the T-CNN. Instead, the Gram matrix is computed for the resulting feature maps as follows:

\[ G_{ij} = \frac{1}{M} \sum_{k=1}^{M} F_{ik} F_{jk}, \]  

where \( F_{ij} \) denotes the activation of the \( i \)-th filter at position \( j \) and \( M \) the size (height x width) of the feature map.

Figure 10 shows this transformation of the raw data to the new space.
Fig. 10. Architecture of R-CNN used for ILD patch texture representation. Each blue cube represents the set of 128 feature maps for each kernel with its respective size of \( f \). The Gram matrix is computed as the new representation of the input (with texture information).

The element-wise mean and the standard deviation for the whole set of Gram matrices obtained from training patches are computed to obtain two signature matrices \( E^k \) and \( S^k \) respectively, \( k = \{1,2\} \) for each of the two classes to be detected. Finally, the classification is made by calculating the discrepancy \( \delta^k \) between the Gram matrix of the test patch \( G_i \) and the signature Gram matrix of the two classes - fibrosis and ground glass, applying the following formula:

\[
\delta^k = \sum_{i} \frac{|G_{ij} - E_{ij}^k|}{1 + S_{ij}^k}.
\]

The minimum value for this computation indicates higher resemblance between textures and dictates the decision making:

\[
\text{Assigned class} = \text{argmin}_k \delta^k. \quad (10)
\]

3. Results and discussion

The 3D test database not being fully annotated, a fully-quantitative evaluation of the classification result was not possible. A semi-quantitative assessment was performed using sparse annotations of the ground truth from an expert radiologist (10 axial images per scan, evenly distributed).

If we analyze the normalized confusion matrix for the whole test database (Fig 11), we notice a good prediction of the proposed method for normal and emphysema, but also a less discriminative power between fibrosis and ground glass patches. This can be explained by the effect of LCF filtering which also “flattens” the grayscale variation of the fibrosis regions making them more similar to ground glass (especially for the most confusing situations). Another explanation of the quantitative result for fibrosis vs. ground glass comes from an effect of fuzzy decision at the interface between fibrosis and normal areas (patches falling on these interface zones will be more affected by a misclassification).

Fig. 11. Normalized confusion matrix for the whole test database for the cascade of CNNs approach.

We have analyzed the performance of the proposed approach in terms of sensitivity (true positive rate, TPR), specificity and accuracy on the test database (10 patients):

\[
\text{TPR} = \frac{[TP]}{[TP] + [FN]} \quad (11)
\]

\[
\text{SPEC} = \frac{[TN]}{[TN] + [FP]} \quad (12)
\]

\[
\text{ACC} = \frac{[TP] + [TN]}{N} \quad (13)
\]

where \([TP], [TN], [FP], [FN]\) denote respectively the true positive, true negative, false positive and false negative cardinal, and \(N\) the size (in pixels) of the lung mask.

The average values obtained for each class are as follows. Sensitivity: 75.7% for normal, 43% for emphysema, 40% for fibrosis and 50% for ground glass. Specificity: 73.5% for normal, 90.6% for emphysema, 95% for fibrosis and 86.4% for ground glass. Accuracy: 75.8% for normal, 90% for emphysema, 91% for fibrosis and 81.5% for ground glass. Fig. 12 summarizes the results per analyzed case.

The lower performance in terms of sensitivity can be understood when investigating the results from a qualitative point of view. Figs. 13-15 illustrate several examples in comparison with the ground truth. The segmentation also includes the airways [7] and the vessels (cf. §2.1). The following qualitative findings are to be mentioned:

- the detected normal regions correspond globally well with the ground truth; the noticed differences in Fig. 13 in the right lung might be explained by a high diffuse aspect of the pathology but also by the very rough delineation of the ground truth. The overestimation of emphysema in Fig. 15 is not unexpected since the image density in this case is indeed low in these regions, and
the effect might be amplified by the patch-filling procedure in the lung border regions;
- we notice a transition zone between fibrosis and normal classes which is defined as ground glass by the algorithm. This might be explained either by a texture appearance similar to GDG (see Fig. 1 for ontological overlapping of classes) or by a patch size effect which has to be investigated in future studies;

Fig. 12. Quantitative results of the classification sensitivity, specificity and accuracy for the test database. Cases 2, 5 and 7 do not have emphysema.

- note that the precision of the ground truth delineation is rather coarse but an accurate definition in the context of diffuse and imbricate pathological tissues would be far more tedious and time consuming for radiologist. This affects also the sensitivity figures that we reported. Ideally, a better ground truth specification should be targeted for a more accurate quantitative assessment, by involving several expert radiologists, which still remains a difficult task to achieve in real practice.

Fig. 13. Qualitative results. From left to right and top to bottom: original image, LCF preprocessing, ground truth, segmentation result. N=normal, F=fibrosis, E=emphysema, GDG=ground glass (right lung is on the left in the image).

Finally, we have investigated the benefit of the prefiltering module in the proposed framework by assessing the system performance with and without LCF on the same training and validation dataset. Overall, an increase of nearly 10% in
sensitivity, 3.5% in specificity and 4.5% in accuracy was assessed when using LCF versus no prefiltering.

Conclusion

This paper presented a deep learning framework consisting of a cascade of two CNNs for quantitative assessment of infiltrative lung diseases. An original image pre-processing framework based on multiscale locally connected filtering was presented to help improving the learning process and boost the performance for lung texture discrimination. By removing the dense vascular network from input images, locally connected filtering unbias the normal texture detection. However, LCF may affect some salient features of fibrosis patterns, making them more difficult to discriminate from ground glass. Future work should consider improving the preprocessing module in order to preserve discriminating features of pathological areas while still removing vascular structures.

References


