Postprint: A Pulmonary Nodule Detection Algorithm Combining Hybrid Loss Joint Optimization and Multi-scale Classification

Authors: Yao Yujin, Zhang Li

Date: 2018-05-24T00:00:00+00:00

Abstract

To address the issues of low detection sensitivity and a large number of false positives in the task of automatic lung nodule detection in CT images, we propose a lung nodule detection method that combines a 3D fully convolutional network based on hybrid loss with an attention-based multi-scale 3D residual network. First, a 3D fully convolutional network is pre-trained based on similarity loss, which is utilized to screen hard example samples, and hard examples are jointly fine-tuned with positive samples based on hybrid loss to obtain a candidate nodule detection network for rapid screening of suspicious nodules. Then, an attention-based multi-scale 3D residual convolutional network is employed to classify suspicious nodules, accurately distinguishing true nodules from candidate nodules. On the LUN16 dataset, the sensitivity of the candidate nodule detection stage reaches 97.18% when the number of false positives per case is 59.1, and the average sensitivity of the detection system is 0.880, demonstrating that the proposed algorithm can improve the sensitivity of lung nodule detection and effectively control false positives, achieving superior performance on the LUNA16 dataset.

Full Text

Preamble

Pulmonary Nodule Detection via Hybrid Loss based Joint Fine-tuning and Multi-scale Classification

Yao Yujin, Zhang Li
(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: To address the problems of low detection sensitivity and numerous false positives in automated pulmonary nodule detection from CT images, this paper proposes an integrated method combining a hybrid loss-based 3D fully convolutional network for candidate detection with an attention-based multi-scale 3D residual network for nodule classification. First, a 3D fully convolutional network is pre-trained using similarity loss to filter hard negative samples, which are then jointly fine-tuned with positive samples based on hybrid loss to obtain a candidate nodule detection network for rapid screening of suspicious nodules. Subsequently, an attention-based multi-scale 3D residual convolutional network classifies these candidates to accurately distinguish true nodules from false positives. On the LUNA16 dataset, the candidate detection stage achieves a sensitivity of 97.18% at 59.1 false positives per scan, while the entire detection system attains an average sensitivity of 0.880. These results demonstrate that the proposed algorithm can improve detection sensitivity while effectively controlling false positives, achieving superior performance on the LUNA16 dataset.

Key words: lung nodule detection; hybrid loss; joint fine-tune; attention; multi-scale

0 Introduction

Lung cancer currently has the highest mortality rate among all cancers worldwide. Early detection and treatment can significantly improve patient survival rates. Early-stage lung cancer often presents as asymptomatic pulmonary nodules, making nodule detection crucial for early diagnosis and treatment. Low-dose chest CT is the most commonly used screening method for lung cancer, capable of detecting even tiny nodules. However, manual screening is labor-intensive and diagnostically challenging, especially with increasing demand for early cancer screening, requiring substantial human resources and time for CT analysis. Therefore, accurate and efficient automated pulmonary nodule detection is highly valuable.

Due to substantial variations in nodule size, shape, and density distribution, as well as diverse lung environments where nodules may adhere to vessels or lung walls or appear independently, automated detection from CT images remains challenging. A typical pulmonary nodule detection system consists of two stages: (a) candidate nodule detection, which screens suspected nodules from CT scans containing numerous false positives; and (b) false positive reduction, which employs more sophisticated classification algorithms to eliminate false positives and produce final detection results.

Traditional detection algorithms [2] primarily use image processing methods based on nodule shape and CT value characteristics to obtain candidates, followed by local feature extraction and classifier construction. Murphy et al. designed shape index and curvature features to select seed voxels, applied hysteresis thresholding for region growing to obtain candidate nodules, and employed a two-stage coarse-to-fine K-nearest neighbor classifier to reduce false positives. The coarse classifier descriptor concatenated 18 geometric and grayscale features from candidate voxel blocks, while the fine classifier descriptor included 135 richer shape and gradient features. Jacobs et al. used dual-thresholding based on CT values to screen candidate nodule voxels, applied morphological operations to reduce partial volume effects, and performed connected component analysis for clustering. They extracted grayscale, texture, shape, and environmental context features from candidate nodule neighborhoods and also adopted a two-stage classification approach using an ensemble GentleBoost classifier. Retico et al. designed filter functions based on the relationship between Hessian matrix eigenvalues and local geometry for multi-scale nodule enhancement, detected local maxima to obtain candidates, computed gradient and Hessian matrices in 3D neighborhoods, used their eigenvalues as descriptors, and trained a three-layer neural network classifier.

These traditional methods rely heavily on descriptor accuracy. However, given the complex and diverse manifestations of pulmonary nodules, hand-crafted descriptors typically capture only superficial features and struggle to extract discriminative features for distinguishing true nodules from similar structures (e.g., vessels, airways). With the rapid development of deep learning, convolutional neural networks have been increasingly applied to medical image processing, automatically extracting discriminative features in a data-driven manner. Recent studies [11][12] generally train fully convolutional networks to directly obtain prediction probability maps for each slice, select candidate nodules by thresholding, and then extract fixed-size 2D patches centered at candidate centroids for binary classification to eliminate false positives. Since CT data is inherently 3D volumetric information, relying solely on 2D cross-sectional information is insufficient. Consequently, multi-view or 3D convolution-based algorithms have been proposed to better utilize volumetric data [15][16][17]. Setio et al. extracted nine different 2D cross-sections from 3D blocks containing candidates, input them into a classification network with nine parallel independent branches, and fused the results at the network' s end to obtain the final 3D block classification. Dou et al. addressed large nodule scale variations by extracting three different-scale 3D patches centered at candidate centroids, processing them through three separate classification networks, and fusing the scale-specific probabilities for final classification. Additionally, some studies have proposed Faster R-CNN-based candidate detection methods [18][19]. Ding et al. added deconvolution layers after the final convolutional layer to restore finer features and designed fixed-size anchors matched to nodule dimensions, addressing the small relative size of nodules.

However, traditional segmentation and classification alone are insufficient for nodule detection. By incorporating morphological and distributional characteristics of pulmonary nodules, novel strategies can be explored to extract more discriminative information from 3D blocks, further improving system performance. This paper proposes an improved pulmonary nodule detection framework that optimizes both system stages. The candidate detection stage is enhanced through a 3D fully convolutional network with hybrid loss-based joint fine-tuning: pre-training with Dice loss, hard example mining, and joint fine-

tuning of hard negatives with positives to improve false positive discrimination. The nodule classification stage employs an attention-based multi-scale residual network that leverages candidate nodule information to extract local attention regions and enhances multi-scale nodule discrimination through multi-scale architecture, thereby improving classification accuracy. The proposed algorithm was validated on the large-scale LUNA16 dataset, demonstrating significant performance improvements. Comparative experiments further verified the effectiveness of each algorithmic improvement.

---

1 Method

This section presents an automated pulmonary nodule detection system whose framework is shown in Figure 1 [Figure 1: see original paper], comprising improved candidate nodule detection based on 3D U-net and attention-based multi-scale nodule classification.

1.1 Improved 3D U-net based Candidate Nodule Detection

1.1.1 Pre-training Model based on Dice Loss The candidate nodule detection stage first pre-screens CT scans to select suspected nodules while filtering out most other voxels, thereby reducing the number of samples input to the subsequent classification network. This pre-screening must ensure high detection sensitivity, preserving as many nodules as possible while eliminating non-nodules. To this end, we designed a segmentation network based on 3D U-net. First, distinguishing vessels from nodules using only 2D images is nearly impossible; therefore, all convolution and pooling operations in the network are 3D to fully utilize spatial information from CT volumetric data. Second, the fully convolutional network outputs the probability distribution of each voxel belonging to a nodule, enabling rapid acquisition of nodule attribute probability distributions across entire CT volumes.

Using entire CT volumes as training input would incur excessive computational cost. Moreover, determining whether a voxel V_i belongs to a nodule depends on its local CT value distribution, with more distant voxels having less influence on the prediction result. Therefore, we extract fixed-size 3D blocks as network input, reducing computational complexity while maintaining detection sensitivity. Due to the small input block size, we simplified the U-net architecture. The specific network structure is shown in Figure 1. The network outputs a lower-resolution 3D probability volume corresponding to the input block. Since the average nodule diameter is less than 10 mm, for a $56\times56\times56$ block containing a nodule, the positive-to-negative sample ratio is approximately 1:200 when treating each voxel as a sample. To address this severe class imbalance, we adopt the improved Dice coefficient as the loss function to prevent model bias toward negative samples. Dice coefficient measures similarity through set overlap. For a given 3D block $\{X, Y, label\}$, where X denotes network input, Y denotes corresponding annotations, and *label* indicates whether the block contains a nodule (1 for positive samples with nodules, 0 for negative samples without nodules), the Dice loss function is:

$$L_{dice} = -\frac{2 \times \sum_{i=1}^{N} y_i \cdot p(\hat{y}_i = 1|X, W)}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p(\hat{y}_i = 1|X, W)}$$

where W represents network parameters, y_i denotes the ground truth label (0 or 1) for each voxel, and $p(\hat{y}_i = 1|X, W)$ denotes the predicted probability for each voxel. For nodule-containing blocks, Dice loss encourages nodule detection, improving sensitivity. However, for nodule-free blocks, Dice loss is constantly zero, meaning all training samples must contain at least one nodule.

1.1.2 Hard Example Mining and Joint Fine-tuning Mechanism While the above algorithm can identify some non-nodules, it lacks discrimination ability for nodule-like negative samples. Most non-nodules are lung background regions easily distinguishable from nodules, but some (e.g., vessels, airways) have CT values and local morphologies similar to nodules. Although these challenging samples constitute a minority, they are difficult to differentiate, and incorporating them into training improves the detection network' s discriminative capability. To further leverage these hard examples, we designed a hard example mining and joint fine-tuning mechanism.

First, the model trained in Section 1.1.1 is applied to all nodule-free negative samples. Voxels with prediction probabilities exceeding a set threshold are considered false positive nodules, and negative samples containing these false positives constitute the mined hard examples. Hard examples and positive samples are jointly fine-tuned. Since Dice loss for negative samples is constantly zero, we employ improved cross-entropy as the loss function for negative samples while retaining Dice coefficient for positive samples, yielding the fine-tuned model' s loss function:

$$L_{ce} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log p(\hat{y}_i = 1|X, W) + (1 - y_i) \log(1 - p(\hat{y}_i = 1|X, W))]$$

$$L_{neg} = 2 \times \text{sigmoid}(L_{ce}) - 2, \quad L_{pos} = L_{dice}$$

$$L = \mathbb{1}(label = 1) \times L_{pos} + \mathbb{1}(label = 0) \times L_{neg}$$

where L_{ce} is cross-entropy loss, sigmoid transformation ensures loss values for positive and negative samples are in comparable ranges, L_{pos} and L_{neg} are positive and negative sample losses respectively, $\mathbb{1}(label = 1)$ is an indicator function applying different losses to positive/negative samples, and L is the hybrid loss for fine-tuning.

During testing, since fully convolutional networks accept arbitrary input sizes, entire CT volumes are input to obtain voxel-wise predictions, enabling direct localization of nodules across the whole CT. False positive nodules are suppressed through extreme value detection on the probability volume.

As the training network only incorporates local information, numerous false positives appear outside lung parenchyma. To address this, we employ lung segmentation algorithms to filter out false positives beyond lung regions, a common preprocessing step in lung-related medical image analysis. Through this candidate detection algorithm, we obtain all suspected nodules with their 3D locations, diameters, and nodule attribute probability distributions.

1.2 Attention-based Multi-scale Nodule Classification Network

After obtaining candidate nodules, this stage aims to accurately distinguish true nodules from false positives. We designed a multi-scale binary classification network based on 3D convolutional neural networks and residual modules. Residual networks address performance degradation in deep networks, demonstrating powerful capabilities across various image tasks. Incorporating residual modules ($\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}$, where $\mathbf{x}$ and $\mathbf{y}$ are module input and output, and $\mathcal{F}$ is a nonlinear transformation comprising convolution, batch normalization (BN), and ReLU) in the classification network facilitates error backpropagation and enhances network expressiveness, aiding model optimization. The local branch structure is shown in Figure 2 [Figure 2: see original paper].

This stage typically extracts fixed-size 3D patches centered at candidate nodule centroids for binary classification, where candidates with prediction probabilities exceeding a threshold are considered detected nodules. Standard approaches face two issues: First, single-scale patches cannot represent nodules of varying sizes. Correct classification requires patches to contain not only the nodule itself but also surrounding parenchymal context, which is crucial for accurate classification. However, due to wide nodule size variations, large nodules require larger patches to encompass complete information and sufficient context, while small nodules in oversized patches would have disproportionately small nodule-to-background ratios. Second, using only image patches as input loses annotation information including nodule location and relative size within the patch, which is vital for classification. If each voxel in the input patch were weighted, those near the nodule center should receive higher weights with influence decreasing with distance. To address these issues, we designed an attention-based multi-scale network architecture.

Different image regions contribute unequally to classification. Visual attention mechanisms identify discriminative regions, which in our case are directly obtainable with the attention center being the nodule centroid. During training, positive sample centroids and diameters are obtained from annotations, while negative samples come from hard example screening by the candidate detection network, yielding false positive centroids and diameters. During testing, candidate nodule 3D locations and diameters are similarly obtained from the detection network.

Given a 3D image patch (X, Y, L), where L represents true or false nodule information (x, y, z, d), the network' s hidden layer contains two parallel independent branches. The local branch extracts a local attention region centered at the nodule centroid with size twice the nodule diameter, resampling it to a fixed size. The global branch uses the complete image patch as input. The two branch structures are shown in Figure 2 (global branch configurations in parentheses, others identical to local branch), producing binary probability distributions $p(\hat{Y}|X, W_{global})$ and $p(\hat{Y}|X, L, W_{loc})$ based on global and local patches respectively. The final network output combines these distributions through weighted combination:

$$p(\hat{Y}=1|X, L, W_{global}, W_{loc}) = \alpha \cdot p(\hat{Y}=1|X, W_{global}) + (1-\alpha) \cdot p(\hat{Y}=1|X, L, W_{loc})$$

The weight α is learned from a hidden layer fed by the two branches' Softmax outputs (Figure 1), enabling flexible weighting based on each branch' s discriminative power. The classification network loss function is:

$$L = -\frac{1}{N}\sum_{i=1}^{N} \log p(Y_i|X_i, W_{global}, W_{loc}) + \lambda(\|W_{global}\|_2^2 + \|W_{loc}\|_2^2)$$

where the first term is nodule classification loss defined by log-likelihood, and the second term is regularization. To prevent overfitting, λ is a weight decay parameter controlling model complexity' s influence on the loss.

2 Experimental Design and Results Analysis

We validated the algorithm on the large-scale public dataset LUNA16 [21], derived from the LIDC-IDRI database through expert screening and annotation. LUNA16 contains 888 low-dose CT scans with annotations of nodule center locations and diameters. For experiments, LUNA16 was randomly divided into 10 subsets for cross-validation. Evaluation metrics include sensitivity and average false positives per scan (FPs/scan). The LUNA16 challenge also uses Competition Performance Metric (CPM) [21], which measures average sensitivity at different FPs/scan levels (1/8, 1/4, 1/2, 1, 2, 4, 8). For the entire system, a detection is considered correct if it falls within the radius of a true nodule.

2.1 Implementation Details

Since CT scans may originate from different devices with varying slice thicknesses, we first resampled all data to 1mm×1mm×1mm for consistency. To increase sample diversity, we performed data augmentation on positive samples including random translation [-8,8] voxels in xyz directions, flipping, random scaling [0.9,1.1], and random rotation [-180°,180°]. These parameters were set to keep nodules within a 30×30×30 cube around the original centroid. Training block size for candidate detection was 56×56×56, while classification stage input blocks were 60×60×30 with local patches of 20×20×20.

Networks were optimized using stochastic gradient descent. Pre-trained 3D U-net weights were randomly initialized from Gaussian distribution $\mathcal{N}(0, 0.01)$ with initial learning rate 0.001. The joint fine-tuning network used initial learning rate 0.0001 with 1:1 positive-negative sample ratio. Hard example mining threshold and extreme value detection threshold were both set to 0.8. Nodule classification network parameters were randomly initialized following [17]. The regularization parameter λ was set to 1e-4. Our algorithm was implemented using Theano in the Lasagne deep learning framework on an Nvidia Titan X GPU.

2.2 Results

2.2.1 Candidate Nodule Detection Results To validate the proposed candidate detection algorithm, we designed three comparative experiments using the same simplified 3D U-net architecture. First, we trained models with cross-entropy loss and Dice loss respectively. Results in Table 1 show that Dice loss significantly improves candidate detection sensitivity compared to cross-entropy, indicating better detection of suspicious nodules albeit with increased false positives. Furthermore, incorporating hard example mining and joint fine-tuning: the dataset contained 551,065 negative samples, of which 35,118 (6.4%) containing false positive nodules were identified as hard examples through screening. Joint fine-tuning with positive samples yielded the final detection model. Table 1 shows this mechanism reduced false positive detection rate from 86.9 to 59.1 FPs/scan while maintaining high sensitivity, demonstrating enhanced discrimination between true and false positives. To further demonstrate the necessity of hard example screening and pre-training, we trained a network directly with mixed loss on unscreened positive-negative samples. Results show this approach yields lower discriminative capability and sensitivity compared to our proposed algorithm.

Table 1 Candidate nodule detection comparison results

Method	Sensitivity	FPs/scan
Cross-entropy loss	82.86%	59.1
Dice loss	97.14%	86.9
Mixed loss (proposed)	95.23%	59.1

Method	Sensitivity	FPs/scan
Proposed algorithm	97.18%	59.1

2.2.2 Nodule Classification Results For the nodule classification stage, to validate the effectiveness of residual modules and multi-scale architecture, we implemented three network structures: conventional convolutional classification network, convolutional network with residual modules, and attention-based multi-scale residual classification network. Performance was quantitatively analyzed using Free-response Receiver Operating Characteristic (FROC) curves and CPM. FROC curves intuitively reflect system sensitivity at different false positive rates, with results shown in Figure 3 [Figure 3: see original paper]. The CPM values for the three networks were 0.832, 0.848, and 0.880 respectively, demonstrating that residual modules aid model optimization and improve performance, while the attention-based multi-scale structure effectively utilizes candidate nodule location and diameter information to further enhance discriminative capability, validating the multi-scale architecture' s effectiveness.

Figure 3 also compares our algorithm with existing pulmonary nodule detection methods [15][17]. The multi-scale residual network achieves the best performance. In the proposed multi-scale classification algorithm, the weight for combining two branch predictions is learned through a set of weight parameters. To validate this self-learned weighting, we conducted comparative experiments with fixed weights $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, with CPM results shown in Figure 4 [Figure 4: see original paper]. Fixed extreme weights underutilize one branch, yielding suboptimal performance. Balanced fusion ratios better leverage both branches, achieving optimal performance. The proposed algorithm' s average sensitivity of 0.880 surpasses all fixed-weight variants, demonstrating its ability to adaptively select weights based on each branch' s representational capacity, thereby more effectively utilizing both information sources.

---

3 Conclusion

This paper proposes an improved method for automated pulmonary nodule detection from CT images. The approach first employs an improved U-net with hard example mining and joint fine-tuning for candidate detection, achieving high sensitivity while reducing false positive rates. An attention-based multi-scale residual network then classifies candidates, leveraging nodule position and size information while addressing large scale variations. Experimental results on LUNA16 validate the improved algorithm' s effectiveness and demonstrate superior performance compared to existing methods. Our method addresses two common problems in medical image analysis: class imbalance and large scale variations, achieving better results. Future work will apply these methods to more medical imaging domains, exploring optimal methods and strategies tailored to specific problems.

References

[1] Sluimer I, Schilham A, Prokop M, et al. Computer analysis of computed tomography scans of the lung: a survey [J]. IEEE Trans on Medical Imaging, 2006, 25 (4): 385-405.

[2] Messay T, Hardie R C, Rogers S K. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery [J]. Medical Image Analysis, 2010, 14 (3): 390-406.

[3] Tan M, Deklerck R, Jansen B, et al. A novel computer-aided lung nodule detection system for CT images [J]. Medical Physics, 2011, 38 (10): 5630-5645.

[4] Setio A A A, Jacobs C, Gelderblom J, et al. Automatic detection of large pulmonary solid nodules in thoracic CT images [J]. Medical Physics, 2015, 42 (10): 5642-5653.

[5] Lopez Torres E, Fiorina E, Pennazio F, et al. Large scale validation of the M5L lung CAD on heterogeneous CT datasets [J]. Medical Physics, 2015, 42 (4): 1477-1489.

[6] Murphy K, Van Ginneken B, Schilham A M, et al. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification [J]. Medical Image Analysis, 2009, 13 (5): 757-770.

[7] Cover T, Hart P. Nearest neighbor pattern classification [J]. IEEE Trans on Information Theory, 1967, 13 (1): 21-27.

[8] Jacobs C, Van Rikxoort E M, Twellmann T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images [J]. Medical Image Analysis, 2014, 18 (2): 374-384.

[9] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting [J]. Annals of Statistics, 2000, 28 (2): 337-374.

[10] Retico A, Delogu P, Fantacci M E, et al. Lung nodule detection in low-dose and thin-slice computed tomography [J]. Computers in Biology & Medicine, 2008, 38 (4): 525-534.

[11] Van Ginneken B, Setio A A A, Jacobs C, et al. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans [C]// Proc of the 12th IEEE International Symposium on Biomedical Imaging. Piscataway, NJ: IEEE Press, 2015: 286-289.

[12] Jin Xinyu, Zhang Yuchen, Jin Qiliang. Pulmonary nodule detection based on CT images using convolution neural network [C]// Proc of the 9th IEEE In-

ternational Symposium on Computational Intelligence and Design. Piscataway, NJ: IEEE Press, 2017: 202-204.

[13] Golan R, Jacob C, Denzinger J. Lung nodule detection in CT images using deep convolutional neural networks [C]// Proc of IEEE International Joint Conference on Neural Networks. Piscataway, NJ: IEEE Press, 2016: 243-250.

[14] Anirudh R, Thiagarajan J, Bremer T, et al. Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data [C]// Proc of SPIE Medical Imaging. 2016: 978532.

[15] Huang Xiaojie, Shan Junjie, Vaidya V. Lung nodule detection in CT using 3D convolutional neural networks [C]// Proc of the 14th IEEE International Symposium on Biomedical Imaging. Piscataway, NJ: IEEE Press, 2017: 761-769.

[16] Setio A A A, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks [J]. IEEE Trans on Medical Imaging, 2016, 35 (5): 1160-1169.

[17] Dou Di, Chen Hao, Yu Lequan, et al. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection [J]. IEEE Trans on Bio-medical Engineering, 2017, 64 (7): 1558-1567.

[18] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.

[19] Zhu Wentao, Liu Chaochun, Fan Wei, et al. DeepLung: deep 3D dual path nets for automated pulmonary nodule detection and classification [C]// Proc of IEEE Workshop on Applications of Computer Vision. Piscataway, NJ: IEEE Press, 2018.

[20] Ding Jia, Li Aoxue, Hu Zhiqiang, et al. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks [C]// Proc of International Conference on Medical Computing and Computer-Assisted Intervention. Berlin: Springer, 2017: 559-567.

[21] Sudre C H, Li Wenqi, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations [C]// Proc of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Berlin: Springer, 2017: 240-248.

[22] Setio A A A, Traverso A, De Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge [J]. Medical Image Analysis, 2017, 42: 1-13.

[23] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]// Proc of International Conference on Medical Computing and Computer-Assisted Intervention. Berlin: Springer, 2015:

234-241.

[24] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 761-769.

[25] Van Rikxoort E, De Hoop B, Viergever M A, et al. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection [J]. Medical Physics, 2009, 36 (7): 2934-2947.

[26] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Identity mappings in deep residual networks [C]// Proc of European Conference on Computer Vision. Berlin: Springer, 2016: 630-645.

[27] Loffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// Proc of International Conference on Machine Learning. New York: ACM Press, 2015: 448-456.

[28] Veit A, Wilber M, Belongie S. Residual networks behave like ensembles of relatively shallow networks [C]// Proc of the 29th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015.

[29] Kamnitsas K, Ledig C, Newcombe V F, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. Medical Image Analysis, 2016, 36: 61-78.

[30] Xiao Tianjun, Xu Yichong, Yang Kuiyuan, et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 842–850.

[31] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention [C]// Proc of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2204–2212.

[32] Armato S G, Mclennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans [J]. Medical Physics, 2011, 38 (2): 915-931.