# LUNG NODULE DETECTION WITH A 3D CONVNET VIA IOU SELF-NORMALIZATION AND MAXOUT UNIT

*Fei Li*[1], *Hongyu Huang*[1], *Yawen Wu*[1], *Congbo Cai*[1,2], *Yue Huang*[1], *Xinghao Ding*[*1]

1. Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Information Science and Engineering, Xiamen University, China
2. Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance,
Department of Electronics Science, Xiamen University, China
*dxh@xmu.edu.cn

## ABSTRACT

The automatic pulmonary nodule detection in thoracic computed tomography (CT) scans plays a crucial role in the early diagnosis of lung cancer. In this paper, we propose a novel framework with a 3D convolutional network (ConvNet) for pulmonary nodule detection. To improve the efficiency and flexibility, we adopt one-stage process without the false positive reduction stage. Specially, the great challenge of the nodule detection is the recall rate of small nodules. We propose two methods to solve this issue. Firstly, we set the classification label by the intersection over union (IoU) self-normalization, which enables to eliminate the loss of regression information caused by misleading classification confidence. Secondly, pulmonary nodules differ in size, shape and density, leading to large intra-class variations. We introduce maxout unit to solve this problem. Overall, we achieve an average FROC score of 0.912 on LUNA16 dataset, outperforming all other one-stage models as far as we know.

***Index Terms***— Lung Nodule Detection, Computer-Aided Detection, Deep Learning, Medical Image Analysis

## 1. INTRODUCTION

Lung cancer is the deadliest cancer worldwide, and it accounts for approximately 27% of cancer-related deaths in the United States.[1] The last decade has seen significant advances in machine learning combined with computer aided diagnosis (CAD). Compared with manual detection, CAD is inevitable because computer vision models can quickly scan everywhere without fatigue and emotions. Recent advances in deep learning have assisted radiologists in the reading process and to make arrangement of early treatment.

This work was supported in part by the National Natural Science Foundation of China under Grants 61571382, 81671766, 61571005, 81671674, 61671309 and U1605252, in part by the Fundamental Research Funds for the Central Universities under Grant 20720160075, 20720180059, in part by the CCF-Tencent open fund, and the Natural Science Foundation of Fujian Province of China (No.2017J01126).

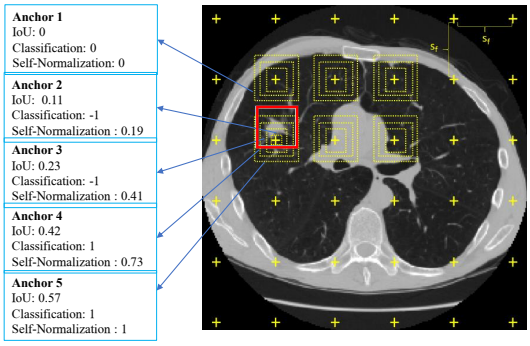


**Fig. 1**. Example IoU self-normalization and classification confidence of anchor boxes. The yellow crosshair "+" indicates the center of the anchor box. The yellow dashed boxes are evenly distributed anchor boxes, and the red solid box represents the target box. $S_f$ is the stride size of the network.

Traditional nodule detection involved hand-crafted features such as form features [2], texture characteristics [3] and wavelet features [4]. Recently, deep learning-based methods have been a major trend in this field. Besides, the emergence of large-scale dataset, LUNA16 [5] facilitated the related research. This detection task is usually divided into two stages: region proposal generation and false positive reduction. For the first subtask, we mainly adopt convolutional networks to generate candidate region proposals. Ding et al.[6] started their model with an ImageNet pre-trained VGG16 model [7] and employed Faster R-CNN [8] to generate candidate bounding boxes. In the second stage, more complex classifiers are used to remove false positive nodules. Zhu et al.[9] proposed extra false positive reduction classifier by deep 3D dual path for detection. Dou et al.[10] designed a novel hybrid-loss 3D ConvNet to improve the lung nodule recognition accuracy.

However, more single shot detectors (SSD) [11] appeared recently, making the two-stage design a bit redundant. In [12], Liao et al presented volumetric one-stage convolutional neu-

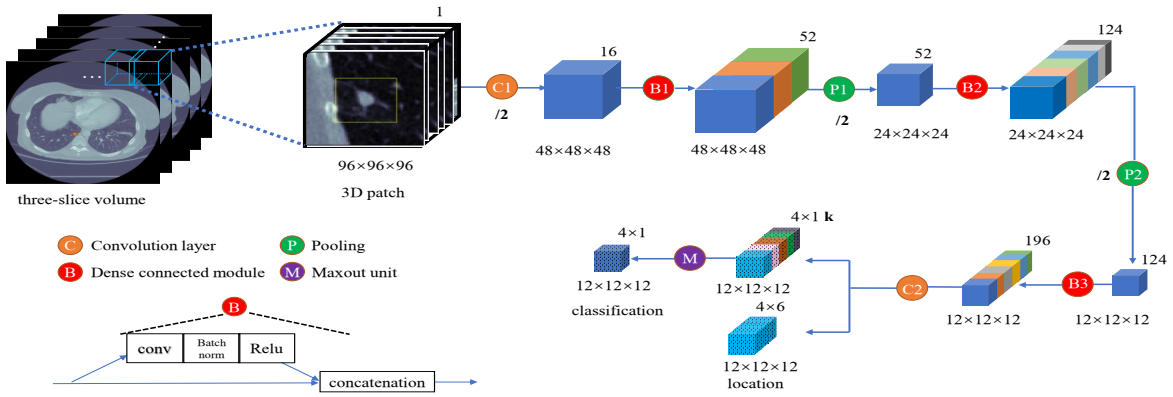


**Fig. 2**. Model overview. Each cuboid in the figure stands for a 4D tensor. "/2" indicates that downsampling is performed with a stride of 2. The number at the upper right corner of each tensor indicates the channel number of the feature map. The dense connected module "B1" contains 3 dense connected units. The "B2" and "B3" contain 6 dense connected units. In addition, the maximum response output module consists of a convolutional layer and a maxout unit.

ral network (CNN) for 3D object detection with the multi-instance learning. To improve the efficiency and highlight variability and flexibility, we adopt a simple one-stage approach. All computations are completely encapsulated in a single network. The relevant works included the comparison between 2D and 3D convolutional networks. Anirudh et al.[13] found that multi-level contextual 3D CNN is superior and more robust than 2D CNN for 3D CT data.

Although the previous results are reassuring, the detection of small nodules is still relatively challenging due to the intensity, size, and shape dissimilarity among different nodules. We propose a novel network leveraging 3D ConvNets via IoU self-normalizaton and maxout unit to improve the recall rate of small nodules. The main contributions in this work can be summarized as follows:

- We propose a novel 3D convolutional network with dense connected modules. It is a single-model and single-stage network for end-to-end detection.
- We use IoU self-normalization to set classification labels without threshold selection. IoU self-normalization enables to regress more information of detection boxes.
- We add maxout unit in the classifier to handle large intra-class variations of pulmonary nodules.
- We modify the classification loss function by incorporating variant of focal loss to dynamically adjust the classification weights of each anchor box.

## 2. METHOD

### 2.1. Network for Detection

The network consists of 19-layers CNN backbones without Region Proposal Network (RPN), as shown in Fig. 2. The dense connected module is used as the basic block of our network. The size of convolution kernel used in the network is 3 $\times$ 3 $\times$ 3. ReLU [14] is used as the activation function. The batch normalization [15] operation is performed after each convolution operation. In the last layer, we design 4 anchors of the size of 8, 16, 32 and 48, as shown in Fig. 1, which are based on the distribution of nodule sizes.

### 2.2. IoU Self-Normalization

In the conventional detection methods, the classification label value $y(A_i)$ is defined as follows,

$$y(A_i) = \begin{cases} 1, & IoU(A_i) > t_H \\ 0, & IoU(A_i) < t_L \\ -1, & \text{otherwise}, \end{cases} \quad (1)$$

where $A_i$ is the index of an anchor, $t_H$ and $t_L$ are thresholds set by hand. An anchor box will be discarded if its IoU is between $t_L(0.05)$ and $t_H(0.3)$, which may lose location information for regression. This mechanism may lead to difficulties in the detection of small-sized nodules and result in over-fitting problem.

We innovatively propose an optimization mechanism, called IoU self-Normalization, to overcome the drawbacks existed in the conventional methods. The improved classification label value $y^*(A_i)$ can be defined as:

$$y^*(A_i) = \frac{IoU(A_i)}{maxIoU(A_i)}, \quad (2)$$

where $maxIoU(A_i)$ is the maximum IoU value of all anchor boxes. The classification label value is changed from discrete to continuous through IoU normalization, which keeps all the information for regression.

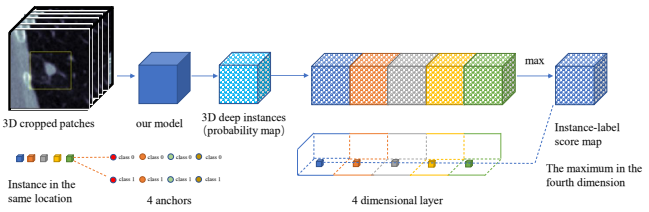


**Fig. 3**. Illustration of maxout unit. We consider the output of our model as 3D deep instances. It is a 3 dimensional probability map, and will produce a $12 \times 12 \times 12 \times K$ scoring feature through a convolution layer. Then, the max pooling operation extracts the final label predictions.

### 2.3. Maxout Unit

In order to cope with large intra-class variations of lung nodules, we introduce a maximum response at the output layer (Fig. 3). This method allows the network to respond differently to various types of nodules and neglects the interaction between nodules. Firstly, we map each point of the last feature vector to a K-dimensional vector. Then we take the maximum value in the fourth dimension of the K-dimensional vector as the final classification output. This mechanism can be regarded as the maxout activation function proposed in [16].

### 2.4. Loss Function

We use the weighted smooth L1 loss function [17] for the regression loss $L_{reg}(t, t^*)$ as follows,

$$
\begin{aligned}
L_{reg}(t, t^*) &= \frac{1}{N_{reg}} \sum_{i \epsilon S_A} y(A_i) \times smooth_{L_1}(t - t^*), \\
t &= (\frac{x - x_a}{d_a}, \frac{y - y_a}{d_a}, \frac{z - z_a}{d_a}, \log(\frac{d}{d_a})), \\
t^* &= (\frac{x^* - x_a}{d_a}, \frac{y^* - y_a}{d_a}, \frac{z^* - z_a}{d_a}, \log(\frac{d^*}{d_a})).
\end{aligned} \tag{3}
$$

Here, $t$ and $t^*$ are the offset of the predicted bounding box and ground truth bounding box respectively. $(x_a, y_a, z_a, d_a)$ is the anchor box and the radius of the box. $(x, y, z, d)$ and $(x^*, y^*, z^*, d^*)$ are the predicted and ground truth coordinates and the radius of nodule respectively. $S_A$ is the collection of all anchor boxes. $N_{reg}$ is normalization factor, which is set as $\sum_{i \epsilon S_A} y(A_i)$.

To address class imbalance by down-weighting easy examples, we incorporate the focal loss function [18] into our binary classification loss $L_{cls}$, defined as follows,

$$
\begin{aligned}
L_{cls} &= \frac{1}{N_{cls}} \sum_{i \epsilon S_A} -(1 - p_t^i)^{\gamma} \log(p_t^i), \\
p_t^i &= y(A_i) p^i + (1 - y(A_i))(1 - p^i),
\end{aligned} \tag{4}
$$

where $p^i$ is the predicted probability for the current anchor i being a nodule; $y(A_i)$ calculated with our IoU normalization method is set by equation 2. $N_{cls}$ is Normalization factor, which is set as $\sum_{i \epsilon S_A} y(A_i)$ . $\gamma$ is the focusing parameter which smoothly adjusts the rate at which easy examples are weakly weighted, and we set $\gamma = 2$ in experiments.

We sum up the regression loss and the classification loss to define the total loss as,

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

where $\lambda$ is a balance parameter and set as 1 in experiments.

## 3. EXPERIMENT

### 3.1. Dataset

LUNA16 dataset consists of 888 low-dose CTs with total of 1186 labeled pulmonary nodules. It is divided into 10 folders. We use nine folders of data for training and the rest for testing. We first clip the raw data into the range of [-1000, 600], then we transform the cropped data to [-1, 1] for preprocessing. Due to the GPU memory limitation, the raw image volume is too large to fed into the 3D CNN directly. We crop train data to a cubic patch size of $96 \times 96 \times 96 \times 1$(depth $\times$ height $\times$ width $\times$ channel). Moreover, the positive patch contains one lung nodule at least.

### 3.2. Implementation

We conduct 10-fold cross validation and data randomisation over all data. The Adam [19] optimization method is utilized for a total of 80,000 training iterations. The batch size is set to 18. The learning rate is decayed by 0.1 from 0.001 every 100 epochs of training data. All experiments were implemented with Tensorflow framework [20] on an Nvidia Titan X GPU.

### 3.3. Results

According to the LUNA16 standard, we evaluate the nodule detection performance by the free-Response Receiver Operating Characteristic (FROC) analysis. Concretely, the FROC-score is defined as the average of the sensitivity at seven predefined false positives (FPs) rates: 0.125, 0.25, 0.5, 1, 2, 4, and 8 FPs per scan.

Our best performance achieves an average FROC score of 0.912, which is visualized in Fig. 4. We can see from the FROC curves that the performance of the D-Net+CL model has been improved after incorporating focal loss, IoU self-normalization, and maxout unit respectively.

The focal loss, a modulating term to the cross entropy loss, focus learning on hard negative examples and naturally handles the class imbalance faced by our one-stage detector. In addition, D-Net+FL+IoU-norm increases the FROC scores a lot comparing to D-Net+FL at the false positive rate as 1/8 FPs, which indicates that IoU Self-normalization is very effective in reducing the leakage rate under the condition of low

| System | 0.125 | 0.25 | 0.5 | 1 | 2 | 4 | 8 | Sensitivity | Candidates/scan |
|---|---|---|---|---|---|---|---|---|---|
| M5L | 0.601 | 0.667 | 0.722 | 0.751 | 0.788 | 0.823 | 0.843 | 0.768 | 22.2 |
| ISICA | 0.652 | 0.723 | 0.864 | 0.924 | 0.942 | 0.942 | 0.942 | 0.856 | 335.9 |
| iDST-VC | 0.755 | 0.821 | **0.887** | 0.918 | **0.968** | 0.976 | 0.987 | 0.897 | - |
| CASED | 0.781 | 0.845 | 0.867 | 0.902 | 0.921 | 0.956 | 0.978 | 0.887 | - |
| Zhu et al.[9] | 0.690 | 0.780 | 0.830 | 0.860 | 0.900 | 0.913 | 0.923 | 0.842 | - |
| Dou et al.[10] | 0.659 | 0.745 | 0.819 | 0.865 | 0.906 | 0.933 | 0.946 | 0.839 | - |
| Ding et al.[6] | 0.750 | **0.854** | 0.882 | 0.928 | 0.931 | 0.936 | 0.946 | 0.890 | 15.0 |
| Ours(K=10) | **0.789** | 0.847 | 0.874 | **0.939** | 0.964 | **0.977** | **0.991** | **0.912** | **13.8** |

**Table 1**. Comparison of performance among our system, other submitted one-stage and one-model approaches (Line 4 and Line 5) on the LUNA16 Challenge and published papers.

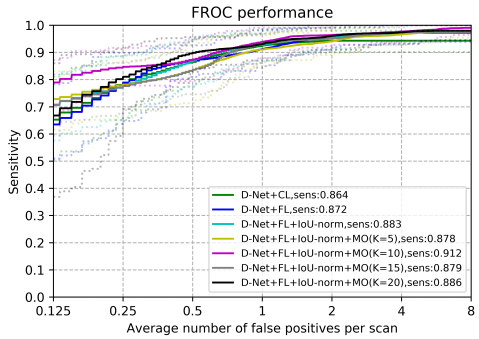


**Fig. 4**. Sensitivity (Recall) rate with respect to FPs. The solid line is the interpolated FROC beads on the prediction. The dash lines are lower bound and upper bound FROC for the bootstrapped FROC performance. D-Net: our model with dense connected modules; CL: binary cross entropy (BCE) loss; FL: focal loss; IoU-norm: IoU Self-normalization; MO: maxout unit; K: dimension of maxout unit.

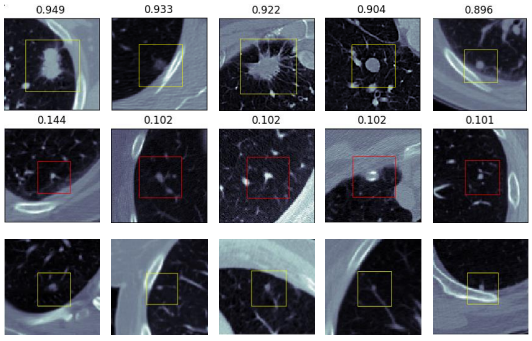


**Fig. 5**. True positives (the first row), false positives (the second row), and false negatives at 2 FPs (the last row). The numbers above stand for the predicted probabilities of the nodule. The yellow and red rectangle boxes are our detected nodules. All are the slice of center z.

error detection. D-Net+FL+IoU-norm+MO(K=5/10/15/20) shows that the results have different improvements with the maxout unit, and when K=10, FROC achieves the best.

We present the comparison among top results of the one-stage paradigm of the leaderboard in LUNA16 Challenge[1] and other submitted papers for equality, which is shown in Table 1. We have achieved the highest score of average sensitivity and the fewest candidates per scan compared with other algorithms. More specifically, our model has the highest recall rate at 1/8 FPs, which indicates that our model has stronger ability to detect nodules with less False positives.

### 3.4. Discussion

We believe that the improvement of performance is mainly due to the improvement of the recall rate of small nodules. We randomly choose nodules from testing fold and visualize our detection results in Fig. 5. The true positives all have high probabilities, while the false positives have low probabilities. At 2 FPs, only 5 nodules are missed throughout the test set. And they are all very small, even experienced doctors cannot reach consensuses in some cases. Therefore, under the condition of low false positives, small-sized nodules are susceptible to false negatives. Yet, the recall rate of our method is optimal in the case of low false positives, indicating that our model has a higher detection performance for small nodules.

## 4. CONCLUSION

This paper proposes a simple and highly effective lung nodule detection model. We utilize the normalized IoU values as classification labels instead of the binary classification confidence, which empowers more bounding box regression. We add maxout unit in the classifier to handle large intra-class variations of pulmonary nodules. In addition, the focal loss is incorporated to effectively tackle the hard/easy sample imbalance problem. Quantitative and qualitative results demonstrate that our contributions lead our network to achieve the state-of-the-art. We hope that our method is extensible to future works on other object detection tasks, and beyond.

[1] https://luna16.grand-challenge.org/results/

## 5. REFERENCES


[1] Cancer facts and figures 2016, "American cancer society web site," 2017.

[2] Kyongtae T Bae, Jin-Sung Kim, Yong-Hum Na, K-wang Gi Kim, and Jin-Hwan Kim, "Pulmonary nodules: automated detection on ct images with morphologic matching algorithmłpreliminary results," *Radiology*, vol. 236, no. 1, pp. 286–293, 2005.

[3] Olga Zinoveva, Dmitriy Zinovev, Stephen A Siena, Daniela S Raicu, Jacob Furst, and Samuel G Armato, "A texture-based probabilistic approach for lung nodule segmentation," in *International Conference Image Analysis and Recognition*. Springer, 2011, pp. 21–30.

[4] Yimo Tao, Le Lu, Maneesh Dewan, Albert Y Chen, Jason Corso, Jianhua Xuan, Marcos Salganicoff, and Arun Krishnan, "Multi-level ground glass nodule detection and segmentation in ct lung images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2009, pp. 715–723.

[5] Xujiong Ye, Xinyu Lin, Jamshid Dehmeshki, Greg Slabaugh, and Gareth Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic ct images," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 7, pp. 1810–1820, 2009.

[6] Jia Ding, Aoxue Li, Zhiqiang Hu, and Liwei Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 559–567.

[7] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] 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.

[9] Wentao Zhu, Chaochun Liu, Wei Fan, and Xiaohui Xie, "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification," *arXiv preprint arXiv:1801.09555*, 2018.

[10] Qi Dou, Hao Chen, Yueming Jin, Huangjing Lin, Jing Qin, and Pheng-Ann Heng, "Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 630–638.

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[12] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu, and Sen Song, "Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network," *arXiv preprint arXiv:1711.08324*, 2017.

[13] Rushil Anirudh, Jayaraman J Thiagarajan, Timo Bremer, and Hyojin Kim, "Lung nodule detection using 3d convolutional neural networks trained on weakly labeled data," in *Medical Imaging 2016: Computer-Aided Diagnosis*. International Society for Optics and Photonics, 2016, vol. 9785, p. 978532.

[14] Vinod Nair and Geoffrey E Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[15] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[16] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio, "Maxout networks," *arXiv preprint arXiv:1302.4389*, 2013.

[17] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," *arXiv preprint arXiv:1708.02002*, 2017.

[19] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[20] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al., "Tensorflow: a system for large-scale machine learning.," in *OSDI*, 2016, vol. 16, pp. 265–283.