An unsupervised semi-automated pulmonary nodule segmentation method based on enhanced region growing

He Ren[1,2#], Lingxiao Zhou[1#], Gang Liu[1#], Xueqing Peng[1], Weiya Shi[1], Huilin Xu[1], Fei Shan[1], Lei Liu[1]

[1]Shanghai Public Health Clinical Center & Institutes of Biomedical Sciences, School of Basic Medical Sciences, School of Data Science, Fudan University, Shanghai 200032, China; [2]Shanghai University of Medicine & Health Sciences, Shanghai 201318 China

[#]These authors contributed equally to this work.

Correspondence to: Lei Liu; Shan Fei. Shanghai Public Health Clinical Center & Institutes of Biomedical Sciences, School of Basic Medical Sciences, School of Data Science, Fudan University, Shanghai 200032, China. Email: liulei_sibs@163.com; shanfei@shphc.org.cn.

Background: Nowadays, computer technology is getting popular for clinical aided diagnosis, especially in the direction of medical images. It makes physician diagnosis of lung nodules more efficient by providing them with reliable and accurate segmentation.

Methods: A region growing based semi-automated pulmonary nodule segmentation algorithm (ReGANS) was developed with three improvements: an automatic threshold calculation method, a lesion area pre-projection method, and an optimized region growing method. The algorithm can quickly and accurately segment a whole lung nodule in a set of computed tomography (CT) images based on an initial manual point.

Results: The average time taken for ReGANS to segment 1 pulmonary nodule was 0.83s, and the probability rand index (PRI), global consistency error (GCE), and variation of information (VoI) from a comparison between the algorithm and the radiologist's 2 manual results were 0.93, 0.06, and 0.3 for the boundary range (BR), and 0.86, 0.06, 0.3 for the precise range (PR). The number of images covered by one pulmonary nodule in a CT image set was also evaluated to compare the segmentation algorithm with the radiologist's results, with an error rate of 15%. At the same time, the results were verified in multiple data sets to validate the robustness.

Conclusions: Compared with other algorithms, ReGANS can segment the lung nodule image region more quickly and more precisely. The experimental results show that ReGANS can assist medical imaging diagnosis and has good clinical application value. It also provides a faster and more convenient method for pre-data preparation of intelligent algorithms.

Keywords: Lung cancer; pulmonary nodule; segmentation; computed tomography (CT)

Introduction

According to reports from the World Health Organization (WHO) and major cancer research centers, lung cancer has the world's highest cancer mortality rate (1). A status report produced by the International Agency for Research on Cancer states that lung cancer remains the leading cause of cancer incidence and mortality, with 2.1 million new lung cancer cases and 1.8 million deaths predicted worldwide in 2018, representing close to 1 in 5 (18.4%) cancer deaths (2). However, early detection with computed tomography (CT) has been shown to help reduce lung cancer-specific mortality (3). In medicine, a lung nodule is defined as a focal opacity whose largest diameter is between 3 mm and 3 cm in length (4). Nodules with diameters less than 3 mm are known as micronodules (5).

In general, the radiologist forms a diagnosis by reading the CT image in digital imaging and communications in medicine (DICOM) format (6). The radiologist quickly searches through all the images to determine the accuracy of suspected pulmonary nodules with obvious features and then combines those images of the previous and subsequent consecutive images to analyze the size, characteristic changes, and signs of the lung nodules in the corresponding images. This kind of reading process takes quite some time and puts intense pressure on radiologists, especially when they have been working all day and must repeatedly switch between various conditions. Hence, computer technology is required to quickly and accurately segment/mark lung nodule images to assist radiologists in diagnosing diseases (7-9). This supplementary approach has been used for applications such as lung parenchyma density analysis (10,11), airway analysis (12,13), and pulmonary nodule detection (14-16). The application of computer-aided detection in pulmonary nodules segmentation with CT images and nodule type and characteristics determination would be helpful for the early detection of lung cancer and tumor diagnosis.

A general process of automatic methods for pulmonary nodule segmentation usually includes image acquisition, preprocessing, pulmonary parenchyma segmentation, focal region extraction, optimization, and feature extraction (17,18).

Many researchers have studied image segmentation (19-22). Some approaches identify lesions based on preset shape-information of the target area and use machine learning-based algorithms to detect lesion features (23). However, many segmentation algorithms require image denoising, which would severely affect the processing speed as considerable manual input is required. Also, some automatic segmentation algorithms require manual operations in preprocessing, such as range determination and threshold testing (24-26). These result in extremely long segmentation time and reduce the accuracy of results. Therefore, these segmentation algorithms have extremely limited practical value. Other studies use images of a single lesion or CT-enhanced scan image for segmentation, and thus the results are not reproducible. Furthermore, those automatic segmentation algorithms also require a large amount of segmented lesion data for prior learning, which takes a lot of pre-work, and the quality of the data used for model training directly affects the quality of the model.

In this paper, a new semi-automated algorithm of lung nodule segmentation is introduced. It links the clinical manifestation of the lesion to image features via an image processing method. The results of ReGANS were compared with other computer-based segmentation techniques and manual segmentation performed by radiologists to validate the efficiency and accuracy, evaluated by probability rand index (PRI) (27), global consistency error (GCE) (28), and variation of Information (VoI) (28). Multiple types of nodules were also randomly selected for an experiment to prove its robustness.

Methods

The study design was approved by the Shanghai University of Medicine and the Health Sciences Ethics Review Board. The need for informed consent was waived.

Dataset

Two datasets were used in this study. The first one was generated at the Shanghai Public Clinical Health Center (SPCHC) and was collected between 2016 and 2017 (*Table 1*). An unenhanced chest CT exam was performed for the entire lung scan of each patient with a United-Imaging 760 CT device (42–126 mA 120 kV, slice thickness of 1 mm) and a Siemens Emotion 16 CT device (34–123 mA, 130 kV, 1 mm) with a 512×512 resolution. A total of 55 sets of image data were collected with 407 CT images. Among them, 25 groups (166 CT images) were used in the analysis of the speed and accuracy of the method, and the remaining 30 groups (241 CT images) were used in further method robustness verification experiments. Two radiologists performed the diagnosis of the pulmonary nodule.

Another dataset was taken from the The Lung Image Database Consortium Image Collection public data set (LIDC-IDRI) (https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI). A total of 30 sets of image data with 234 CT images were randomly selected.

The snake (29), special processing named selective binary Gaussian filtering regularized level set (SBGFRLS) (30), and level-set (31) methods were compared to the ReGANS. These methods were partially modified with reference to the experimental environment and data. As the number of iterations was increased further, the time taken by the algorithm increased substantially without a significant gain in optimization performance. The Snake and level-set used 250 iterations. For the SBGFRLS method, the image was first cut into a small square centered on and fully containing the lung nodule, which in turn increased the speed of the

Table 1 Case information

Serial number	Sex	Detection time	Diagnosis
1	F	2017/2/14	MIA
2	M	2016/12/16	IAC
3	F	2016/8/17	MIA
4	F	2016/12/14	MIA
5	M	2016/12/13	MIA
6	F	2017/1/9	MIA
7	F	2016/6/16	MIA
8	F	2016/5/3	MIA
9	M	2016/7/12	MIA
10	F	2016/7/18	MIA
11	M	2016/12/23	MIA
12	F	2016/8/26	AIS
13	F	2017/2/7	MIA
14	M	2016/9/17	MIA
15	M	2016/9/17	MIA
16	F	2016/12/15	MIA
17	F	2016/4/28	AIS
18	F	2016/7/12	MIA
19	F	2016/11/29	MIA
20	F	2016/7/28	MIA
21	F	2016/5/21	MIA
22	M	2016/10/11	MIA
23	M	2016/12/26	MIA
24	F	2016/11/15	MIA
25	F	2016/8/16	MIA
26	F	2016/8/26	AIS
27	F	2016/7/20	AIS
28	M	2017/1/3	AIS
29	M	2016/12/6	AIS
30	F	2016/4/16	AIS
31	F	2016/12/1	AIS
32	M	2016/5/19	AIS
33	M	2016/9/9	AIS
34	F	2016/10/24	AIS
35	F	2016/11/4	AIS

Table 1 (*continued*)

Table 1 (*continued*)

Serial number	Sex	Detection time	Diagnosis
36	F	2016/7/11	MIA
37	M	2016/9/9	MIA
38	M	2016/7/22	MIA
39	F	2016/1/19	MIA
40	F	2016/5/4	MIA
41	F	2016/12/16	MIA
42	F	2016/5/17	MIA
43	F	2016/7/21	MIA
44	F	2016/3/11	MIA
45	F	2016/8/16	MIA
46	F	2016/11/14	IAC
47	M	2016/7/18	IAC
48	M	2016/10/31	IAC
49	F	2016/8/4	IAC
50	F	2016/9/20	IAC
51	F	2016/6/27	IAC
52	F	2016/4/27	IAC
53	F	2016/5/25	IAC
54	F	2016/11/1	IAC
55	M	2016/6/6	IAC

AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma cancer.

algorithm. The region-growing algorithm in the ReGANS method involved the evaluation of the 8 neighboring voxels surrounding an initial seed point. These surrounding voxels were evaluated according to their similarity to the seed voxel. They were accepted or rejected based on a combination of an artificially set change rate from the starting point, the voxel value change rate of the region, and the initial voxel values of the region (see step "Pulmonary nodule segmentation" for details).

Running environment

All the algorithms in the experiment were implemented in MATLAB 2017a on a personal computer machine with a 2.6 GHz Intel (R) Core (TM) i7-5600 processor and 8 GB of RAM.

236

Ren et al. Segmentation method for lung nodule
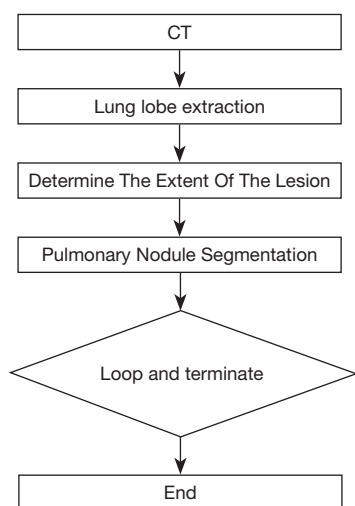
Figure 1 Overview of the steps in the pretreatment approach and pulmonary nodule.

Proposed method

The main steps of ReGANS are: (I) extraction and optimization of lung lobes; (II) determination of the extent of lesions; (III) segmentation of pulmonary nodules; (IV) and determination and continuation (*Figure 1*).

Lung lobe extraction

The goal of this step was to separate the lung lobes from the CT image and link the values of the image pixels to the medical clinical depiction. The parameters used to segment the lung lobes from the image were based on both the optimal threshold algorithm and the medical clinical description.

The gray value was converted to a CT value using information in the DICOM format file, and the abnormal point in the image was removed (6). The threshold was by an iterative algorithm (32), and the lung cavity region was extracted and recorded as $IM2$. Then, the threshold was calculated in $IM2$, combined with the location of the lobe in $IM2$ and other features to separate the lung lobe region $C1$. The boundaries of $C1$ were optimized using image processing techniques (33), and a new image was acquired ($M3$).

Determination of the extent of the lesion

Locating the pulmonary nodules is the most important step for lung nodule segmentation. A fuzzy positioning method was employed to identify the nodule range.

The probability density map $IM3$ was created using

0.45 as the threshold to obtain the upper limit of the lung parenchymal CT values of the current group from the graph. $IM3$ and the upper limit were used as the thresholds to obtain the binary image of the lung lobe ($BL1$). Eight rays crossing the positioned coordinate points with angles of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° from the horizontal were extended on $BL1$ until they intersected the horizontal axis. The maximum and minimum lengths of the 8 rays were removed, and the average length of the other 6 rays was calculated $AvgL$. The "slice_Location" in the DICOM file was used to determine the spacing of each layer to calculate the number of sheets ($Ser1$) that the nodule may cover. A maximum intensity projection (34) was performed on Ser1 to obtain the image $IM4$. A square using positioned coordinate points as the center and $2 \times AvgL$ as the side lengths was drawn, and the square region in $IM4$ was defined as $S1$. The maximum value and the minimum values of $S1$ were extracted and used to perform a "bilateral filter" (35) to obtain the image ($IM5$). An iterative threshold algorithm was used ($IM5$) to obtain the threshold ($T2$). $IM5$ was binarized using the T2 threshold. The "imclose" (33) operation was subsequently performed on the image, and it was filled to obtain the lung nodule range (LNR).

Pulmonary nodule segmentation

The purpose of this step was to segment the pulmonary nodule, and it was completed after image normalization, lung lobe extraction, and lesion extension. In this step, a single image was processed, and the radiologist specified a coordinate point. The subsequent images were defined by the result of the previous image (see step D for details) with the starting image being in range of the LNR.

To calculate the range of pixel variable values within the LNR, 8 rays originating from the specified coordinate points were drawn, and the pixel values were recorded. The absolute value of the gradient value between the pixels before and after was calculated and uniquely represented by Gr. The valid data were extracted and the average ($AvgS$) was calculated. To calculate the center point of the pixel variation range in the LNR region, all pixel values in the were sorted from large to small, and the optimal value (EP) was extracted.

EP, $AvgS$, and the original image were used as inputs, and a modified region-growing method was used to obtain the putative pulmonary nodule region, PN. Based on the original Region-Growing method, our method optimized the pixel growth judgment conditions so that it could be automatically performed under the influence of EP, $AvgS$,

and initial coordinate points. This can solve the problem of the pixel growth range in the original method needing to be manually set each time which leads to inaccurate segmentation.

Further optimization is required because of the influence of factors such as blood vessel penetration in the PN region. An "imclose" operation PN was performed to obtain a series of regions. The largest region was selected and reported as LR. Then, LR was projected onto the LNR. The region was denoted as LNI, and the regions of the LNR outside of the projection were denoted as LNO. The areas of LNI and LNO were calculated, and the average pixel values of LNI and LNO were denoted as AI and AO, and PI and PO. The following were calculated: $P = PI - PO$; $AS = AO - AI$. Using AS and P as the decision conditions, the various possible situations were analyzed, and the target area (TIM) of the image was calculated.

Loop and terminate

Because the CT was based on a series of images, the nodule segmentation continued to the next image. The average CT value of the TIM was then calculated. If it was not less than –775, and the difference in variance between this image and the previous image was less than 0.002 (the variance of the 0th TIM was defined to be same as the 1st TIM), the current image segmentation was considered to be valid, and the next image was processed. Otherwise, the loop was terminated. The starting coordinate point in the IM_{ai} step was defined as the gravity center of the TIM in IM_{ai-1}.

Results

The performance of the newly developed ReGANS method was evaluated and compared with Snake, SBGFRLS, and level-set method, taking two types of radiologist's segmentation results as a benchmark. The lesion boundary delineated by radiologists was defined as the boundary range (BR), while the precise range (PR) was BR without blood vessels and bronchi. The performance of the ReGANS was evaluated in aspects of accuracy, speed, and robustness. In addition, detailed information on the pulmonary nodules used in the first dataset was also recorded (*Table S1*, supplementary material).

The accuracy and robustness of the ReGANS were evaluated by PRI, GCE, and VoI (19,36). Using the MATLAB runtime environment, the "compare_segmentations.m" toolkit was downloaded to perform calculations (https://www.dssz.com/778476.html).

Accuracy

The planar segmentation results of the four algorithms were compared based on the second dataset. The accuracies of these algorithms are shown in *Table 2*.

The average PRI of ReGANS, when compared to BR, was 0.93 (95% CI, 0.90–0.96), while that of the other three algorithms was lower (0.59–0.82), which indicated that ReGANS could get more accurate results close to BR. Meanwhile, the average GCE between ReGANS and BR was 0.058 (95% CI, 0.05–0.07), which suggested that the global error of ReGANS was negligible. Consistently, the average VoI between ReGANS and BR was 0.30 (95% CI, 0.26–0.34). The average GCE and VoI ofReGANS were significantly lower than those of the other three algorithms (for all comparison, P<0.0001). The average PRI, GCE, and VoI between the ReGANS and PR were 0.86 (95% CI, 0.84–0.88), 0.06 (95% CI, 0.052–0.068), and 0.30 (95% CI, 0.27–0.33), respectively. These results indicate that the segmentation performance of the ReGANS was comparable to that of PR.

The supero-inferior coverage difference between the PR and ReGANS in each sample was compared; the radiologist diagnosed 166 images, 28 of the ReGANS results were not in the radiologist's diagnosis, and two correct results were not calculated. The result showed that the error and loss rate of the range of tuberculosis detected by the ReGANS was less than 18% (*Figure 2A*). Most of the missing images were terminal in each sample. The CT values of the nodule range were close to the lung parenchyma, causing it to be ignored by the ReGANS (*Figure 2B*).

The performance of the ReGANS on special types of pulmonary nodules was also adequate. The ReGANS could segment lung nodules with special signs, like the vacuole sign (*Figure 2C*). For special lung nodules with blood vessels or bronchial passages, the ReGANS could also be effectively segmented, although the complete effective area of the pulmonary nodules is not fully depicted in *Figure 2D*. Collectively, the ReGANS exhibited a better accuracy than the other algorithms.

Speed

The segmentation time of each algorithm was recorded as the number of images of lung nodules in each group of data ranging from 4–22. Eight images (average) were used to calculate the overall segmentation time. The average segmentation time of a single CT image was 0.83 s (95%

Table 2 Accuracy of each segmentation result

Variable	ReGANS	Snake	SBGFRLS	Level-set	BR*	PR*
Difference						
Speed	Fast	Not fast	normal	Not fast		
Input	One point	One closed curve	The rectangular area in the image	One closed curve		
Output	An entire nodule in a set of images	A nodule section in one image	A nodule section in one image	A nodule section in one image		
PRI						
BR	0.93	0.59	0.82	0.60	1	0.86
PR	0.86	0.61	0.67	0.59	0.86	1
GCE						
BR	0.06	0.16	0.16	0.13	0	0.06
PR	0.06	0.15	0.13	0.12	0.06	0
VoI						
BR	0.30	0.74	0.41	0.65	0	0.30
PR	0.30	0.73	0.64	0.64	0.30	0

*The BR/PR was compared with BR/PR, thus the three comparison parameters reach the extreme value. PRI, probability rand index; GCE, global consistency error; VoI, variation of information; BR, boundary range; PR, precise range.
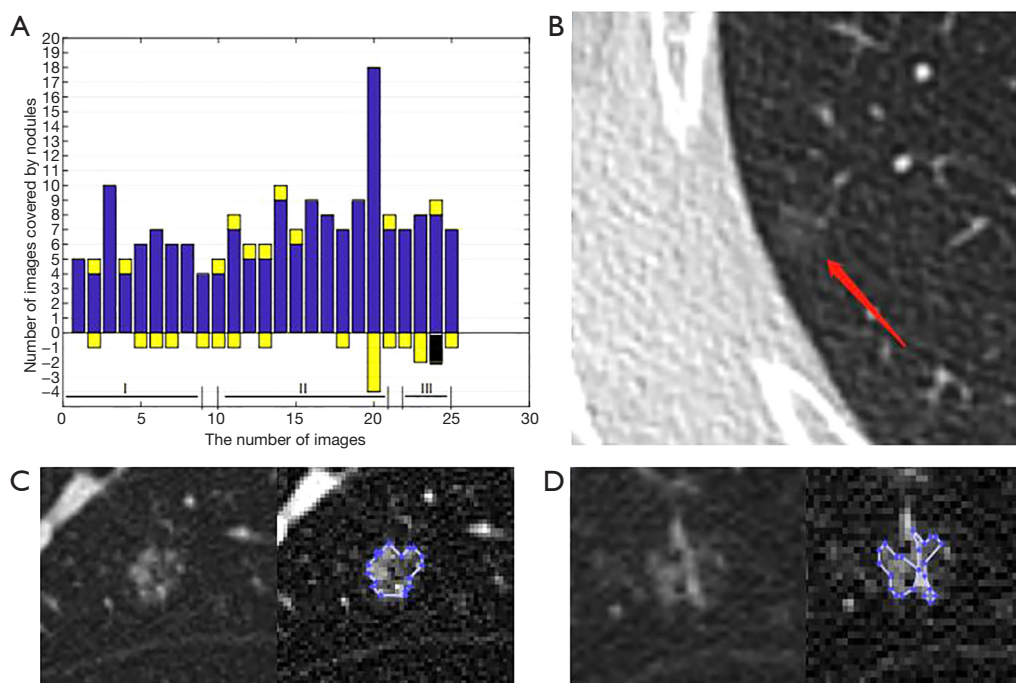
Figure 2 Algorithm detection results compared with actual nodule coverage. In (A), the blue area represents the intersection between ReGANS and the manual annotation superoinferiorly; the yellow area represents the images neglected by ReGANS; the black area represents extra images calculated by ReGANS; I indicates the nodules in this set of data that have no other tissue interference; II represents one blood vessel/bronchus passing through the lung nodules; III represents multiple blood vessels/bronchi passing through the lung nodules. (B) represents a missing image due to the CT value of the lung nodule area in the image being low. (C) represents an extremely irregular pulmonary nodule and depicts the results of the ReGANS algorithm. (D) shows a pulmonary nodule with blood vessels passing through and depicts the results of segmentation by the ReGANS method.

Table 3 Time of segmentation of each method

Variable	ReGANS	Snake	SBGFRLS	Level-set	Manual
Single image (s)	0.83±0.17	13.04±3.24	2.42±0.93	19.02±3.71	45±15
8 images (s)	6.64±1.36	104.32±19.44	19.36±7.44	152.16±29.68	360±120
P value	–	<2.2e–16	<9.141e–11	<2.2e–6	<2.2e–16

Table 4 Method robustness comparison with two datasets

Variable	First dataset (n=30)		Second dataset	
	ReGANS	BR	ReGANS	BR
Time, single image	0.843		0.836	
Accuracy				
PRI				
BR	0.936	1	0.947	1
PR	0.854	0.904	0.899	0.931
GCE				
BR	0.069	0	0.064	0
PR	0.075	0.042	0.069	0.038
VoI				
BR	0.256	0	0.202	0
PR	0.387	0.237	0.282	0.164

PRI, probability rand index; GCE, global consistency error; VoI, variation of information; BR, boundary range; PR, precise range.

CI, 0.829–0.840) and segmentation of all the lung nodule images took 6.64 s (*Table 3*). The average segmentation times of the Snake, SBGFRLS, and level-set methods for a single image were 13.04, 2.42, and 19.02 s, respectively. The radiologists' manual segmentation time was 60 s, and thus, a minimum of 480 s (8 min) was required to process these images.

The time required to segment the image with the ReGANS was generally consistent. Images with special types of nodules did not significantly increase the time required. As shown in *Figure 2D*, the time required by the ReGANS was only 0.92 s. Thus, the speed of the ReGANS was significantly higher than those of the other algorithms.

Robustness

ReGANS was also used to split 30 sets of images in the first dataset and images in the third dataset. The accuracy

and speed of the ReGANS were evaluated for those two datasets, as shown in *Table 4*.

At the same time, the serial number of the lung nodule image in a set of case images calculated by REGAN was also recorded and compared with the diagnostic results of the radiologist. In the 30 images in the first dataset, the radiologist diagnosed 241 images, 17 of the ReGANS results were not in the radiologist's diagnosis, and 12 correct results were not calculated. In the 30 images in the third dataset, the radiologist diagnosed 234 images, 14 of the ReGANS results were not in the radiologist's diagnosis, and 9 correct results were not calculated. The total error rate was 10.95%

Further inspection

To further evaluate the segmentation accuracy of the ReGANS method in other diseases of the lung. We collected some CT images of pulmonary fibrosis for segmentation experiments. The segmentation results are shown in the appendix.

Discussion

Most currently reported algorithms are not satisfactory for clinical application. First, automatic segmentation operations usually must be performed on 5–10 images, sometimes 20, and the time spent for published algorithms is several minutes (29-31). Second, other methods have many complicated pre-processing steps and cumbersome manual operations (37), including determining approximate ranges, building similar shapes, and enhancing image contrast. Moreover, in some algorithms, iterative analysis is required, and professional training is necessary. The automatic segmentation algorithms require a substantial amount of data and manual targets to establish models through algorithms. These models are affected by data volume, data classification, and artificial results. Lastly, in most cases, the reported algorithms cannot process a set of images as a single sample. In other words, the supero-inferior images of a patient cannot be processed automatically. A few researchers have developed automatic

supero-inferior algorithms, but at a low speed with more than 10 min needed.

ReGANS requires only 6–8 s, even for complicated lung nodules. Only one-click operation is required for the ReGANS, and the radiologist does not need to pre-process the images and select the pulmonary nodule one image at a time. ReGANS is a very simple operation that offers a radiologist with a continuous feature image of a pulmonary nodule quickly and accurately. This can effectively help radiologists by providing them with reliable diagnostic information. To a certain extent, they are freed from a whole day of tedious work.

Segmentation algorithms based on machine learning or deep learning do have significant advantages. A mature model that completes training can also quickly and accurately segment the lesion area in practical applications. In some cases where hardware devices are better, they can even do segmentation faster than ReGANS, but, limited by the design of the model itself, the quality of the training image directly affects the capabilities of the model. With high speed and accuracy, ReGANS can provide high-quality annotation images for model training and effectively compress the pre-preparation time.

ReGANS also has its limitations. The threshold calculation process for image binarization is a bit conservative, which leads to the missing of terminal images when segmenting. The automatic boundary calculation of the lung nodule starts from an initial manual point. In some extremely complicated situations, such as multiple blood vessels passing through the lung nodules, to avoid the blockage of blood vessels and other tissues, the segmentation of the lung nodules will always leave some areas behind. In addition, determining the extent of pulmonary nodules for radiologists is only part of the diagnostic process, and subsequent analysis of signs such as lobulation, bronchus sign, and pleural indentation requires a more accurate diagnosis. This requires the subsequent segmentation method to be able to not only accurately determine the lesion area, but also to identify each sign. What is more, the segmentation results of pulmonary fibrosis also reflect the limitations of ReGANS. Although the "maximum intensity projection" method applied in ReGANS can define the lesion area of the lung nodule image well, it can also cause overlapping features of multiple lesions in some more complex images, which instead interferes with the algorithm's determination of the suspected area of the lesion and also interferes with the "segmentation threshold calculation"

and "segmentation result optimization" in the later stages of the algorithm.

An evolved method is then required to convert the image information into medical clinical description, so as to identify the lesion signs from the perspective of image semantics, and finally, provide more accurate diagnostic information for radiologists. This is also the future research direction of ReGANS.

Conclusions

The experiments showed that this semi-automatic lung nodule segmentation method can accurately and quickly segment all the lung nodule images in a set of CT data while also preserving the original characteristics of lesions. It can effectively assist the radiologist's diagnosis and provide reliable training data for intelligent algorithms.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The study design was approved by the Shanghai University of Medicine and the Health Sciences Ethics Review Board. The need for informed consent was waived.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.
2. 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 Cancer J Clin 2018;68:394-424.
3. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, et al.. The National Lung

Screening Trial: overview and study design. Radiology 2011;258:243-53.

4. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology 2008;246:697-722.

5. Raad RA, Suh J, Harari S, Naidich DP, Shiau M, Ko JP. Nodule characterization: subsolid nodules. Radiol Clin North Am 2014;52:47-67.

6. Shin HB, Sheen H, Lee HY, Kang J, Yoon DK, Suh TS. Digital Imaging and Communications in Medicine (DICOM) information conversion procedure for SUV calculation of PET scanners with different DICOM header information. Phys Med 2017;44:243-8.

7. Zhao Y, de Bock GH, Vliegenthart R, van Klaveren RJ, Wang Y, Bogoni L, de Jong PA, Mali WP, van Ooijen PM, Oudkerk M. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. Eur Radiol 2012;22:2076-84.

8. Jin S, Zhang B, Zhang L, Li S, Li S, Li P. Lung nodules assessment in ultra-low-dose CT with iterative reconstruction compared to conventional dose CT. Quant Imaging Med Surg 2018;8:480-90.

9. Mao L, Chen H, Liang M, Li K, Gao J, Qin P, Ding X, Li X, Liu X. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. Quant Imaging Med Surg 2019;9:263-72.

10. Theilig D, Doellinger F, Kuhnigk JM, Temmesfeld-Wollbrueck B, Huebner RH, Schreiter N, Poellinger A. Pulmonary lymphangioleiomyomatosis: analysis of disease manifestation by region-based quantification of lung parenchyma. Eur J Radiol 2015;84:732-7.

11. Rausch SM, Martin C, Bornemann PB, Uhlig S, Wall WA. Material model of lung parenchyma based on living precision-cut lung slice testing. J Mech Behav Biomed Mater 2011;4:583-92.

12. Perez-Rovira A, Kuo W, Petersen J, Tiddens HA, de Bruijne M. Automatic airway-artery analysis on lung CT to quantify airway wall thickening and bronchiectasis. Med Phys 2016;43:5736.

13. Mumcuoğlu EU, Long FR, Castile RG, Gurcan MN. Image analysis for cystic fibrosis: computer-assisted airway wall and vessel measurements from low-dose, limited scan lung CT images. J Digit Imaging 2013;26:82-96.

14. Wang B, Tian X, Wang Q, Yang Y, Xie H, Zhang S, Gu L. Pulmonary nodule detection in CT images based on shape constraint CV model. Med Phys 2015;42:1241-54.

15. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, Gingras M, Atkar-Khattra S, Berg CD, Evans K, Finley R, Yee J, English J, Nasute P, Goffin J, Puksa S, Stewart L, Tsai S, Johnston MR, Manos D, Nicholas G, Goss GD, Seely JM, Amjadi K, Tremblay A, Burrowes P, MacEachern P, Bhatia R, Tsao MS, Lam S. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 2013;369:910-9.

16. Cavalcanti PG, Shirani S, Scharcanski J, Fong C, Meng J, Castelli J, Koff D. Lung nodule segmentation in chest computed tomography using a novel background estimation method. Quant Imaging Med Surg 2016;6:16-24.

17. Shen S, Bui AAT, Cong J, Hsu W. An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. Comput Biol Med 2015;57:139-49.

18. Lee SLA, Kouzani AZ, Hu EJ. Automated detection of lung nodules in computed tomography images: a review. Machine Vision Applications 2012;23:151-63.

19. Jiang X, Marti C, Irniger C, Bunke H. Distance measures for image segmentation evaluation. Eur J Applied Signal Processing 2006;(35909).

20. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001;20:490-8.

21. Brown MS, McNitt-Gray MF, Mankovich NJ, Goldin JG, Hiller J, Wilson LS, Aberle DR. Method for segmenting chest CT image data using an anatomical model: Preliminary results. IEEE Trans Med Imaging 1997;16:828-39.

22. Wong LM, Shi L, Xiao F, Griffith JF. Fully automated segmentation of wrist bones on T2-weighted fat-suppressed MR images in early rheumatoid arthritis. Quant Imaging Med Surg 2019;9:579-89.

23. Zhao L, Jia K. Multiscale CNNs for Brain Tumor Segmentation and Diagnosis. Comput Math Methods Med 2016;2016:8356294.

24. Zhu Y, Tan Y, Hua Y, Zhang G, Zhang J. Automatic Segmentation of Ground-Glass Opacities in Lung CT Images by Using Markov Random Field-Based Algorithms. J Digit Imaging 2012;25:409-22.

25. Zhou S, Cheng Y, Tamura S. Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images. Biomed Signal Processing Control

242

Ren et al. Segmentation method for lung nodule

2014;13:62-70.

26. Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 2010;14:390-406.

27. Lin J, Peng B, Li T, Chen Q. A Learning-Based Framework for Image Segmentation Evaluation. In: International Conference on Intelligent Networking & Collaborative Systems; 2013 2013: IEEE Computer Society; 2013:691-6.

28. Bellare M, Namprempre C, Neven G. Security Proofs for Identity-Based Identification and Signature Schemes. J Cryptol 2009;22:1-61.

29. Kass M, Witkin A, Terzopoulos D. Snakes: Active Contour Models. Int J Computer Vision 1987;1:321-31.

30. Zhang K, Zhang L, Song H, Zhou W. Active contours with selective local or global segmentation: A new formulation and level set method. Image Vision Computing 2010;28:668-76.

31. Osher SP, Fedkiw R. Level Set Methods and Dynamic Implicit Surfaces. In: Applied Mathematical Sciences

2004:xiv+273.

32. Perez A, Gonzalez RC. An Iterative Thresholding Algorithm for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 1987;9:742-51.

33. Gonzalez RC, Woods RE. Digital Image Processing. 1977.

34. Sato Y, Shiraga N, Nakajima S, Tamura S, Kikinis R. Local maximum intensity projection (LMIP): a new rendering method for vascular visualization. J Comput Assist Tomogr 1998;22:912-7.

35. Morillas S, Gregori V, Sapena A. Fuzzy Bilateral Filtering for Color Images. Sensors (Basel) 2011;11:3205-13.

36. Monteiro FC, Campilho AC. Performance evaluation of image segmentation. In: Campilho A, Kamel M. editors. ICIAR 2006. Lecture Notes Computer Science 2006:248-59.

37. Li B, Chen Q, Peng G, Guo Y, Chen K, Tian L, et al.. Segmentation of pulmonary nodules using adaptive local region energy with probability density function-based similarity distance and multi-features clustering. Biomed Eng Online 2016;15:49.

Proposed method

The main steps of the ReGANS are depicted in *Figure S1*. In the first step, the gray values of the CT image are converted into CT values, and the pulmonary lobe is segmented and optimized. The maximum projection area of the pulmonary nodules on the horizontal plane is then determined, and the pulmonary nodules are subsequently segmented using the projection area. Finally, the parameters are calculated to decide whether to continue to the next image.

Lung lobe extraction

The goal of this step was to separate the lung lobes from the CT image and link the values of the image pixels to the

medical clinical depiction. The parameters used to segment the lung lobes from the image were based on both the optimal threshold algorithm and the medical clinical description.

Step 1: image normalization

DICOM is a standard format for the communication and management of medical imaging information and related data. The "RescaleSlope" and "RescaleIntercept" parameters were used to convert the calculation of the gray value back to the CT values using the following formula:

$$HU = pixel\_value \times RescaleSlope + RescaleIntercept \qquad [1]$$

where "RescaleSlope" and "RescaleIntercept" are the DICOM tags that specify the linear transformation for
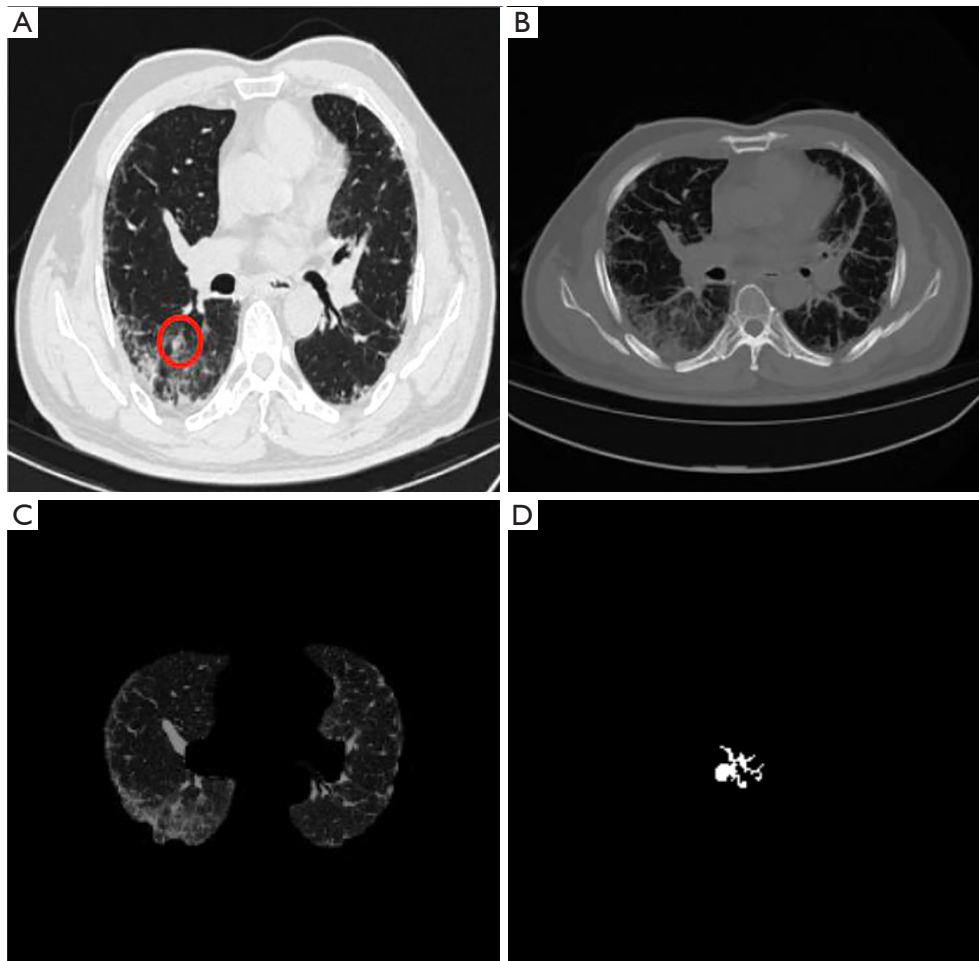
Figure S1 Segmentation results of ReGANS in CT images of pulmonary fibrosis lesions.

the pixels to be converted from their stored on-disk representation to their in-memory representation.

Extreme data were subsequently removed. To prevent interference from extreme data points such as those resulting from metal objects in the image, highlighted pixels in the image were removed. The full graph histogram was obtained, using 0.0039 as the interval, because the abnormal values were generally in the last few intervals. The pixel value distribution of each interval from the back to the previous time was analyzed, and extreme values were eliminated using thresholds. The processed images (denoted IM1) were used for further analyses.

Step 2: extraction of the lung lobes

This step involves pulmonary cavity extraction and lung lobe extraction. Pulmonary cavity extraction is necessary for lung lobe extraction because interference of invalid information, including machine shadows and the background, must be removed from the image. There are typically few background shadows, and their CT values are usually consistent. Here, an Iterative thresholding algorithm was used. The zero gray values were not used to obtain the threshold (T) of the image, which allowed it to be binarized (the zero values were removed, and the image was processed by the iterative algorithm to obtain the threshold for the pulmonary cavity extraction). The boundary was defined, holes inside the regions were filled, and the largest closed area yielding the novel image (IM2) was obtained. This was used for further lung lobe extraction. IM1 was binarized with T to obtain the binary image BIM1. The following formula was used: $C1 = BIM1 - IM2$. The number of closed areas of C1 was calculated and denoted H1. If H1 > 2, the three largest closed areas were extracted, and the left and right areas on the X-axis were selected as lung lobes. If H1 = 2, the centers were labeled as (x1, y1) and (x2, y2). If $|x1-x2|>|y1-y2|$, both areas were retained. Otherwise, the larger area was retained as a merged lung lobe. If H1 = 1, the area was identified as a merged lung lobe.

Step 3: boundary optimization

The boundaries of the lungs were optimized to facilitate the extraction. An ellipse with 10 and 30 pixels as the minor and major axes, respectively, was created to accommodate the shape of the lung. The left and right lungs were separated into two independent images. An "imopen" operation on the two images was performed with the previous ellipse (if the left and right lungs in the previously extracted lung lobe area were connected at the top, the image erosion

algorithm was used to separate and restore them through the expansion algorithm). The processed images were subsequently combined into a new binary image (BIM2) representing the extracted lung lobes, and the following calculation was performed: $IM3 = BIM2 - IM1$.

Determination of the extent of the lesion

Locating the pulmonary nodules is the most crucial step for lung nodule segmentation. Most nodules grow in approximate symmetry. A fuzzy positioning method was employed to identify the nodule.

First, a histogram of IM4 was plotted. The probability density map was created using 0.45 as the threshold to obtain the upper limit of the lung parenchymal CT values of the current group from the graph. IM4 and the upper limit were used as the thresholds to obtain the binary image of the lung lobe (BL1). Eight rays crossing the positioned coordinate points with angles of 0°, 90°, 45°, 135°, 180°, 225°, 270°, and 315° from the horizontal were extended on BL1 until they intersected the horizontal axis. The maximum and minimum lengths of the 8 rays were removed, and the average length of the other 6 rays was calculated as AvgL. The DICOM image information "slice_Location" was used to determine the spacing of each layer (the slice location values of the current image (SL1) and the previous image (SL2) were obtained by parsing the DICOM file) to calculate the number of sheets that the nodule may cover (AvgL/(|SL1–SL2|)). The maximum area of the nodules was also in the middle. To avoid projecting excessive sheets, if the estimated image number exceeded the threshold (5), the number of sheets was reduced to the threshold as Series 1 (Ser1). A maximum intensity projection (35) was performed on Ser1 to obtain an image, IM3. The previous "lung lobe extraction" step process, M1, was used. A square using positioned coordinate points as the center and $2 \times AvgL$ as the side lengths was drawn, and the square region was defined as S1. The maximum value and the minimum values in S1 were extracted and used to perform a "bilateral filter" (36) to obtain the image (IM5). An iterative threshold algorithm was used on IM5 to obtain the threshold (T2). IM5 was binarized using a threshold T2. The "imclose" operation was subsequently performed on the image, and it was filled to obtain the lung nodule range (LNR).

Pulmonary nodule segmentation

After the image normalization, lung lobe extraction, and

lesion extension, the purpose of this step was to segment the pulmonary nodule. In this step, a single image was processed, and the radiologist specified a coordinate point. The subsequent images were defined by the result of the previous image (see step D for details). The starting image was in the range of the LNR.

(I) Eight rays originating from the specified coordinate points were drawn, and the pixel values were recorded.

(II) The unique absolute gradient values between the pixels on these 8 lines were calculated and denoted as Gr.

(III) The largest 20% of all gradients were extracted and denoted as S.

(IV) If the number of S was less than 4, Gr was used as S.

(V) The maximum of S was removed, and the remaining average of S was calculated and denoted as AvgS.

(VI) If RA was less than 0.031 and S was greater than 6, the average of the largest 6 values of S replaced AvgS.

(VII) If AvgS was greater than 0.1, AvgS was replaced with 0.08.

(VIII) All the pixel values of the LNR were sorted, and the pixel values between 50% and 85% were extracted and denoted as EP.

(IX) The 3rd quartile of EP was extracted.

(X) EP, AvgS, and IM1 were used as inputs, and a modified region-growing method was used to obtain the putative pulmonary nodule region, PN.

(XI) PN was optimized. A closing operation on PN was performed to obtain a series of regions. The largest region was selected and denoted as LR.

(XII) LR was projected onto the LNR. The region was denoted as LNI, and the regions of the LNR outside of the projection were denoted as LNO.

(XIII) The area of LNI and LNO were calculated, and the average pixel values of LNI and LNO were denoted as AI and AO, and PI and PO. The following were calculated: P=PI-PO, AS=AO/AI.

(XIV) If AS was greater than 0.1 and P was greater than 0 LNO was iterated to obtain a threshold and omit pixels larger than the threshold in PN as the target region of this image (TIM); if AS was in the 0.1–0.3 range, and P was less than 0, an open operation on PN was performed, and the TIM was obtained. If AS was greater than 0.3 and P was less than 0, PN was projected onto the LNR, and the TIM was obtained. For the other conditions, PN was defined as the TIM.

Loop and terminate

Because the CT was based on a series of images, the nodule segmentation continued to the next image. The average CT value of the TIM was then calculated. If it was not less than –775, and the difference in variance between this image and the previous image was less than 0.002 (the variance of the 0th TIM was defined to be same as the 1st TIM), the current image segmentation was considered to be valid, and the next image was processed. Otherwise, the loop was terminated. The starting coordinate point in the IM_{ai} step was defined as the gravity center of the TIM in IM_{ai-1}.

Table S1 Details of each lung nodule in the first dataset

No.	ReGANS				BR				Volume ratio
	LD (mm)	SD (mm)	Thickness (mm)	Volume (mm^3)	LD (mm)	SD (mm)	Volume (mm^3)	Thickness (mm)	
1	10.33	5.31	9	296.71	7.31	5.91	255.53	10	1.16
2	8.82	7.61	7	293.91	9.69	8.13	301.39	8	0.98
3	11.64	8.69	9	450.22	12.51	8.15	392.65	8	1.15
4	12.33	6.79	7	188.61	11.96	3.57	225.58	9	0.84
5	10.95	8.21	10	328.54	9.87	6.41	326.66	10	1.01
6	7.87	4.55	8	174.56	7.68	3.19	160.06	8	1.09
7	16.87	9.04	18	1,236.46	15.93	8.64	1,132.56	22	1.09
8	9.90	6.37	6	149.29	10.00	6.54	160.52	7	0.93
9	6.05	2.01	4	62.29	6.07	2.52	84.24	5	0.74
10	9.95	6.75	9	251.78	9.33	6.51	230.72	9	1.09
11	9.23	6.66	6	114.66	7.33	5.40	102.49	7	1.12
12	10.53	6.29	7	172.69	9.12	5.49	207.32	8	0.83
13	8.13	4.88	6	116.53	9.47	4.74	110.92	6	1.05
14	9.14	6.82	9	232.13	8.52	6.37	203.58	9	1.14
15	9.10	4.10	7	101.09	8.95	2.95	107.17	8	0.94
16	6.52	4.99	4	69.73	6.85	5.84	71.14	5	0.98
17	11.13	9.41	10	446.00	11.24	9.76	417.46	9	1.07
18	4.70	3.58	4	40.25	5.65	4.41	47.27	6	0.85
19	9.19	8.49	7	181.58	9.90	5.99	174.10	8	1.04
20	10.58	7.54	7	207.79	10.33	7.28	212.47	9	0.98
21	16.88	10.32	8	237.28	17.00	10.42	253.66	10	0.94
22	8.02	4.91	5	130.57	7.95	3.25	123.55	7	1.06
23	6.44	5.88	5	92.66	6.30	5.40	98.28	6	0.94
24	7.77	6.17	5	84.24	7.22	5.39	94.07	5	0.90
25	5.38	3.88	4	57.10	6.81	4.10	71.14	6	0.80
26	8.50	7.04	5	156.78	8.72	6.95	174.10	6	0.90
27	9.02	6.71	9	282.67	10.01	6.28	288.29	7	0.98
28	11.13	5.92	8	328.54	10.93	5.65	353.34	8	0.93
29	7.18	6.73	5	73.48	7.18	6.73	75.82	4	0.97
30	5.94	5.71	5	74.41	5.73	4.70	82.84	5	0.90
31	9.95	5.27	6	185.33	9.28	5.84	188.60	5	0.98
32	9.42	6.17	8	193.28	9.51	6.41	199.84	8	0.97
33	8.22	6.01	8	153.50	7.78	6.22	152.10	6	1.01
34	7.76	5.78	10	183.92	7.71	6.03	181.58	10	1.01
35	9.17	7.02	7	285.95	9.59	7.05	298.12	6	0.96
36	11.19	7.35	8	353.81	12.21	8.13	398.74	7	0.89
37	8.13	6.14	7	160.52	8.19	6.58	190.01	7	0.84
38	10.37	8.46	6	292.03	10.62	7.96	296.24	6	0.99
39	9.17	8.64	11	299.52	9.75	6.63	374.87	11	0.80
40	8.38	6.18	8	267.70	8.93	6.93	285.95	8	0.94
41	9.48	7.40	8	223.24	9.75	5.69	240.08	8	0.93
42	8.54	5.81	8	305.60	8.46	6.18	330.41	8	0.92
43	11.48	7.54	5	341.17	11.63	6.63	342.58	7	1.00
44	8.02	3.67	6	208.73	7.89	6.55	218.09	8	0.96
45	6.85	5.10	9	58.97	8.33	7.12	98.75	8	0.60
46	14.08	7.24	11	344.45	13.43	4.38	346.79	15	0.99
47	7.89	5.27	10	97.34	8.95	5.57	113.72	10	0.86
48	14.59	10.68	8	585.00	16.66	11.19	845.68^3	7	0.69
49	13.54	6.98	9	326.20	13.91	8.27	380.48	11	0.86
50	9.61	5.26	11	245.70	11.72	6.07	357.08	11	0.69
51	10.83	8.43	10	446.00	12.13	9.17	508.72	11	0.88
52	9.29	6.99	9	235.40	9.56	6.31	258.34	10	0.91
53	8.67	6.39	6	250.85	8.86	6.86	281.74	9	0.89
54	20.79	9.26	7	1,242.54	20.29	8.64	1,243.48	7	1.00
55	14.19	11.21	6	601.85	14.17	11.94	647.71	5	0.93

LD, long diameter of lung nodules; SD, short diameter of lung nodules.