Deep patch-based priors under a fully convolutional encoder-decoder architecture for interstitial lung disease segmentation

Maria Vakalopoulou, Guillaume Chassagnon, Nikos Paragios, Marie-Pierre Revel, Evangelia Zacharaki

DEEP PATCH-BASED PRIORS UNDER A FULLY CONVOLUTIONAL ENCODER-DECODER ARCHITECTURE FOR INTERSTITIAL LUNG DISEASE SEGMENTATION

*M. Vakalopoulou[†], *G. Chassagnon,[† ‡], N. Paragios[† §], M.-P. Revel[‡], E. I. Zacharaki[⋆ †]

[†] CVN, CentraleSupélec, Université Paris-Saclay, § TheraPanacea, Paris, France
[⋆] Department of Computer Engineering & Informatics, University of Patras, Greece
[‡] Department of Radiology, Groupe Hospitalier Cochin-Hotel Dieu, Université Paris Descartes, France

ABSTRACT

Interstitial lung diseases (ILD) encompass a large spectrum of diseases sharing similarities in their physiopathology and computed tomography (CT) appearance. In this paper, we propose the adaption of a deep convolutional encoder-decoder (CED) that has shown high accuracy for image segmentation. Such architectures require annotation of the total region with pathological findings. This is difficult to acquire, due to uncertainty in the definition and extent of disease patterns and the need of significant human effort, especially for large datasets. Therefore, often current methods use patch-based implementations of convolutional neural networks, which however tend to produce spatially inhomogeneous segmentations due to their local contextual view. We exploit the advantages of both architectures by using the output of a patch-based classifier as a prior to a CED. Our method could advance the state-of-the-art in lung tissue segmentation using only a small number of newly annotated images.

Index Terms— Encoder-decoder, probabilistic priors, pixelwise classification

1. INTRODUCTION

Disease monitoring is a major challenge for both patient care and approval of new treatments for many chronic lung diseases. It mainly relies on physiological data such as pulmonary function testing (PFT), which are often used as primary endpoint in clinical trials. However, PFT measurements are imperfect and only reflect lung function, not necessarily disease activity.

Among imaging techniques, CT is the gold standard for in vivo morphological assessment of lung parenchyma and for the detection of early ILD [1]. Several visual scoring systems have been proposed but they only allow basic quantification of ILD severity and they suffer from poor-to-moderate inter-observer agreement [2]. Our aim is to develop an imaging-based quantification tool for assessment of ILD severity and monitoring the disease progression. This requires segmentation of the diseased tissue in an automatic and reproducible fashion. Delineation of disease extent on CT images is often difficult and the manual outlining in 3D is extremely labor intensive. This is a common problem in medical imaging, which explains the fact of missing ground truth or partially labeled databases. Indeed, to the best of our knowledge, the only available ILD dataset [3] is sparsely annotated; *i.e.* it includes only some parts of selected image slices out of the whole volumetric dataset. Until now, such databases are exploited within slice [4] or patch-based [5, 6] classification paradigms. However, patch-based classification of lung tissue is not very appropriate for disease segmentation since it fails to capture accurate and smooth boundaries between different tissue types. Intra-patient classifiers [7] could be used to estimate the disease in the remaining part (out of patch), but such patient-specific models are sensitive to outliers due to their limited view. ILD segmentation, *i.e.* voxel classification, has been pursued using a Markov Random Field, Gaussian Mixture Model and Mean Shift algorithm in [8], or watershed segmentation algorithm and Fuzzy C-Means in [9]. These methods however are strongly based on pre- or post-processing steps, like morphological operations, and do not incorporate prior knowledge from available databases failing to achieve robust segmentation results.

In this paper, we introduce a novel framework that uses a patch-based implementation of convolutional neural networks (CNN), exploits sparsely annotated regions of high confidence and learns local textural patterns, followed by a deep CED architecture, trained on dense (whole image) annotations. The patch-based CNN learns local textural patterns while the CED learns also high-level relationships and thus ensures spatial homogeneity of the produced segmentation. The combination of the two architectures allows to transfer the learned features across different datasets and thus enrich the deployed feature space, so that the labor intensive annotation of whole image slices can be limited to only a few examples. The method achieved high accuracy for ILD segmentation in CT images of patients with scleroderma, as quantified by overlap scores. The different architectures are compared, and their advantages and possible limitations are discussed shedding light into their performance for tissue characterization in medical images.

* indicates equal contribution.

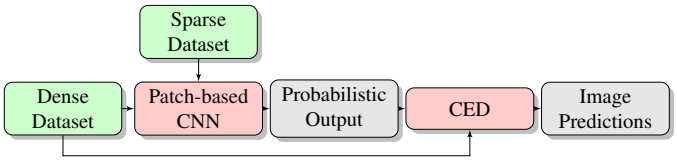


Fig. 1. The proposed combination model with the two training datasets.

2. METHODOLOGY

The idea of the proposed framework is to combine local and global learning through the use of patch-based and fully convolutional dense networks, respectively. Specifically, we implemented a patch-based deep CNN in order to exploit the available partial datasets and introduced its probabilistic output as an additional layer into a subsequent dense CED network, in order to increase the spatial homogeneity of segmentation. Although the proposed deep ensemble is currently implemented in a sequential fashion, the optimization function in the individual architectures is the same, thus the results do not contradict each other. Due to unavailability of disease annotation on whole slices to train the encoder-decoder, manual annotations on 3D CT images have been created by a medical expert for a small number of subjects. The whole framework is illustrated in Figure 1, whereas more details on the individual steps are provided next.

Patch-based classification: A relatively simple convolutional architecture (*ConvNet*) is chosen as it offers a good trade-off between computational time and accuracy. In particular, the network consists of 2 convolutional, 2 fully connected layers, and a softmax layer. Each convolutional layer (with filters of size 5×5) is followed by a *tanh* activation function and a 3×3, stride 2, max-pooling operation which downsamples the input patch. Similarly, each of the 2 fully-connected layers is followed by a *tanh* function, the output of which is introduced to a softmax layer for the final classification. For classification of new images, a patch is extracted around each pixel of the CT slices and the predicted label for the patch is inherited to the central pixel. For our experiments, we used patches of size 29×29. The model was trained for 50 epochs with learning rate for the stochastic gradient descent (SGD) 0.5 while every 3 epochs the learning rate was reduced to half. The momentum was set to 0.9 and the weight decay to $5{\cdot}10^{-4}$.

Deep CED architecture: Motivated by the state-of-the-art performance of fully convolutional networks in a variety of problems we adapt them for dense labelling of CT image slices. We use the *SegNet* deep learning network [10] which performs pixelwise classification and is composed of an encoder and a decoder network. It consists of 13 convolutional layer groups, similar to the ones of *VGG16* network, which have been initialised with the *VGG16* pretrained weights. The entire architecture consists of repetitive blocks of convolutional, batch normalization, rectified-linear units (ReLU) and max-pooling layers. For more details we refer to the original publication. For our experiments, we used image slices of the original dimension (512×512) and we first pre-trained the model for 60 epochs (without incorporating probability maps) with learning rate for the SGD starting at 0.01 and then reducing every 20 epochs by $2.5{\cdot}10^{-3}$. Similarly to the patch-based classifier, the momentum was set to 0.9 and the weight decay to $5{\cdot}10^{-4}$.

Combination of architectures: There are different approaches to fuse or transfer knowledge between different deep learning architectures. In this approach, we combine datasets as well as network outputs. In particular, after training a patch-based CNN we calculate the probability maps for each class and use them as input to a CED, additionally to the gray level image (*i.e.* extra channels). Through this transfer, the second network can exploit during the learning procedure the prior knowledge about the spatial distribution of different tissue classes. We trained the CED architecture for 70 epochs, with combined input (CT images and probability maps), initialising the encoder weights with the ones calculated from the pre-training phase.

3. DATASET & IMPLEMENTATION DETAILS

We performed experiments using two different datasets. The first dataset includes (volumetric) CT images from patients with scleroderma disease, fully annotated with 3 different classes: healthy lung, lung with disease, and the rest (body tissue outside the lung and background). Lung segmentation was performed using Myrian XP-lung software (Intrasense, Montpellier, France) and subsequently manually corrected (whenever required). ILD segmentation was performed by a medical expert, by tracing the disease boundaries in axial view over all slices. Assessment of the method was performed on 20 additional patients with full annotations only on selected CT slices ($n = 20$) resulting to 400 slices in total for the testing set.

The second dataset is the publicly available ILD dataset [3] which contains 905 selected image slices from 120 patients including in total 13 ILD labels (such as ground glass, fibrosis, micronodules, consolidation, *etc*). This dataset is sparsely annotated, and thus could only be used for the patch-based model. Since our goal was to segment tissue with ILD from scleroderma patients, we were not interested in the subcategorization of individual tissue abnormalities, thus we merged all the indicated disease manifestations into a single class.

For the patch-based CNN, 80000 patches were extracted for each class containing samples from both datasets (ILD and scleroderma). For training of the CED we used around 3400 CT slices and for the training of the CED with the combined input 150 CT slices. Moreover, for training the CED and the combination models, we performed median frequency

	Specificity		Sensitivity		Precision	
Method	Disease	Healthy	Disease	Healthy	Disease	Healthy
Patch-based	0.948	**0.990**	**0.902**	0.743	0.489	**0.924**
CED	0.971	0.983	0.906	0.889	0.598	0.923
CED retrain	0.981	0.980	0.819	0.932	0.639	0.891
Proposed	**0.989**	0.985	0.827	**0.943**	**0.671**	0.904

Table 1. Different evaluation metrics for the testing dataset.

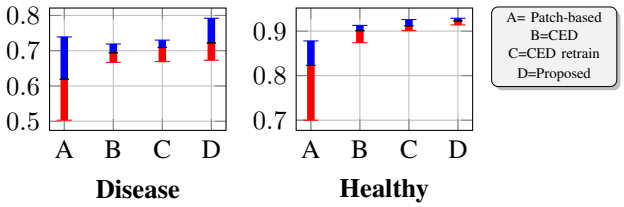


Fig. 2. The average, maximum and minimum values for the calculated Dice coefficient for diseased and healthy tissue and all employed strategies. Blue and red indicate values above and below the average.

balancing [10] to balance the data, as the samples with disease were considerably fewer than the rest of the samples. In such a setup, the assigned weight to a class in the loss function was the ratio of the median of class frequencies computed on the entire training set divided by the class frequency.

4. RESULTS & DISCUSSION

We evaluated the detection accuracy of each class against any-other class, using four metrics, namely sensitivity, specificity and precision (Table 1) and the Dice coefficient (Figure 2). Since the metrics were calculated for every slice, we averaged the results across slices by weighting the values based on the coverage of each class within each slice. This should be equivalent to the calculation of the metrics in 3D. For comparison with others, we report results for each architecture individually to approximate the performance of the methods in [6] and [11] by the patch-based CNN (Table 1, 1st row) and [10] by the CED (Table 1, 2nd row). Moreover, in order to assess the importance of information fusion, we also retrained the CED without using the probability maps from the patch-based network (Table 1, row 3) using the same 150 slices as in the proposed scheme (Table 1, row 4).

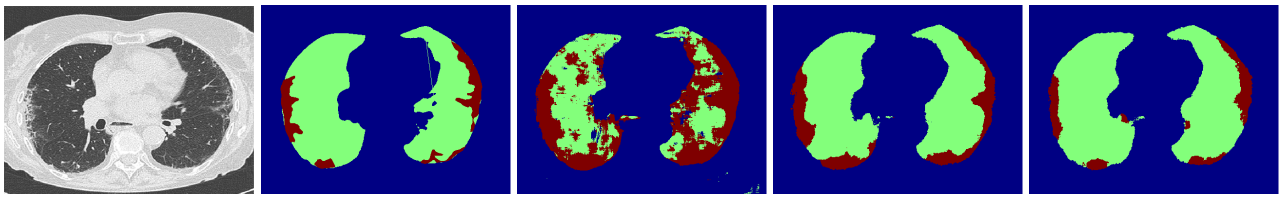
From Table 1 it can be observed that the proposed methodology presents higher precision and specificity for disease tissue but lower sensitivity compared to the other three methods. The high sensitivity of the patch-based strategy is accompanied by low precision indicating that it is not appropriate for dense classification. Additionally, one can observe that retraining the CED without information fusion, boosts the accuracies without, however, outperforming the proposed method. Moreover, Figure 2 shows that the proposed strategy outperforms the other two in respect to Dice coefficient for both diseased and healthy tissue, with mean values for the diseased tissue 0.619, 0.694, 0.709 and 0.722 for the patch-based, CED, CED with retrain and proposed strategy, respectively.

The most similar studies [6, 11] targeting the classification of ILD patterns with patch-based CNNs achieved 85.5% and 92.8% classification accuracy, respectively. Binary classification of ILD, non-ILD and healthy (normal) subjects (15 in each group) was performed in [12] using linear regression and texture features. Detection rates for the patients with ILD were 73% when compared to non-ILD and 67% when compared to normal. Although not directly comparable, our method, which is equivalent to a 3-class voxel-wise classification problem, achieved a total accuracy of 89.3% (percentage of correctly classified voxels) and sensitivity 82.7% and 94.3% for ILD and healthy lung respectively. ILD segmentation was investigated also by other groups, but to the best of our knowledge only in respect to selected patterns. In [13], watershed and FCM segmentation methods were applied to capture microcystic patterns or larger cysts (present in ILD of grade 2 or 3) and assessed only visually on data of a single patient.

Visual assessment of the segmentation is illustrated for a testing subject in Figure 3. In agreement with the quantitative assessment the patch-based approach produces noisy detections with healthy lung and diseased tissue often intermixed. The other two approaches detect the different categories with higher consistency (indicated by higher lower bounds of Dice), with the proposed strategy reporting less false detections for the disease class than the CED.

Additionally to high detection accuracy, the proposed method produced spatially homogeneous regions within slices (over axial sections) which is attributed to the fully convolutional architecture. For this reason, there was no need to impose spatial consistency constraints for example through the use of Conditional Random Fields [11]. Moreover accurate boundary localization was achieved due to the design of the decoder network. The hierarchical decoders use the max-pooling indices received from each corresponding encoder to perform non-linear upsampling of the low resolution feature maps. Mapping the low resolution feature maps to the resolution of the original CT images allowed to overcome coarse labellings and localize boundaries. Spatial homogeneity and accurate localization was also observed across slices (over sagittal and coronal sections), thus, although not guaranteed, we do not expect to have a significant gain from the substitution of our 2D networks by 3D architectures. Furthermore, collection of good quality 3D ground truth is difficult, because the manual outlining is performed over in one plane (usually over axial slices), thus often lacks smooth boundaries across the other two planes.

For our experiments we used a Tesla K40c GPU with both Torch and Caffe deep learning libraries. The training of the patch-based model required a bit less than an hour as the ar-



(a) CT image (b) Ground Truth (c) Patch-based (d) CED (e) Proposed

Fig. 3. One testing subject with the classification of the different employed deep learning strategies. With red color the lung with disease is depicted, with green the healthy lung and with blue the rest.

chitecture was relatively small. In that strategy, the extraction of patches was the most computationally demanding task. This is a limitation in the testing phase where the number of voxels to be annotated is particularly large. On the other hand, the CED required around 24 hours for the training and another 6 hours for the final training by the combined input. However, for both of them, the testing phase lasted only a few seconds for segmentation of one lung image.

5. CONCLUSIONS

In this work, we have introduced a novel framework that integrates deep, patch-based (trained on available databases) priors with a fully convolutional encoder-decoder network (trained on a small number of images), to improve generalization performance. The combination of the two architectures allowed to enrich the deployed feature space and thus increase segmentation accuracy. The method was applied for ILD segmentation in CT images, but generalizes to other tissue or texture characterization problems as well. It does not incorporate any shape or deformation priors or physiological models that guide detection, therefore it is not specialized to the detection of localized structures, such as organs or uniform masses (although it can handle also this type of problems given enough training data). On the other hand, by learning only from image content, the method is modular and easy to use. Future work includes the investigation of 3D convolution and the trade-off between possible gain in volumetric smoothness over the additional computational cost required to minimize discontinuities of annotations over sagittal and coronal views.

6. ACKNOWLEDGEMENTS

The work has been partially supported from the Fondation pour la Recherche Médicale (FRM) and the GE healthcare.

7. REFERENCES

[1] G. Bussone and L. Mouthon, "Interstitial lung disease in systemic sclerosis," *Autoimmun Rev*, vol. 10, no. 5, 2011.

[2] G. Camiciottoli et al., "Lung ct densitometry in systemic sclerosis: Correlation with lung function, exercise testing, and quality of life," *Chest*, vol. 131, no. 3, 2007.

[3] 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, Apr 2012.

[4] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans Med Imaging*, vol. 35, no. 5, May 2016.

[5] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, Y. Wang, M. J. Fulham, and M. Chen, "Large margin local estimate with applications to medical image classification," *IEEE Trans on Med Imaging*, vol. 34, no. 6, 2015.

[6] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," *IEEE Trans Med Imaging*, vol. 35, no. 5, 2016.

[7] R. Verma et al., "Multi-parametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images," *Academic radiology*, vol. 15, no. 8, 2008.

[8] J. K. Dash, V. Madhavi, S. Mukhopadhyay, N. Khandelwal, and P. Kumar, "Segmentation of interstitial lung disease patterns in hrct images," *Proc SPIE*, 2015.

[9] N. M. Noor, R. Rosid, M. H. Azmi, O. M. Rijal, R. M. Kassim, and A. Yunus, "Comparing watershed and fcm segmentation in detecting reticular pattern for interstitial lung disease," in *2012 IEEE-EMBS*, 2012.

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE PAMI*, 2017.

[11] M. Gao, Z. Xu, L. Lu, A. Wu, I. Nogues, R. M. Summers, and D. J. Mollura, "Segmentation label propagation using deep convolutional neural networks and dense conditional random field," in *ISBI*, 2016.

[12] M. F. M. Hamzah, R. M. Kasim, A. Yunus, O. M. Rijal, and N. M. Noor, "Detection of interstitial lung disease using correlation and regression methods on texture measure," in *icIVPR*, 2017.

[13] J. T. C. Ming, O. M. Rijal, R. M. Kassim, A. Yunus, and N. M. Noor, "Texture-based classification for reticular pattern and ground glass opacity in high resolution computed tomography thorax images," in *IECBES*, 2016.