

Deep Learning for Automatic Real-time Pulmonary Nodule Detection  
and Quantitative Analysis

by

Chenyang Liu

Duke Kunshan and Duke University  
Graduate Program in Medical Physics

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Fang-Fang Yin, Supervisor

\_\_\_\_\_  
Chunhao Wang

\_\_\_\_\_  
David Huang

\_\_\_\_\_  
Xin Chen

Thesis submitted in partial fulfillment of  
the requirements for the degree of Master of Science in the  
Medical Physics Graduate Program of  
Duke Kunshan University

2019

ABSTRACT

Deep Learning for Automatic Real-time Pulmonary Nodule Detection

and Quantitative Analysis

by

Chenyang Liu

Duke Kunshan and Duke University  
Graduate Program in Medical Physics

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Fang-Fang Yin, Supervisor

\_\_\_\_\_  
Chunhao Wang

\_\_\_\_\_  
David Huang

\_\_\_\_\_  
Xin Chen

An abstract of a thesis submitted in partial fulfillment of  
the requirements for the degree of Master of Science in the  
Medical Physics Graduate Program of  
Duke Kunshan University

2019

Copyright by  
Chenyang Liu  
2019

## Abstract

**Purpose:** To develop a novel computer-aided diagnosis (CAD) pulmonary nodule detection system that can not only perform real-time detection but also characterize quantitative nodule information based on deep learning methods.

**Method:** We constructed a convolutional neural network (CNN) for automated pulmonary nodule detection and characterization. Nodule detection was accomplished by customizing a detection algorithm (YOLO v3), which comprised of a feature extractor and a bounding box generator. The feature extractor had 19 convolutional layers with 7 residual shortcut connections to extract features on input images at three different down-sampling scales (i.e. 4, 8, and 16). The bounding box generator had 7 convolutional layers to determine the location and size of each detected nodule. A python-based characterization system was then developed to characterize size, diameter, and central coordinates of each detected nodule within the generated bounding box. This characterization system applied a non-maximum suppression algorithm to exclude nodules below true positive probability threshold. The system was trained and validated using ten-fold cross-validation with 300 CT scans from XCAT simulation and 888 patient CT scans from LIDC-IDRI public dataset, separately. System performance was evaluated using Free-Response Receiver Operating Characteristic (FROC) analysis, competition performance metric (CPM) score, as well as precision analysis of central coordinates and diameters.

**Result:** The developed CAD system achieved CPM scores of 0.99 in the simulation image study and 0.873 in the public database study. The average performance time per image was less than 0.1 second. Compared with ground truth data, the detection precision in diameter were 0.26 mm using simulated images and 1.05 mm using public database, while the precision in central coordinate were 0.76 mm and 1.44 mm, respectively.

**Conclusion:** Preliminary evaluation showed that our proposed CAD system using deep learning methods was robust and achieved real-time nodule detection with high accuracy and characterization with high precision.

# Contents

Abstract .....	iv
List of Tables .....	viii
List of Figures .....	ix
Acknowledgements .....	x
Introduction .....	1
1.1 Lung cancer .....	1
1.2 Computer-aided detection and diagnosis.....	1
1.2.1 Conventional CAD systems.....	3
1.2.2 Deep learning CAD system .....	4
1.2.2.1 Pulmonary nodule classification .....	5
1.2.2.2 Pulmonary nodule detection.....	7
1.2.2.3 Pulmonary nodule segmentation.....	8
1.2.3 Luna 16 pulmonary nodule detection competition.....	8
1.2.4 Limitations of current CAD systems .....	10
1.3 Proposed method for pulmonary nodule detection .....	11
1.4 Research Objective.....	12
2 Materials and Methods.....	13
2.1 Data acquisition .....	13
2.1.1 CT images simulation.....	13
2.1.2 Public database CT images acquisition.....	13
2.1.3 Image pre-processing.....	14

2.2	CAD system design.....	14
2.2.1	Nodule detection system.....	16
2.2.1.1	Feature extractor .....	17
2.2.1.2	Bounding box generator .....	19
2.2.2	Characterization system.....	26
2.3	Evaluation method .....	27
3	Results.....	29
3.1	XCAT simulation study results .....	29
3.2	Public database study results.....	31
4	Discussion.....	34
4.1	CAD system design.....	34
4.2	Results discussion.....	35
4.2.1	Results discussion in simulated image and public database .....	35
4.2.2	Results comparison between different architectures .....	36
4.2.3	Results comparison between different CAD systems .....	37
4.3	Limitations and future works .....	41
5	Conclusion.....	43
	References .....	44

## List of Tables

Table 1: Review of conventional CAD systems .....	4
Table 2: Review of nodule classification CAD systems.....	6
Table 3: Review of nodule detection CAD systems .....	7
Table 4: Feature extractor structure .....	18
Table 5: Bounding box structure .....	19
Table 6: Anchor box .....	21
Table 7 Four architectures design.....	26
Table 8 Ten-fold cross-validation results in simulation study .....	29
Table 9 : Performance time, and detection precision in central coordinate and diameter in simulation image study .....	30
Table 10: Ten-fold cross-validation results in public database study.....	31
Table 11: Performance time, and detection precision in central coordinate and diameter in public database study .....	32
Table 12: Architecture comparison.....	34
Table 13: Result comparison in different architectures .....	36
Table 14: Implementation platform comparison .....	40

## List of Figures

Figure 1: Research workflow .....	15
Figure 2: Nodule detection system.....	16
Figure 3: Example of output array.....	20
Figure 4: Anchor box prediction.....	22
Figure 5 Non-maximum suppression .....	26
Figure 6: XCAT phantom simulated image study result example.....	29
Figure 7: Public database study result example .....	32
Figure 8: FROC curve for different architectures .....	37
Figure 9: FROC curve comparison with conventional CAD systems .....	38
Figure 10: FROC curve comparison with 2D deep learning-based CAD systems .....	39
Figure 11: FROC curve comparison with 3D based deep learning CAD systems.....	40
Figure 12: CPM comparison .....	41

## **Acknowledgements**

I would like to express my great appreciation to my supervisor, Dr. Fang-Fang Yin, for his invaluable support of my master thesis study. Also, I would like to thank the rest of my thesis committee: Dr. David Huang, Dr. Chunhao Wang, and Dr. Xin Chen, for their encouragement and insight question. My sincere thanks to Claire Luo for her support. I thank all my classmates, especially the machine learning group, for the day and night we worked together. Last but not least, I would like to express my deepest gratitude to my family.

## **Introduction**

### ***1.1 Lung cancer***

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related fatalities in the world. It accounts for 2.1 million (11.6%) new diagnoses and 1.8 million (18.4%) cancer deaths, respectively.<sup>1</sup> In 2018, over 150,000 people have succumbed to lung cancer, representing about 25 percent of all cancer deaths in the United States (US).<sup>2</sup> The five-year survival rate for lung cancer is 18.6% in the US, which is lower than other leading sites, such as colorectal (64.5%), breast (89.6%), and prostate (98.2%).<sup>3</sup> Based on data from American Cancer Society, the five-year survival rate in localized non-small cell lung cancer (NSCLC) is 60% and 29% in small cell lung cancer (SCLC). However, majority of lung cancer cases are diagnosed in advanced stages (stage III or stage IV), when the five-year survival rate is reduced to less than 10%.<sup>4</sup> Nonobvious symptoms of lung cancer in early stages make diagnosis challenging. Low-dose computed tomography (LDCT) screening among high-risk populations is an effective way of early detection. Studies have found that it can decrease lung cancer mortality by almost 20%.<sup>5,6</sup>

### ***1.2 Computer-aided detection and diagnosis***

LDCT screen aims to detect pulmonary nodule, which is the small rounded or irregular opacity growth inside the lung. Pulmonary nodule can be caused by many reasons, including benign tumors, lung cancers, lung infections, autoimmune disorders, and vascular malformations. LDCT screen present all suspicious nodules. Diagnostic

testing or biopsy may be necessary to determine the type of nodule. Previously, pulmonary nodule detection on CT images heavily relied on manual counting, which is labor-intensive, time-consuming, and inaccurate. Since the 1980s, several studies have been performed to develop techniques for automatic diagnosis of pulmonary nodules, which are known as computer-aided detection and diagnoses (CAD). CAD systems are classified into two categories: computer-aided detection system (CADe) and computer-aided diagnosis system (CADx). The CADe system is responsible for detecting abnormal lesions in medical images. The CADx system provides medical aids by characterizing disease, such as type, severity, stage, or progression.<sup>7,8</sup> Some CAD systems have both detection and diagnosis function.

Developing a CAD system includes five steps: (1) data acquisition, (2) data pre-processing, (3) lung segmentation, (4) nodule detection, and (5) false positive reduction. Some public databases are used in the development of CAD system. One such database called LIDC-IDRI (Lung Image Database Consortium and the Image Database Resource Initiative).<sup>9</sup> Data pre-processing performs techniques (e.g. median filter, morphological hat, enhancement filter, etc.) on the CT images to exclude confusing structures inside the lung. Lung segmentation defines the regions of interest by using methods like thresholding and region growing. Nodule detection is to distinguish between true nodule from vessels, bronchi, and ribs inside regions of interest. Usually, nodule detection is followed by a false positive reduction step, which is a classifier to

distinguish the real nodule from nodule candidates based on the extracted features information.<sup>10</sup>

Some published CAD techniques in recent years are reviewed in the following sections, which are divided into two categories: conventional CAD systems (Section 1.2.1) and deep learning CAD systems (Section 1.2.2).

### **1.2.1 Conventional CAD systems**

Conventional CAD systems are highly dependent on features designing.<sup>11</sup> Nodule candidates are detected by methods such as Hessian matrix<sup>12-14</sup>, Stable 3D Mass-Spring Models<sup>15</sup>, thresholding<sup>16</sup>, and 3D template matching<sup>17</sup>. Features - the useful information that characterize these candidates - are extracted from these candidates for false positive reduction. A summary of conventional CAD systems is shown in Table 1. Although several studies have shown remarkable progress such as high sensitivity<sup>12,14,16,17</sup>, they are subjective to high false positive rates, intermediate results bias, intense image pre-processing, as well as high feature annotation cost, which all impedes their clinical application.

**Table 1: Review of conventional CAD systems**

<i>Author</i>	<i>Year</i>	<i>Method</i>	<i>Database</i>
Choi and Choi <i>et al.</i>	2014	Feature descriptor guided by 3D shape of the object	LIDC
Santos <i>et al.</i>	2014	Gaussian mixture models, Tsallis entropy and SVM	LIDC
Cascio <i>et al.</i>	2012	Stable 3D Mass-Spring Models	LIDC
Chen <i>et al.</i>	2012	an algorithm based on local intensity structure analysis and surface propagation in 3D chest CT images	Private database and LIDC
Ozekes <i>et al.</i>	2008	3D template matching and thresholding based on fuzzy logic	16 exams with 16 nodules in LIDC
Suiyuan and Junfeng	2012	Using shape features to segment pulmonary parenchyma and detect nodule by thresholding and region growing	Private database

### 1.2.2 Deep learning CAD system

Deep learning has attracted great attention in clinical decision supporting approaches due to its encouraging performance. Convolutional neural network (CNN) is one deep learning technique widely used in image analysis.<sup>18</sup> The most fundamental advantage of CNN is automatic features extraction. CNN extracts features – the unique and useful pattern in input images - in every location of input images by using shift-invariant trainable filters. Meanwhile, it exploits interactions and hierarchical relation among features within the deep neural network by multi-layer convolution.<sup>19</sup> It is less dependent on image pre-processing and independent on feature annotation. In both academia and the industry, significant progress has been made to apply deep learning techniques, especially CNN, to nodule classification, detection, and segmentation. There

is a rapid surge of interest in deep learning-based CAD studies. Less than 10 papers were available about deep learning-based pulmonary nodule diagnosis in 2015. However, the numbers of papers increased to over 30 in 2016 and reached over 100 in 2017.<sup>20</sup>

Early deep learning-based studies focused on nodule classification (Section 1.2.2.1) and latest studies incorporated nodule classification into nodule detection (Section 1.2.2.2). Meanwhile, several studies achieved semantic nodule segmentation (Section 1.2.2.3).

#### **1.2.2.1 Pulmonary nodule classification**

Nodule classification is to distinguish between the nodules and non-nodules (N/NN), or more specifically, between malignant and benign nodules (M/B), in either cropped CT image patches or entire CT images. Table 2 summarizes the latest published papers. Hua *et al.* applied 2D CNN for pulmonary nodule classification.<sup>19</sup> Setio *et al.* proposed a 2D multi-view CNN to reduce false positive.<sup>21</sup> The network had a parallel architecture with multiple inputs corresponding to multiple view of candidate CT patch. Gruetzemacher *et al.* designed hierarchical CNN model for malignant nodule diagnosis in three steps: regions of interest generation, nodule classification, and malignancy determination.<sup>22</sup> More recently, Kang *et al.* proposed a multi-view design that cropped CT volume into different sizes that offering different view areas.<sup>23</sup> The generated nodule volumes were then fed into 3D CNN with residual and inception design. Zhu *et al.* used

3D CNN with dual path design to extract deep features. These deep features then were combined with nodule size information and nodule cropped image for classification.<sup>24</sup>

Jin *et al.* used 3D CNN with residual network to function as a false positive reducer in their CAD system.<sup>25</sup>

**Table 2: Review of nodule classification CAD systems**

<i>Authors</i>	<i>Year</i>	<i>Methods</i>	<i>Inputs</i>	<i>Aim</i>
Hua <i>et al.</i>	2015	2D CNN	Cropped CT image	2D N/NN
Setio <i>et al.</i>	2016	2D multi-view CNN	Cropped CT image	2D N/NN
Gruetzemacher <i>et al.</i>	2016	2D hierarchical CNN	Not mentioned	N/NN and M/B
Kang <i>et al.</i>	2017	3D multi-view CNN with residual and inception design	Cropped CT image	3D N/NN and M/B
Zhu <i>et al.</i>	2017	3D CNN with dual path design	Cropped CT image	3D N/NN
Jin <i>et al.</i>	2018	3D CNN with residual design	Cropped CT image	3D N/NN
Annotation:	N/NN:	Nodule or non-nodule		
	M/B:	Malignant or benign		

Some nodule classification methods have shown promising results: classification accuracies reach to 90%<sup>24,25</sup> or have both sensitivities and specificities exceed 90%.<sup>23</sup>

However, classification systems are incapable of localizing nodules within the images.

They primarily serve as false positive reducers – the fifth step in a CAD design – and are not complete CAD systems.

### 1.2.2.2 Pulmonary nodule detection

Several studies have achieved whole image automatic nodule detection by utilizing CNN algorithms in object detection. These CAD systems can be designed in one stage system that direct detects and localizes nodules or two stage system first generates nodule candidates and then proceeds false positive reduction. Table 3 summarizes the latest public papers related nodule detection.

**Table 3: Review of nodule detection CAD systems**

<i>Author</i>	<i>Year</i>	<i>Method</i>	<i>Type</i>
Dou <i>et al.</i>	2017	3D CNN with an online sample filtering scheme to acquire the coordinate of suspicious nodule, following by a Hybrid-loss 3D CNN with residual block for false positive reduction.	Two-stage
Xie <i>et al.</i>	2019	2D Faster R-CNN with two region proposal network and one deconvolutional layer.	One-stage
Ding <i>et al.</i>	2017	2D Faster R-CNN with three neighboring axial slices as input, following by a 3D CNN for false positive reduction.	Two-stage
Zhu <i>et al.</i>	2017	3D Faster R-CNN with U-net-like encoder-decoder design.	Two-stage

Dou *et al.* used 3D CNN with an online sample filtering scheme to acquire the coordinates of suspicious nodules, following by a Hybrid-loss 3D CNN with residual block for false positive reduction.<sup>26</sup> Faster regional-based convolutional neural network (faster R-CNN), proposed by Ren *et al.*, is one of the most widely used object detection algorithms.<sup>27</sup> It designed a region proposal network for generating regions of interest and then applied precise object localization and classification to these regions of interest.

Many faster R-CNN-based CAD systems are available now. Ding *et al.* used 2D faster R-CNN as nodule candidate generator, which combined three neighbouring slices in the axial direction as inputs.<sup>28</sup> With generated nodule candidate, a 3D CNN that had six convolutional layers was used for false positive reduction. Xie *et al.* also used 2D faster R-CNN for detection.<sup>29</sup> Its faster R-CNN structure has two region proposal networks and one deconvolutional layer. Zhu *et al.* used 3D faster R-CNN with a U-net encoder-decoder structure for nodule detection.<sup>24</sup>

### 1.2.2.3 Pulmonary nodule segmentation

Nodule segmentation aims to achieve pixel-wised segmentation that provides morphological information. Ronneberger *et al.* proposed encoder-decoder U-net shape CNN for image segmentation.<sup>30</sup> Lan *et al.* combined U-Net and residual network for nodule segmentation, which achieved dice coefficient of 71.9%.<sup>31</sup> Similarly, Nam *et al.* used CT patches to achieve mean segmentation dice coefficient of 78.78%.<sup>32</sup>

### 1.2.3 Luna 16 pulmonary nodule detection competition

LUNA 16 is a CAD competition that use LIDC-IDRI database.<sup>33</sup> Since this study also used this database and the result of our proposed CAD system was compared with some CAD systems submitted to this competition, a brief introduction of these systems are as follows:

**ZNET:** ZNET used U-net for every axial slice. The nodule candidates were generated based on probability map from U-net output. Candidate masks were obtained

by thresholding. Morphological erosion was applied to candidate mask to remove partial volume effects. With nodule candidate generated, a 3D residual network was used for false positive reduction.

**Aidence:** Aidence was a company that aiming for developing computer-aided diagnosis tools for radiologists based on deep learning (<http://aidence.com/>). The training database was National Lung Screening Trial (NLST) and LIDC-IDRI database was only used for validation.

**Jianpei CAD:** JianPei was a CAD system developed by Hangzhou JianPei Technology company (<http://www.jianpeicn.com>). They proposed a two-stage nodule detection framework. 3D U-net and 2D U-net were both used for candidate detection. The inputs were 2D cropped axial slices (128x128) and 3D cropped CT volume (128x128x128). A deep 3D residual CNN was used for false positive reduction. The detection system was training on their in-house resource and LIDC-IDRI dataset was only used for validation.

**MOT\_M5L:** the multi-opening and Thresholding (MOT) was a CAD system to be combined into M5L system.<sup>34</sup> 3D region growing was used to segment lung volume. Multiple grey level thresholding and morphological processing were used to detect candidate nodule.<sup>35</sup> Candidate nodule was separated from vascular structures based on segmentation method.<sup>36</sup>

**VISIACTLung:** an FDA approved commercially available lung CAD system (MeVis Medical Solutions AG, Bremen, Germany)

**ETROCAD:** a CAD system that first segmented nodule based on nodule and vessel enhancement filters and then located the center of nodule by computing divergence feature.<sup>37</sup>

**PAtech:** This system was developed by Ping An Technology Company. It designed a 3D CNN with feature pyramid architecture for nodule detection and a following 3D CNN for false positive reduction. The inputs of the training set were 128x128x128 cropped images.

**Resnet:** a CAD system developed by Dou et al.<sup>26</sup>

**zhongliu\_xie:** a CAD system developed by Zhongliu Xie (Imperial College London). It used 3D region proposal U-net with dense and residual learning for lung nodule detection.

#### **1.2.4 Limitations of current CAD systems**

Although previous studies have proposed feasible methods for nodule detection and have achieved promising results. They still have the following limitations:

- (1) High false positive rate: some CAD systems have high false rate to reach an acceptable sensitivity;
- (2) Requirement of false positive reduction: some two-stage detection system need a false positive reduction design to reduce false positive rate.

- (3) Intense image processing: some CAD systems require image pre-processing including lung segmentation and regions of interest generation.
- (4) High feature annotation costs.
- (5) High computation cost: some CAD systems were designed as deep 3D CNN, which is low efficient and occupies computation space. Furthermore, few papers mentioned their CAD systems' capabilities in real-time detection.
- (6) Complex two-stage detection process: some CAD systems need both candidate generation and false positive reduction to guarantee the detection accuracy.
- (7) Fail to characterize quantitative nodule information, such as number, diameter, and location of nodules.

### **1.3 Proposed method for pulmonary nodule detection**

Redmon *et al.* designed an object detection algorithm called You Only Look Once (YOLO).<sup>38</sup> Contrary to faster R-CNN, YOLO combined regions of interest generation, object localizations, and classification into one step, which achieve one-stage detection with low false positive rate. YOLO had 24 convolutional layers and 2 fully connected layers. This algorithm divided image into grids and then predicted bounding boxes and class probability for each grid cell. Ramachandran *et al.* have achieved real-time nodule detection and localization by using YOLO-based deep learning model.<sup>39</sup> This model achieved a sensitivity of 89% with six false positive per image, which was comparable to

the result reported by recent studies. Later, Redmon and Farhadi forwarded a new version of YOLO algorithm that modified the method of predicting bounding box.<sup>40</sup> Instead of direct predicting bounding boxes, the algorithm will predict the offsets of predetermined sets of anchor boxes. The latest version, YOLO v3, presented by Redmon and Farhadi, incorporated the residual block into network.<sup>41</sup> Furthermore, YOLO v3 makes prediction across three different scales, which improved capabilities in detecting small object. However, no study has been performed to evaluate the performance of YOLO v3 in pulmonary nodule detection, which has the potential to achieve high accuracy one-stage real-time pulmonary nodule detection.

#### **1.4 Research Objective**

This study aims to develop a novel one-stage CAD system that can achieve automatic real-time pulmonary nodule detection in high sensitivity with minimized false positive rate, minimized image pre-processing requirement, and minimized computation cost. More specifically, this study aims to customize YOLO v3 algorithm for pulmonary nodule detection. Meanwhile this study aims to develop a characterization system that can be integrated to the CAD system for generating quantitative nodule information (i.e. number of nodule, size of nodules, and location).

## **2 Materials and Methods**

### **2.1 Data acquisition**

#### **2.1.1 CT images simulation**

300 transverse CT scans with nodules in different diameters and locations were simulated by using Cardiac-torso (XCAT) digital phantoms. XCAT phantom can produce realistic 3D images with detailed whole-body anatomical information.<sup>42</sup> Simulation image resolution is 512x512 pixels and pixel width is 0.836 mm. Nodules were designed as a spherical shape because of well-circumscribed appearance of lung nodule.<sup>43</sup> Nodules were randomly located within the lung tissue. The sizes of nodules ranged from 6mm to 20 mm with 1 mm interval, which was based on the study that only 1% of the nodules are smaller than 5 mm and 80% of the nodules with a diameter up to 20 mm.<sup>44</sup>

#### **2.1.2 Public database CT images acquisition**

The public database CT images came from LIDC-IDRI. 888 CT scans were included in this study based on the criteria that scan slice thickness no greater than 2.5 mm. Each CT scan has nodule annotation information from four experienced radiologists. 1186 nodules that marked by at least three out of four radiologists were regarded as positive examples.

### **2.1.3 Image pre-processing**

Image pre-processing was minimized to keep the simplicity of the CAD system. All CT images pixel intensity value were rescaled into its Hounsfield Unit (HU) value. Pixels that have HU value larger than 1000 HU will be reassigned HU value equal to 1000, while values outside the patient will be set to -1025 HU. All pixel intensities were normalized into the range of zeros to one, where zero stood for HU values equal to -1025 and one stood for HU values equal to 1000.

2D transverse CT slices containing the central slices of these nodules were extracted to form as dataset, which was then divided into ten subsets for ten-fold cross-validation.

## **2.2 CAD system design**

This CAD system was constructed by two subsystems: a nodule detection system and a characterization system. The nodule detection system was based on an object detection algorithm YOLO v3. It was comprised by a feature extractor for identifying nodules and a bounding box generator to determine the size of the detected nodule. Section 2.2.1 gives a detailed explanation of nodule detection system. A following characterization system took the generated nodule bounding boxes as input and characterized all detect nodules based on lesion number, size, diameter, and central coordinates. Section 2.2.2 gives a detailed explanation of the characterization system.

The overall workflow of designing this CAD system can be seen in Figure 1. Data processing provided training data, validation data, and testing data to the nodule detection system. The detected nodule in nodule detection system then went through characterization for result evaluation.

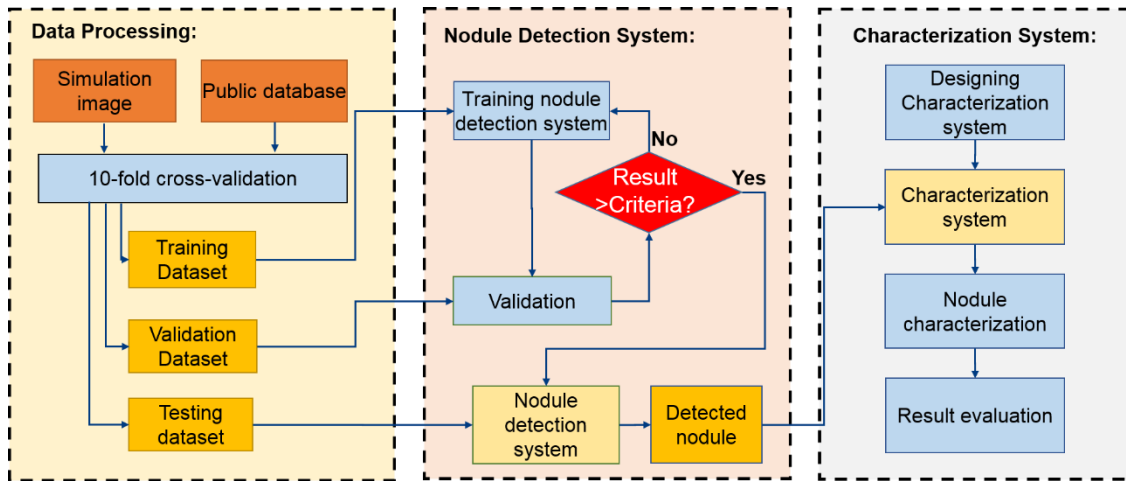


Figure 1: Research workflow

## 2.2.1 Nodule detection system

Nodule detection system contained two parts: feature extractor and bounding box generator, which is displayed in Figure 2.

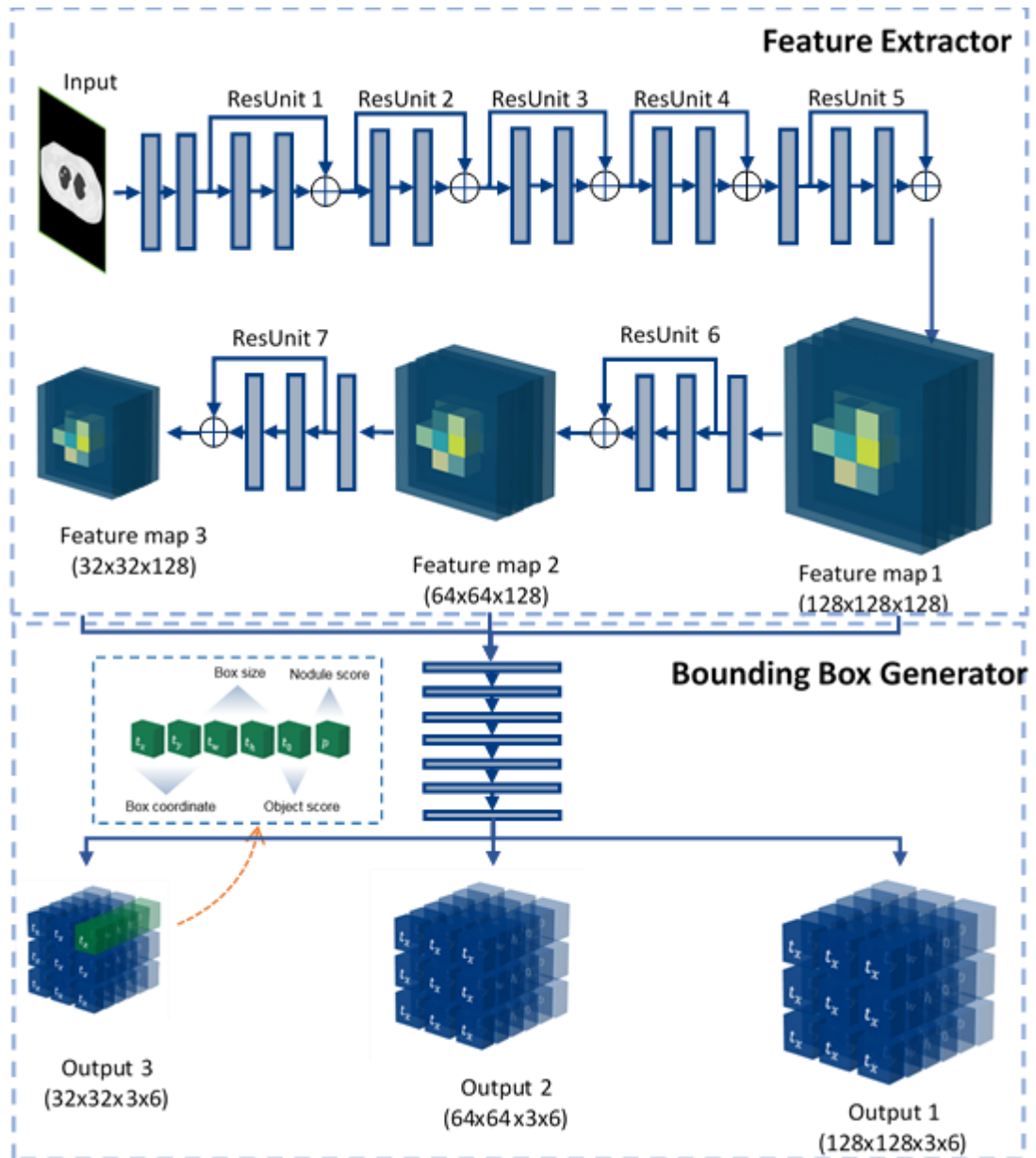


Figure 2: Nodule detection system

### 2.2.1.1 Feature extractor

The feature extractor was a residual network that contained seven residual units (ResUnit). The detailed structure of feature extractor can be seen in Table 4. The residual network, proposed by He *et al.*, had a shortcut operation that connected two non-adjacent layers.<sup>45</sup> The output of shallower layer skipped one middle layer and was summed with the output of the deeper layer. Residual network addressed vanishing or exploding gradients problem in the deep neural network. One ResUnit had three or two convolutional layers and a shortcut operation. Three-layer ResUnit has an extra convolutional layer to down-sample the inputs by setting stride equals to two. In each convolutional layer, the activation function used was leaky Relu (L-Relu). The slope of L-Relu in negative part was set to 0.1. Each convolutional layer followed by a batch normalization operation. The shortcut operation in this network was after Relu activation function, which was different from the original residual network. Padding mode in each convolutional layer was same or valid. Same mode was used in the layers that have stride equals to one to keep the image dimension. Valid mode, which means no padding applied, was used in layers that have stride equals to two to avoid introducing useless boarder information during down-sampling.

The feature extractor extracted features in three scales. The input images were down-sampled by a factor of 4, 8 or 16. Correspondingly, the three outputs of feature extractor divided the input image into grid with dimensions 128×128, 64×64, or 32×32.

Feature maps grid cells in the same position are responsible for predicting the nodule located within the corresponding grid cell in the original image.

**Table 4: Feature extractor structure**

	Layer index	Type	Filters	Size	Stride	Padding mode	Activation	Output
	1	Convolution	32	3x3	1	Same	L-Relu	512x512x32
ResUnit 1	2	Convolution	64	3x3	2	Valid	L-Relu	256x256x64
	3	Convolution	32	1x1	1	Same	L-Relu	256x256x32
	4	Convolution	64	3x3	1	Same	L-Relu	256x256x64
	5	Residual: Adding layer 2 and layer 4					-	256x256x64
ResUnit 2	6	Convolution	32	1x1	1	Same	L-Relu	256x256x32
	7	Convolution	64	3x3	1	Same	L-Relu	256x256x64
	8	Residual: Adding layer 5 and layer 7					-	256x256x64
ResUnit 3	9	Convolution	32	1x1	1	Same	L-Relu	256x256x32
	10	Convolution	64	3x3	1	Same	L-Relu	256x256x64
	11	Residual: Adding layer 8 and layer 10					-	256x256x64
ResUnit 4	12	Convolution	32	1x1	1	Same	L-Relu	256x256x32
	13	Convolution	64	3x3	1	Same	L-Relu	256x256x64
	14	Residual: Adding layer 11 and layer 14					-	256x256x64
ResUnit 5	15	Convolution	128	3x3	2	Valid	L-Relu	128x128x128
	16	Convolution	64	1x1	1	Same	L-Relu	128x128x64
	17	Convolution	128	3x3	1	Same	L-Relu	128x128x128
	18	Residual: Adding layer 15 and layer 17					-	128x128x128
ResUnit 6	19	Convolution	128	3x3	2	Valid	L-Relu	64x64x128
	20	Convolution	64	1x1	1	Same	L-Relu	64x64x64
	21	Convolution	128	3x3	1	Same	L-Relu	64x64x128
	22	Residual: Adding layer 19 and layer 21					-	64x64x128
ResUnit 7	23	Convolution	128	3x3	2	Valid	L-Relu	32x32x128
	24	Convolution	64	1x1	1	Same	L-Relu	32x32x64
	25	Convolution	128	3x3	1	Same	L-Relu	32x32x128
	26	Residual: Adding layer 23 and layer 25					-	32x32x128

### 2.2.1.2 Bounding box generator

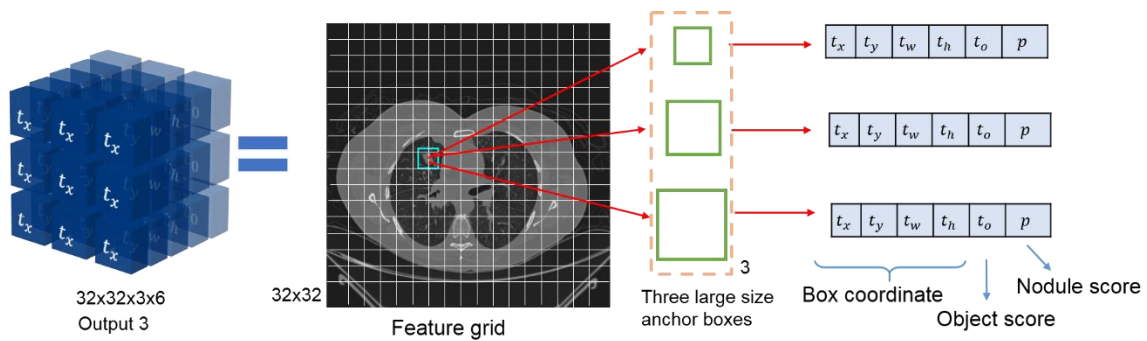
The precise location of the nodule is determined by the bounding box generator.

Bounding box is a rectangular that can contour the nodule in the image. The bounding box generator was composed of seven convolutional layers and one reshape layer. The structure of bounding box generator is shown in Table 5.

**Table 5: Bounding box structure**

Layer index	Type	Filters	Size	Stride	Padding mode	Activation	Output	
1	Convolution	128	1x1	1	Same	L-Relu	?x?x128	
2	Convolution	256	3x3	1	Same	L-Relu	?x?x256	
3	Convolution	128	1x1	1	Same	L-Relu	?x?x128	
4	Convolution	256	3x3	1	Same	L-Relu	?x?x256	
5	Convolution	128	1x1	1	Same	L-Relu	?x?x128	
6	Convolution	256	3x3	1	Same	L-Relu	?x?x256	
7	Convolution	18	1x1	1	Same	L-Relu	?x?x18	
8	Reshape	Reshape the array's last dimension form 18 to 3x6						?x?x3x6

The outputs of the bounding box generator were three four-dimension arrays with a size of  $128 \times 128 \times 3 \times 6$ ,  $64 \times 64 \times 3 \times 6$ , or  $32 \times 32 \times 3 \times 6$ . The first two dimensions are those of the feature grid. The third dimension stands for the three anchors boxes in each grid cell. The fourth dimension stores the bounding boxes information. An illustration of output 3 array is shown in Figure 3.



**Figure 3: Example of output array**

#### 2.2.1.2.1 Anchor box

Anchor boxes stands for the pre-defined bounding boxes that locate in upper left corner of every grid cell. Most of the pulmonary nodules have a similar shape and size: well-circumscribed and relatively fixed diameter ranges (0.3 cm to 3 cm). For this reason, in this study, nine anchor boxes in three size scales were determined, which are shown in Table 6. The varied length in widths and heights can cover all nodule's morphological variation. Grids with high resolution small size anchor boxes because they were used to detect nodule with small size. Thus, small, medium, and large size of anchor boxes correspond to  $128 \times 128$ ,  $64 \times 64$ , and  $32 \times 32$  grids, respectively.

**Table 6: Anchor box**

	Anchor box index	Heights (pixels)	Width (pixels)
Small Size (128x128 grid)	1	8	8
	2	10	10
	3	12	12
Medium Size (64x64 grid)	4	15	15
	5	20	20
	6	25	25
Large size (32x32 grid)	7	30	30
	8	35	35
	9	40	40

### 2.2.1.2.2 Bounding box prediction

The bounding box information that reflects nodules' coordinates and diameters is stored in the fourth dimension of the array. It contains six trainable elements, namely  $t_x$ ,  $t_y$ ,  $t_h$ ,  $t_w$ ,  $t_o$ , and  $p$ . As demonstrated in Figure 4, the training process is to let the detection system learn to shift the anchor box to fix the nodule into a box.

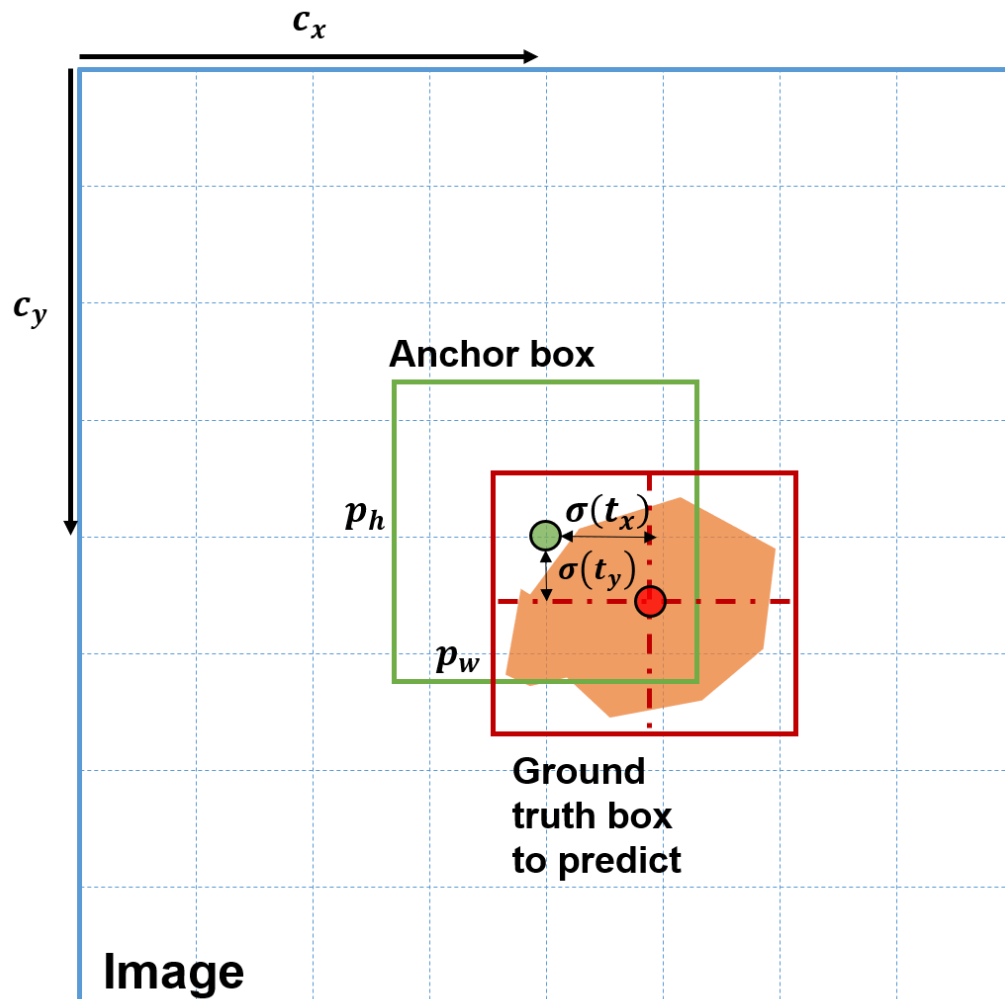


Figure 4: Anchor box prediction

$t_x$  and  $t_y$  are central coordinate shifts from anchor box to predicted box. Thus, the central coordinates of the nodule are calculated by equation (1) and equation (2).

$$b_x = \sigma(t_x) + c_x \quad (1)$$

$$b_y = \sigma(t_y) + c_y \quad (2)$$

Here, sigmoid activation is used to constraint the output value ( $t_x$  and  $t_y$ ) between 0 to 1.  $c_x$  and  $c_y$  are the offset from the top left corner of the grid. Using offsets instead of absolute coordinates will reduce the training complexity.<sup>40</sup>

$t_h$  and  $t_w$  are height and width factors that correspond to the change of height and width between anchor box and predicted box. Thus, the predicted box's height and width are expressed in equation (3) and equation (4).

$$b_w = p_w \cdot e^{t_w} \quad (3)$$

$$b_h = p_h \cdot e^{t_h} \quad (4)$$

Here,  $p_w$  and  $p_h$  are the width and height of the anchor box.

$t_o$  stands for object score: it is a binary value that one indicates presence of an object inside the grid cell and zero indicate absence.  $p_i$  stands for the probability that the object inside the grid cell belong to object category  $i$ . In this study, only one category is used: nodule. Number of  $p$  value equals to number of detection category. Thus, only one  $p$  value is stored in the final output. The confident score ( $P$ ) is determiend by equation (5), which reflects the detection systems' confidence of detection result.

$$P = t_o \cdot p_i \quad (6)$$

### 2.2.1.2.3 Loss computation

The loss function contains three parts: (1) object loss; (2) classification loss (3) bounding box loss, as shown in equation (7).

$$L = L_{object} + L_{classification} + L_{coordinate} \quad (7)$$

The loss calculation methods are binary cross entropy (BCE) and square error (SE), which are defined in equation (8) and (9), respectively.

$$BCE(x, \hat{x}) = -\hat{x} \cdot \log(x) - (1 - \hat{x})\log(1 - x) \quad (8)$$

$$SE(x, \hat{x}) = (x - \hat{x})^2 \quad (9)$$

Object loss accounts for object prediction loss. It both calculates the existence object loss and the non-existence object loss, which is defined in equation (10).

$$L_{objectness} = \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{i,j}^{obj} [-\log(t_o)] + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{i,j}^{non-obj} [-\log(1 - t_o)] \quad (10)$$

Classification loss calculates the error in predicting the object's category, as shown in equation (11). Only one class category is used in this study, this loss equation is reserved for future multi-category study.

$$L_{classification} = \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{i,j}^{obj} \left[ \sum_{c \in classes} BCE(p(c), \hat{p}(c)) \right] \quad (11)$$

Coordinate loss calculates the error in predicting bounding box. It compares the predicted bounding box coordinate  $(t_x, t_y, t_h, t_w)$  with ground truth bounding box  $(\hat{t}_x, \hat{t}_y, \hat{t}_h, \hat{t}_w)$ , which is shown in equation (12).

$$L_{coordinate} = \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{i,j}^{obj} \{BCE(t_x, \hat{t}_x) + BCE(t_y, \hat{t}_y) + SE(t_h, \hat{t}_h) + SE(t_w, \hat{t}_w)\} \quad (12)$$

For equation abovementioned, the notations are shown in the table below.

<i>Notation</i>	<i>Meaning</i>
$t_x, t_y$	predicted central location coordinate of bounding box
$t_h, t_w$	predicted bounding box height and width
$t_o$	predicted object score
$p(c)$	predicted bounding box category
$\lambda_{box\ loss\ scale}$	weighting factor of coordinate prediction loss
$\mathbb{1}_{i,j}^{obj}$	whether $j$ th anchor box in cell $i$ has object, it is defined as $\begin{cases} 1 & \text{if object exist} \\ 0 & \text{if object dose not exist} \end{cases}$
$\mathbb{1}_{i,j}^{non-obj}$	is denoted as whether $j$ th anchor box in cell $i$ has not object, it is defined as $\begin{cases} 0 & \text{if object exist} \\ 1 & \text{if object dose not exist} \end{cases}$ . This is determined by a threshold value. Only predicted boxes which have an IoU of less than the threshold with any object are considered in the non-object loss
$(\hat{x})$	The symbol $(\hat{x})$ refers to the corresponding ground truth value of predicted value ( $x$ )

#### 2.2.1.2.4 The influence of number of ResUnits

To explore the influence of the number of ResUnits in nodule detection accuracy. Four architectures with different number of ResUnits in feature extractor were designed, which are shown in Table 7. Due to the computation memory limitation of the graphic card used in this study, the maximum number of ResUnits was set to seven. These four architectures were validated with same dataset in public database.

**Table 7 Four architectures design**

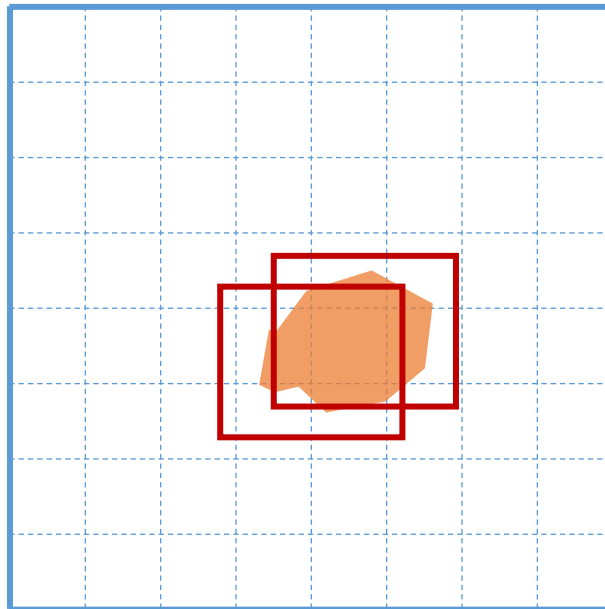
	Number of ResUnits	Number of convolutional layers
Architecture 1	ResUnit 1, 2, 3, 4, 5, 6, and 7.	19
Architecture 2	ResUnit 1, 2, 3, 4, 6, and 7.	16
Architecture 3	ResUnit 1, 2, 3, 6, and 7.	14
Architecture 4	ResUnit 1, 2, 6, and 7.	12

#### **2.2.1.2.5 Implementation details**

The training was based on Keras with TensorFlow as backend. Each training has 500 epochs and training time is around 15 hours.

#### **2.2.2 Characterization system**

The characterization system applied non-maximum suppression algorithm to filter out nodules with low confident scores.<sup>46</sup> As illustrated in Figure 5, when several



**Figure 5 Non-maximum suppression**

bounding boxes detect the same nodule simultaneously, on the bounding box that has higher confident score is kept.

This characterization system gives a summarized report containing nodules' quantitative information. The trained nodule detection system model is loaded in characterization system. Input images pass through the detection model and generate the output that containing detected nodules. A confident threshold value is set to filter out nodules with confident score lower than this value. The quantitative information was extracted from the output, includes nodules' number, central point coordinate, diameter, and the confident score. The final report is a summary of nodules within the input images

### **2.3 Evaluation method**

The accuracy of the nodule detection system was evaluated by Free-Response Receiver Operating Characteristic (FROC) curve and competition performance matrix (CPM). FROC curve plots sensitivity against false positive detection per image (FPPI) at various threshold setting.<sup>47</sup> Sensitivity, or true positive rate, is defined in equation (13), which is the number of true positives (TP) nodules detection divided by total number of ground truth nodules should be detected (sum of true positive detection with false negative).

$$sensitivity = \frac{TP}{All\ ground\ truths} = \frac{TP}{TP + FN} \quad (13)$$

FPPI is defined in equation (14), which is the sum of false positive detection in each image ( $FPD_i$ ) divided by total number of images.

$$FPPI = \frac{\sum_i^n FPD_i}{n} \quad (14)$$

CPM is defined as averaged sensitivity at seven FPPI values: 1/8, 1/4, 1/2, 1, 2, 4, and 8.

Average mean square error was applied to evaluate the detection precision in diameter, which can be seen in equation (15).

$$Error_d = \frac{1}{n} \sum_i^n \sqrt{(d_i - \hat{d})^2} \quad (15)$$

Here  $d_i$  stands for predict diameter,  $\hat{d}$  stands for ground truth diameter, and n stands for total number of nodules. Central coordinate precision was evaluated by equation (16).

$$Error_c = \frac{1}{n} \sum_i^n \sqrt{(x_i - \hat{x})^2 + (y_i - \hat{y})^2} \quad (16)$$

Here  $(x_i, y_i)$  stands for predict coordinate,  $(\hat{x}, \hat{y})$  stands for ground truth coordinate, and n stands for total number of nodules.

### 3 Results

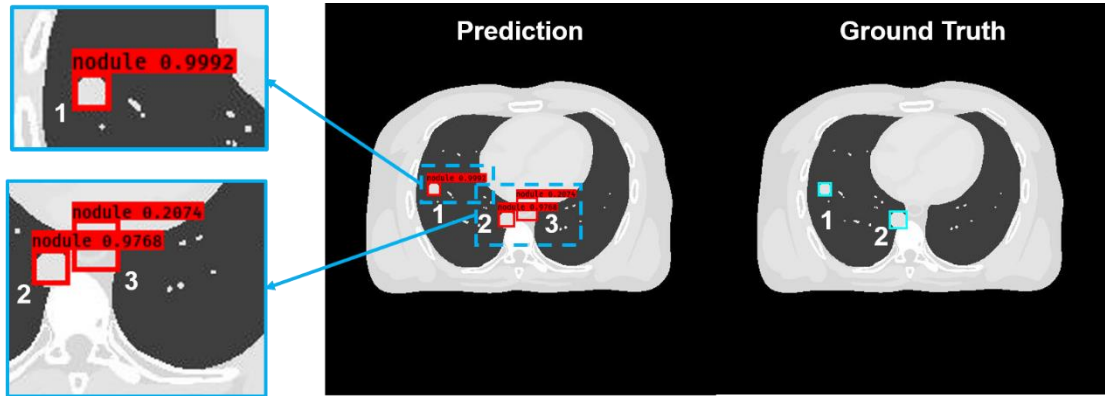
#### 3.1 XCAT simulation study results

The ten-fold cross-validation results using XCAT simulated images is displayed in Table 8. The average CPM scores is 0.99.

**Table 8 Ten-fold cross-validation results in simulation study**

	FROC analysis							CPM score
	1/8	1/4	1/2	1	2	4	8	
<b>validation 1</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<b>validation 2</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<b>validation 3</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<b>validation 4</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<b>validation 5</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<b>validation 6</b>	0.97	0.98	0.98	0.98	1.00	1.00	1.00	0.99
<b>validation 7</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<b>validation 8</b>	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.99
<b>validation 9</b>	0.98	0.98	0.98	0.98	0.98	1.00	1.00	0.99
<b>validation 10</b>	0.96	0.98	0.98	0.98	0.98	1.00	1.00	0.98
<b>average</b>	0.99	0.99	0.99	0.99	1.00	1.00	1.00	<b>0.99</b>

Examples of detection results from CAD systems is shown in Figure 6. This image has two ground truth nodules, as indicated in two blue boxes on the right image.



**Figure 6: XCAT phantom simulated image study result example**

Three nodules were detected, as indicated in the red box on the left. Nodule 1 and nodule 2 are ground truth nodules. The automatic generated confident scores are attached to each detected bounding box. It can be seen that two ground truth nodules have high confident scores (0.9992 and 0.9768) and the false positive detection has low confident score (0.2074).

The ten-fold cross-validation results in performance time, and detection precision in central coordinate and diameter in simulation image study is shown in Table 9. The average detection time per image is 0.070 second. The average precision in central coordinates is 0.759 mm and the precision in diameter is 0.261 mm.

**Table 9 : Performance time, and detection precision in central coordinate and diameter in simulation image study**

	Time per Image (s)	Precision in central coordinates (mm)	Precision in diameters (mm)
<b>validation 1</b>	0.070	0.799	0.331
<b>validation 2</b>	0.070	0.872	0.340
<b>validation 3</b>	0.066	0.734	0.292
<b>validation 4</b>	0.070	0.772	0.204
<b>validation 5</b>	0.070	0.710	0.219
<b>validation 6</b>	0.070	0.678	0.198
<b>validation 7</b>	0.070	0.703	0.238
<b>validation 8</b>	0.070	0.812	0.305
<b>validation 9</b>	0.070	0.767	0.218
<b>validation 10</b>	0.070	0.744	0.268
<b>average</b>	0.070	0.759	0.261

### 3.2 Public database study results

The detection performance on the public database was validated using ten-fold cross-validation. FROC analysis and CPM score results are shown in Table 10. The average CPM score is 0.873. Table 11 shows the ten-fold cross-validation results in performance time, and detection precision in central coordinate and diameter. The average detection time per image is 0.077 second. The average precision in central coordinates is 1.438 mm and the precision in diameter is 1.049 mm.

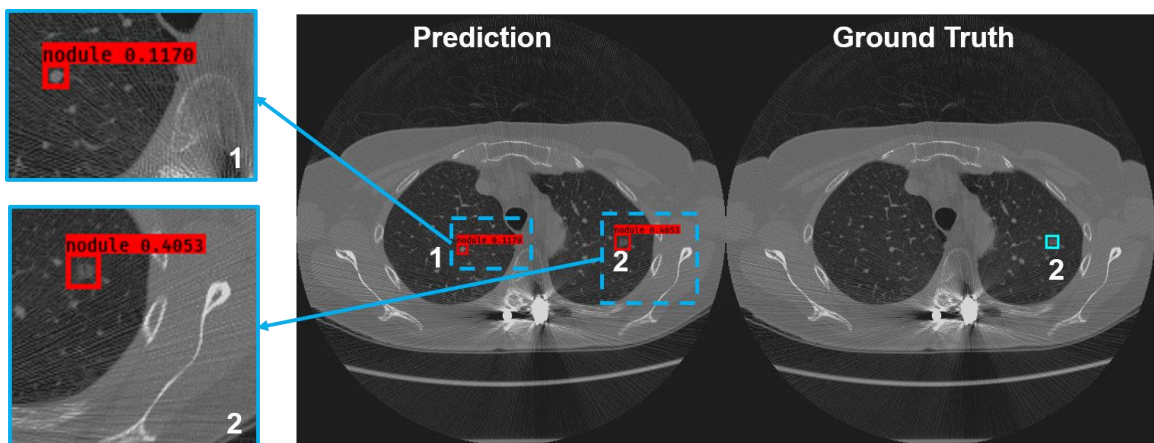
**Table 10: Ten-fold cross-validation results in public database study**

	FROC analysis							CPM score
	1/8	1/4	1/2	1	2	4	8	
<b>validation 1</b>	0.760	0.846	0.882	0.911	0.960	0.976	0.984	0.903
<b>validation 2</b>	0.626	0.799	0.889	0.918	0.952	0.960	0.968	0.873
<b>validation 3</b>	0.681	0.755	0.833	0.895	0.933	0.959	0.976	0.862
<b>validation 4</b>	0.681	0.773	0.850	0.869	0.921	0.931	0.961	0.855
<b>validation 5</b>	0.657	0.768	0.825	0.880	0.919	0.951	0.992	0.856
<b>validation 6</b>	0.638	0.745	0.852	0.880	0.902	0.934	0.951	0.843
<b>validation 7</b>	0.689	0.827	0.861	0.900	0.919	0.932	0.968	0.871
<b>validation 8</b>	0.719	0.797	0.846	0.921	0.960	0.991	1.000	0.890
<b>validation 9</b>	0.761	0.832	0.899	0.927	0.963	0.974	1.000	0.908
<b>validation 10</b>	0.689	0.827	0.861	0.900	0.919	0.932	0.968	0.871
<b>average</b>	0.690	0.797	0.860	0.900	0.935	0.954	0.977	<b>0.873</b>

**Table 11: Performance time, and detection precision in central coordinate and diameter in public database study**

	Time per Image (s)	Precision in central coordinates (mm)	Precision in diameters (mm)
<b>validation 1</b>	0.077	1.435	1.100
<b>validation 2</b>	0.077	1.541	0.930
<b>validation 3</b>	0.077	1.323	1.198
<b>validation 4</b>	0.078	1.680	1.121
<b>validation 5</b>	0.077	1.398	1.099
<b>validation 6</b>	0.077	1.460	1.002
<b>validation 7</b>	0.077	1.449	1.087
<b>validation 8</b>	0.077	1.236	0.912
<b>validation 9</b>	0.079	1.492	1.145
<b>validation 10</b>	0.078	1.362	0.896
<b>average</b>	0.077	1.438	1.049

Figure 7 shows an example of detection result in the public database. This image has one ground truth nodule, as indicated in the blue box on the right image. Two nodules were detected, as indicated in red box on the left. Nodule 2 is ground truth



**Figure 7: Public database study result example**

nodule. It is accurately detected with high confident score (0.4053). Nodule 1 is false positive detection that has low confident score (0.1170).

## 4 Discussion

### 4.1 CAD system design

As mentioned in the introduction, popular CNN architectures such as faster R-CNN and U-Net have demonstrated their capabilities in nodule detection. Studies has shown YOLO v3 has better detection accuracy when compared with faster R-CNN and U-Net. However, no study has been performed to evaluate the performance of YOLO v3 in nodule detection. The results of this study demonstrated that YOLO v3 architecture was competent in nodule detection. Detailed performance evaluation is discussed in Section 4.2.

Multiple adjustments for YOLO v3 have been made. A summary of modification is shown in Table 12.

**Table 12: Architecture comparison**

	<i>Original YOLO v3</i>	<i>Our CAD system</i>
Feature extractor design	Darknet (53 convolutional layers)	7 ResUnits (26 convolutional layers)
Total number of layers	106	50
Up sampling	Yes	No
Output dimension	16x16; 32x32; 64x64	32x32; 64x64; 128x128

The original YOLO v3 algorithm was designed for multi-classes detection. It had over 100 convolutional layers to meet the requirement of extracting a high variety of features. Such a deep convolutional neural network is not necessary for pulmonary nodule detection as it only focuses on a single object. As such, in this study, the number

of convolutional layers in the feature extractor was reduced to 26; total number of layers, correspondingly, was reduced to 50. In this way, the implementation requirement of the algorithm is reduced. Simplifying the algorithm's structure makes applying the algorithm in clinic more feasible. A low-performance GPU, such as GTX1060Ti in this experiment, is sufficient to load the algorithm for prediction. Meanwhile, the algorithm was designed to extract feature in two dimensions instead of in three dimensions in order to keep the real-time efficiency in low-performance GPU. Furthermore, the original YOLO v3 was designed for detecting both large and small objects. For these studies, the algorithm was customized to only focus on detecting objects that have a reasonable size (from several millimetres to several centimeter). One such change was to increase the output dimension. Another change was adjusting anchor boxes to the sizes specified in Table 6.

## **4.2 Results discussion**

### **4.2.1 Results discussion in simulated image and public database**

Results demonstrated that this CAD system achieved promising performance in both simulated images and public database images. The reason of choosing FROC curve and CPM score for detection accuracy evaluation is because false positive detection cannot be fully eliminated for every detection model. Using FROC curve and CPM score can fully evaluate the performance in both low threshold setting and high threshold setting. Specifically, CAD systems had better performance in the simulated images. One

reason is the simplicity of phantom image: the well-circumscribed shape of simulated nodules, no noise added to images, and homogeneous lung tissue. The function of the simulated image study is to examine the reliability of the detection system because nodules are all pre-defined with ground truth indication. The high precision in predicting diameter and central coordinates demonstrated this CAD system's ability to precisely localize nodule. In public database study, although the average errors were slightly higher, the accuracy was still clinically acceptable. The detection model trained by XCAT simulated data has poor performance in detecting public data patient image. This is also caused by the simplicity of the simulated nodules.

#### 4.2.2 Results comparison between different architectures

Detection results with different architectures specified in Table 7 are shown in Table 13 and visualized in Figure 7. Result shows that the detection accuracy increases with number of ResUnit increases. Further work is required to explore the relationship between number of ResUnits and detection accuracy.

**Table 13: Result comparison in different architectures**

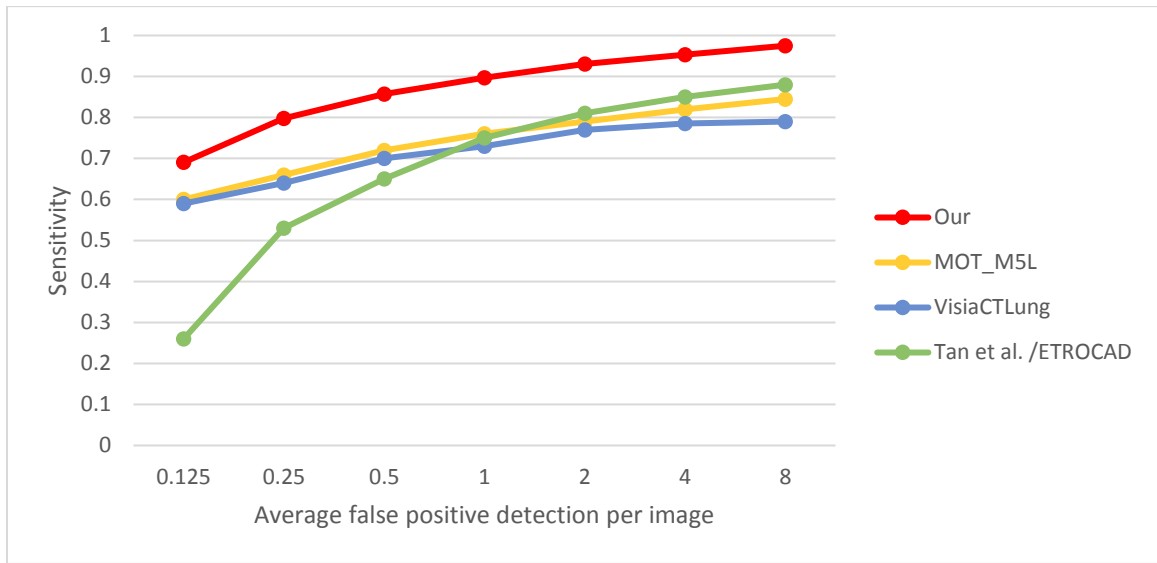
	False positive per image							CPM
	0.125	0.25	0.5	1	2	4	8	
Architecture 1	0.760	0.846	0.882	0.911	0.960	0.976	0.984	0.903
Architecture 2	0.778	0.838	0.887	0.921	0.936	0.952	0.952	0.895
Architecture 3	0.717	0.825	0.860	0.866	0.897	0.935	0.972	0.867
Architecture 4	0.771	0.801	0.856	0.879	0.879	0.890	0.953	0.861



**Figure 8: FROC curve for different architectures**

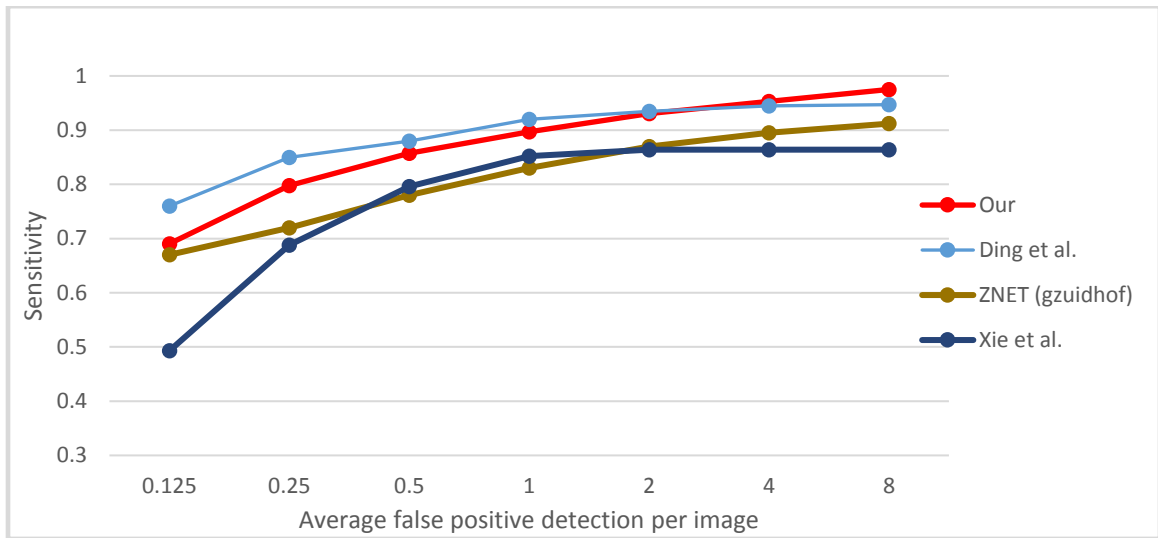
### 4.2.3 Results comparison between different CAD systems

The CAD system’s detection performance in public database was compared with other CAD systems that used the same database. Figure 9 shows the FROC curves among our CAD system and other conventional CAD systems that submitted on LUNA 16 challenge. It shows that our CAD system accomplished the best performance.



**Figure 9: FROC curve comparison with conventional CAD systems**

Our detection result was also compared with other deep learning-based CAD systems. Figure 10 shows the comparison between our system with other 2D CNN systems. Our CAD system has better performance than CAD systems using 2D faster R-CNN Xie, Yang, Sun, Chen, Zhang<sup>29</sup> and 2D U-Net (ZNET). It also gets comparable results with two-stage CAD systems<sup>28</sup> that has a 3D CNN architecture for false positive reduction after nodule detection.



**Figure 10: FROC curve comparison with 2D deep learning-based CAD systems**

Many deep learning based-CAD systems perform 3D convolution operation to extract features within CT volumes instead CT slice. Our CAD system performance was also compared with these systems, which is shown in Figure 11. Results show that our CAD system out-performs several 3D based CAD systems (Dou et al., Zhu et al., CCElarfeCubeCnn). It also achieves the comparable results with the three best perform CAD systems (Pingan, JianPei, and Zhongliu\_Xie). Using 3D convolution operation results in high computation cost. Table 14 lists the implementation platform used in abovementioned CAD. It can be seen all these CAD systems used expensive graphic card that cost thousands dollar. It is not feasible for a hospital to buy such expensive facilities. In contrast, our CAD system can be implemented in a normal desktop, which is more practical in clinic.

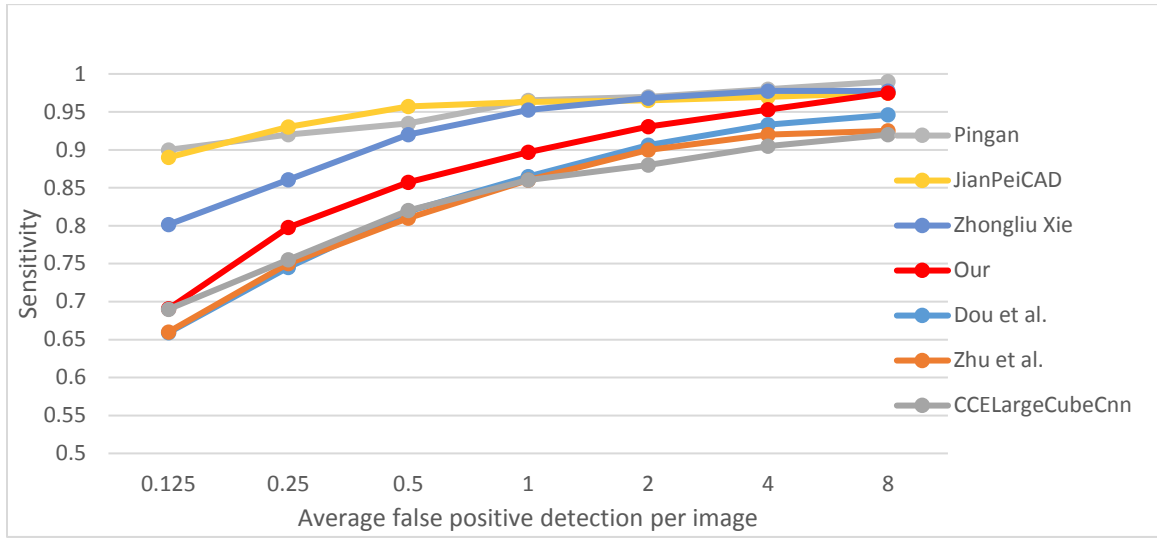
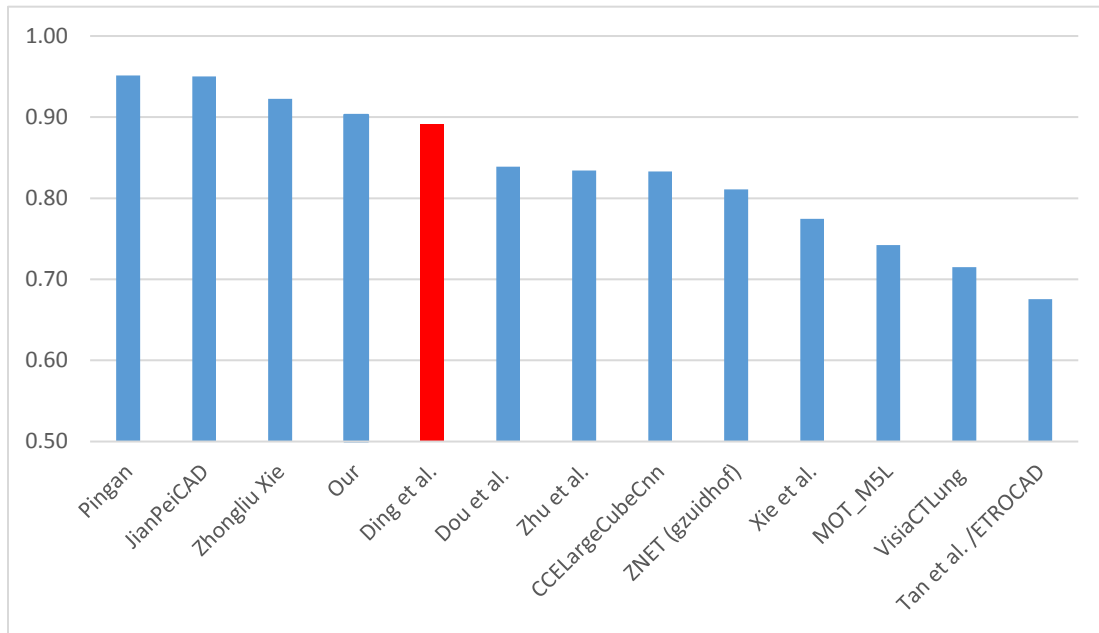


Figure 11: FROC curve comparison with 3D based deep learning CAD systems

Table 14: Implementation platform comparison

	<i>Implementation platform</i>	<i>Cost (\$)</i>
Pingan	4 NVIDIA Tesla K80	7100
JianPeiCAD	2 NVIDIA GTX Titan-X	4998
Zhongliu Xie	2 NVIDIA Tesla K80	3550
Xie et al.	1 NVIDIA Tesla K80	1775
Dou et al.	1 NVIDIA Titan X	1200
<b>Our</b>	<b>1 NVIDIA GTX 1060 Ti</b>	<b>400</b>

The overall CPM score comparison is shown on Figure 12. Our CAD system ranked 6 out of 13 CAD systems.



**Figure 12: CPM comparison**

### **4.3 Limitations and future works**

The detection accuracy in low FPPI should be improved. Increasing number of ResUnits in the feature extractor is one method and adapting the algorithm into 3D CNN that extracts features in 3D CT volumes instead of 2D CT slices is another method. Increased number of ResUnits and 3DD CNN may result in real-time detection latency because of the increased computation cost. Further exploration should be performed to balance the detection accuracy and efficiency.

Due to database limitation, nodule size smaller than 3 mm is not involved. The ability to detect nodule smaller than 3 mm may beneficial to nodule early diagnose. More dataset that includes nodule size smaller than 3 mm should be used in future

study. Furthermore, simulation image can be more realistic. The future study can simulate nodule with morphological variation and add noise to the simulated images.

The effect of organ motion is not evaluated in this study. In future study, 4D CT dataset should be used to test the influence of organ motion in real-time nodule tracking. Furthermore, the detection accuracy may dependent on location of the nodule. Future study should evaluate the influence of nodule position in nodule detection.

This study achieves the nodule detection and simple characterization. The algorithm can be modified for more specific nodule characterization, such as characterizing the type of nodule (benign tumor, lung cancers, infections, etc). It requires dataset with detailed annotations.

## 5 Conclusion

Results evaluation showed that our proposed CAD system using deep learning methods was robust. It achieved real-time nodule detection with high accuracy and characterization with high precision. The performance time is within 0.1 seconds and the localization precision is about 1 mm. This CAD system can be directly implemented in a clinically available computer. It has potential clinical applications in real-time nodule detection and tracking, as well as nodule quantitative analysis.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6):394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7-30.
3. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: A Cancer Journal for Clinicians*. 2014;64(1):9-29.
4. Lemjabbar-Alaoui H, Hassan OU, Yang Y-W, Buchanan P. Lung cancer: Biology and treatment options. *Biochimica et biophysica acta*. 2015;1856(2):189-210.
5. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England journal of medicine*. 2011;365(5):395-409.
6. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Annals of internal medicine*. 2014;160(5):311-320.
7. Valente IRS, Cortez PC, Neto EC, Soares JM, de Albuquerque VHC, Tavares JMRS. Automatic 3D pulmonary nodule detection in CT images: A survey. *Computer Methods and Programs in Biomedicine*. 2016;124:91-107.
8. Firmino M, Angelo G, Morais H, Dantas MR, Valentim R. Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. *BioMedical Engineering OnLine*. 2016;15(1):2.
9. Armato SG, 3rd, 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. *Medical physics*. 2011;38(2):915-931.
10. Ginneken Bv, Setio AAA, Jacobs C, Ciompi F. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. Paper presented at: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI); 16-19 April 2015, 2015.

11. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Medical physics*. 2019;46(1):e1-e36.
12. Choi W-J, Choi T-S. Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Computer Methods and Programs in Biomedicine*. 2014;113(1):37-54.
13. Santos AM, de Carvalho Filho AO, Silva AC, de Paiva AC, Nunes RA, Gattass M. Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. *Engineering Applications of Artificial Intelligence*. 2014;36:27-39.
14. Chen B, Kitasaka T, Honma H, et al. Automatic segmentation of pulmonary blood vessels and nodules based on local intensity structure analysis and surface propagation in 3D chest CT images. *International Journal of Computer Assisted Radiology and Surgery*. 2012;7(3):465-482.
15. Cascio D, Magro R, Fauci F, Iacomi M, Raso G. Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models. *Computers in Biology and Medicine*. 2012;42(11):1098-1109.
16. Suiyuan W, Junfeng W. Pulmonary Nodules 3D Detection on Serial CT Scans. Paper presented at: 2012 Third Global Congress on Intelligent Systems; 6-8 Nov. 2012, 2012.
17. Ozekes S, Osman O, Ucan ON. Nodule Detection in a Lung Region that's Segmented with Using Genetic Cellular Neural Networks and 3D Template Matching with Fuzzy Rule Based Thresholding. *Korean J Radiol*. 2008;9(1):1-9.
18. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. Paper presented at: Proceedings of 2010 IEEE International Symposium on Circuits and Systems; 30 May-2 June 2010, 2010.
19. Hua K-L, Hsu C-H, Hidayati SC, Cheng W-H, Chen Y-J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*. 2015;8:2015-2022.
20. Yang Y, Feng X, Chi W, et al. Deep learning aided decision support for pulmonary nodules diagnosing: a review. *Journal of thoracic disease*. 2018;10(Suppl 7):S867-S875.

21. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Transactions on Medical Imaging*. 2016;35(5):1160-1169.
22. Gruetzemacher R, Gupta A. Using deep learning for pulmonary nodule detection & diagnosis. 2016.
23. Kang G, Liu K, Hou B, Zhang N. 3D multi-view convolutional neural networks for lung nodule classification. *PloS one*. 2017;12(11):e0188290.
24. Zhu W, Liu C, Fan W, Xie X. DeepLung: 3D Deep Convolutional Nets for Automated Pulmonary Nodule Detection and Classification. *ArXiv e-prints*. 2017. <https://ui.adsabs.harvard.edu/#abs/2017arXiv170905538Z>. Accessed September 01, 2017.
25. Jin H, Li Z, Tong R, Lin L. A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. *Medical physics*. 2018;45(5):2097-2107.
26. Dou Q, Chen H, Jin Y, Lin H, Qin J, Heng P-A. Automated Pulmonary Nodule Detection via 3D ConvNets with Online Sample Filtering and Hybrid-Loss Residual Learning. *ArXiv e-prints*. 2017. <https://ui.adsabs.harvard.edu/#abs/2017arXiv170803867D>. Accessed August 01, 2017.
27. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *ArXiv e-prints*. 2015. <https://ui.adsabs.harvard.edu/#abs/2015arXiv150601497R>. Accessed June 01, 2015.
28. Ding J, Li A, Hu Z, Wang L. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks. *ArXiv e-prints*. 2017. <https://ui.adsabs.harvard.edu/#abs/2017arXiv170604303D>. Accessed June 01, 2017.
29. Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition*. 2019;85:109-119.
30. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv e-prints*. 2015.

- <https://ui.adsabs.harvard.edu/#abs/2015arXiv150504597R>. Accessed May 01, 2015.
31. Lan T, Li Y, Kimani Murugi J, Ding Y, Qin Z. RUN:Residual U-Net for Computer-Aided Detection of Pulmonary Nodules without Candidate Selection. *ArXiv e-prints*. 2018. <https://ui.adsabs.harvard.edu/#abs/2018arXiv180511856L>. Accessed May 01, 2018.
  32. Nam C-M, Kim J, Lee KJ. Lung nodule segmentation with convolutional neural network trained by simple diameter information. 2018.
  33. Arindra Adiyoso Setio 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. *ArXiv e-prints*. 2016. <https://ui.adsabs.harvard.edu/#abs/2016arXiv161208012A>. Accessed December 01, 2016.
  34. Torres EL, Fiorina E, Pennazio F, et al. Large scale validation of the M5L lung CAD on heterogeneous CT datasets. *Medical physics*. 2015;42(4):1477-1489.
  35. Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med Image Anal*. 2010;14(3):390-406.
  36. Kuhnigk JM, Dicken V, Bornemann L, et al. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging*. 2006;25(4):417-434.
  37. Tan M, Deklerck R, Jansen B, Bister M, Cornelis J. A novel computer-aided lung nodule detection system for CT images. *Medical physics*. 2011;38(10):5630-5645.
  38. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. *ArXiv e-prints*. 2015. <https://ui.adsabs.harvard.edu/#abs/2015arXiv150602640R>. Accessed June 01, 2015.
  39. Ramachandran S. S, George J, Skaria S, V. V. V. Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans. Paper presented at: Medical Imaging 2018: Computer-Aided Diagnosis; February 01, 2018, 2018.

40. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. *ArXiv e-prints*. 2016. <https://ui.adsabs.harvard.edu/#abs/2016arXiv161208242R>. Accessed December 01, 2016.
41. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. *ArXiv e-prints*. 2018. <https://ui.adsabs.harvard.edu/#abs/2018arXiv180402767R>. Accessed April 01, 2018.
42. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Medical physics*. 2010;37(9):4902-4915.
43. Tanoue LT, Tanner NT, Gould MK, Silvestri GA. Lung Cancer Screening. *American Journal of Respiratory and Critical Care Medicine*. 2015;191(1):19-33.
44. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *The Lancet*. 1999;354(9173):99-105.
45. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *ArXiv e-prints*. 2015. <https://ui.adsabs.harvard.edu/#abs/2015arXiv151203385H>. Accessed December 01, 2015.
46. Rothe R, Guillaumin M, Van Gool L. Non-maximum suppression for object detection by passing messages between windows. Paper presented at: Asian Conference on Computer Vision2014.
47. Receiver Operating Characteristic Analysis in Medical Imaging: Contents. *Journal of the International Commission on Radiation Units and Measurements*. 2008;8(1):NP-NP.