DeepEM: Deep 3D ConvNets with EM for Weakly Supervised Pulmonary Nodule Detection

Wentao Zhu[1]([✉]), Yeeleng S. Vang[1], Yufang Huang[2], and Xiaohui Xie[1]

[1] University of California, Irvine, USA
{wentaoz1,ysvang,xhx}@uci.edu
[2] Lenovo Research, Beijing, China
huangyf8@lenovo.com

Abstract. Recently deep learning has been witnessing widespread adoption in various medical image applications. However, training complex deep neural nets requires large-scale datasets labeled with ground truth, which are often unavailable in many medical image domains. For instance, to train a deep neural net to detect pulmonary nodules in lung computed tomography (CT) images, current practice is to manually label nodule locations and sizes in many CT images to construct a sufficiently large training dataset, which is costly and difficult to scale. On the other hand, electronic medical records (EMR) contain plenty of partial information on the content of each medical image. In this work, we explore how to tap this vast, but currently unexplored data source to improve pulmonary nodule detection. We propose DeepEM, a novel deep 3D ConvNet framework augmented with expectation-maximization (EM), to mine weakly supervised labels in EMRs for pulmonary nodule detection. Experimental results show that DeepEM can lead to 1.5% and 3.9% average improvement in free-response receiver operating characteristic (FROC) scores on LUNA16 and Tianchi datasets, respectively, demonstrating the utility of incomplete information in EMRs for improving deep learning algorithms (https://github.com/uci-cbcl/DeepEM-for-Weakly-Supervised-Detection.git).

Keywords: Deep 3D convolutional nets
Weakly supervised detection · DeepEM (deep 3D ConvNets with EM)
Pulmonary nodule detection

1 Introduction

Lung cancer is the most common cause of cancer-related death in men. Low-dose lung computed tomography (CT) screening provides an effective way for early diagnosis and can sharply reduce the lung cancer mortality rate. Advanced computer-aided diagnosis (CAD) systems are expected to have high sensitivities while maintaining low false positive rates to be truly useful. Recent advance in

deep learning provides new opportunities to design more effective CAD systems to help facilitate doctors in their effort to catch lung cancer in their early stages.

The emergence of large-scale datasets such as the LUNA16 [12] has helped to accelerate research in nodule detection. Typically, nodule detection consists of two stages: nodule proposal generation and false positive reduction. Traditional approaches generally require hand-designed features such as morphological features, voxel clustering and pixel thresholding [6,9]. More recently, deep convolutional architectures were employed to generate the candidate bounding boxes. Setio et al. proposed multi-view convolutional network for false positive nodule reduction [11]. Several work employed 3D convolutional networks to handle the challenge due to the 3D nature of CT scans. The 3D fully convolutional network (FCN) was proposed to generate region candidates and deep convolutional network with weighted sampling was used in the false positive reduction stage [3,8,13,14]. CASED proposed curriculum adaptive sampling for 3D U-net training in nodule detection [7,10]. Ding et al. used Faster R-CNN to generate candidate nodules, followed by 3D convolutional networks to remove false positive nodules [2]. Due to the effective performance of Faster R-CNN [14], Faster R-CNN with a U-net-like encoder-decoder scheme was proposed for nodule detection [14].
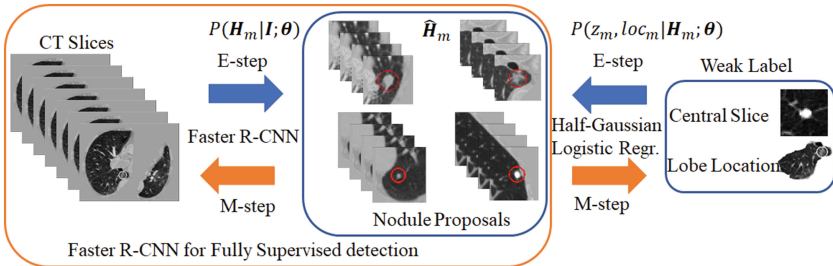

Fig. 1. Illustration of DeepEM framework. Faster R-CNN is employed for nodule proposal generation. Half-Gaussian model and logistic regression are employed for central slice and lobe location respectively. In the E-step, we utilize all the observations, CT slices, and weak label to infer the latent variable, nodule proposals, by maximum a posteriori (MAP) or sampling. In the M-step, we employ the estimated proposals to update parameters in the Faster R-CNN and logistic regression.

A prerequisite to utilization of deep learning models is the existence of an abundance of labeled data. However, labels are especially difficult to obtain in the medical image analysis domain. There are multiple contributing factors: (a) labeling medical data typically requires specially trained doctors; (b) marking lesion boundaries can be hard even for experts because of low signal-to-noise ratio in many medical images; and (c) for CT and magnetic resonance imaging (MRI) images, the annotators need to label the entire 3D volumetric data, which can be costly and time-consuming. Due to these limitations, CT medical image

datasets are usually small, which can lead to over-fitting on the training set and, by extension, poor generalization performance on test sets [16].

By contrast, medical institutions have large amount of weakly labeled medical images. In these databases, each medical image is typically associated with an electronic medical report (EMR). Although these reports may not contain explicit information on detection bounding box or segmentation ground truth, it often includes the results of diagnosis, rough locations and summary descriptions of lesions if they exist. We hypothesize that these extra sources of weakly labeled data may be used to enhance the performance of existing detector and improve its generalization capability.

There are previous attempts to utilize weakly supervised labels to help train machine learning models. Deep multi-instance learning was proposed for lesion localization and whole mammogram classification [15]. Different pooling strategies were proposed for weakly supervised localization and segmentation respectively [1,4]. Self-transfer learning co-optimized both classification and localization networks for weakly supervised lesion localization [5]. Different from these works, we consider nodule proposal as latent variable and propose DeepEM, a new deep 3D convolutional nets with Expectation-Maximization optimization, to mine the big data source of weakly supervised label in EMR as illustrated in Fig. 1. Specifically, we infer the posterior probabilities of the proposed nodules being true nodules, and utilize the posterior probabilities to train nodule detection models.

2   DeepEM for Weakly Supervised Detection

Notation. We denote by $\boldsymbol{I} \in \mathbb{R}^{h \times w \times s}$ the CT image, where h, w, and s are image height, width, and number of slices respectively. The nodule bounding boxes for $\boldsymbol{I}$ are denoted as $\boldsymbol{H} = \{\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_M\}$, where $\boldsymbol{H}_m = \{x_m, y_m, z_m, d_m\}$, the (x_m, y_m, z_m) represents the center of nodule proposal, d_m is the diameter of the nodule proposal, and M is the number of nodules in the image $\boldsymbol{I}$. In the weakly supervised scenario, the nodule proposal $\boldsymbol{H}$ is a latent variable, and each image $\boldsymbol{I}$ is associated with weak label $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M\}$, where $\boldsymbol{X}_m = \{loc_m, z_m\}$, $loc_m \in \{1, 2, 3, 4, 5, 6\}$ is the location (right upper lobe, right middle lobe, right lower lobe, left upper lobe, lingula, left lower lobe) of nodule $\boldsymbol{H}_m$ in the lung, and z_m is the central slice of the nodule.

For fully supervised detection, the objective function is to maximize the log-likelihood function for observed nodule ground truth $\boldsymbol{H}$ given image $\boldsymbol{I}$ as

$$\mathcal{L}(\boldsymbol{\theta}) = \log P(\boldsymbol{H} \cup \bar{\boldsymbol{H}} | \boldsymbol{I}; \boldsymbol{\theta}) = \frac{1}{M} \sum_{m=1}^{M} \log P(\boldsymbol{H}_m | \boldsymbol{I}; \boldsymbol{\theta}) + \frac{1}{N} \sum_{n=1}^{N} \log P(\bar{\boldsymbol{H}}_n | \boldsymbol{I}; \boldsymbol{\theta}),$$

(1)

where $\bar{\boldsymbol{H}} = \{\bar{\boldsymbol{H}}_1, \bar{\boldsymbol{H}}_2, \ldots, \bar{\boldsymbol{H}}_N\}$ are hard negative nodule proposals [14], $\boldsymbol{\theta}$ is the weights of deep 3D ConvNet. We employ Faster R-CNN with 3D Res18 for the fully supervised detection because of its superior performance.

For weakly supervised detection, nodule proposal $\boldsymbol{H}$ can be considered as a latent variable. Using this framework, image $\boldsymbol{I}$ and weak label $\boldsymbol{X} = \{(loc_1, z_1), (loc_2, z_2), \ldots, (loc_M, z_M)\}$ can be considered as observations. The joint distribution is

$$
\begin{aligned}
P(\boldsymbol{I}, \boldsymbol{H}, \boldsymbol{X}; \boldsymbol{\theta}) &= P(\boldsymbol{I}) \prod_{m=1}^{M} \big(P(\boldsymbol{H}_m | \boldsymbol{I}; \boldsymbol{\theta}) P(\boldsymbol{X}_m | \boldsymbol{H}_m; \boldsymbol{\theta}) \big) \\
&= P(\boldsymbol{I}) \prod_{m=1}^{M} \big(P(\boldsymbol{H}_m | \boldsymbol{I}; \boldsymbol{\theta}) P(loc_m | \boldsymbol{H}_m; \boldsymbol{\theta}) P(z_m | \boldsymbol{H}_m; \boldsymbol{\theta}) \big).
\end{aligned}
\tag{2}
$$

To model $P(z_m | \boldsymbol{H}_m; \boldsymbol{\theta})$, we propose using a half-Gaussian distribution based on nodule size distribution because z_m is correct if it is within the nodule area (center slice of $\boldsymbol{H}_m$ as $z_{\boldsymbol{H}_m}$, and nodule size σ can be empirically estimated based on existing data) for nodule detection in Fig. 2(a). For lung lobe prediction $P(loc_m | \boldsymbol{H}_m; \boldsymbol{\theta})$, a logistic regression model is used based on relative value of nodule center $(x_{\boldsymbol{H}_m}, y_{\boldsymbol{H}_m}, z_{\boldsymbol{H}_m})$ after lung segmentation. That is

$$
P(z_m, loc_m | \boldsymbol{H}_m; \boldsymbol{\theta}) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{|z_m - z_{\boldsymbol{H}_m}|^2}{2\sigma^2} \big) \frac{\exp(\boldsymbol{f}(\boldsymbol{H}_m)\boldsymbol{\theta}_{loc_m})}{\sum_{loc_m=1}^{6} \exp(\boldsymbol{f}(\boldsymbol{H}_m)\boldsymbol{\theta}_{loc_m})},
\tag{3}
$$

where $\boldsymbol{\theta}_{loc_m}$ is the associated weights with lobe location loc_m for logistic regression, feature $\boldsymbol{f}(\boldsymbol{H}_m) = (\frac{x_{\boldsymbol{H}_m}}{x_{\boldsymbol{I}}}, \frac{y_{\boldsymbol{H}_m}}{y_{\boldsymbol{I}}}, \frac{z_{\boldsymbol{H}_m}}{z_{\boldsymbol{I}}})$, and $(x_{\boldsymbol{I}}, y_{\boldsymbol{I}}, z_{\boldsymbol{I}})$ is the total size of image $\boldsymbol{I}$ after lung segmentation. In the experiments, we found logistic regression converges quickly and is stable.

The expectation-maximization (EM) is a commonly used approach to optimize the maximum log-likelihood function when there are latent variables in the model. We employ the EM algorithm to optimize deep weakly supervised detection model in Eq. 2. The expected complete-data log-likelihood function given previous estimated parameter $\boldsymbol{\theta}'$ in deep 3D Faster R-CNN is

$$
\begin{aligned}
Q(\boldsymbol{\theta}; \boldsymbol{\theta}') = \frac{1}{M} \sum_{m=1}^{M} &\mathbb{E}_{P(\boldsymbol{H}_m | \boldsymbol{I}, z_m, loc_m; \boldsymbol{\theta}')} \big[\log P(\boldsymbol{H}_m | \boldsymbol{I}; \boldsymbol{\theta}) \\
&+ \log P(z_m, loc_m | \boldsymbol{H}_m; \boldsymbol{\theta}) \big] + \mathbb{E}_{Q(\bar{\boldsymbol{H}}_n | \boldsymbol{z})} \big[\log P(\bar{\boldsymbol{H}}_n | \boldsymbol{I}; \boldsymbol{\theta}) \big],
\end{aligned}
\tag{4}
$$

where $\boldsymbol{z} = \{z_1, z_2, \ldots, z_m\}$. In the implementation, we only keep hard negative proposals far away from weak annotation $\boldsymbol{z}$ to simplify $Q(\bar{\boldsymbol{H}}_n | \boldsymbol{z})$. The posterior distribution of latent variable $\boldsymbol{H}_m$ can be calculated by

$$
P(\boldsymbol{H}_m | \boldsymbol{I}, z_m, loc_m; \boldsymbol{\theta}') \propto P(\boldsymbol{H}_m | \boldsymbol{I}; \boldsymbol{\theta}') P(z_m, loc_m | \boldsymbol{H}_m; \boldsymbol{\theta}').
\tag{5}
$$

Because Faster R-CNN yields a large number of proposals, we first use hard threshold (-3 before sigmoid function) to remove proposals of small confident probability, then employ non-maximum suppression (NMS) with intersection over union (IoU) as 0.1. We then employ two schemes to approximately infer the latent variable $\boldsymbol{H}_m$: maximum a posteriori (MAP) or sampling.

Algorithm 1. DeepEM for Weakly Supervised Detection

Input: Fully supervised dataset $D_F = \{(\boldsymbol{I}, \boldsymbol{H})_i\}_{i=1}^{N_F}$, weakly supervised dataset $D_W = \{(\boldsymbol{I}, \boldsymbol{X})_i\}_{i=1}^{N_W}$, 3D Faster R-CNN and logistic regression parameters $\boldsymbol{\theta}$.

1: *Initialization*: Update weights $\boldsymbol{\theta}$ by maximizing Eq. 1 using data from D_F.
2: *for epoch = 1 to #TotalEpochs*:
        ▷▷▷ **Weakly supervised training**
3:      Use Faster R-CNN model $\boldsymbol{\theta}'$ to obtain proposal probability $P(\boldsymbol{H}_m|\boldsymbol{I}; \boldsymbol{\theta}')$ for weakly supervised data sampled from D_W.
4:      Remove proposals with small probabilities and NMS.
5:     *for m = 1 to M*:   ▷▷▷ Each weak label
6:         Calculate $P(z_m, loc_m|\boldsymbol{H}_m; \boldsymbol{\theta})$ for each proposal by Eq. 3.
7:         Estimate posterior distribution $P(\boldsymbol{H}_m|\boldsymbol{I}, z_m, loc_m; \boldsymbol{\theta}')$ by Eq. 5 with normalization.
8:         Employ MAP by Eq. 6 or Sampling to obtain the inference of $\boldsymbol{H}_m$.
9:      Obtain the expect log-likelihood function by Eq. 4 using the estimated proposal (MAP) or by Eq. 7 (Sampling).
10:     Update parameter by equation 8.
        ▷▷▷ **Fully supervised training**
11:     Update weights $\boldsymbol{\theta}$ by maximizing Eq. 1 using fully supervised data D_F.

DeepEM with MAP. We only use the proposal of maximal posterior probability to calculate the expectation.

$$\hat{\boldsymbol{H}}_m = \arg\max_{\boldsymbol{H}_m} P(\boldsymbol{H}_m|\boldsymbol{I}; \boldsymbol{\theta}') P(z_m, loc_m|\boldsymbol{H}_m; \boldsymbol{\theta}') \tag{6}$$

DeepEM with Sampling. We approximate the distribution by sampling $\hat{M}$ proposals $\hat{\boldsymbol{H}}_m$ according to normalized Eq. 5. The expected log-likelihood function in Eq. 4 becomes

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}') = \frac{1}{M\hat{M}} \sum_{m=1}^{M} \sum_{\hat{\boldsymbol{H}}_m}^{\hat{M}} \big(\log P(\hat{\boldsymbol{H}}_m|\boldsymbol{I}; \boldsymbol{\theta}) + \log P(z_m, loc_m|\hat{\boldsymbol{H}}_m; \boldsymbol{\theta}) \big) \\ + \mathbb{E}_{Q(\bar{\boldsymbol{H}}_n|\boldsymbol{z})} \big[\log P(\bar{\boldsymbol{H}}_n|\boldsymbol{I}; \boldsymbol{\theta}) \big]. \tag{7}$$

After obtaining the expectation of complete-data log-likelihood function in Eq. 4, we can update the parameters $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}} = \arg\max Q(\boldsymbol{\theta}; \boldsymbol{\theta}'). \tag{8}$$

The M-step in Eq. 8 can be conducted by stochastic gradient descent commonly used in deep network optimization for Eq. 1. Our entire algorithm is outlined in Algorithm 1.

3 Experiments

We used 3 datasets, LUNA16 dataset [12] as fully supervised nodule detection, NCI NLST[1] dataset as weakly supervised detection, Tianchi Lung Nodule Detection[2] dataset as holdout dataset for test only. LUNA16 dataset is the largest publicly available dataset for pulmonary nodules detection [12]. LUNA16 dataset removes CTs with slice thickness greater than 3 mm, slice spacing inconsistent or missing slices, and consist of 888 low-dose lung CTs which have explicit patient-level 10-fold cross validation split. NLST dataset consists of hundreds of thousands of lung CT images associated with electronic medical records (EMR). In this work, we focus on nodule detection based on image modality and only use the central slice and nodule location as weak supervision from the EMR. As part of data cleansing, we remove negative CTs, CTs with slice thickness greater than 3 mm and nodule diameter less than 3 mm. After data cleaning, we have 17,602 CTs left with 30,951 weak annotations. In each epoch, we randomly sample $\frac{1}{16}$ CT images for weakly supervised training because of the large numbers of weakly supervised CTs. Tianchi dataset contains 600 training low-dose lung CTs and 200 validation low-dose lung CTs for nodule detection. The annotations are location centroids and diameters of the pulmonary nodules, and do not have less than 3 mm diameter nodule, which are the same with those on LUNA16 dataset.

Parameter Estimation in $P(z_m|\boldsymbol{H}_m; \boldsymbol{\theta})$**.** If the current z_m is within the nodule, it is a true positive proposal. We can model $|z_m - z_{\boldsymbol{H}_m}|$ using a half-Gaussian distribution shown as the red dash line in Fig. 2(a). The parameters of the half-Gaussian is estimated from the LUNA16 data empirically. Because LUNA16 removes nodules of diameter less than 3 mm, we use the truncated half-Gaussian to model the central slice z_m as $\max(|z_m - z_{\boldsymbol{H}_m}| - \mu, 0)$, where μ is the mean of related Gaussian as the minimal nodule radius with 1.63.

Performance Comparisons on LUNA16. We conduct 10-fold cross validation on LUNA16 to validate the effectiveness of DeepEM. The baseline method is Faster R-CNN with 3D Res18 network denoted as **Faster R-CNN** [14]. Then we employ it to model $P(\boldsymbol{H}_m|\boldsymbol{I}; \boldsymbol{\theta}')$ for weakly supervised detection scenario. Two inference scheme for $\boldsymbol{H}_m$ are used in DeepEM denoted as **DeepEM (MAP)** and **DeepEM (Sampling)**. In the proposal inference of DeepEM with Sampling, we sample two proposals for each weak label because the average number of nodules each CT is 1.78 on LUNA16. The evaluation metric, Free receiver operating characteristic (FROC), is the average recall rate at the average number of false positives at 0.125, 0.25, 0.5, 1, 2, 4, 8 per scan, which is the official evaluation metric for LUNA16 and Tianchi [12].

From Fig. 2(b), DeepEM with MAP improves about 1.3% FROC over Faster R-CNN and DeepEM with Sampling improves about 1.5% FROC over Faster R-CNN on average on LUNA16 when incorporating weakly labeled data from NLST. We hypothesize the greater improvement of DeepEM with Sampling over

---

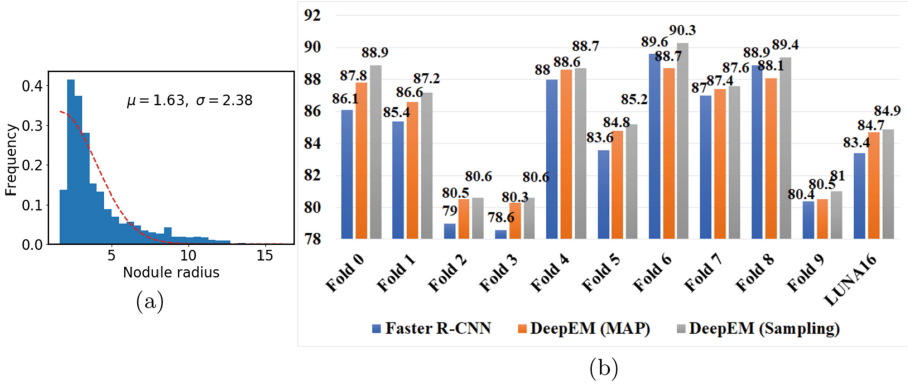(a)   (b)

Fig. 2. (a)Empirical estimation of half-Gaussian model for $P(z_m|\boldsymbol{H}_m;\boldsymbol{\theta})$ on LUNA16. (b) FROC (%) comparison among Faster R-CNN [14], DeepEM with MAP, DeepEM with Sampling on LUNA16.

DeepEM with MAP is that MAP inference is greedy and can get stuck at a local minimum while the nature of sampling may allow DeepEM with Sampling to escape these local minimums during optimization.

Performance Comparisons on Holdout Test Set from Tianchi. We employed a holdout test set from Tianchi to validate each model from 10-fold cross validation on LUNA16. The results are summarized in Table 1. We can see DeepEM utilizing weakly supervised data improves 3.9% FROC on average over Faster R-CNN. The improvement on holdout test data validates DeepEM as an effective model to exploit potentially large amount of weak data from electronic medical records (EMR) which would not require further costly annotation by expert doctors and can be easily obtained from hospital associations (Fig. 3).

Table 1. FROC (%) comparisons among Faster R-CNN with 3D ResNet18 [14], DeepEM with MAP, DeepEM with Sampling on Tianchi.

Fold	0	1	2	3	4	5	6	7	8	9	Average
Faster R-CNN	72.8	70.8	69.8	71.9	76.4	73.0	71.3	74.7	72.9	71.3	72.5
DeepEM (MAP)	77.2	75.8	75.8	74.9	77.0	75.5	77.2	75.8	76.0	74.7	76.0
DeepEM (Sampling)	77.4	75.8	75.9	75.0	77.3	75.0	77.3	76.8	77.7	75.8	76.4

Visualizations. We compare Faster R-CNN with the proposed DeepEM visually in Fig. 3. We randomly choose nodules from Tianchi. From Fig. 3, DeepEM yields better detection for nodule center and tighter nodule diameter which demonstrates DeepEM improves the existing detector by exploiting weakly supervised data.

Fig. 3. Detection visual comparison among Faster R-CNN [14], DeepEM with MAP and DeepEM with Sampling on nodules randomly sampled from Tianchi. DeepEM provides more accurate detection (central slice, center and diameter) than Faster R-CNN.

4   Conclusion

In this paper, we have focused on the problem of detecting pulmonary nodules from lung CT images, which previously has been formulated as a supervised learning problem and requires a large amount of training data with the locations and sizes of nodules precisely labeled. Here we propose a new framework, called DeepEM, for pulmonary nodule detection by taking advantage of abundantly available weakly labeled data extracted from EMRs. We treat each nodule proposal as a latent variable, and infer the posterior probabilities of proposal nodules being true ones conditioned on images and weak labels. The posterior probabilities are further fed to the nodule detection module for training. We use an EM algorithm to train the entire model end-to-end. Two schemes, maximum a posteriori (MAP) and sampling, are used for the inference of proposals. Extensive experimental results demonstrate the effectiveness of DeepEM for improving current state of the art nodule detection systems by utilizing readily available weakly supervised detection data. Although our method is built upon the specific application of pulmonary nodule detection, the framework itself is fairly general and can be readily applied to other medical image deep learning applications to take advantage of weakly labeled data.

References

1. Bilen, H., et al.: Weakly supervised deep detection networks. In: CVPR (2016)
2. Ding, J., Li, A., Hu, Z., Wang, L.: Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 559–567. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_64

3. Dou, Q., Chen, H., Jin, Y., Lin, H., Qin, J., Heng, P.-A.: Automated pulmonary nodule detection via 3D ConvNets with online sample filtering and hybrid-loss residual learning. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 630–638. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_72

4. Feng, X., Yang, J., Laine, A.F., Angelini, E.D.: Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 568–576. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_65

5. Hwang, S., Kim, H.-E.: Self-transfer learning for weakly supervised lesion localization. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 239–246. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_28

6. Jacob, C., et al.: Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. Med. Image Anal. **18**, 374–384 (2014)

7. Jesson, A., Guizard, N., Ghalehjegh, S.H., Goblot, D., Soudan, F., Chapados, N.: CASED: curriculum adaptive sampling for extreme data imbalance. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 639–646. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_73

8. Liao, F., et al.: Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network. arXiv preprint (2017)

9. Lopez Torres, E., et al.: Large scale validation of the M5L lung cad on heterogeneous ct datasets. Med. Phys. **42**, 1477–1489 (2015)

10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

11. Setio, A.A.A., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE TMI **35**, 1160–1169 (2016)

12. Setio, A.A.A., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. Med. Image Anal. **42**, 1–13 (2017)

13. Tang, H., Kim, D., Xie, X.: Automated pulmonary nodule detection using 3D deep convolutional neural networks. In: ISBI (2018)

14. Zhu, W., Liu, C., Fan, W., Xie, X.: Deeplung: deep 3d dual path nets for automated pulmonary nodule detection and classification. In: IEEE WACV (2018)

15. Zhu, W., Lou, Q., Vang, Y.S., Xie, X.: Deep multi-instance networks with sparse label assignment for whole mammogram classification. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 603–611. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_69

16. Zhu, W., et al.: Adversarial deep structured nets for mass segmentation from mammograms. In: IEEE ISBI (2018)