Simultaneous Accurate Detection of Pulmonary Nodules and False Positive Reduction Using 3D CNNs

Yulei Qin, Hao Zheng, Yue-Min Zhu, Jie Yang

SIMULTANEOUS ACCURATE DETECTION OF PULMONARY NODULES AND FALSE POSITIVE REDUCTION USING 3D CNNS

Yulei Qin[1], *Hao Zheng*[1], *Yue-Min Zhu*[2], *Jie Yang*[1]

[1]Department of Automation, Shanghai Jiao Tong University,
800 Dongchuan RD. Minhang District, Shanghai, China
[2]University Lyon, INSA Lyon, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, France

ABSTRACT

Accurate detection of nodules in CT images is vital for lung cancer diagnosis, which greatly influences the patient's chance for survival. Motivated by successful application of convolutional neural networks (CNNs) on natural images, we propose a computer-aided diagnosis (CAD) system for both accurate pulmonary nodule detection and false positive reduction. To generate nodule candidates, we build a full 3D CNN model that employs 3D U-Net architecture as the backbone of a region proposal network (RPN). We adopt multi-task residual learning and online hard negative example mining strategy to accelerate the training process and improve the accuracy of nodule detection. Then, a 3D DenseNet-based model is presented to reduce false positive nodules. The densely connected structure reuses nodules' feature and boosts feature propagation. Experimental results on LUNA16 datasets demonstrate the superior effectiveness of our approach over state-of-the-art methods.

Index Terms— Pulmonary nodule detection, computer-aided diagnosis, false positive reduction

1. INTRODUCTION

Pulmonary cancer has been the second most common cancer in both men and women and annually causes 1.3 million deaths worldwide [1]. As such, early treatment is vital to greatly increase patients' chances of survival. Small masses found inside lung are called pulmonary nodules, which have potential to become cancerous. In recent years, computed tomography (CT) has been adopted widely as a screening tool. However, due to high demand for CT scanning and similarity of nodule's density to some lung tissue, radiologists often find it challenging to find nodules in huge volumes of CT images.

Doctors typically diagnose nodules based on experiential knowledge [2] and misdiagnosis may occur since nodules vary greatly in size (diameter ranging from 3 to 30 mm), shape, sharpness, brightness and compactness [3]. Study shows only 68% of retrospectively detected pulmonary nodules were detected originally by one radiologist and 82% by two radiologists. To relieve radiologists from this difficult work, computer-aided diagnosis (CAD) systems were developed for pulmonary nodule detection. In general, a CAD system consists of three steps: The first is to do basic preprocessing of lung CT images, which usually utilizes morphological operation and thresholding to standardize input data. The second is to find suspicious nodule candidates. Usually nodule detection is performed with high sensitivity to search for nodules as much as possible, which spontaneously produces many false positives. The ratio of false positive nodules to true positive ones is not negligible. Hence, the third step aims to reduce false positives by classifying the detected candidates as true nodules or not.

Traditional CAD systems use assumptions and experience [4][5][6] (e.g. pulmonary nodules are bright, circular masses on the walls of the chest cavity). They detect nodules based on discriminative characteristics. Messay et al. [7] detect nodule candidates with a feature combination descripter of shape, intensity and gradient features. Torres et al. [8] first segment nodule ROIs and propose some low-level descripters such as spatial features and HU intensity for classification. Jacobs et al. [9] use features based on intensity, shape, texture characteristics and structure information to find nodules. All these methods achieve a high detection sensitivity, but fail to adapt to variable shape, size and texture of nodules. Therefore, their performance may degrade drastically on large datasets.

With the development of convolutional neural networks (CNNs) [10][11][12][13] and availability of large labeled data [3][14], researchers tend to employ CNNs for pulmonary nodule diagnosis. Setio et al. [15] use 2D multiview CNNs to learn nodules' characteristics and filter out false positive candidates. They extract many patches from different oriented planes at each nodule's position and use adjacent slices' spatial information as much as possible. Their method achieves a sensitivity of 85.4% at 1.0 false positive per subject on LIDC dataset [3]. Sakamoto et al. [16] use cascaded CNNs which perform as selective classifiers to solve the class imbalanced problem. Dou et al. [17] propose a 3D CNN framework for false positive reduction and achieve a CPM score [3] of 0.827. This method directly encodes multilevel contextual informa-

This research is partly supported by Committee of Science and Technology, Shanghai, China (No. 17JC1403000).

tion of 3D samples to meet the challenge of dealing with large variation of nodules.

In this paper, we propose a CAD system for simultaneous nodule candidates detection and false positive reduction. The main contributions are as follows: (1) We propose a full 3D CNN framework that accurately detects pulmonary nodules. To generate nodule candidates, we introduce the 3D U-Net [18] structure to the region proposal network (RPN) [19] for region of interests (ROI) classifier and regression. The 3D architecture aims at better utilizing spatial contextual knowledge. Multi-task residual learning and online hard negative example mining strategy are employed to accelerate training and improve accuracy. (2) We propose a DenseNet [20] based 3D CNN for false positive reduction. The structure of densely connected net strengthens nodule feature propagation and encourages feature reuse. Besides, the model shares early convolution layers with the detection model for feature extraction. (3) The proposed CAD system achieves accurate detection of pulmonary nodules while reducing false positives on LUNA16 datasets.

This paper is organized as follows. In section 2, the relation to previous work is presented. In section 3, we describe our framework for nodule detection and false positive reduction. In section 4, we report detailed experimental results. Finally, our conclusion is presented in section 5.

2. RELATION TO PREVIOUS WORK

The 2D models proposed by Setio et al. [15] and Sakamoto et al. [16] are mainly devoted to false positive reduction. Only 3 consecutive slices are used due to the limitation of 2D architecture. To effectively learn representations from 3D contexts, 3D CNN model is considered a good choice. Dou et al. [17] use 3D CNNs to incorporate different levels of volumetric information, which lays foundation for our work. But Dou et al. do not propose to generate nodule candidates directly from slices. Therefore we use RPN [19] with ROI regression and classification to obtain nodule candidates. While our study is related to approaches mentioned above, it is a full 3D framework that integrates both nodule detection and false positive reduction, which was not considered in previous work.

3. METHODS

The CAD system we proposed is illustrated in Fig. 1. We use 3D U-Net as a backbone to build a RPN model, which generates nodule candidates with high sensitivity. Then we design a 3D DenseNet-based model for false positive reduction.

3.1. Nodule Candidate Detection

We establish a 3D model to detect nodule candidates. The region proposal network is based on a U-Net structure. We use 3 sizes of anchors to obtain nodule candidates, with each anchor corresponding to a ROI classifier and a ROI regression. The ROI classifier classifies the object inside ROI as nodules or not. And the ROI regression makes the prediction of ROI's position closer to ground truth position.

Architecture: This network is fed with cropped cubes of pulmonary CT images. A 3D U-Net is developed to effectively leverage rich spatial contextual information for feature extraction and generate predictions in a volume-to-volume manner. For the contracting (forward) path, we define 5 blocks of residual unit:

$$y = \mathcal{F}(x, \{W_i\}) + x, \tag{1}$$

where x and y are respectively the input and output of the layer. $\mathcal{F}(x, W_i)$ is a 3D residual mapping illustrated in Fig. 2. The use of residual units could boost gradient flow and benefit optimization process. Each residual unit (except the first one) is followed by a max pooling ($2 \times 2 \times 2$) layer to downsample the feature cube.

For the expansive (backward) path, we upconvolute the feature cube and concatenate it with the corresponding feature cube from contracting path, then pass it through a residual unit and a dropout layer. The final feature output is fed into the RPN for ROI regression and classification.

Multi-task learning with weighted loss: Each voxel in the final feature output cube is parameterized with respect to the reference anchors. According to the size distribution of nodules, 3 anchors are designed with different sizes: $10 \times 10 \times 10, 20 \times 20 \times 20, 40 \times 40 \times 40$. For each anchor, we calculate its Intersection-over-Union (IoU) overlap with the nearest ground-truth box. Anchors that have IoU higher than 0.5 are assigned positive labels. If an anchor's IoU overlap is lower than 0.02, then it is labeled as negative. Anchors that are neither positive nor negative will be neglected and not involved in the training process.

The loss function of multi-task learning is composed of classification loss and regression loss. We define the weighted cross-entropy (WCE) loss for classification problem with the imbalanced nodule datasets. Given a set of training pairs $\{(x_i, y_i)\}_{i=1,2,...,N_{cls}}$, the WCE loss between the label target y_i and prediction output o_i is:

$$\begin{aligned} \mathcal{L}_{cls} &= -\frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} w y_i log(o_i) + (1 - y_i) log(1 - o_i), \\ w &= \frac{N_{cls} - \sum_{i=1}^{N_{cls}} o_i}{\sum_{i=1}^{N_{cls}} o_i}, \end{aligned} \tag{2}$$

where w is the weight attributed to the nodule class. Such weight forces the model to focus on positive samples to improve detection sensitivity.

The regression loss of location information is given by:

$$\mathcal{L}_{reg} = \frac{1}{N_{reg}} \sum_{i=1}^{N_{reg}} \sum_{\gamma \in \{x,y,z,d\}} y_i R(t_\gamma - t_\gamma^*), \tag{3}$$

where $R(\cdot)$ represents the smooth ℓ_1 loss defined in [12]. t_γ and t_γ^* designate 4 parameters of the predicted box and ground truth box (associated with a positive anchor), respectively. The 4 parameters are defined as:

$$\begin{aligned} t_x = (x - x_a)/d_a, t_y = (y - y_a)/d_a, \\ t_z = (z - z_a)/d_a, t_d = log(d/d_a), \end{aligned} \tag{4}$$

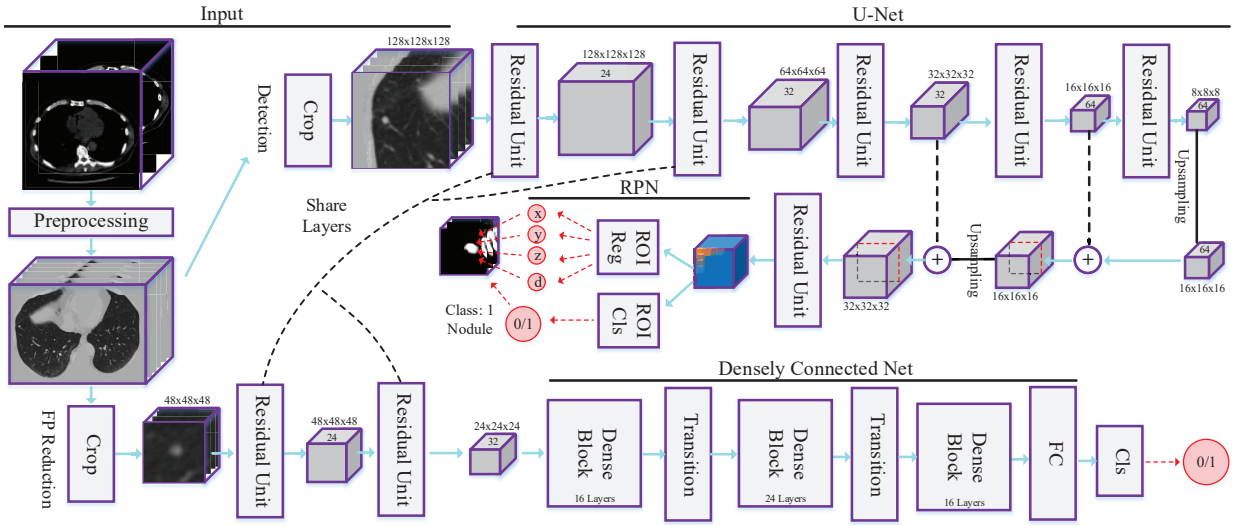

Fig. 1: An overview of the proposed CAD system.

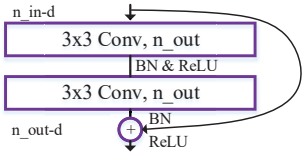

Fig. 2: Residual Unit.

where x, y, z are coordinates and d is edge length of the predicted box. The variables x_a, y_a, z_a, d_a are parameters of the anchor cube. And $t_x^*, t_y^*, t_z^*, t_d^*$ are defined similarly. This loss function penalizes to make a ROI regression from a fixed anchor box to a nearby ground truth box. Thus the total multi-task loss is given by:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{reg}, \tag{5}$$

where λ is a weight balancing the two loss terms. Such multi-task loss function can output the probability of being nodule and its position at the same time. It also improves detection accuracy by considering the inner-relation between the task of locating nodule candidates and the task of classification.

Online hard negative example mining: In medical datasets, most negative examples could be easily discriminated by the network except hard ones, which usually contribute to high loss. Since hard examples contain more valuable information than simple ones, we adopt an online hard negative example mining strategy (OHNEM) during training [21]. In implementation, we first process training samples with a large batch size. After forward propagation, N negative samples are selected randomly. Then, we sort these samples by their loss. Top 10% samples on which current model performs worst are selected as hard negatives and others are discarded. Such strategy allows us to reduce correlation between negative samples and accelerate convergence. In the present study, N is set as 2 times batch size.

3.2. False Positive Reduction

After extracting nodule candidates, we build a 3D CNN model to reduce false positive nodules.

Architecture: The cropped candidate cubes are first fed into 2 blocks of residual unit which share the same parameters with early convolution layers in the detection network. Then the feature map is pooled by a $2 \times 2 \times 2$ kernel with stride = 2, followed by 3 blocks of densely connected net with a transition layer between two adjacent blocks. The l^{th} layer in the dense block [20] is illustrates in Fig. 3. It reuses all preceding layers as input:

$$x_l = H_l([x_0, x_1, ..., x_{l-1}]), \tag{6}$$

where H_l is a bottleneck layer defined as the combination of BN-ReLU-Conv($1 \times 1 \times 1$) and BN-ReLU-Conv($3 \times 3 \times 3$), followed by a dropout layer. The transition layer consists of a batch normalization (BN) layer, a ReLU layer, an $1 \times 1 \times 1$ convolution layer and a $2 \times 2 \times 2$ average pooling layer. At the end of the last Dense Block, a fully connected net is attached to classify candidates as true nodules or false positive ones.

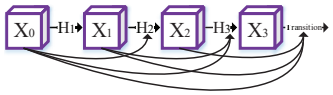

Fig. 3: Dense Block (4 layers).

The loss function of false positive reduction is WCE loss defined by Equation 2. Since the overwhelming easy negatives will give rise to degenerative models, OHNEM is also used to avoid the situation where training is dominant by easy negative examples.

4. EXPERIMENTS

We use LUNA16 datasets [3] to validate the proposed CAD framework. Besides, ablation studies on OHNEM are presented to confirm its validity. The datasets are divided into 10 subsets for 10-fold cross validation and they contain 888 scans in total. Each scan has annotations of nodule centroids' positions in world coordinates and their diameters.

Preprocessing: We use the lung segmentation images provided by [3] to obtain lung region masks. For each CT slice, each lung mask is replaced by its convex hull image if

the convex hull area is over 1.5 times the size of the original area. Then each lung is dilated to include more margin space. The lung masks are multiplied with raw images and the masked regions are normalized to $[0, 255]$. Voxels outside lung regions are padded with 170 to imitate the intensity of body tissue. All slices are resampled to the same spacing of $1 \times 1 \times 1(mm)$ and cropped into lungs' minimum circumscribed cubes for lower computation cost.

Implementation details: For nodule detection stage, positive samples are augmented via flipping, swapping, random scaling between $[0.8, 1.2]$ and random rotation between $[0°, 180°]$. Besides, to balance the distribution of nodules' sizes, large nodules are upsampled. Compared to small nodules, the sampling frequency of nodules larger than 20 (or 30) mm is increased 2 (or 4) times. Due to the limitation of GPU memory, 3D patch-based training is adopted with patch size of $128 \times 128 \times 128$. For false positive reduction stage, positive samples are augmented via translation of $\pm 1(mm)$ along each axis, flipping, swapping, random scaling and rotation. The cropped input size is $48 \times 48 \times 48$. Both models are initialized from a Gaussion distribution $\mathcal{N}(0, 0.01)$ and λ in the multi-task loss function is 0.5. The dropout rate is set as 0.2 to reduce oscillation of parameters and speed up convergence. We implement our framework with PyTorch using 4 NVIDIA GTX 1080 GPUs.

Evaluation metrics: The Free-Response Receiver Operating Characteristic (FROC) analysis [3] is used to evaluate the performance of CAD systems. In the FROC curve, sensitivity is ploted as a function of false positives per scan (FPs/scan). A detection is viewed as true positive if it locates within the radius of a nodule's centroid and the sensitivity is defined as the total number of true positives divided by the number of all detected nodules. The CPM score is also calculated as average sensitivity at 7 FPs/scan rates: $1/8, 1/4, 1/2, 1, 2, 4, 8$.

Candidate detection results: Detection results are compared with other methods [3] as shown in Table 1 and Fig. 4. Our method achieves the highest sensitivity of **96.7%** with 25.97 candidates per scan on average. The detection sensitivity (achieved by ETROCAD) is 92.9% and it is improved by **4.09%**. The CPM score of our system is **0.834** and it outperforms state-of-the-art methods descriped in [3]. For FROC curves at 1, 2, 4, 8 FPs/scan, we obtain the highest sensitivity of **87.2%, 91.9%, 94.6%, 96.4%** respectively, which proves the superiority of our system and its great clinical value.

Table 1: Comparison of CAD systems for nodule detection

System name	Sensitivity	Avg. candidates/scan
ISICAD	0.856	335.9
SubsolidCAD	0.361	290.6
LargeCAD	0.318	47.6
M5L	0.768	22.2
ETROCAD	0.929	333.0
Our Method	**0.967**	25.9

False positive reduction results: To evaluate the performance of our system, we also compare our results with top CAD systems descriped in [3]. As shown in Fig. 5, although CUMedVis system achieves comparable sensitivity, our method performs better at 0.125, 0.25 FPs/scan and yields the highest CPM score of **0.909**. At 4, 8 FPs/scan, we achieve the sensitivity of **97.8%**, **98.0%**, respectively. The superior effectiveness of our method proves that the 3D DenseNet architecture can utilize contextual information and make full use of nodule's feature to filter out false positives.

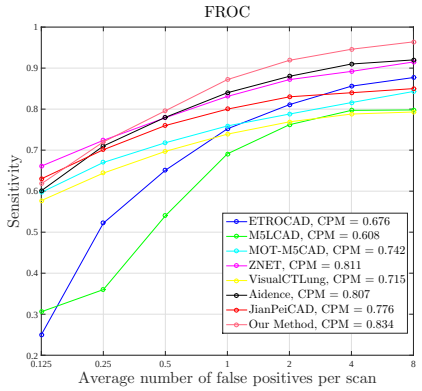

Fig. 4: FROC curves of CAD systems for nodule detection.

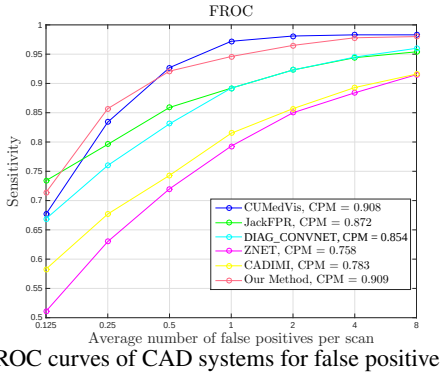

Fig. 5: FROC curves of CAD systems for false positive reduction.

Ablation Studies: The results on ablation are shown in Table 2 and they validate that using OHNEM can force the model to learn effectively from hard negative examples and thus improve detection sensitivity.

Table 2: Ablation Study of OHNEM

Stage	Nodule Detection	
Experiment	without OHNEM	OHNEM
Sensitivity	0.925	**0.967**
Stage	False Positive Reduction	
Experiment	without OHNEM	OHNEM
Sensitivity	0.931	**0.982**

5. CONCLUSION

We have proposed a full 3D framework for pulmonary nodule detection. A RPN model is built based on U-Net backbone to generate nodule candidates. Then a 3D DenseNet-based model is developed for false positive reduction. Experimental results on LUNA16 dataset demonstrate that the proposed method achieves accurate detection of pulmonary nodules while reducing false positives, thus suggesting its potential for clinical applications.

6. REFERENCES

[1] Lindsey A. Torre, Rebecca L. Siegel, and Ahmedin Jemal, *Lung Cancer Statistics*, pp. 1–19, Springer International Publishing, 2016.

[2] S. C. B. Lo, S. L. A. Lou, Jyh-Shyan Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, 1995.

[3] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas de Bel, et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.

[4] Yang Liu, Zhian Xing, C. Deng, Ping Li, and M. Guo, "Automatically detecting lung nodules based on shape descriptor and semi-supervised learning," in *2010 International Conference on Computer Application and System Modeling*, 2010, pp. V1–647–V1–650.

[5] Y. Guan, Y. Wang, Y. Zou, Y. Li, and M. Liu, "Computer-aided detection for pulmonary nodules base on the morphological and spatial features," in *2010 International Conference on Biomedical Engineering and Computer Science*, 2010, pp. 1–4.

[6] N. Miyake, H. Kim, Y. Itai, J. K. Tan, S. Ishikawa, and S. Katsuragawa, "Automatic detection of lung nodules in temporal subtraction image by use of shape and density features," in *2009 Fourth International Conference on Innovative Computing, Information and Control*, 2009, pp. 1288–1292.

[7] Temesguen Messay, Russell C Hardie, and Steven K Rogers, "A new computationally efficient cad system for pulmonary nodule detection in ct imagery," *Medical Image Analysis*, vol. 14, no. 3, pp. 390–406, 2010.

[8] E Lopez Torres, E Fiorina, F Pennazio, C Peroni, M Saletta, et al., "Large scale validation of the m5l lung cad on heterogeneous ct datasets," *Medical Physics*, vol. 42, no. 4, pp. 1477–1489, 2015.

[9] Colin Jacobs, Eva M van Rikxoort, Thorsten Twellmann, Scholten, et al., "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis*, vol. 18, no. 2, pp. 374–384, 2014.

[10] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[12] Ross Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[13] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 1626–1630.

[14] Bram Van Ginneken, Samuel G Armato, Bartjan de Hoop, et al., "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study," *Medical Image Analysis*, vol. 14, no. 6, pp. 707–722, 2010.

[15] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Gerke, et al., "Pulmonary nodule detection in ct images: false positive reduction using multiview convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[16] M. Sakamoto and H. Nakano, "Cascaded neural networks with selective classifiers and its evaluation using lung x-ray ct images," *arXiv preprint arXiv:1611.07136*, 2016.

[17] Q. Dou, H. Chen, L. Yu, J. Qin, and P. A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2017.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[19] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[20] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.

[21] Abhinav Shrivastava, Gupta Abhinav, and Girshick Ross, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 761–769.