Deep Active Learning for Effective Pulmonary Nodule Detection

Jingya Liu[1], Liangliang Cao[2,3], and Yingli Tian[1,*]

[1] The City College of New York, New York, NY 10031
[2] UMass CICS, Amherst, MA 01002
[3] Google Inc, New York, NY, 10011

Abstract. Expensive and time-consuming medical imaging annotation is one of the big challenges for the deep learning-based computer-aided diagnosis (CAD) on the low-dose computed tomography (CT). To address this problem, we propose a novel active learning approach to improve the training efficiency for a deep network-based lung nodule detection framework as well as reduce the annotation cost. The informative CT scans, such as the samples that inconspicuous or likely to produce high false positives, are selected and further annotated for the nodule detector network training. A simple yet effective schema suggests the samples by ranking the uncertainty loss predicted by multi-layer feature maps and the Region of Interests (RoIs). The proposed framework is evaluated on a public dataset DeepLesion and achieves results that surpass the active learning baseline schema at all the training circles.

Keywords: Lung Nodule Detection · Active Learning · Low-dose CT · Deep Learning

1 Introduction

Lung cancer is the leading cause of cancer death in the United States [12]. To better assist the clinic diagnosis, the low-dose computed tomography (CT) screening is recommended as the most effective lung nodule diagnostic tool. The automatic computer-aided diagnosis (CAD) systems based on CT scans have been widely exploited for automatic lung nodule classification, segmentation, and detection.

In recent years, many researchers have devoted to applying the deep learning-based frameworks to medical imaging analysis [15]. Since deep learning algorithms are data-hungry, they may suffer from the expensive data acquisition and annotation of medical images. To label the medical images, the oracle requires clinical and biomedical background knowledge while the nature scene labeling needs mainly common sense. Time-consuming pulmonary nodule annotation brings difficulty to acquire large annotated datasets. With the limited data, the deep learning-based models is prone to overfitting. Recently, semi-supervised

* Corresponding author. Email: ytian@ccny.cuny.edu

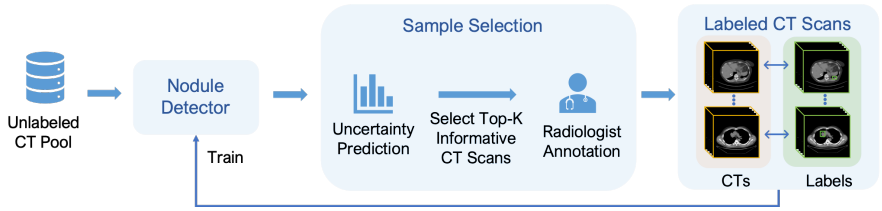

Fig. 1. The flowchart of the proposed active learning nodule detection framework. A group of randomly sampled data is selected to train the nodule detection network. In each of the training cycle, the trained network evaluates the remaining unlabeled CT scans in the unlabeled CT pool by predicting the uncertainty for each CT scan. Based on the uncertainty scores, the top-k informative CT scans are selected in the current training circle considered as the most informative samples for labeling. The newly annotated CTs are further joined the labeled CTs for the detector training. The steps are iterated until the network achieves good performance.

and unsupervised learning-based methods utilize partial or unlabeled data to avoid the annotation cost [5, 6]. These methods mainly focus on extracting good features, however, the data annotations are still required for the downstream tasks, such as classification, segmentation, and detection.

Existing approaches consider active learning methods to reduce data annotation cost by selecting valuable representatives from unlabeled data [2, 4, 9, 10, 13]. For medical imaging, the active learning strategy has shown high potential to reduce the annotation cost [1], such as in biomedical image segmentation [3, 14, 18] and pathology image classification [11, 13]. Most of the active learning methods for object detection tasks are based on the uncertainty scores predicted by image-level representation extracted from backbone network to select informative samples. Recently, Yoo *et al.* [19] proposed a state-of-the-art learning-loss schema to predict the loss value of unlabeled data based on the target network, which can be applied to the object detection task. Multi-level features from the target model are fused to map the scalar value for loss prediction. Since the loss prediction network only considers the model loss and despite the task, the loss prediction strategy is robust to various tasks and has high potential to apply to nodule detection tasks. However, the performance of object detection task is not guaranteed with small objects or 3D volumetric data. The image-level distribution is not sufficient and not detailed for pulmonary nodule detection.

There are two major challenges in implementing the existing active learning methods for pulmonary nodule detection. First, unlike natural scene images, CT scans contain relatively small volume of lung nodules relative to the whole 3D CT. Second, the feature similarity between normal tissues and nodules may mislead the sample selection. The image-level uncertainty prediction cannot be directly applied to the 3D features. Although the uncertainty of the whole CT scan can be learned, the informative features that distinguish nodules and tissues are tended to be ignored.

To address the above challenges, this paper makes the following contributions: 1) A deep active learning-based nodule detection framework is proposed to achieve comparable performance with less data and annotations, as shown in Fig. 1. The unlabeled CT scans are assessed by a trained model with predicted uncertainties. By sorting the uncertainty scores of CTs, the top-K informative candidates are selected, further annotated as the training data. 2) A simple yet effective uncertainty selection method predicts the loss through the features extracted from the basic blocks of the backbone network which evaluates the learning-loss uncertainty, and the features from region proposals for an RoI-level prediction. 3) The proposed framework is conducted on the large public DeepLesion dataset [17] and the proposed method surpasses the active learning baseline [19] at all the training cycles.

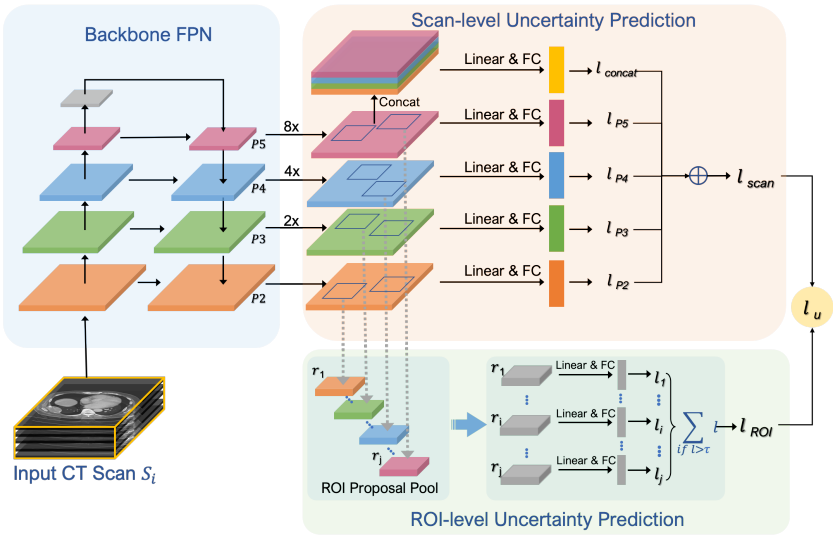

Fig. 2. The proposed active learning schema for pulmonary nodule detection. A trained 3DFPN detector takes the input of 3D CT volume and obtains the learning-loss features maps through multiple layers $[P2, P3, P4, P5]$. The learning-loss uncertainty is computed by aggregating the loss of four-layer feature maps and a global loss predicted by concatenated four-layer features. The RoI-level uncertainty prediction considers the local features of the region proposals prediction and is obtained by summing predicted scores greater than the threshold τ. The uncertainty of sample selection combines learning-loss loss with RoI-level loss for each unlabeled CT scan.

2 Method

The informative CT scans with higher uncertainty values are defined as hard samples, while the non-informative CT scans with lower scores are defined as easy samples. The key insight of our method is to learn from the hard samples of CT scans for pulmonary nodule detection. More specifically, we consider the following two scenarios. 1) The CT scans contain normal tissues with similar features of nodules. The high probability to be detected as nodules leads to a high false-positive rate. 2) The CT scans include the nodules that are difficult to detect, with a small size or low intensity. In this section, we introduce a simple yet effective active learning scheme for pulmonary nodule detection network.

2.1 Nodule Detection Framework

By appending the low-level texture features with higher-level strong-semantic features, the feature pyramid networks (FPNs) [7] achieved good performance on detecting small objects by combining both local and global features through the multi-layered feature extraction. Since the original FPN is designed to process 2D images, considering the spatial information of nodules in consecutive CT slices, we follow the nodule detection network of 3DFPN [8] by applying 3D ResNet-18 as the backbone network and detecting nodule locations in 3D CT volumes. The feature pyramid network consists of four layers of $[P2, P3, P4, P5]$, which integrates low-level features through a top-down pathway by lateral connections. The feature maps are further applied to evaluate the uncertainty. Currently, two-stage object detection methods such as RCNN, are widely used in small object detection combined with the FPN backbone framework. Yan *et al.* [16] proposed a 3D Context Enhanced (3DCE) RCNN model for lesion detection, however, the spatial information is not guaranteed by the three-channel images. In this paper, we employ the 3D volume CT scan as input to extract multi-layer features from 3D Feature Pyramid ConvNet ($3DFPN$) with the region of interest (RoI) proposal selection. The prediction includes a confidence score of each nodule candidate, as well as the corresponding position and nodule size, as $[x, y, z, d]$, where $[x, y]$ are the spatial coordinates of the candidate, z is the CT slice index number, d is the diameter of the nodule.

2.2 Active Learning with Nodule Detection

This section introduces the proposed active learning framework for lung nodule detection. During the initial network training, a group of CT scans is randomly selected from all unlabeled CT pool for annotation and defined as S_{train}. In the first learning cycle, a deep learning framework is trained by S_{train} and predicts uncertainties l for each unlabeled CT in the unlabeled pool $S_{unlabel}$. As the CT scans are not equally contributed to the performance of the model, the higher predicted uncertainty loss indicates the greater difficulty for the nodule prediction, with the high false-negative rate for the missing nodules or high false-positive prediction for tissues detected as nodules. We define informative CT

scans with high uncertainty value as hard samples and non-informative CT scans with lower scores as easy samples. The hard samples aim to efficiently improve the performance of the nodule detection network. By sorting the predicted loss of l in descending order, top-k CT scans are selected as the hard samples, which would be annotated by radiologists and aggregated to the training data S_{label}. The data augmentation methods of the random flip, crop, and rotate are applied to the labeled k CT scans to avoid overfitting. The augmented data are joined as the input to train the lung nodule detector. The trained model is applied to predict the loss for unlabeled data to select another set of top-k CT scans for annotation. Repeat the active learning schema until the nodule detector achieves good performance.

As illustrated in Fig. 2, an active learning-based sample selection approach is proposed to predict loss through global features and RoIs. 3DFPN is applied as the backbone network by the multi-layer feature maps extracting by a top-down path. The feature maps of $\{P2, P3, P4, P5\}$ extracted from the last four convolution layers represent the global features, which are normalized by 3D global average pooling and a fully connected (FC) layer. Following the learning-loss schema [19], a scan-level prediction consists of multi-layer uncertainty losses of the four feature scalars and a loss predicted by the concatenated four feature maps. The objective function for the learning-loss uncertainty is obtained by the five uncertainty scores, shown as Eq. 1:

$$l_{scan} = l_{P2} + l_{P3} + l_{P4} + l_{P5} + l_{P_{concat}}, \tag{1}$$

where i indicates the current feature layers, P_{concat} concatenates the feature maps of $\{P2, P3, P4, P5\}$.

To obtain the detailed local features for CT scans contained nodules with the small scale and tissues with the similar feature as nodules, the learning-loss schema is not sufficient to select the most valuable CT scans by predicting the loss through the feature maps, where the informative features of the true negative and false positives candidates are not statistically significant shown. As the region proposal network predicts nodule candidate location and the region of interest (RoI) cropped from the multi-layer feature maps, in particular, we introduce a simple yet effective sample selection method based on uncertainty prediction by statistically selecting by RoI level uncertainty prediction, as shown in Fig. 2. We aim to select CT scans with a large number of high uncertainty RoI regions as additional criteria for CT scan sample selection.

In order to select the CT scans with the most false-positive samples, a set of uncertainty scores is predicted for each RoI region of the entire 3D CT scan. For each CT scan s_i, a set of RoI regions r_{s_i} is obtained. RoI-level loss prediction L_{RoI} sums the scores of the region proposal l_i for the value greater than the threshold τ, shown as Eq. 2:

$$L_{RoI} = \begin{cases} \sum_{}^{i} l_i & if \quad l_i > \tau \\ 0 & otherwise. \end{cases} \tag{2}$$

Therefore, the objective function is combined with learning-loss, RoI-level loss, and detector loss as Eq. 3:

$$l_{final} = l_{target}(\hat{y}, y) + \lambda_1 \cdot L_{scan}\{\hat{l}, l\} + \lambda_2 \cdot L_{RoI}\{\hat{l}, l\}, \tag{3}$$

where λ_1, λ_2 are the weights of learning-loss and RoI-level prediction respectively. ($\hat{l}$, l) are the two samples selected from different batches followed the batch strategy of [19] by comparing two samples from different batches.

3 Experiments

3.1 Experimental Settings

Dataset and Evaluation The NIH DeepLesion dataset [17] contains 32,000 annotated lesions on the CT scans acquired from 4,400 patients. In this paper, 1,281 CT scans with 2,592 lung nodules are conducted in the evaluation. The 1,000 CT scans are for training and 281 CT scans from the official split are for testing. To assess the proposed framework, we assume all the CT scans are unlabeled and in each active learning cycle, 10% of the unlabeled dataset are selected and annotated. We follow the evaluation method of the baseline detector 3DCE [16] and employ the sensitivity at certain false positives per image similar to the Free-Response Receiver Operating Characteristic (FROC). In this paper, we present the results for sensitivity at 4 false positives per image for a fair comparison with the baseline detector 3DCE [16] for the performance of lung nodule detection.

Training In the first active learning circle, the detector is trained by a random selected 10% CT scans from the unlabeled data pool with annotations (100 CT scans). The trained model is applied to predict the loss of the remaining 90% unlabeled data. By sorting the loss prediction, the top 100 CTs, which is 10% of the initial unlabeled CT pool, are annotated and added to the training data to finetune the trained model. A total of 30 epochs are applied for each training cycle. The learning rate is initialized as 0.01 and decreased by 1/10 at the 280 epoch. 300 epochs are conducted for the framework. The 18-layer residual network (ResNet-18) is applied as the backbone network of the 3DFPN detector. Following the same implementation of Yoo *et. al* [19], loss prediction module of learning-loss schema is conducted for scoring the four feature maps of the backbone network. The RoIs with the prediction score greater than 0.5 are applied to obtain the RoI-level loss. To avoid overfitting in the training process, the random flip, rotation and crop are also applied for the data augmentation.

3.2 Results

Performance Evaluation Table 1 shows the results compared to the baseline detector. By using 80% of all the training data, the sensitivity at 4 false positives per image of our proposed framework is comparable to the baseline 3DCE

Table 1. Comparison between 3DCE [16] and our proposed active learning-based nodule detection network. With the 80% annotated training data, the proposed method achieves a comparable result with the baseline 3DCE network.

Methods	3DCE [16]	Ours-60%	Ours-70%	Ours-80%	Ours-90%	Ours-100%
Sensitivity	0.910	0.875	0.896	0.908	0.915	0.921

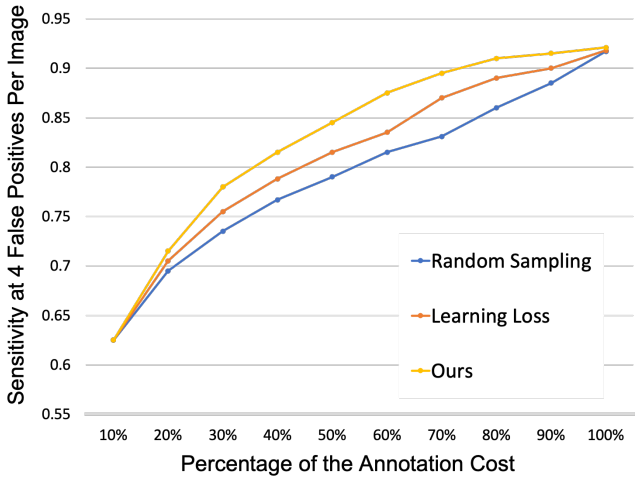

Fig. 3. The sensitivities at 4 false positives per image for the proposed active learning-based framework compared with random sampling and learning-loss sample selection.

model with the full training data. By only trained with 60% annotated data, the result of our model is approaching to the sensitivity of 3DCE trained by the full training data with only 3.5% less. With all the training data annotated, our model surpasses the baseline detector by 1.1% sensitivity.

In addition, we compare the sensitivity of the proposed active learning method with random sampling and learning-loss sampling schema [19]. The sensitivities of our proposed method and the baseline active learning methods at 4 false positives per image are shown in Figure 3. Our active learning strategy surpasses learning-loss prediction [19] and random sampling at all active learning cycles. The learning-loss features may fail to capture the features of the nodule location, while the RoI feature region prediction may lose the global information. As shown in Fig. 3, by training with 60% annotated data, the proposed method obtains an 87.5% sensitivity at 4 false positives per image which is 6% higher than the random sampling baseline, and 4% higher than the learning-loss sample selection strategy. The performance of the nodule detector is greatly improved by combining the learning-loss and RoI-level prediction shown the effectiveness of the proposed active learning sample selection strategy.

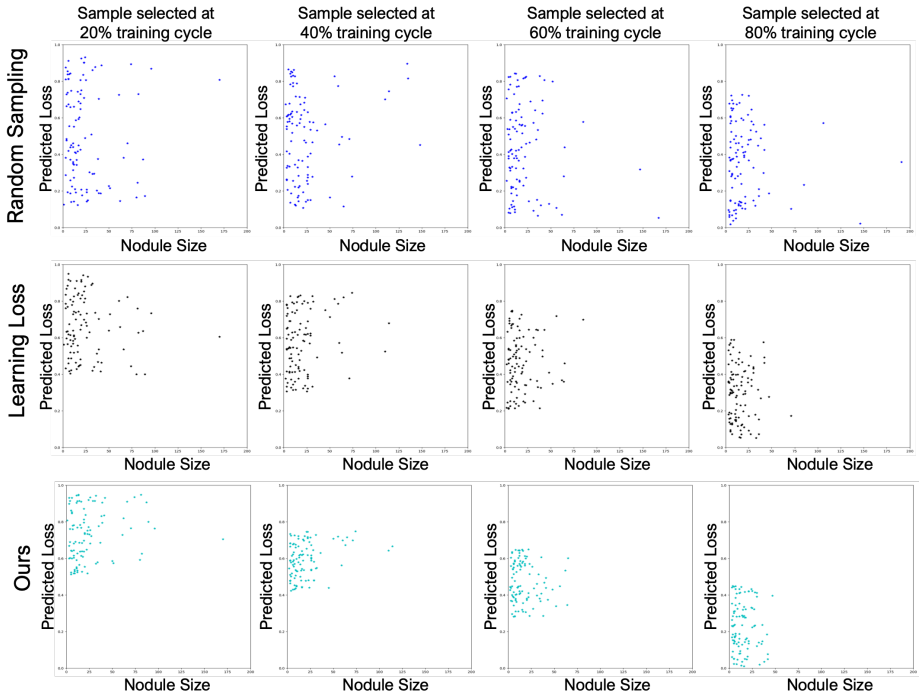

Fig. 4. Visualization of the selected nodule samples on 20%,40%,60%,80% active learning circle respectively. The x-axis of the scatter map indicates the nodule diameter and y-axis correspond to the uncertainty loss prediction. We compare the samples for the random sampling (marked as blue dots), and learning-loss sampling (marked with black dots) with the proposed active learning schema (marked as cyan dots).

Selected Sample Analysis The samples selected by the proposed method, learning-loss prediction, and random sampling within 20%, 40%, 60%, 80% active learning cycles are compared in Fig 4. X-axes and Y-axes indicate the nodule diameter and the predicted loss value for the corresponding CT scan. As the RoI regions selected by the proposed detector are the potential candidate regions, by selecting the candidates containing RoI regions with higher loss in the object function, the algorithm aims at finding the candidates with the highest potential to be predicted as miss-detected candidates (negative candidates) and the false detected candidates (positive candidates). For the samples selected from the 20% training cycle, we can observe that most of the predicted losses from the proposed active learning strategy (marked as cyan dots) are higher than other methods especially random sampling (marked as blue). With the learning circle increasing from 40% to 60%, the proposed method leans to focus on the nodules with the small size, while the random sampling still selects the large nodule with low loss values. The learning-loss (marked as black) predicts higher and clusters are more sparse. When the training cycle increases to 80%, it is worth noting that

the predicted loss values of the active learning-based sample selection strategies gradually decreases due to the priority to select informative samples, while the remaining 20% of the unlabeled data contains majority easy samples.

4 Conclusion

To eliminate expensive annotation costs and effort to acquire large datasets, we have proposed an active learning schema to select the valuable CT scans to train a deep learning-based pulmonary nodule detector. A simple yet effective active learning schema predicts loss from multi-layer features and RoIs for sample selection. The experimental results have demonstrated that the proposed framework has great potential in accelerating clinic diagnosis.

Acknowledgement

This material is based upon work supported by the National Science Foundation under award number IIS-1400802.

References

1. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. arXiv preprint arXiv:1910.02923 (2019)
2. Budd, S., Sinclair, M., Khanal, B., Matthew, J., Lloyd, D., Gomez, A., Toussaint, N., Robinson, E.C., Kainz, B.: Confident head circumference measurement from ultrasound with real-time feedback for sonographers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 683–691. Springer (2019)
3. Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., Zheng, Y.: Learning active contour models for medical image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11632–11640 (2019)
4. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 1183–1192. JMLR. org (2017)
5. Jing, L., Tian, Y.: Self-supervised visual feature learning with deep neural networks: A survey. arXiv preprint arXiv:1902.06162 (2019)
6. Károly, A.I., Fullér, R., Galambos, P.: Unsupervised clustering for deep learning: A tutorial survey. Acta Polytechnica Hungarica **15**(8), 29–53 (2018)
7. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2117–2125 (2017)
8. Liu, J., Cao, L., Akin, O., Tian, Y.: 3DFPN-HS^2: 3D Feature Pyramid Network Based High Sensitivity and Specificity Pulmonary Nodule Detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2019)

9. Lowell, D., Lipton, Z.C., Wallace, B.C.: Practical obstacles to deploying active learning. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 21–30 (2019)
10. Mahapatra, D., Bozorgtabar, B., Thiran, J.P., Reyes, M.: Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 580–588. Springer (2018)
11. Nalisnik, M., Gutman, D.A., Kong, J., Cooper, L.A.: An interactive learning framework for scalable classification of pathology images. In: 2015 IEEE International Conference on Big Data (Big Data). pp. 928–935. IEEE (2015)
12. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2020. CA: A Cancer Journal for Clinicians **70**(1), 7–30 (2020)
13. Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L.: Cost-effective active learning for deep image classification. IEEE Transactions on Circuits and Systems for Video Technology **27**(12), 2591–2600 (2016)
14. Wen, S., Kurc, T.M., Hou, L., Saltz, J.H., Gupta, R.R., Batiste, R., Zhao, T., Nguyen, V., Samaras, D., Zhu, W.: Comparison of different classifiers with active learning to support quality control in nucleus segmentation in pathology images. AMIA Summits on Translational Science Proceedings **2018**, 227 (2018)
15. Wu, J., Qian, T.: A survey of pulmonary nodule detection, segmentation and classification in computed tomography with deep learning techniques. J. Med. Artif. Intell **2** (2019)
16. Yan, K., Bagheri, M., Summers, R.M.: 3d context enhanced region-based convolutional neural network for end-to-end lesion detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 511–519. Springer (2018)
17. Yan, K., Wang, X., Lu, L., Summers, R.M.: Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. Journal of medical imaging **5**(3), 036501 (2018)
18. Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z.: Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 399–407. Springer (2017)
19. Yoo, D., Kweon, I.S.: Learning loss for active learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 93–102 (2019)