# Deep Active Self-paced Learning for Accurate Pulmonary Nodule Segmentation

Wenzhe Wang[1], Yifei Lu[1], Bian Wu[2], Tingting Chen[1], Danny Z. Chen[3], and Jian Wu[1(✉)]

[1] College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
wujian2000@zju.edu.cn

[2] Data Science and AI Lab, WeDoctor Group Limited, Hangzhou 311200, China

[3] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

**Abstract.** Automatic and accurate pulmonary nodule segmentation in lung Computed Tomography (CT) volumes plays an important role in computer-aided diagnosis of lung cancer. However, this task is challenging due to target/background voxel imbalance and the lack of voxel-level annotation. In this paper, we propose a novel deep region-based network, called Nodule R-CNN, for efficiently detecting pulmonary nodules in 3D CT images while simultaneously generating a segmentation mask for each instance. Also, we propose a novel Deep Active Self-paced Learning (DASL) strategy to reduce annotation effort and also make use of unannotated samples, based on a combination of Active Learning and Self-Paced Learning (SPL) schemes. Experimental results on the public LIDC-IDRI dataset show our Nodule R-CNN achieves state-of-the-art results on pulmonary nodule segmentation, and Nodule R-CNN trained with the DASL strategy performs much better than Nodule R-CNN trained without DASL using the same amount of annotated samples.

## 1 Introduction

Lung cancer is one of the most life-threatening malignancies. A pulmonary nodule is a small growth in the lung which has the risk of being a site of cancerous tissue. The boundaries of pulmonary nodules have been regarded as a vital criterion for lung cancer analysis [1], and lung Computed Tomography (CT) is one of the most common methods for examining the presence and boundary features of pulmonary nodules. Automated segmentation of pulmonary nodules in CT volumes will promote early diagnosis of lung cancer by reducing the need for expensive human expertise.

Deep learning has become a powerful tool for a variety of medical imaging applications. However, to achieve good performance on 3D image segmentation, sufficient voxel-level annotations are commonly needed to train a deep


A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11071, pp. 723–731, 2018.
https://doi.org/10.1007/978-3-030-00934-2_80

network, which is both time-consuming and costly to obtain. As an effort to tackle this predicament, recent studies [2,3] conducted pulmonary nodule segmentation using weakly labeled data. However, restricted by rough annotations in lung CT volumes, these methods did not perform very well, usually producing rough boundary segmentation of pulmonary nodules and incurring considerable false positives. On the other hand, a deep active learning framework [4] was proposed to annotate samples during network training. Although being able to make good use of fully-annotated samples, this approach did not utilize abundant unannotated samples in model training.

In this paper, we propose a novel deep region-based network, called Nodule R-CNN, for volumetric instance-level segmentation in lung CT volumes, and a novel Deep Active Self-paced Learning (DASL) strategy to reduce annotation effort based on bootstrapping [4,5]. Due to the sparse distribution of pulmonary nodules in CT volumes [6], employing 3D fully convolutional networks (e.g., [7,8]) to semantically segment them may suffer the class imbalance issue. Built on 3D image segmentation work [8,9] and Mask R-CNN [10], our 3D region-based network provides an effective way for pulmonary nodule segmentation. Further, to alleviate the lack of fully-annotated samples and make use of unannotated samples, we propose a novel DASL strategy to improve our Nodule R-CNN by combining Active Learning (AL) [11] and Self-Paced Learning (SPL) [12] schemes. To our best knowledge, this is the first work on pulmonary nodule instance segmentation in 3D images, and the first work to train 3D CNNs using both AL and SPL.

Figure 1 outlines the main steps of our framework. Starting with annotated samples, we train our Nodule R-CNN, and use it to predict on unannotated samples. After ranking the confidence and uncertainty of each test sample, we utilize high-confidence and high-uncertainty samples in self-paced and active annotation learning [13], respectively, and add them to the training set to fine-tune Nodule R-CNN. The testing and fine-tuning of Nodule R-CNN repeat until Active Learning process is terminated.

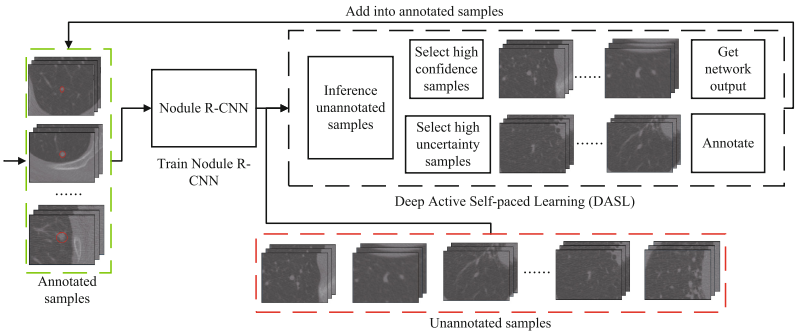


**Fig. 1.** Our weakly-supervised pulmonary nodule segmentation model.

Experimental results on the LIDC-IDRI dataset [6] show (1) our Nodule R-CNN can achieve state-of-the-art pulmonary nodule segmentation performance, and (2) our weakly-supervised segmentation approach is more effective than common fully supervised methods [3] and other weakly labeled methods [2,3].

## 2 Method

Our framework consists of two major components: (1) a novel region-based network (Nodule R-CNN) for pulmonary nodule instance segmentation; (2) a Deep Active Self-paced Learning (DASL) strategy for 3D CNN training.

### 2.1 Nodule R-CNN

Building on recent advances of convolutional neural networks such as Region Proposal Networks (RPN) [14], Feature Pyramid Networks (FPN) [15], Mask R-CNN [10], and DenseNet [16], we develop a novel deep region-based network for pulmonary nodule instance segmentation in 3D CT images.

Figure 2 illustrates the detailed architecture of our proposed Nodule R-CNN. Like Mask R-CNN [10], our network has a convolutional backbone architecture for feature extraction, a detection branch that outputs class labels and bounding-box offsets, and a mask branch that outputs object masks. In our backbone network, we extract diverse features in different levels by exploring an FPN-like architecture, which is a top-down architecture with lateral connections to build an in-network feature pyramid from a single-scale input. Three 3D DenseBlocks [9] with a growth rate 12 are used to ease network training by preserving maximum information flow between layers and to avoid learning redundant feature maps by encouraging feature reuse. Deconvolution is adopted to ensure that the size of the feature map is consistent with the size of the input volume. Our model employs an RPN-like architecture to output classification results and bounding-box regression results. The architecture provides three anchors for each detected location. We use a patch-based training and testing strategy instead of using RoIAlign to extract feature maps from RoIs due to limited GPU memory. In the mask branch, we utilize RoIPool to extract a small feature map from each RoI, and a Fully Convolutional Network (FCN) to generate the final label map of pulmonary nodule segmentation. In the final label map, the value of each voxel $a$ represents the probability of $a$ being a voxel of a pulmonary nodule.

We define a multi-task loss on each sampled RoI as $L = L_{cls} + L_{box} + L_{mask}$, where the classification loss $L_{cls}$ and the bounding-box loss $L_{box}$ are defined as in [17]. We define our segmentation loss $L_{mask}$ as Dice loss (since the output of the models trained with Dice loss is almost binary, it appears visually cleaner [8,18]). Specifically, the Dice loss is defined as:

$$L_{Dice} = -\frac{2\sum_i p_i y_i}{\sum_i p_i + \sum_i y_i}, \tag{1}$$

where $p_i \in [0,1]$ is the $i$-th output of the last layer in the mask branch passed through a sigmoid non-linearity and $y_i \in \{0,1\}$ is the corresponding label.

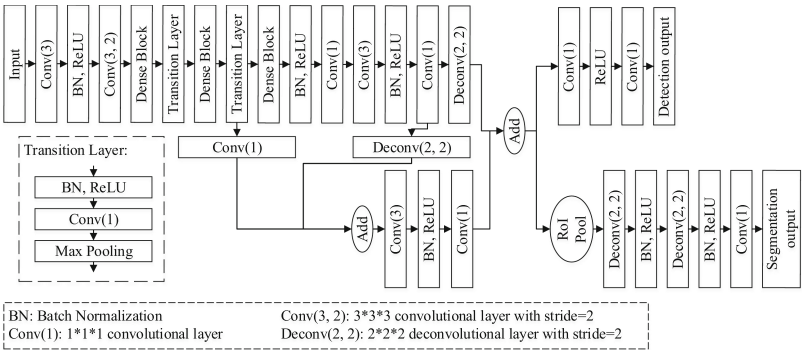


**Fig. 2.** The detailed architecture of our Nodule R-CNN.

### 2.2 The Deep Active Self-paced Learning Strategy

**Active Learning Scheme.** Active Learning attempts to overcome the annotation bottleneck by querying the most confusing unannotated instances for further annotation [11]. Different from [4] which applied a set of FCNs for confusing sample selection, we utilize a straightforward strategy to select confusing samples during model training. To do so, a common approach is to calculate the uncertainty of each unannotated sample, and filtrate the most uncertain ones. The calculation of uncertainty is defined as:

$$U_d = \frac{1}{n}\sum_{i=1}^{n}(1 - \max(p_i, 1 - p_i)), \tag{2}$$

where $U_d$ denotes the uncertainty of the $d$-th sample and $n$ denotes the voxel number of the $d$-th sample.

Note that the initial training set is often too small to cover the entire population distribution. Thus, there are usually a lot of samples which a deep learning model is not (yet) trained with. It is not advisable to extensively annotate those samples of similar patterns in one iteration. As in [4], we use cosine similarity to estimate the similarity between volumes. Therefore, the uncertainty of the $d$-th volume is defined as:

$$U_d = \frac{1}{n}\sum_{i=1}^{n}(1 - \max(p_i, 1 - p_i)) \times \left(\frac{\sum_{j=1}^{D}\mathrm{sim}(P_d, P_j) - 1}{D - 1}\right)^{\beta}, \tag{3}$$

where $D$ denotes the number of unannotated volumes, sim() denotes cosine similarity, $P_d$ and $P_j$ denote the output of the $d$-th and $j$-th volumes, respectively, and $\beta$ is a hyper-parameter that controls the relative importance of the similarity term. Note that when $\beta = 0$, this definition degenerates to the *least confident* uncertainty as defined in Eq. (2). We set $\beta = 1$ in our experiments.

In each iteration, after acquiring the uncertainty of each unannotated sample, we select the top $N$ samples for annotation and add them to the training set for further fine-tuning.

**Self-paced Learning Scheme.** Self-Paced Learning (SPL) was inspired by the learning process of humans/animals that gradually incorporates easy-to-hard samples into training [12]. It utilizes unannotated samples by considering both prior knowledge known before training and the learning progress made during training [13].

Formally, let $L(\mathbf{w}; \mathbf{x}_i, p_i)$ denote the loss function of Nodule R-CNN, where $\mathbf{w}$ denotes the model parameters inside the model, $\mathbf{x}_i$ and $p_i$ denote the input and output of the model, respectively. SPL aims to optimize the following function:

$$\min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \mathbb{E}(\mathbf{w}, \mathbf{v}; \lambda, \Psi) = C \sum_{i=1}^{n} v_i L(\mathbf{w}; \mathbf{x}_i, p_i) + f(\mathbf{v}; \lambda), \quad s.t. \quad \mathbf{v} \in \Psi \tag{4}$$

where $\mathbf{v} = [v_1, v_2, \ldots, v_n]^T$ denotes the weight variables reflecting the samples' confidence, $f(\mathbf{v}; \lambda)$ is a self-paced regularization term that controls the learning scheme, $\lambda$ is a parameter for controlling the learning pace, $\Psi$ is a feasible region that encodes the information of predetermined curriculum, and $C$ is a standard regularization parameter for the trade-off of the loss function and the margin. We set $C = 1$ in our experiments.

Note that a self-paced function should satisfy three conditions [19]. (1) $f(\mathbf{v}; \lambda)$ is convex with respect to $\mathbf{v} \in [0, 1]^n$. (2) The optimal weight of each sample $v_i^*$ should be monotonically decreasing with respect to its corresponding loss $l_i$. (3) $\|\mathbf{v}\|_1 = \sum_{i=1}^{n} v_i$ should be monotonically increasing with respect to $\lambda$.

To linearly discriminate the samples with their losses, the regularization function of our learning scheme is defined as follows [19]:

$$f(\mathbf{v}; \lambda) = \lambda \left( \frac{1}{2} \|\mathbf{v}\|_2^2 - \sum_{i=1}^{n} v_i \right), \tag{5}$$

With $\Psi = [0, 1]^n$, the partial gradient of Eq. (4) using our learning scheme is equal to

$$\frac{\partial \mathbb{E}}{\partial v_i} = C l_i + v_i \lambda - \lambda = 0, \tag{6}$$

where $\mathbb{E}$ denotes the objective in Eq. (4) with a fixed $\mathbf{w}$, and $l_i$ denotes the loss of the $i$-th sample. The optimal solution for $\mathbb{E}$ is given by Eq. (8) below. Note that since the labels of unannotated samples are unknown, it is challenging to calculate their losses. We allocate each "pseudo-label" by Eq. (7).

$$y_i^* = \underset{y_i=\{0,1\}}{\operatorname{argmin}} \, l_i, \tag{7}$$

$$v_i^* = \begin{cases} 1 - \dfrac{C l_i}{\lambda}, & C l_i < \lambda \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

For pace parameter updating, we set the initial pace as $\lambda^0$. For the $t$-th iteration, we compute the pace parameter $\lambda^t$ as follows:

$$\lambda^t = \begin{cases} \lambda^0, & t = 0 \\ \lambda^{(t-1)} + \alpha \times \eta^t, & 1 \leq t < \tau \\ \lambda^{(t-1)}, & t \geq \tau, \end{cases} \tag{9}$$

where $\alpha$ is a hyper-parameter that controls the pace increasing rate, $\eta^t$ is the average accuracy in the current iteration, and $\tau$ is a hyper-parameter for controlling the pace update. Note that based on the third condition defined above, $\|\mathbf{v}\|_1 = \sum_{i=1}^{n} v_i$ should be monotonically increasing with respect to $\lambda$. Since $\mathbf{v} \in [0, 1]^n$, the updating of the parameter $\lambda$ should be stopped after a few iterations. Thus, we introduce the hyper-parameter $\tau$ to control the pace updating.

## 3 Experiments and Results

We evaluate our proposed approach using the LIDC-IDRI dataset [6]. Our experimental results are given in Table 1.

The LIDC-IDRI dataset contains 1010 CT scans (see [6] for more details of this dataset). In our experiments, all nodules are used except those with a diameter $<$ 3mm, and each scan is resized to 512 $\times$ 512 $\times$ 512 voxels by linear interpolation. The inputs of our model are 3D patches in the size of 128 $\times$ 128 $\times$ 128 voxels, which are cropped from CT volumes. 70% of the input patches contain at least one nodule. For this part of the inputs, segmentation masks are cropped to 32 $\times$ 32 $\times$ 32 voxels with nodules centering in them. We obtain the rest of the inputs by randomly cropping scans that very likely contain no nodule. The output size of the detection branch is 32 $\times$ 32 $\times$ 32 $\times$ 3 $\times$ 5, where the second last dimension represents 3 anchors and the last dimension corresponds to the classification results and bounding-box regression results. In our experiments, 10% of the whole dataset are randomly selected as the validation set. We use a small subset of the remaining scans to train the initial Nodule R-CNN and the rest samples are gradually added to the training set during the DASL process.

First, we evaluate our Nodule R-CNN for pulmonary nodule instance segmentation. As shown in Table 1, its Dice achieves 0.64 and TP Dice (Dice over truly detected nodules) achieves 0.95, both of which are best results among state-of-the-art methods.

We then evaluate the combination of Nodule R-CNN and the DASL strategy. In our experiments, $\alpha$ is set to 0.002 and $\lambda^0$ is set to 0.005, due to the high confidence of positive prediction. To verify the relationship between AL and SPL in DASL, we use a sequence of "SPL-AL-SPL" to fine-tune Nodule R-CNN. To verify the impact of different amounts of initial annotated samples, we conduct three experiments with 50, 100, and 150 initial annotated samples, respectively. Figure 3 summarizes the results. We find that, in DASL, when using less initial annotated samples to train Nodule R-CNN, SPL tends to incorporate more unannotated samples. This makes sense since the model trained with less

samples does not learn enough patterns and is likely to allocate high confidence to more unseen samples. One can see from Fig. 3 that although the amount of samples selected by AL is quite small ($N = 20$ in our experiments), AL does help achieve higher Dice. Experimental results are shown in Table 1. We find that more initial annotated samples bring better results, and the experiment with 150 initial annotated samples gives the best results among our experiments on DASL, which is comparable to the performance of Nodule R-CNN trained with all samples.

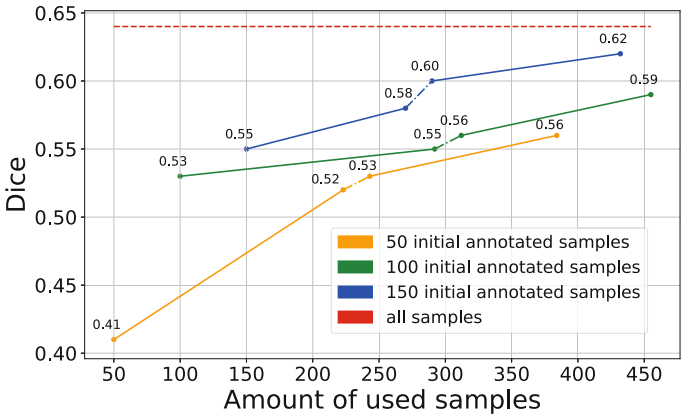


**Fig. 3.** Comparison using different amounts of initial annotated inputs for DASL: The solid lines are for the SPL process, the dotted lines are for the AL process, and the dashed line is for the current state-of-the-art result using full training samples.

**Table 1.** Results on the LIDC-IDRI dataset for pulmonary nodule segmentation.

| Method | Dice mean ± SD | TP Dice mean ± SD |
|---|---|---|
| Method in [3] | 0.55(±0.33) | 0.74(±0.14) |
| Nodule R-CNN (full training samples) | **0.64(±0.44)** | **0.95(±0.12)** |
| Nodule R-CNN with DASL (50 initial annotated samples) | 0.56(±0.45) | 0.87(±0.09) |
| Nodule R-CNN with DASL (100 initial annotated samples) | 0.59(±0.45) | 0.90(±0.05) |
| Nodule R-CNN with DASL (150 initial annotated samples) | 0.62(±0.43) | 0.92(±0.03) |

## 4 Conclusions

We have developed a novel Deep Active Self-paced Learning framework for pulmonary nodule instance segmentation in 3D CT images by combining our proposed Nodule R-CNN and Deep Active Self-paced Learning. Our new approach makes two main contributions: (1) A Nodule R-CNN model that attains state-of-the-art pulmonary nodule segmentation performance; (2) a weakly-supervised method that can make good use of annotation effort as well as information of unannotated samples.

**Acknowledgement.** The research of D.Z. Chen was supported in part by NSF Grant CCF-1617735.

## References

1. Gonçalves, L., Novo, J.: Hessian based approaches for 3D lung nodule segmentation. Expert Syst. Appl. **61**, 1–15 (2016)
2. Messay, T., Hardie, R.C.: Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset. Med. Image Anal. **22**(1), 48–62 (2015)
3. Feng, X., Yang, J., et al.: Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. In: MICCAI, pp. 568–576 (2017)
4. Yang, L., Zhang, Y., et al.: Suggestive annotation: a deep active learning framework for biomedical image segmentation. In: MICCAI, pp. 399–407 (2017)
5. Li, X., Zhong, A., et al.: Self-paced convolutional neural network for computer aided detection in medical imaging analysis. In: International Workshop on Machine Learning in Medical Imaging, pp. 212–219 (2017)
6. Armato, S.G., McLennan, G.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phy. **38**(2), 915–931 (2011)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
8. Milletari, F., Navab, N., et al.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 4th IEEE International Conference on 3D Vision, pp. 565–571 (2016)
9. Yu, L., et al.: Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 287–295. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_33
10. He, K., Gkioxari, G., et al.: Mask R-CNN. In: ICCV, pp. 2980–2988 (2017)
11. Settles, B.: Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin - Madison (2009)
12. Kumar, M.P., Packer, B., et al.: Self-paced learning for latent variable models. In: NIPS, pp. 1189–1197 (2010)

13. Lin, L., Wang, K., et al.: Active self-paced learning for cost-effective and progressive face identification. IEEE TPAMI **40**(1), 7–19 (2018)
14. Ren, S., He, K., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS, pp. 91–99 (2015)
15. Lin, T.Y., Dollár, P., et al.: Feature pyramid networks for object detection. In: CVPR, vol. 1, p. 4 (2017)
16. Huang, G., Liu, Z., et al.: Densely connected convolutional networks. In: CVPR, vol. 1, p. 3 (2017)
17. Girshick, R.: Fast R-CNN. In: ICCV, pp. 1440–1448 (2015)
18. Drozdzal, M., Vorontsov, E., et al.: The importance of skip connections in biomedical image segmentation. In: Deep Learning and Data Labeling for Medical Applications, pp. 179–187 (2016)
19. Jiang, L., Meng, D.: Self-paced curriculum learning. In: AAAI, vol. 2, p. 6 (2015)