SEGMENTATION LABEL PROPAGATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS AND DENSE CONDITIONAL RANDOM FIELD

Mingchen Gao, Ziyue Xu, Le Lu, Aaron Wu, Isabella Nogues, Ronald M. Summers, Daniel J. Mollura

Department of Radiology and Imaging Sciences, National Institutes of Health (NIH), Bethesda, MD 20892

ABSTRACT

Availability and accessibility of large-scale annotated medical image datasets play an essential role in robust supervised learning of medical image analysis. Missed labeling of regions of interest is a common issue on existing medical image datasets due to the labor intensive nature of the annotation task which requires high levels of clinical proficiency. In this paper, we present a segmentation based label propagation method to a publicly available dataset on interstitial lung disease [3], to address the missing annotation challenge. Upon validation from an expert radiologist, the amount of available annotated training data is largely increased. Such a dataset expansion can can potentially increase the accuracy of Computer-aided Detection (CAD) systems. The proposed constrained segmentation propagation algorithm combines the cues from the initial annotations, deep convolutional neural networks and a dense fully-connected Conditional Random Field (CRF) that achieves high quantitative accuracy levels.

Index Terms— Segmentation Label Propagation, Dense Conditional Random Field, Convolutional Neural Network, Interstitial Lung Disease, Multi-class Labeling

1. INTRODUCTION

Interstitial lung disease (ILD) is a class of more than 150 chronic lung disorders. Most of them cause progressive scarring of lung tissues and eventually affect breathing. High resolution computed tomography (HRCT), as a non-invasive imaging method with sub-millimeter resolutions, is considered the gold standard imaging protocol is ILD diagnosis. The appearance and quantification of the types of lung tissue patterns in HRCT are very informative for the diagnosis of an ILD. The visual illustration of the most common lung tissue patterns are depicted in Fig. 1. However, interpreting pulmonary HRCT images can still be a challenge for trained chest radiologists and lung specialists. The diagnosis may result in errors by omission or confusion of diverse pathological lung tissue types. In order to improve the accuracy and efficiency of diagnosis, computer-aided detection (CAD) systems are actively studied for the quantitative identification of ILD patterns. Most of the previous work assign a single label to an image patch (usually of size 32×32) extracted from manually drawn Regions of Interests (ROI) from lung CT slices [2, 3, 9, 10]. Recently, a holistic image classification method is presented to address this problem from a more clinically practical setting [4]. Studies generally assume that there is only one label for every image patch or image slice even multiple ILD can commonly appear in a single CT slice.

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 [7]. The publicly available and widely used ILD dataset [3] is arguably the most comprehensive ILD datasets to-date. As shown in Fig. 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.

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. In this paper, 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. 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 [6]. The proposed method is applicable to other problem-

This research effort is supported by NIH intramural research program. Acknowledgment to Nvidia Corp. for donation of K40 GPUs.

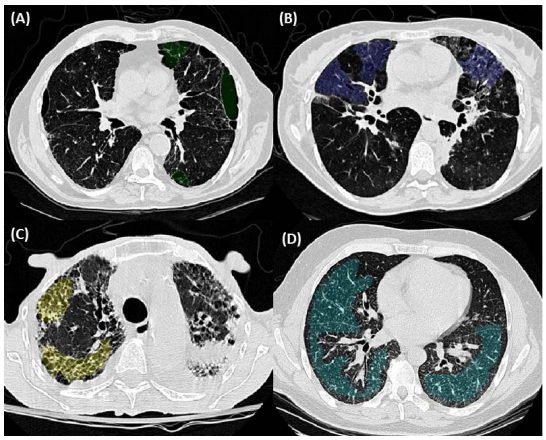

Fig. 1. Visual aspects of the most common lung tissue patterns in HRCT axial slices. Infected regions are annotated with different colors in the publicly available dataset [3]. (A) Emphysema (EM). (B) Gound Glass (GG). (C) Fibrosis (FB). (D) Micronodules (MN).

s as a generic semi-supervised image segmentation solution. This work is partially inspired by interactive graph-cut image segmentation [1] and automatic population of pixelwise object-background segmentation from manual annotations on ImageNet database [5].

Specifically, we explore the possible ways to **propogate** 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 [3] and the segmentation/labeling results are validation by an expert radiologist.

2. METHOD

In this work, 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 [6].

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 $\{X_1, ..., X_N\}$, with $X_i \in X$ being associated with every pixel $i \in V$ and taking a value from the label set $L = \{l_1, ..., l_K\}$ of label categories. The labeling of X from images is obtained with a maximum a posterior (MAP) estimation of the following conditional log-likelihood:

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j), \tag{1}$$

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 [4]. The pairwise potentials in our model have the form

$$\begin{aligned} \psi_p(x_i, x_j) &= u(x_i, x_j) \sum_{m=1}^{K} k(f_i, f_j) \\ &= u(x_i, x_j) \sum_{m=1}^{K} \omega^{(m)} k^{(m)}(f_i, f_j). \end{aligned} \tag{2}$$

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

$$k^{(m)}(f_i, f_j) = exp(-\frac{1}{2}(f_i - f_j)^T \Lambda^{(m)}(f_i - f_j)), \tag{3}$$

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 [4]) and positions p_i and p_j:

$$\begin{aligned} k(f_i, f_j) = \omega^{(1)} exp(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}) \\ + \omega^{(2)} exp(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}). \end{aligned} \tag{4}$$

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 kernel, which removes small isolated regions. The parameters θ_α, θ_β and θ_γ 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.

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 handcrafted [2, 9, 10] or CNN learned features [4]. We use the CNN based CRF unary classifier because of its state-of-the-art performance: classification accuracy of 87.9% reported in [4]. To facilitate comparison, five common ILD patterns are studied in this work: healthy, emphysema, ground glass, fibrosis and micronodules (Fig. 3). Image patches of size 32×32 pixels within the

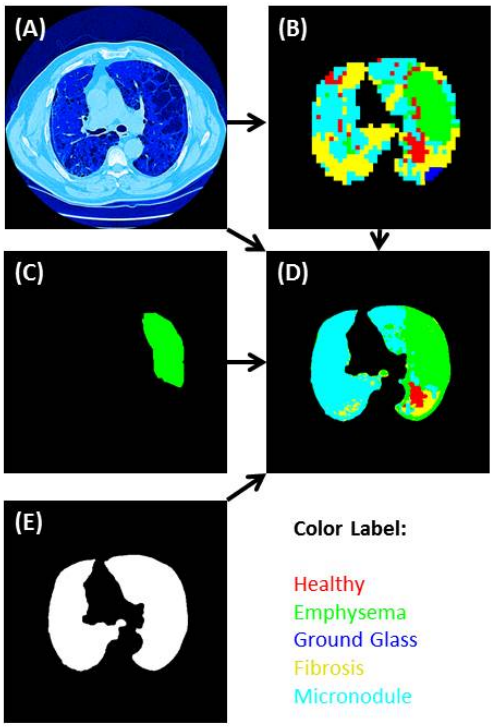

Fig. 2. Major intermediate results. (A) Three channels of different HU windows illustrated as RGB values. (B) CNN classifier at a spatial interval of 10 pixels. (C) Annotated ROI. (D) Final result integrating image features, unary prediction and hard constraints. (E) Annotated lung mask.

ROI annotations of these five classes, are extracted to train a deep CNN classifier. The well known CNN AlexNet model [7] trained on ImageNet is used to fine-tune on our image patch dataset. 32×32 pixel images patches are rescaled to 224×224 and three channels of different HU windows [4] are generated to accommodate the CNN model.

Hard Constraints: The image labels given by radiologists from the dataset [3] 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 equation 4, the parameters θ_α, θ_β and θ_γ 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.

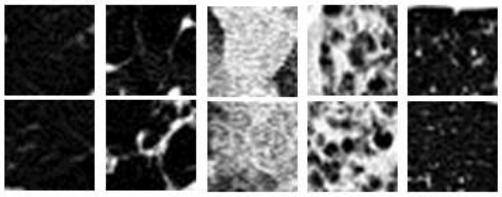

Fig. 3. Examples of 32×32 patches for each ILD category. From left to rights columns: Healthy, Emphysema, Ground Glass, Fibrosis and Micronodules.

Ground Truth	Prediction				
	NM	**EM**	**GG**	**FB**	**MN**
NM	**0.9792**	0.0067	0.0029	0.0020	0.0092
EM	0.2147	**0.7389**	0	0.0170	0.0294
GG	0	0	**1.0000**	0	0
FB	0.0414	0.0271	0.0118	**0.8046**	0.1151
MN	0.0007	0.0013	0.0174	0.0058	**0.9748**
precision	0.9500	0.9320	0.8175	0.9060	0.9666
recall	0.9792	0.7389	1	0.8046	0.9748
F-scall	0.9644	0.8243	0.8996	0.8523	0.9707

Table 1. Confusion matrix, precision, recall and F-score of ILD pattern labeling.

3. EXPERIMENTS & CONCLUSION

The publicly available ILD dataset [3] is used for training and validation under two-fold cross validation. ROIs of total 17 different lung patterns and lung masks are also provided along with the dataset. Fig. 4 shows the annotation provided by the dataset, the labeling obtained from our algorithm and the ground 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 [8] can be used to automate this process.

Quantitative evaluation is given in Table 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 [3]. Thus the labeled training dataset [3] 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 CNN classification is implemented in Matlab using MatConvNet package [11], and is run on a PC with 3.1GHz CPU, 32 GB memory and a Nvidia Tesla K40 GPU. 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 bilinearly interpolated and later refined by the CRF pairwise constraints. In our implementation, we predict the labels of

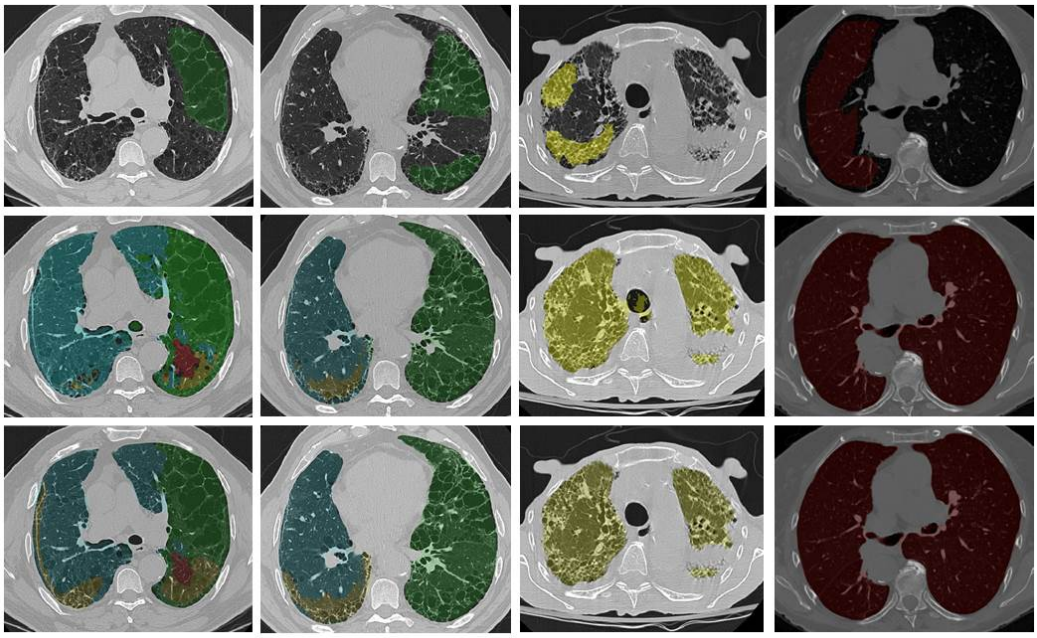

Fig. 4. First Row: Annotated region of interest provided by the original dataset [3] where missing regions of interest are evident. Second Row: Results produced from our algorithm. Third Row: Ground truth labeled by experienced radiologists.

image patches at a spatial interval of 10 pixels. Parameters θ_α, θ_β and θ_γ 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.

In this paper, we present segmentation label propagation method that 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 [3] is significantly expanded and will be publicly shared upon publication [1] .

References

[1] Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *ICCV*, volume 1, pages 105–112. IEEE, 2001.

[2] A. Depeursinge, D. Van 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. *TITB*, 16(4):665–675, 2012.

[3] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller. Building a reference multimedia database for interstitial lung diseases. *CMIG*, 36(3):227–238, 2012.

[4] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, et al. Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *1st MICCAI Workshop on Deep Learning in Medical Image Analysis*, 2015.

[5] M. Guillaumin, D. Küttel, and V. Ferrari. Imagenet auto-annotation with segmentation propagation. *IJCV*, 110(3):328–348, 2014.

[6] P. Krähenbühl and V. Koltun. Efficient inference in fully connected CRFs with gaussian edge potentials. In *NIPS*, pages 109–117, 2011.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[8] A. Mansoor, U. Bagci, Z. Xu, B. Foster, K. N. Olivier, J. M. Elinoff, A. F. Suffredini, J. K. Udupa, and D. J. Mollura. A generic approach to pathological lung segmentation. *TMI*, 33(12):2293–2310, 2014.

[9] Y. Song, W. Cai, H. Huang, Y. Zhou, D. Feng, Y. Wang, M. Fulham, and M. Chen. Large margin local estimate with applications to medical image classification. *TMI*, 34(6):1362–1377, 2015.

[10] Y. Song, W. Cai, Y. Zhou, and D. D. Feng. Feature-based image patch approximation for lung tissue classification. *TMI*, 32(4):797–808, 2013.

[11] A. Vedaldi and K. Lenc. Matconvnet-convolutional neural networks for matlab. *arXiv preprint:1412.4564*, 2014.

[1] http://www.research.rutgers.edu/ minggao