Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network

Chen Zhao
College of Computer Science and Technology
Xi'an University of Posts and Telecommunications
Xi'an, China
1603210019@stu.xupt.edu.cn

Jungang Han
College of Computer Science and Technology
Xi'an University of Posts and Telecommunications
Xi'an, China
hjg@xupt.edu.cn

Yang Jia
College of Computer Science and Technology
Xi'an University of Posts and Telecommunications
Xi'an, China
jiayang@xupt.edu.cn

Fan Gou
College of Computer Science and Technology
Xi'an University of Posts and Telecommunications
Xi'an, China
poppy_fan@126.com

***Abstract*—Lung cancer is the most common cause of cancer death worldwide. Lung nodule detection based on CT image is the most prevailing method for detecting lung cancer. In this paper, we propose a patch-based 3D U-Net and contextual convolutional neural network (CNN) to automatically segment and classify lung nodule and help the radiologists read CT images. Typically, lung nodule detection task could be divided into three stages, including lung segmentation, nodule detection or segmentation and false positive reduction. In lung segmentation stage, we use morphological methods to segment pulmonary parenchyma from raw CT images. To segment lung nodule, 3D U-Net is employed to extract suspicious nodule from preprocessed CT images. In order improve model accuracy, we use Generative Adversarial Network (GAN) to boost model training. To further enhance model performance, we use online sampling strategy to augment data and use 3D contextual CNN with Inception blocks to determine whether the volume is malignant nodule or not. Experimental results demonstrate that the proposed method could effectively detect the cancerous nodule from the CT scans.**

Keywords- Lung Nodule Detection; 3D U-Net; 3D Inception CNN; Generative Adversarial Network

I. INTRODUCTION

Lung cancer is one of the most rapidly growing rates of morbidity and mortality and one of the most serious malignancy for human's health and life [1]. Early detection of lung nodules is crucial for diagnosing and clinical managing of lung cancer. At initial stage, the lung nodules are still in a treatable stage and it will give more chance to survive if it could be detected successfully. However, detecting malignant lung nodule often needs experienced clinical knowledge and the process is time-consuming and tedious. In addition, the nodules are not easy to recognize since they vary considerably in shape and intensity [2]. Under these circumstances, it is necessary to develop an algorithm to automatically detect and segment nodules from CT images. Automatic lung nodule detection is beneficial for both patients and radiologists.

In the field of lung nodule detection, the key problem is how to segment candidate lung nodules from raw CT radiographs or how to localize malignant nodules. In general, the largest lung nodule has only 30×30×30 mm^3 volume. Compared with the entire lung 3D CT radiographs with a 400×400×400 mm^3 volume, detecting lung nodule can be defined as a small object searching task. Examples of malignant lung nodules are depicted in Fig. 1.

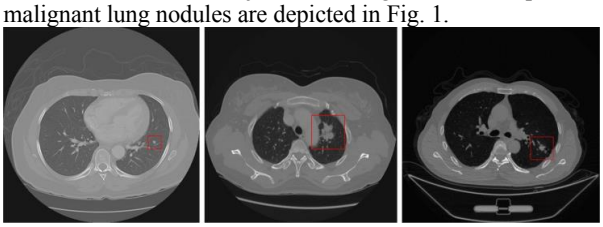

Figure 1. A close-up of a malignant nodules.

In recent years, deep learning technique, especially the use of convolutional neural networks (CNNs), achieves impressive performance in medical images classification [16], segmentation [9] and localization tasks [17]. It motivates us to develop an automatic lung nodule detection and segmentation framework. In this paper, we propose a fully automated lung nodule detection algorithm, including pulmonary parenchyma segmentation, candidate nodule segmentation and contextual CNN classifier, as illustrated in Fig. 2. For a CT image, we first use morphological methods such as dilation and erosion to segment lung masks. Next, we design a patch-based 3D U-Net to segment suspicious nodule from lung parenchyma. Since there exist a huge number of false positive candidate nodules, we employ a deep CNN to classify the detected nodules into either malignant or benign. Considering the computation ability and memory limitation, it is impossible to feed a whole 3D CT image to CNN. In this situation, to incorporate context of nodule Region of Interest (RoI), we propose contextual CNN to classify nodules.

978-1-5386-8303-3/18/$31.00 ©2018 IEEE
DOI 10.1109/NaNA2018.2018.00070

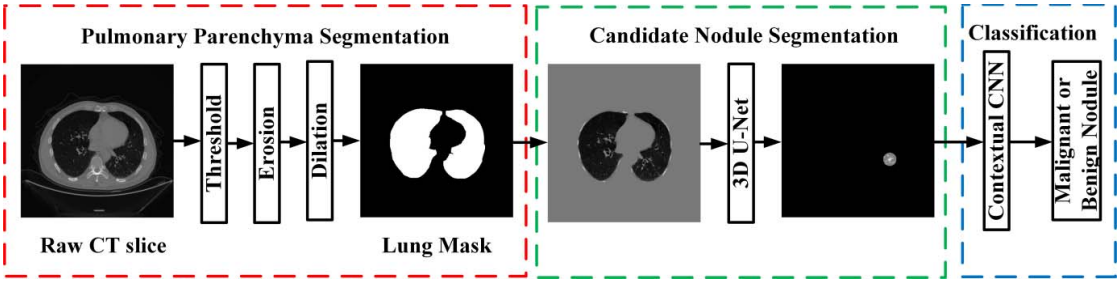

Figure 2. Overview of proposed lung nodule detection framework.

Our main contributions are as follow:

1. To localize candidate nodule, a patch-based 3D U-Net is proposed to segmented nodule from raw CT scans;
2. Due to the difficulty of training 3D U-Net segmentation network, we employed Generative Adversarial Network (GAN) to boost neural work training and enhance model performance;
3. Our classification network incorporates the context information and improves the model accuracy.

II. Related Work

In this section, we will review some proposed lung nodule detection and classification methods.

Traditional nodule detection task requires manually extracted features, such as morphological features, voxel clustering and pixel thresholding [3]. Recently, several approaches have been proposed to use deep CNNs to automatically extracted image features, which is proven to be much more accurate and effective than hand-crafted features. Motivated by the achievement of deep CNNs, various computer aided diagnosis systems for lung nodule detection have been proposed: in [4], selective search is employed to generate candidate nodule RoIs and an Inception classification network is used to reduce false positive samples. In [5], the Faster R-CNN is used to generate candidate nodules followed by a 3D CNN to remove false positive nodules. In [6], a novel 3D Faster R-CNN with dual path blocks is proposed and a U-Net-like encoder-decoder scheme is employed to learn nodule features.

Nodule classification is an effective approach to remove false positive candidate nodules and has been widely adopted in several works. In [7], a cascaded support vector machine (SVM) is built to classify suspicious nodule. In [8], a four-channel CNN model is designed to learn the knowledge of radiologists for detecting nodules for four level. The results demonstrate that multi-group learning system is efficient to improve the performance of lung nodule detection. In [9], a 3D CNN styled computer aided system incorporates sliding window scanning technique is designed to search candidate nodule and achieves impressive performance.

In general, it is important to detect lung nodule RoI so that the searching space could be significantly reduced. In addition, to eliminate a large amount false positive candidate nodules, a powerful classifier is designed to determine the possibility of candidate nodule is a malignant one or not. In our approach, we use lung nodule segmentation technique to localize candidate nodules. To exploit contextual information, we propose contextual CNN with Inception blocks to remove false positive nodules.

III. Methodology

Our fully automated lung nodule detection framework consists of three parts: lung parenchyma segmentation, candidate nodule detection and classification.

A. Pulmonary Parenchyma Segmentation

In order to reduce searching space and easily to find volumes surrounding suspicious nodules, it is necessary to segment lung parenchyma from raw CT radiographs. The value in CT radiographs is represented by HU value with range from -2048 to 400. Each value in the HU represents different tumors, for example, 1000 indicates the air and 400 represents the bone. Based on several experiments, we found that it works with a window range from -1000 to +400 to crop relevant RoI firstly. To get the binary parenchyma mask, we set the threshold of -400 to coarsely segment raw CT radiographs.

However, the threshold method has limitation to precisely generate parenchyma mask due to the variability of CT radiographs. In order to further improve the segmentation results, we firstly select two largest connected volumes from processed binary make because a patient has two lungs. Through the experiments, we discovered that under segmentation occurs frequently, the binary mask often larger than real lung mask. To tackle this problem, we need to shrink bright regions, i.e. processed lung mask, and enlarge dark regions, i.e. background of mask. We use erosion and dilation method to adjust binary lung mask.

The overview of lung parenchyma segmentation approach is demonstrated in left part of Fig. 2.

B. Candidate Nodules Segmentation

2D U-Net [9] has shown compelling accuracy and impressive convergence behaviors on medical image segmentation tasks. Meanwhile, skip connections from the down sampling to the up sampling path is usually adopted to recover spatially detailed information by reusing feature maps [10]. In order to leverage the powerful capability of U-Net and tackle the nodule segmentation task effectively and efficiently, we refine the original U-Net architecture by alternating 2D CNN layers with 3D CNN layers. Considering the computation ability, we use 3 down sampling layers

corresponded with 3 up sampling layers. The architecture of 3D U-Net is depicted in Fig. 3.

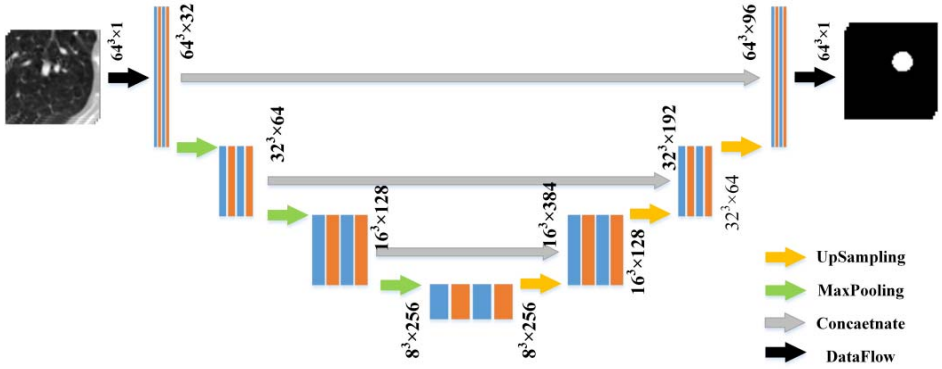

Figure 3. The architecture of refined 3D U-Net.

In the research field of image semantic segmentation, a pixel-wise loss function is usually used to penalize the distance between the ground truth and the predicted probability map. Often, the pixel-wise loss function is defined by a cross entropy as follows,

$$L_{pixel\_wise}=\sum_i -y_i \log(\hat{y}_i) -(1-y_i) \log(1-\hat{y}_i). \quad (1)$$

where y_i is a binary value of the corresponding pixel i and $\hat{y}_i$ is a predicted probability for the pixel.

DICE coefficient is another useful metric to evaluate the quality of segmentation, since it considers the overlapping between segmented result and ground truth. What's more important, in medical image segmentation, the border continuity can be improved for models with DICE loss [10]. The DICE coefficient is defined as follows:

$$DICE=2\ (|g(\hat{y})\cap y|)/(|g(\hat{y})|+|y|) \quad (2)$$

where y is a binary vessel map of the corresponding image and $g(\hat{y})$ is the post-processed binary hand mask on predicted probability map with OTSU algorithm [11].

The higher the DICE coefficient between the ground truth and the segmented result, the more powerful the deep neural network and the more accurate the segmented result. In order to optimize neural networks by DICE coefficient, we choose $L_{DICE} = 1 - DICE$ as a penalty function and minimize it. Finally, our loss function for 3D U-Net is:

$L_{loss} = L_{pixel-wise} + L_{DICE}$

$$= \sum_i \left(-y_i \log(\hat{y}_i) -(1-y_i) \log(1-\hat{y}_i)\right) + \left(1-2\frac{|g(\hat{y})\cap y|}{|g(\hat{y})|+|y|}\right). \quad (3)$$

Even though U-Net contains a strong generalization ability, it is still difficult to converge on candidate nodule detection task since the CT images vary considerably in contrast, brightness and sharpness. What's even severe, the lung nodules have diverse intensity and a huge number of morphological features such as solid, non-solid [2]. To further enhance model performance, we train our 3D U-Net with GAN.

GAN is a framework that enables to create as realistic outputs as the gold standard [12]. Usually, a GAN consist of two networks, a discriminator and a generator. The discriminator tries to distinguish whether the image is from gold standard or outputs generated by generator, while the generator tries to generate as realistic outputs as the discriminator cannot differentiate from the gold standard. In our task, we define the generator G be a map from a patch of CT image x to a segmented nodule candidate y, formally, $G: x \rightarrow y$. In our network setting, G is a 3D U-Net depicted in Fig. 3. The discriminator D maps a pair of $\{x, y\}$ to binary classification $\{0, 1\}$, where 0 and 1 indicate whether y is gold standard or generated by G. The relationship between G and D is shown in Fig. 4.

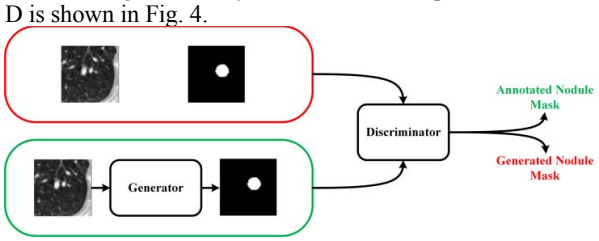

Figure 4. Overview of 3D U-Net with GAN.

Due to the image segmentation task can be defined as an image generation task, we adopt (3) as the loss function of G. Then, the objective function of GAN for nodule segmentation task can be formulated as:

$$L_{GAN}(G, D) = E_{x,y\sim p_{data(x,y)}}\left[\log(D(x, y))\right] + E_{x\sim p_{data(x)}}\left[1 - \log(D(x, G(x)))\right] \quad (4)$$

Note that in conditional GAN [13], G takes a random noise to generate images, while in our task, G takes 3D patched CT images to generate enhanced images. Then, the optimization problem can be defined as:

$$G^* = \arg\min_G \left[\max_D E_{x,y\sim p_{data(x,y)}}\left[\log(D(x, y))\right] + E_{x\sim p_{data(x)}}\left[1 - \log\left(D(x, G(x))\right)\right]\right] \quad (5)$$

For D, it goal is to correctly distinguish the volume is a generated or a gold standard, the optimization objective for D is:

$$D^* = \arg\max_D E_{x,y\sim p_{data(x,y)}}\left[\log\left(D(x,y)\right)\right] + E_{x\sim p_{data(x)}}\left[1-\log\left(D\left(x,G(x)\right)\right)\right] \quad (6)$$

C. *Contextual CNN for False Positive Reduction*

By using the 3D U-Net to segment candidate nodules from CT radiographs, there exists a large amount of false positive nodules. In order to reduce them, a classification network is employed to classify the detected nodules into either malignant or benign. For CT data, advanced method should be effective to extract 3D volume feature [14]. Inspired by the Inception structure [15], we design the 3D Inception with residual convolution blocks as the candidate nodule classification network (N). The architecture of proposed network is shown in Fig. 5. The detail of spatial reduction blocks and residual convolutional blocks are illustrated in [4].

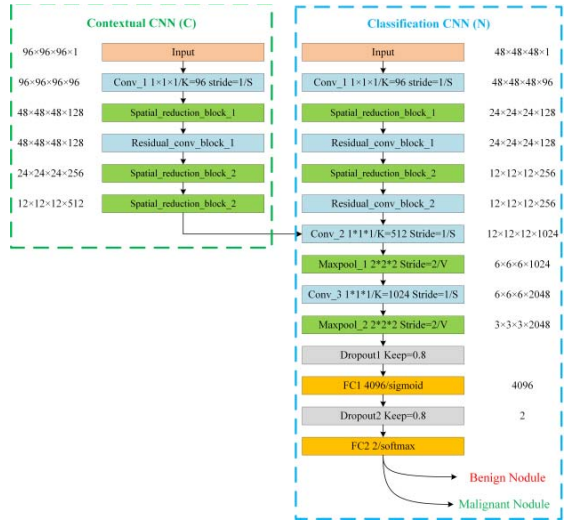

Figure 5. Architecture of contextual CNN for nodule classification.

For 3D neural networks, the balance between memory consumption and feature representation ability should be considered. Many off-the-shell 2D deep CNN take a whole 2D slice as input and can capture features in a large receptive field. However, feeding a whole 3D volume as input will consume a lot of GPU memory. As a trade-off, patch-based 3D CNN is used in 3D image classification in general. Because of the memory confinement, the limits of resolution and number of features in the networks often lead to a low representation ability. In this situation, we propose the Contextual CNN (C). The 'contextual' indicates that the proposed network could use contextual information of candidate nodule to generate final prediction. In general, the receptive field of contextual CNN is larger than candidate nodule classification network so that the whole discrimination network could incorporate contextual information. We down sample the feature maps of C so that the feature maps of both C and N have same size in the middle layer of network N and we can concatenate them to perform final prediction, as illustrated in Fig. 5.

IV. EXPERIMENT

In this section, we describe the experimental setting and results of lung parenchyma segmentation, nodule segmentation and nodule classification. The experiments are based on a relevant large dataset contains 800 CT scans and the dataset has been manually divided into training set with 600 samples and test set with 200 samples. The training and test set include 975 and 269 malignant nodules, respectively. Each nodule is annotated with the position in full image and diameter.

A. *Lung Parenchyma Segmentation*

The examples of segmented lung parenchyma binary masks are depicted in Fig. 6. Each column in Fig. 6 indicates an extracted CT slice with predicted mask and corresponding segmentation result. Through the experiments, we found that by using proposed lung parenchyma segmentation method, the segmented lung mask is both accurate and precise.

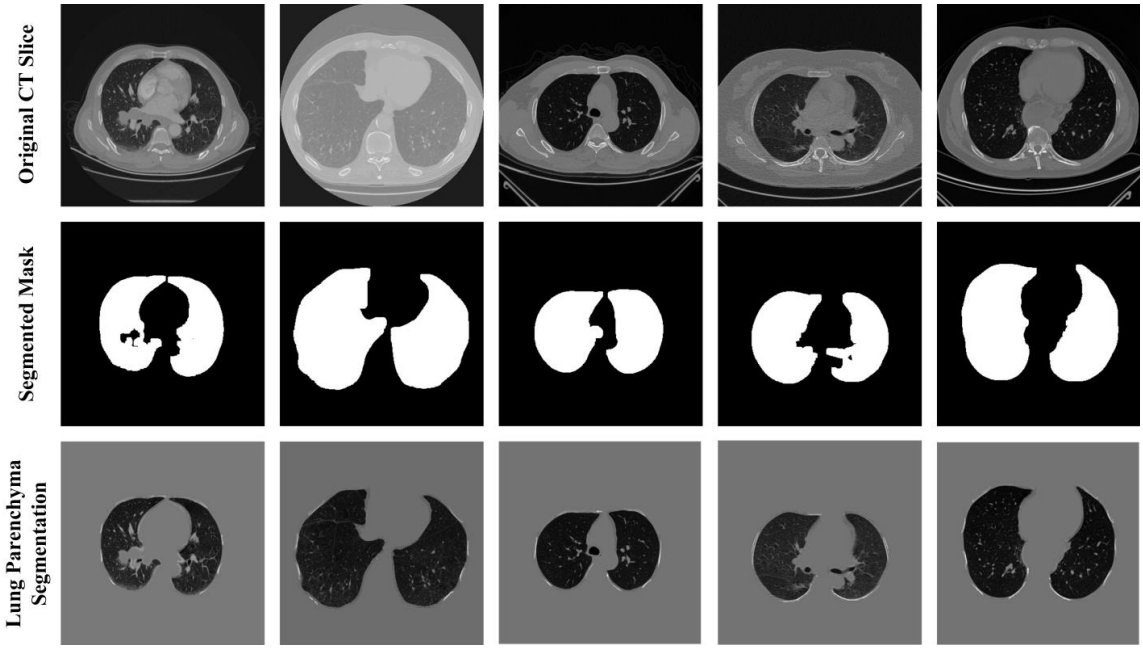

Figure 6. Examples of lung parenchyma segmentation.

B. Candidate Nodule Segmentation

Considering our computation ability, we set the size of 3D patch volume as 64, and the tensor size of 3D U-Net is shown in Fig. 3. In training stage, we save the model with best DICE score and use the model to evaluate the CT images in test set. To train 3D U-Net, we randomly crop 3D volume near the center of lung nodules. For the label of segmentation network training, we define the nodule as a sphere with the annotated diameter and center. At the training stage, we augment data by using online sampling strategy to prevent overfitting. Online sampling indicates that we randomly select training nodule sample and randomly flip cropped volume. Detected nodules are visualized as 2D slices in Fig. 7. From the top to bottom in Fig. 7. represents cropped CT volumn, nodule segmentation probability map, post processed binary mask and ground thuth. Through Fig. 7., we could find the nodule segmentation network achieves a powerful performance.

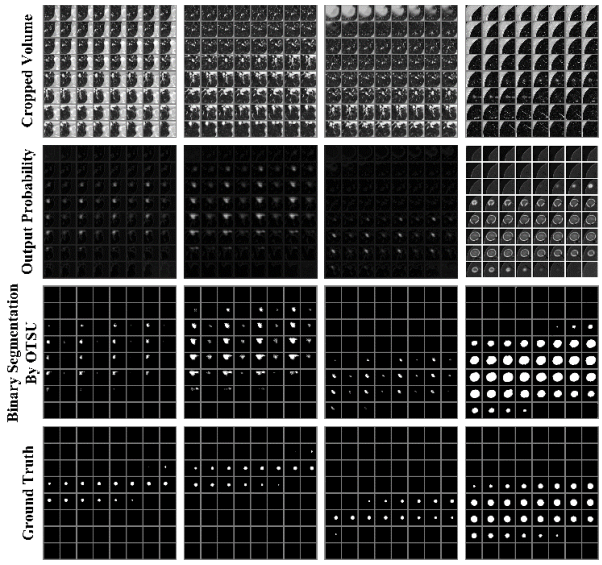

Figure 7. Examples of candidate nodule segmentation.

C. Contextual CNN for False Positive Reduction

Within the candidate nodules segmented by the 3D U-Net, there exists a huge number of benign nodules and we need to reduce the false positive samples. Like segmentation stage, we adopt online sampling strategy and crop the nodule volume and its context with size of 48 and 96 respectively to train nodule classification network. We also visualize the nodule classification results from test set in Fig. 8. The probability below the nodule indicates the predicted malignant probabilities and the red rectangle represents RoI of nodule in

2D slice. In our experiments, if the probability is larger than 0.5, it predicts malignant.

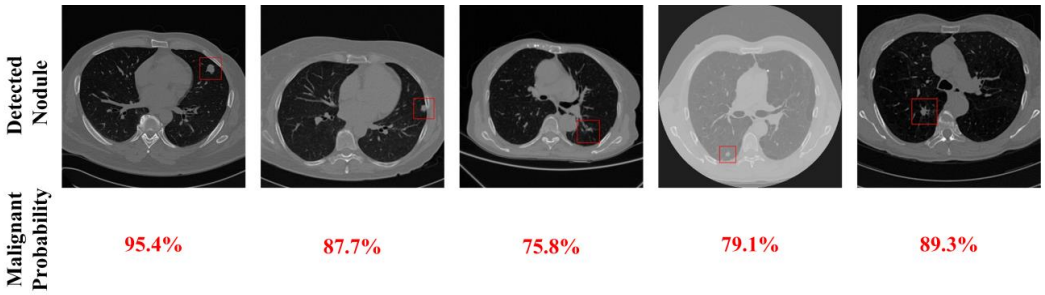

Figure 8. Visualization of nodule classification results on test set.

V. CONCLUSION

In this paper, we propose a complete automatic algorithm for lung nodule detection. First, we set a threshold on HU value and use morphological methods to segment lung parenchyma. Then we use patched 3D U-Net with adversarial training to localize candidate lung nodule in whole CT radiographs. Finally, we propose a contextual CNN to classify candidate nodules into either malignant or benign to reduce false positive candidate nodules. The experiments results demonstrate that the proposed method are accurate and effective to detect malignant nodule.

ACKNOWLEDGMENT

This work was supported by the Graduate Innovation Foundation in Xi'an University of Posts and Communications under Grant CXJJ2017005.

REFERENCES

[1] Overman V P. American Cancer Society[J]. International Journal of Dental Hygiene, 2006, 4(2):109-109.

[2] Ciompi F, Chung K, Riel S J V, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning[J]. Scientific Reports, 2017, 7:46479.

[3] Murphy, K.; van Ginneken, B.; Schilham, A. M.; De Hoop, B.; Gietema, H.; and Prokop, M. 2009. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. Medical image analysis 13(5):757–770. Jacobs et al. 2014.

[4] Zhao C, Han J, Jia Y. 3D Inception Convolutional Neural Networks for Automatic Lung Nodule Detection[C]// International Conference on Computational Science and Computational Intelligence. 2017.

[5] Ding J, Li A, Hu Z, et al. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks[J]. 2017:559-567.

[6] Zhu, Wentao, Liu, Chaochun, Fan, Wei, et al. DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification[J]. 2018:673-681.

[7] Wiemker R, Klinder T. Pulmonary nodule detection using a cascaded SVM classifier[C]// SPIE Medical Imaging. 2016:978513.

[8] Jiang H, He M, Wei Q, et al. An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network[J]. IEEE Journal of Biomedical & Health Informatics, 2017, PP(99):1-1.

[9] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015:234-241.

[10] Drozdzal M, Vorontsov E, Chartrand G, et al. The Importance of Skip Connections in Biomedical Image Segmentation[J]. 2016:179-187.

[11] OTSU. A Threshold Selection Method from Gray-Level Histograms

[12] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// International Conference on Neural Information Processing Systems. MIT Press, 2014:2672-2680.

[13] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.

[14] Yan X, Pang J, Qi H, et al. Classification of Lung Nodule Malignancy Risk on Computed Tomography Images Using Convolutional Neural Network: A Comparison Between 2D and 3D Strategies[C]// Asian Conference on Computer Vision. Springer, Cham, 2016:91-101.

[15] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]. 2016.

[16] Zhang J, Xia Y, Wu Q, et al. Classification of Medical Images and Illustrations in the Biomedical Literature Using Synergic Deep Learning[J]. 2017.

[17] van Sloun R J, Demi L, Postema A W, et al. Ultrasound-contrast-agent dispersion and velocity imaging for prostate cancer localization[J]. Medical Image Analysis, 2017, 35:610-619.