Erasmus MC
University Medical Center Rotterdam

DEEP LEARNING FROM LABEL PROPORTIONS FOR EMPHYSEMA QUANTIFICATION

Bortsova G.[1], Dubost F.[1], Ørting S.[2], Katramados I.[3], Hogeweg L.[3], Thomsen L.[4], Wille M.[5], De Bruijne M.[1,2]

[1] Biomedical Imaging Group Rotterdam, Erasmus MC, The Netherlands [2] Department of Computer Science, University of Copenhagen, Denmark [3] COSMONiO, The Netherlands [4] Department of Diagnostic Imaging, Bispebjerg Hospital, Denmark [5] Department of Respiratory Medicine, Gentofte Hospital, Denmark

Abstract

We propose a novel weakly supervised deep learning method that learns to segment and estimate the extent of emphysema from visually estimated image-level proportions of emphysematous tissue. Our method is a combination of an architecture specialized for learning from proportions and a custom loss. We outperform a commonly used regression network, traditional and recently published machine-learning based methods for emphysema quantification by a large margin. This work is accepted to MICCAI 2018.

Learning from Label Proportions for Emphysema Quantification

- Dataset: Danish Lung Cancer Screening Trial (DLCST) [1]
- Lungs are automatically segmented and divided into 6 regions (3 per lung)
- Emphysema extent is scored as intervals of a percentage of the affected tissue
- There are six extent categories: 0: 0%, 1: 1-5%, 2: 6-25%, 3: 26-50%, 4: 51-75%, 5: 76-100%
- The problem of learning from these labels can be viewed as learning from label proportions (LLP) [2] problem, in which only proportions of positive voxels or patches is known for every training region

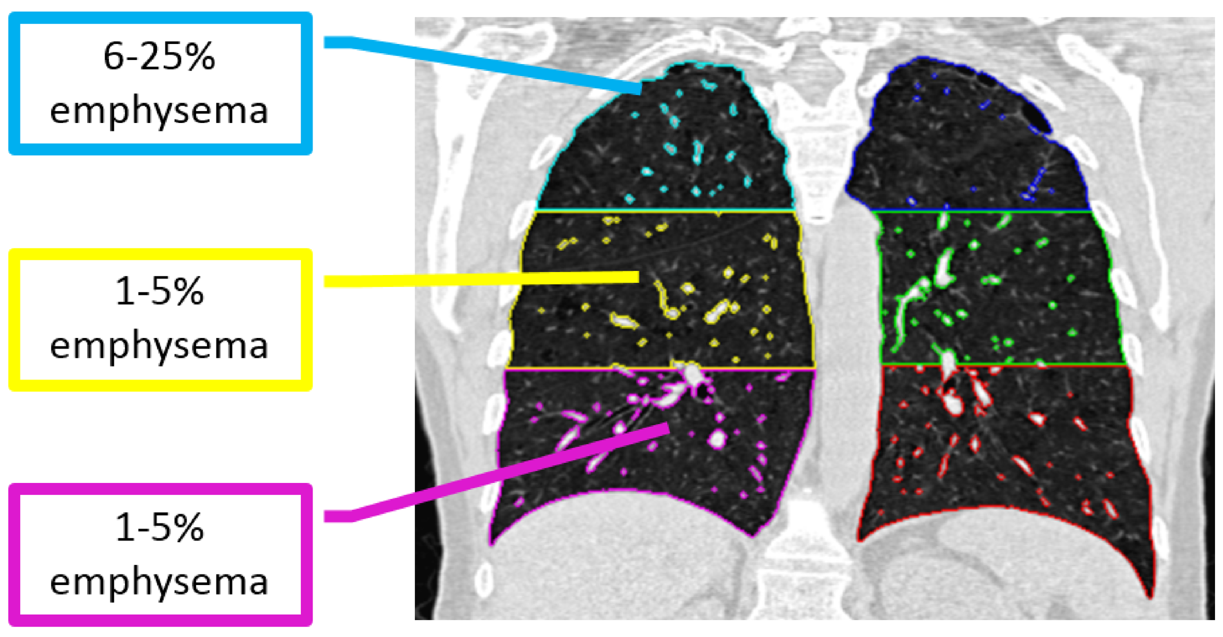

Figure 1: *Every lung region is scored using a six point grading system in which grades correspond to intervals of percentage of region tissue affected by the disease.*

Method Overview

- The main point of the method is to incorporate prior knowledge about the nature of labels into the model
- Our architecture reflects the assumption that the label is related to the volume of the abnormality
- Our loss is specifically designed for learning from interval-based ground truth

Architectures: Baseline and Proposed

- Baseline GAPNet represents the traditional CNN approach, which is blind to that the label represents the area of pathology
- Proposed ProportionNet:
  - More regularized, since it reflects the meaning of the label
  - More interpretable, since it can localize (i.e. segment) emphysema
- Both architectures have same number of parameters, convolution and pooling layers. The only difference is how final pooling is performed

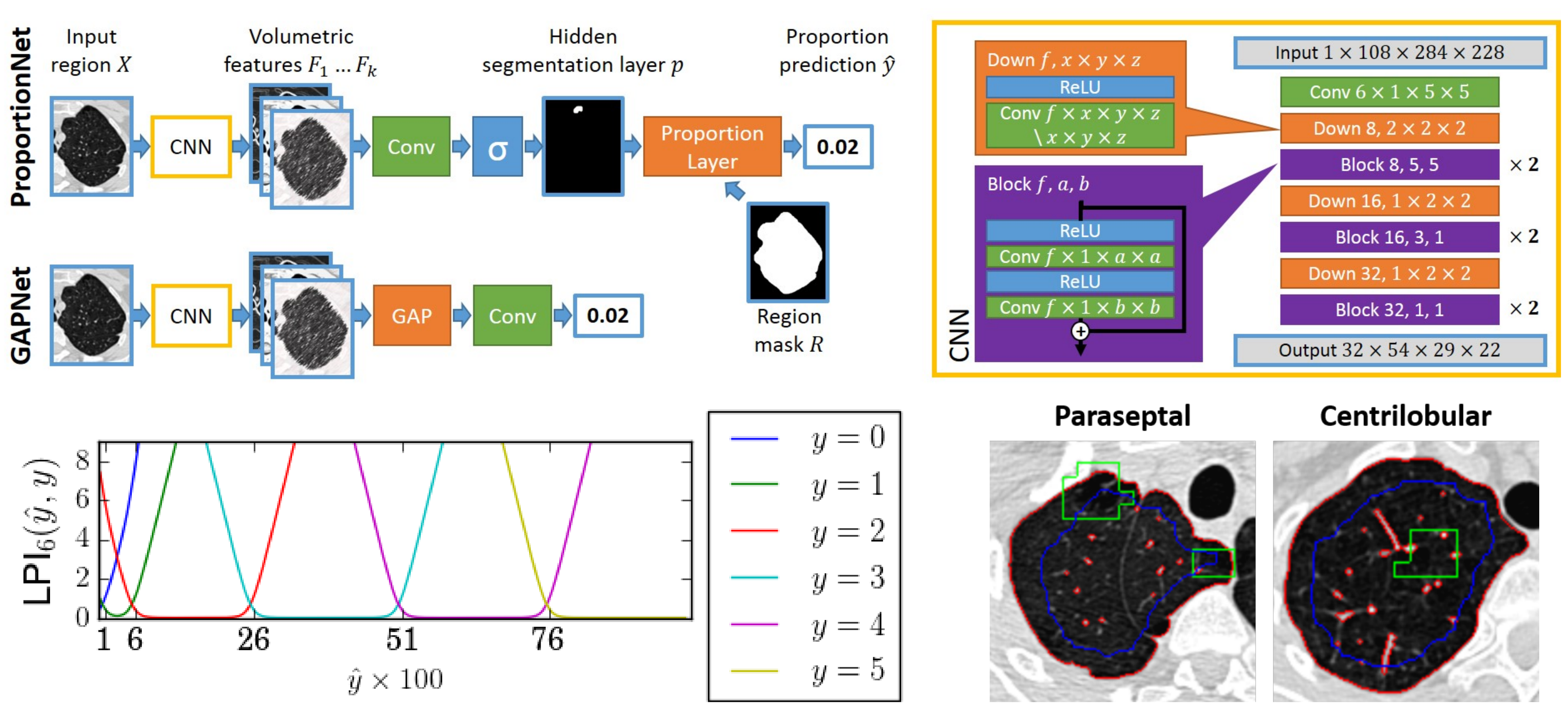

Figure 2: *Top: The proposed ProportionNet and baseline GAPNet architectures. Bottom left: The proposed LPI_6 loss with all $w_c = 1$, $\alpha = 120$. Bottom right: Images with different predominant emphysema patterns. Green: ProportionNet segmentations; red: region mask; blue: the 10px margin for separating near-boundary detections from the rest.*

A Loss for Learning from Proportion Intervals (LPI)

- A good LPI loss would be near-constant when the predicted proportion $\hat{y}$ is inside the ground truth interval $[\mathrm{thresh}_{y+1}, \mathrm{thresh}_{y+2})$ and would increase as $\hat{y}$ goes further away from the interval's boundaries. We propose a differentiable loss that approximates those properties:

$$\mathrm{LPI}_{\mathrm{ncat}}(\hat{y}, y) = \sum_{c=1}^{\mathrm{ncat}-1} w_c \mathrm{CrossEntropy}(\sigma_\alpha(\hat{y} - \mathrm{thresh}_{c+1}), \mathbb{I}(y \geq c))$$

  where $\sigma_\alpha(x) = (1 + e^{-\alpha x})^{-1}$ is a sharper version of the sigmoid function, w are tunable weights and $y \in [0, \mathrm{ncat} - 1]$ is the ground truth grade
- LPI_1 can be used for multiple instance learning (MIL) when only emphysema presence labels are available
- LPI_6 with $\mathrm{thresh} = (0, 0.005, 0.055, 0.255, 0.505, 0.755, 1)$ is used in our emphysema proportion prediction (thresh is defined by the grading system). See Fig. 2 (bottom left) for an illustration

Results

- We experimented with two scenarios: MIL (only binary emphysema presence labels are available) and LLP (emphysema grades are available)
- We compared GAPNet and ProportionNet in MIL (both nets trained using LPI_1) and LLP (trained with LPI_6) scenarios

Table 1: *Performance of emphysema presence detection and extent estimation (measured in average AUC over multiple test sets) of networks trained on sets of different size (in patients) and using different labels (MIL: only presence labels, LLP: extent labels).*

	Architecture:	**GAPNet**		**ProportionNet**	
	Training set size\Task:	Presence	Extent	Presence	Extent
MIL	small sets (50, 75, 100)	0.87 ± 0.05	0.68 ± 0.06	$\mathbf{0.95 \pm 0.01}$	$\mathbf{0.74 \pm 0.02}$
	medium sets (150, 200, 300)	$\mathbf{0.96 \pm 0.01}$	0.72 ± 0.02	$\mathbf{0.96 \pm 0.01}$	$\mathbf{0.74 \pm 0.02}$
	large set (700)	**0.96**	0.76	**0.96**	**0.79**
LLP	small sets (50, 75, 100)	0.90 ± 0.04	0.74 ± 0.06	$\mathbf{0.94 \pm 0.01}$	$\mathbf{0.79 \pm 0.02}$
	medium sets (150, 200, 300)	$\mathbf{0.96 \pm 0.01}$	0.80 ± 0.02	$\mathbf{0.96 \pm 0.01}$	$\mathbf{0.84 \pm 0.01}$
	large set (700)	0.96	0.79	**0.97**	**0.86**

- We compared our method to an established automatic method (densitometry) and recently published methods using the same dataset [3] and [4]; we also compare our custom loss LPI_6 to a conventional regression loss (RMS)

Table 2: *Comparison of our networks with other approaches. "RU" and "LU" stand for right and left upper regions.*

Labels:	**LLP**								**MIL**	
Training set size:	**100 subjects**				**700 subjects**				**700 subjects**	
Region:	**RU**		**LU**		**RU**		**LU**		**RU**	**LU**
Metric:	ICC	r_s	ICC	r_s	ICC	AUC	ICC	AUC	AUC	AUC
Densitometry	-	0.23	-	0.14	-	0.59	-	0.54	0.59	0.54
[3] and [4]	0.72	-	0.63	-	-	-	-	-	0.89	0.87
GAPNet+RMS	-	-	-	-	0.79	0.93	0.76	0.90	-	-
GAPNet+LPI_6	0.77	0.62	0.74	0.52	0.82	0.96	0.76	0.94	**0.96**	**0.94**
ProportionNet	**0.87**	**0.73**	**0.81**	**0.66**	**0.87**	**0.97**	**0.85**	**0.95**	**0.96**	**0.94**

- We validated ProportionNet segmentations indirectly by evaluating how predictive they are of emphysema patterns. The most common emphysema patterns are centrilobular and paraseptal. Paraseptal emphysema is located adjacent to lung pleura, whereas centrilobular can be anywhere in the lungs. We designed a simple feature to discriminate between the two, given an emphysema segmentation: a ratio between the foreground volume near the boundary and inside the region (see Fig. 2, bottom right). We computed this feature using segmentations of ProportionNet and obtained AUC 0.89 and this performance is on a par with human raters.

References

[1] M. M. W. Wille, L. H. Thomsen, A. Dirksen, J. Petersen, J. H. Pedersen, and S. B. Shaker, "Emphysema progression is visually detectable in low-dose CT in continuous but not in former smokers," *European Radiology*, vol. 24, no. 11, pp. 2692–2699, 2014.

[2] G. Patrini, R. Nock, P. Rivera, and T. Caetano, "(Almost) No Label No Cry," *NIPS 2014*, no. c, pp. 1–9, 2014.

[3] S. N. Ørting, J. Petersen, M. M. W. Wille, L. H. Thomsen, and M. de Bruijne, "Quantifying Emphysema Extent from Weakly Labeled CT Scans of the Lungs using Label Proportions Learning," in *Proc. of Sixth International Workshop on Pulmonary Image Analysis*, 2016.

[4] S. N. Ørting, J. Petersen, L. H. Thomsen, M. M. W. Wille, and M. de Bruijne, "Detecting Emphysema with Multiple Instance Learning," in *ISBI*, 2018.

 g.bortsova@erasmusmc.nl