A-LugSeg: Automatic and Explainability-Guided Multi-Site Lung Detection in Chest X-ray images

Tao Peng[a, 1], Yidong Gu[b, 1], Zhenyu Ye[c, 1], Xiuxiu Cheng[d, 1], Jing Wang[a, *]

[a] Department of Radiation Oncology, Medical Artificial Intelligence and Automation Laboratory, University of Texas Southwestern Medical Center, 2280 Inwood Road, Dallas, TX, United States of America

[b] Department of Medical Ultrasound, the Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Suzhou 215002, Jiangsu, China

[c] Department of Radiology, the Second Affiliated Hospital of Soochow University, 1055 San Xiang Road, Suzhou, Jiangsu, China

[d] Department of Ultrasound, the Affiliated Children's Hospital of Soochow University, Suzhou 215003, Jiangsu, China

Abstract

Large variations in anatomical shape and size, too much overlap between anatomical structures, and inconsistent anatomical shapes make accurate lung segmentation in chest x-rays (CXR) a challenging problem. In this paper, we propose an automatic method called A-LugSeg that consists of two subnetworks for lung segmentation in CXRs. The first is a segmentation subnetwork based on a deep learning model (i.e., Mask-RCNN), which completes a coarse segmentation for each input CXR image. The second is a refinement subnetwork designed to optimize the coarse segmentation result by combining an improved closed principal curve method and an enhanced machine learning, where the lung contour's explainability-guided mathematical model is expressed by the machine learning's parameters. The performance of the proposed method is evaluated on three public datasets, namely the ShenZhen hospital Chest X-ray dataset (SZCX), Japanese Society of Radiological Technology dataset (JSRT), and Montgomery County chest x-ray dataset (MC), which contain the 662 CXRs, 247 CXRs, and 138 CXRs, respectively. We used different datasets for training/validation (SZCX) and testing (SZCX/JSRT/MC). Furthermore, we used six evaluation metrics to evaluate the performance of our proposed method, including Dice Similarity Coefficient (DSC), Jaccard Similarity Coefficient (Ω), Accuracy (ACC), Precision, Sensitivity, and Specificity. The obtained results (DSC=0.973, Ω=0.958, ACC=0.972, and p-value for DSC <0.001) for JSRT, (DSC=0.971, Ω=0.955, ACC=0.97, and p-value for DSC <0.001) for MC, (DSC=0.972, Ω=0.956, and ACC=0.97) for hybrid datasets (JSRT+MC), and (Precision, Sensitivity, and Specificity are higher than 0.98) show the superior performance of the proposed dual subnetwork segmentation algorithm compared to the existing state of the art approaches.

1 Introduction

Chest radiography remains the most commonly used imaging modality for diagnosing various pulmonary diseases such as tuberculosis and lung cancer (Geetha et al., 2021). Because of its low cost, low dose of radiation, and widespread availability, hundreds to thousands of chest x-ray images (CXRs) are generated in a typical

---

[1] All authors contributed equally to this work.
[*] Corresponding author.
E-mail addresses: sdpengtao401@gmail.com (Tao Peng), guyidong850820@aliyun.com (Yidong Gu),
yezhenyu2006@sina.com (Zhenyu Ye), xiuxiu001465@163.com (Xiuxiu Cheng),
Jing.Wang@utsouthwestern.edu (Jing Wang).

hospital daily, which creates significant diagnostic workloads (Chandra et al., 2020). Using a Computer-Aided Diagnosis system (CAD) to automatically interpret these CXRs can significantly reduce the diagnosis time and greatly help clinical radiologists in effective decision making. Accurate lung segmentation is often a necessary step for quantitative lung image analysis in CAD, as it is important for identifying lung diseases in clinical evaluations (Vidal et al., 2021; Chakraborty, & Mali, 2021).

Although lung segmentation in CXRs has been researched extensively over the last four decades (Chondro et al., 2018), it is still a challenging task due to the following reasons. (1) There are large variations in anatomical shape and size between different patients due to gender, age, and genetic variations (Candemir et al., 2014). (2) There are overlapping parts between anatomical structures, such as the heart, rib cage, and clavicle bones. (3) Inconsistencies in anatomical shape between different individuals make segmenting small costophrenic angles challenging. Fig. 1 shows two sample CXR images that illustrate these challenges associated with accurate CXR segmentation.
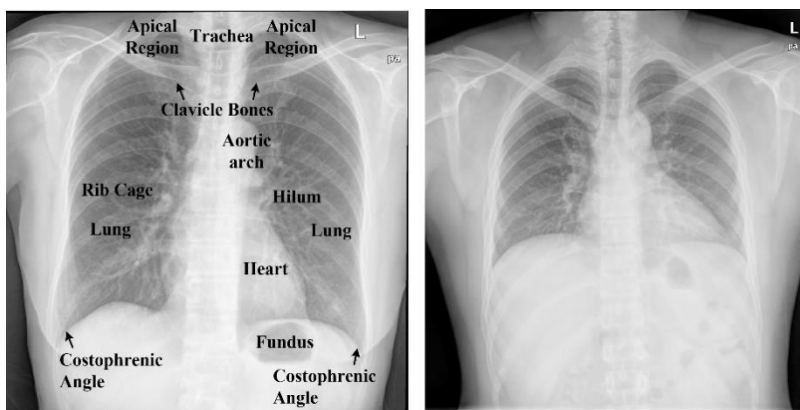

Fig. 1. Two randomly selected CXRs on Posterior-Anterior view and their differences. Variable shapes of the lung, visible heart shape, strong edges of the rib cage, and sharp corner at the costophrenic angle make it challenging for automatic lung segmentation methods.

Many different segmentation methods have been explored, including (i) feature classification methods (Shu et al., 2021), (ii) region segmentation methods (Zeng et al., 2019; Zeng, Li, Wang, et al., 2021), and (iii) contour detection methods (Novikov et al., 2018). Shen et al., (2015) presented an automated lung segmentation method that uses bidirectional chain codes to minimize the over-segmentation of neighboring regions. The average over-segmentation rate was as low as 0.3%, but it sometimes failed to re-include the juxtapleural nodules sitting in consolidation regions. Wan Ahmad et al., (2015) proposed a Content-Based Medical Image Retrieval System (CBMIRS) for lung segmentation, but the Jaccard Similarity Coefficient (Ω) was only 0.87 when they tested the method on a public dataset. Alternatively, the shape of anatomical structures can be represented in the form of points by the contour detection method (Tu et al., 2018). Compared with the other methods, the contour detection method easily obtains the shape of the anatomical structure with less time complexity (Peng et al., 2019).

The main purpose of the contour detection method is to approximate the contour of organs by using region representation (Wang et al., 2015) or curve approximation (Afzali et al., 2018). Souza et al., (2019) investigated a Deep Convolutional Neural Network method (DCNN) for automatic lung segmentation with good performance. However, this method decreased some of the performance metrics during the reconstruction step. Yang et al., (2018) presented a structured edge detector (SED) for lung field segmentation in chest radiographs with high accuracy, but accurately segmenting abnormal portions of the lungs with this method is still challenging. Farhangi et al., (2017) proposed a lung nodule segmentation method based on Sparse Linear Combination of Training Shapes (SCoTS), but this method's performance depends on the diversity and accuracy of the training shapes.

The principal curve algorithm has been widely used for separating abnormal anatomical structures from other

neighboring normal organs, because it handles noisy input well and obtains robust results (Peng et al., 2019). The principal curve is a smooth curve that passes through the middle of an n-dimensional dataset, providing a nonlinear description of the n-dimensional dataset (Hastie, & Stuetzle, 1989). The principal curve has been shown to be effective in contour detection. Ataer-Cansizoglu et al., (2013) proposed a contour-based shape representation method based on the principal curve and tested it on a lung dataset with high accuracy. Khedher et al., (2015) combined the Partial Least Squares, principal curve and Support Vector Machine for medical imaging segmentation. Meanwhile, machine learning has been used widely in a variety of medical image applications. In the field of disease detection, combining machine learning with the principal curve-based method has shown promising results for disease detection (Peng, Xu, Wang, Zhou, et al., 2020; Zeng, Li, & Peng, 2021).

In the present work, we propose a novel automatic lung segmentation framework called A-LugSeg for multi-site CXR datasets. The proposed framework consists of two subnetworks: 1) the Mask Region-based Convolutional Neural Network algorithm (Mask-RCNN) (He et al., 2017) to complete the coarse segmentation of the lung and 2) a hybrid refinement step to optimize the final segmentation results. In the hybrid refinement step, we combine the improved principal curve and enhanced machine learning methods. Our contributions/novelties are summarized as follows:

- We present a coarse-to-fine cascade segmenting framework that automatically segments the lung contour on CXRs while minimizing the impact of other neighboring organs.
- In our segmenting framework, the coarse segmentation is achieved using Mask-RCNN, while the hybrid method combines an Adaptive Closed Polyline Searching algorithm (ACPS) with a Fractional-Order Backpropagation Learning algorithm (FOBL) being used as the refinement step.
- Compared with the traditional Polyline Searching algorithm (PS) (Kégl et al., 2000; Kégl, & Krzyzak, 2002), our previous works (Peng et al., 2019; Peng, Xu, Wang, Zhou et al., 2020; Peng et al., 2018) propose an improved Closed Polyline Searching algorithm (CPS) for the first time by adding several limitation conditions. Based on our previous works, inspired by Ref. (Zhang et al., 2013; Yang et al., 2014), we proposed the ACPS by adding certain improvements, including the improved projection step, improved vertex optimization step, and adaptive radius selection method.
- Compared with the traditional Back Propagation algorithm (BP), inspired by Ref. (Wang et al., 2017; Chen et al., 2020), we proposed the FOBL, which used the Caputo-type fractional gradient descent algorithm-based backpropagation step.
- We proposed a smooth explainability-guided mathematical model to express the lung contour, where the mathematical model is the first time to be denoted by the parameters of the FOBL.

The rest of the paper is organized as follows. Section 2 introduces our method, including Mask-RCNN, ACPS, FOBL, and the explainability-guided mathematical model. In Section 3, we introduce the datasets and evaluation metrics. In Section 4, we present the experimental results. The last section concludes the paper.

2. Methods

Our method aims to achieve accurate lung segmentation to aid in lung disease diagnosis. Fig. 2 shows the framework of our proposed method for segmenting lungs from CXRs, including a flowchart of the segmentation pipeline. The pipeline consists of a preprocessing step (stage 1) and a refinement step (stage 2).
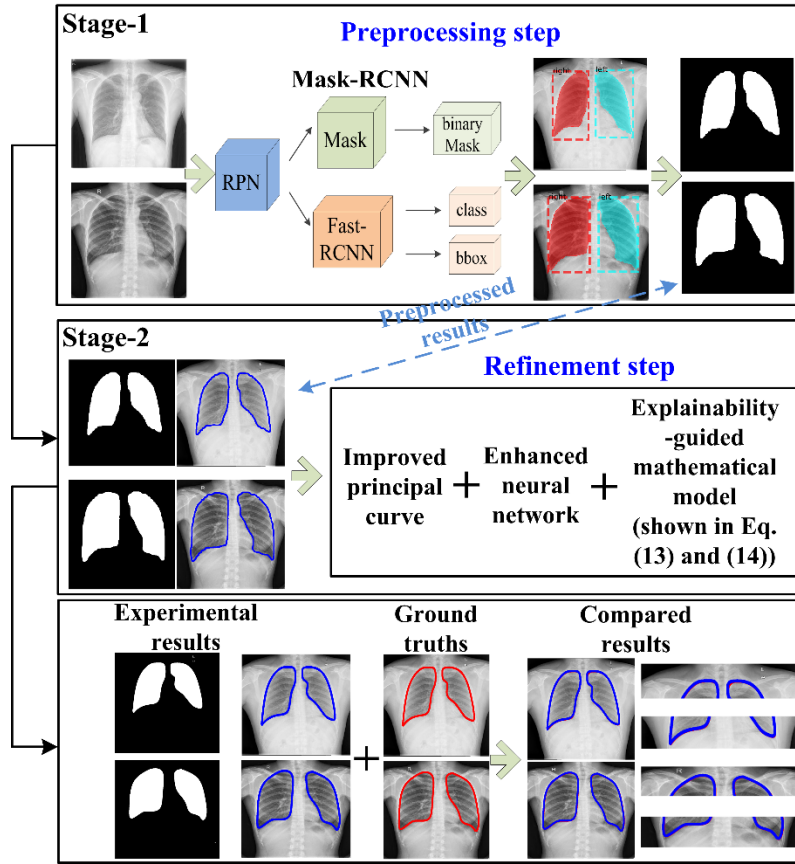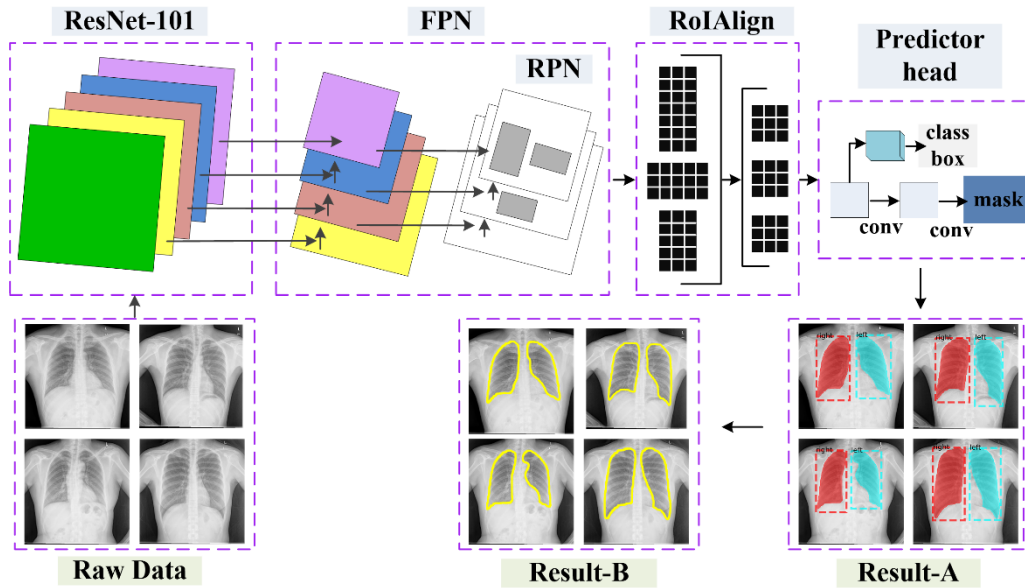
Fig. 2. The framework of our proposed method.

2.1 Preprocessing step

The preprocessing step completes the coarse segmentation by using the Mask-RCNN, which is a recently proposed general framework for object detection and segmentation. The details of the Mask-RCNN are illustrated in Fig. 3. The network layout of the Mask-RCNN is detailed below.

Note:
Result-A shows the experimental results consisting of contour, box, mask, and annotated name
Result-B shows the experimental results which only consists of the contour

Fig. 3. Illustration of the Mask-RCNN for the coarse segmentation.

We used Mask-RCNN as a coarse segmentation step mainly due to the following reason: Mask-RCNN was the COCO 2016 challenge winner on single-model entry, which achieved good accuracy and robustness (Cai et al., 2020) for object detection and has been generalized to other tasks such as medical imaging segmentation, achieving very promising results (Privalov et al., 2021; Vuola et al., 2019; Gamage et al., 2019; Bagheri et al., 2021).

The Mask-RCNN was presented in 2017 and extends the Faster RCNN (Ren et al., 2017) for objection localization and object instance segmentation of natural images. Mask-RCNN's main improvements over the Faster RCNN include: 1) Mask-RCNN uses a Region of Interest Align (RoIAlign) operation to replace the less precise ROI-Pooling operation used in the Faster RCNN, which results in more accurate segmentation masks; 2) Mask-RCNN extends a predictor head with a fully convolutional mask prediction branch to produce the desired instance segmentation (Fig. 3); and 3) Mask-RCNN uses decoupled mask and class predictions, where the mask network head predicts the mask independently from the network head that predicts the class (Fig. 3).

Mask-RCNN is built on a backbone convolutional neural network architecture for feature extraction. In principle, the backbone network could be any convolutional neural network (CNN) designed for image analysis (i.e., ResNet-50 or ResNet-101) (He et al., 2016), but it has been demonstrated that combining a feature pyramid network (FPN) with a network (i.e., ResNet-50 or ResNet-101) as the backbone of the Mask-RCNN increases both its accuracy and its speed (He et al., 2017).

The properties of natural images differ from those of medical images, but given the effectiveness of Mask-RCNN for general-purpose object segmentation, we investigated its performance for coarse lung segmentation in CXRs. Because combining ResNet-101 with FPN achieves higher accuracy than ResNet-50 (He et al., 2016), we used the FPN and ResNet-101 for the backbone network of Mask-RCNN.

In this work, we set the configuration of Mask-RCNN as follows: we used a pre-trained weight on the COCO dataset (Lin et al., 2014) to handle the limited training data. In Region Proposal Network (RPN), if a ROI has an Intersection Over Union (IoU) with ground truth less than 0.5, it is regarded as negative, otherwise it is regarded as positive. Each mini-batch has 2 images per GPU and each image has 256 ROIs, where the ratio of positive to negative is 1:3. Furthermore, we set the anchor ratio scale at (16, 32, 64, 128, 256). In total, we trained for 100 epochs using stochastic gradient descent with momentum of 0.9, where we set the constant learning rate of 0.001, and a weight decay of 0.0001. Furthermore, Mask-RCNN produces the predicted detection outputs as the bounding boxes. Each bounding box is associated with a detection confidence score. All the boxes below the detection confidence threshold parameter are ignored for further processing. At last, we accepted the detection confidence level threshold of 0.7.

2.2 Refinement step

The goal in this step is to increase the accuracy of lung contours by fine-tuning the coarse segmentation result via an improved ACPS coupled with the improved FOBL.

2.2.1 ACPS

Principal curves were firstly defined by Hastie, & Stuetzle (1989) as "self-consistent" smooth curves which pass through the "middle" of a n-dimensional probability distribution or data cloud, whereas the medial axis is a set of smooth curves that go equidistantly from the contours of a character. To improve the efficiency and robustness of our previous work, we proposed the ACPS, where the workflow of the ACPS is shown in Fig. 4, respectively.
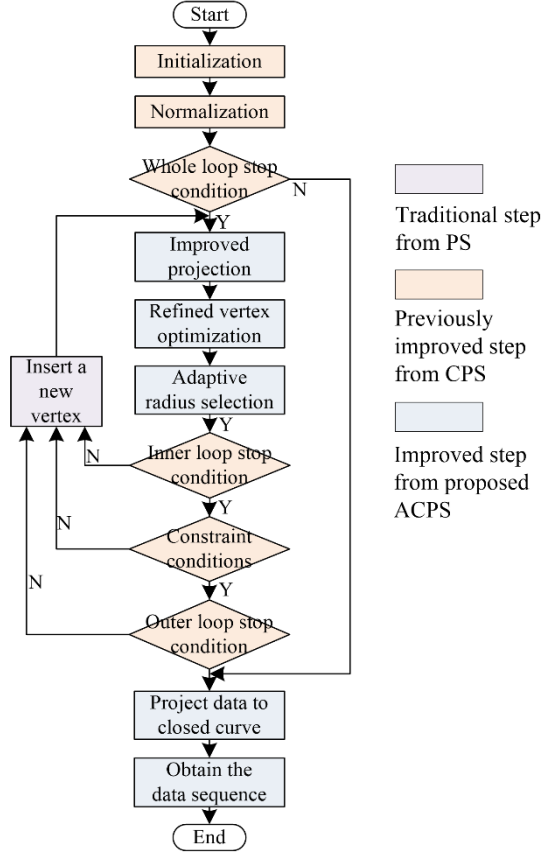
Fig. 4. The workflow of the ACPS (purple box shows the traditional steps from PS, orange box shows the improvement steps by our previously proposed CPS, and light blue box shows the improvements steps by this work).

A. Closed Polyline Searching algorithm (CPS)

Kégl et al., (2000) first proposed the traditional PS to find the principal curve from the data cloud. However, the main challenge for the actual principal curve is to correctly describe the projection index of the dataset to produce a closed and smooth curve (Peng et al., 2019). To handle this issue, we previously proposed an improved CPS (Peng et al., 2019; Peng, Xu, Wang, Zhou et al., 2020; Peng et al., 2018). Compared with the traditional PS, there are several improvements in CPS, such as initialization, normalization, stop conditions, and constraint conditions, as shown below,

Initialization and normalization steps: To deal with the closed dataset, in the improved initialization step, we use a closed square as the first principal component line, where the four vertices of the closed square are (0.05, 0.05), (-0.05, 0.05), (-0.05, -0.05), and (0.05, -0.05). Then, the initial step of the PS is removed, and the new initial step is added, as shown in Fig. 4. In the added normalization step, the coordinates of data points x_i are unified into the range of {(-1,1)}.

Stop conditions: The stop conditions are mainly used in the whole loop, outer loop, and inner loop. The whole loop is used as the global constraint loop and executes after the normalization step. The whole loop will exit when k satisfies,

$$k > c\left(n, \Delta_n(f_{k,n})\right) = \beta n^{1/3} \Delta_n(f_{k,n})^{-1/2} r \tag{1}$$

where the robustness of the principal curve method is determined by the selection of the number of segments k, which depends on the average squared distance $\Delta_n(f_{k,n})$ (Kégl, & Krzyzak, 2002). The parameter β is set as 0.3

determined by experimental results (Kégl et al., 2000), n shows the number of data points, f denotes the principal curve, and r is the radius of data points.

Both the inner and outer loops need to satisfy the condition that the difference in value between the current distance and the last loop distance is smaller than the maximum distance deviation $\Delta s=0.002$. The distance deviation Δs used in both the outer and inner loops was determined by several trial runs.

Constraint conditions: The innovations of this part mainly focus on the optimal selection of vertices, line segments, and the shape of the contour curve.

The main steps in selecting the optimal vertices are the following: 1) a new vertex is inserted, and the whole dataset is chosen to project the segments or the determined vertices; 2) the distance function from data points to the curve is calculated; and 3) the position of each vertex is adjusted only when the value of the distance function becomes smaller; otherwise, another new vertex is selected to start a new test.

In selecting the line segments, the longest segment will be chosen because the line segment needs to contain as many projection points as possible. The obtained contour curve stays closed all the time.

B. ACPS

When dealing with complex data, PS and CPS often have low robustness. In this work, inspired by the works (Zhang et al., 2013; Yang et al., 2014), we propose a new ACPS by adding three improvements based on the CPS, including, 1) improved projection step, 2) improved vertex optimization step, and 3) adaptive selection of radius of data points.

Improved projection step: In the traditional projection step, the data points are classified according to which segment or vertex they project (Peng et al., 2019). During this process, the entire principal curve is scanned to partition each data point x_i into the nearest neighbor vertices or segments (Kégl et al., 2000). Due to too many required scans, this process is time-consuming when dealing with a large number of complex data. In this work, we only scan certain regions of the principal curve around the data point x_i and partition it into the nearest neighborhood region to take the place of the entire principal curve. If the data point x_i finds the suitable project region, then the scanning process of this data point ends.

Improved vertex optimization step: In our previously proposed CPS, the curvature penalty P(f) is denoted by,

$$P(f) = \frac{1}{m} \sum_{i=1}^{m} P_v(v_i) \tag{2}$$

where m is the number of vertices, the curvature penalty function $P_v(v_i)$ on the vertex v_i is a triangle-based function (Zhang et al., 2013). Furthermore, we set the average squared distance $\Delta(f)$ is from data point c to the nearest segments or vertices of the principal curve f, shown as below,

$$\Delta(f) = \frac{1}{n} \sum_{i=1}^{n} \Delta(c_i, f) \tag{3}$$

where n is the number of data points. The penalized distance function E(f) in the vertex optimization step is shown as below:

$$E(f) = \Delta(f) + \lambda P(f) \tag{4}$$

where λ is the penalty factor.

Considering that P(f) uses triangle functions for calculating that decrease the efficiency of the method, this work uses a new constraint function D(f) to take the place of P(f) in the vertex optimization step, where D(f) only uses the addition and average for calculating as follows:

$$D(f) = \frac{1}{m}\sum_{i=1}^{m}\Delta(c,v_i) \tag{5}$$

where $\Delta(c, v)$ denotes the distance between a data point c to the nearest vertex v. Then, the new penalized distance function $E_{new}(f)$ can be denoted as below,

$$E_{new}(f) = \Delta(f) + \lambda D(f) \tag{6}$$

Adaptive radius selection method: As shown in Eq. (1), the data points' radius r determines the value of the number of segments k, where k is used to balance between the accuracy and efficiency of the method. In this work, we proposed an adaptive radius selection method to improve the adaptive ability of the proposed method. Detailed steps are summarized below:

(i) the vertices (v_i, i=1,2,..,num) are first determined, and vertices' number is *num*.

(ii) through scanning all the data points, the minimum distance DS_{min} and maximum distance DS_{max} from data points to the principal curve are calculated, respectively.

(iii) if the *num* is larger than 3, it turns to the next step.

(iv) a middle point *cm* is found by simple averaging (Yang et al., 2014), and the distance DS_{cm} from the middle point *cm* to the principal curve f is calculated.

(v) if the coordinate of *cm* is different from the coordinate of any determined vertex, it turns to the next step.

(vi) the distance DS_{i-1i} between vertex v_{i-1} and vertex v_i is calculated.

(vii) we set the initial interval radius within the range of $[r_d, r_u]$, where initial r_d and r_u equal to DS_{i-1i} and $2DS_{i-1i}$, respectively. Furthermore, we suppose the suitable radius $r=r_d+(r_u-r_d)/2$.

(viii) if it meets the condition that $DS_{min}<= DS_{cm} <=DS_{max}$, *cm* is regarded as a newly added vertex and r is determined; or it turns to the next step.

(ix) if $DS_{cm}<DS_{min}$, we set $r_d=r$; or if $DS_{cm}>$ or $= DS_{min}$, we set $r_u=r$.

(x) the data points' radius r is obtained.

2.2.2 FOBL using the Fractional-order gradient descent learning with Caputo derivative

The Caputo-type fractional gradient descent algorithm has good memory and heredity (Wang et al., 2017). To improve the performance of traditional BP, inspired by Ref. (Wang et al., 2017; Chen et al., 2020), we proposed the FOBL using the Fractional-order gradient descent learning with Caputo derivative, which was used to replace the gradient descent learning of the BP and update the weights vector in the backpropagation step. Furthermore, the forward propagation step of both methods (BP and FOBL) is the same, where the details can be found in the work (Chen et al., 2020).

A. Overall

In this section, we introduce the FOBL with three layers, including the input layer, hidden layer, and output layer, as a fully connected neural network. The FOBL mainly contains two steps, including the forward and back propagation steps. *First*, in the forward propagation step, the data sequence $d=\{d_i=(t_i, c_i(x,y)),i=1,2\cdots,n; t_1<t_2<\cdots<t_n\}$ obtained by ACPS is used by the FOBL, where projection index t is used as the input of the FOBL's input layer, and the coordinates of the data point c(x,y) are used as the expected output to calculate the total error E. Meanwhile, Sigmoid activation function $g_1=1/(1+e^{-x})$ and Tanh activation function $g_2=(e^x-e^{-x})/(e^x+e^{-x})$ are used in the hidden layer and output layer, respectively. *Secondly*, we use the improved back propagation step, where the main improvements are shown in the next section. Table 1 shows the key parameters of the FOBL, where # denotes the parameters optimized in this work, and * denotes the selection of parameters, discussed in the work (Peng et al., 2018).

Table 1 The parameters of the FOBL.

Type of the layer of FOBL	Parameters of each layer	Parameters of the whole model
Input Layer	#Input: projection index (t)	
	#Number of hidden layers	
	*Hidden neurons (h)	
Hidden Layer	*Epoch	
	*Learning rate	* weights vector (w₁, w₂)
	*Momentum	
	#Threshold (a)	
	#Activation function: g₁(x)	
	#Output (x, y)	
Output Layer	#Threshold (b)	
	#Activation function: g₂(x)	

B. Improvements of proposed FOBL based on BP

Compared with the BP, we introduce the improvements of the FOBL, which are mainly in the backpropagation step. The details of differences between these two methods (BP and FOBL) are summarized below.

In the traditional BP, to minimize the total error E, the weights vector $\{w_1, w_2\}$ are updated at the (r+1)-th training iteration by the gradient descent method in the backpropagation step, shown as the following rule:

$$w_2(r+1) = w_2(r) - \eta_1 \times \frac{\partial E(w_2)}{\partial v(r)}$$
$$= w_2(r) - \eta_1 \times \sum_{s=1}^{N} g_k'(\sum_{k=1}^{O} w_2(r) \times g_1(w_1(r) \times t_s - a_j) - b_k) \times g_1(w_1(r) \times t_s - a_j) \tag{7}$$

$$w_1(r+1) = w_1(r) - \eta_2 \times \frac{\partial E(w_1)}{\partial w(r)}$$
$$= w_1(r) - \eta_2 \times \sum_{s=1}^{N} g_k'(\sum_{k=1}^{O} w_2(r) \times g_1(w_1(r) \times t_s - a_j) - b_k) w_2(r) \times g_1'(w_1(r) \times t_s - a_j) t_s \tag{8}$$

where the projection index t is the input of input layer and N is the number of points of coarse segmentation result. w denotes the weight from the i-th input neurons to j-th hidden neurons and v denotes the weight from the j-th hidden neurons to k-th output neurons. η_1 and η_2 are the learning rates from input to the hidden layer and from hidden to the output layer, respectively, where both η_1 and η_2 are in the range of [0, 1]. Furthermore, constructed function $g_k(\bullet)$ is the composite function of $g_1(\bullet)$ and $g_2(\bullet)$, shown as $g_k(r) = \frac{1}{2N}(\exp - g_2(r))^2$, where *exp* is the expected results. $a_j(j=1,2,…,h)$ and $b_k(k=1,2)$ are the thresholds of the j-th hidden neuron and the k-th output neuron, respectively.

In our proposed FOBL, the backpropagation step is trained by the improved fractional-order gradient descent method with the Caputo derivative (Wang et al., 2017). The weights vector $\{w_1, w_2\}$ are updated as follows,

$$w_2(r+1) = w_2(r) - \eta_1 \times Caputo(w_2(r)) \times \frac{\partial E(w_2)}{\partial v(r)} \tag{9}$$

$$w_1(r+1) = w_1(k) - \eta_2 \times Caputo(w_1(r)) \times \frac{\partial E(w_1)}{\partial w(r)} \qquad (10)$$

The Caputo($\cdot$) function is defined as:

$$Caputo(w_1(r)) = \frac{1}{(1-\alpha) \times \Gamma(1-\alpha)} \times w_2(r) \times sum'(w_1(r) \times t_s) \times t_s \times (w_1(r) - ap)^{1-\alpha} \qquad (11)$$

$$Caputo(w_2(r)) = \frac{1}{(1-\alpha) \times \Gamma(1-\alpha)} \times sum(w_1(r) \times t_s) \times (w_2(r) - ap)^{1-\alpha} \qquad (12)$$

where α is the fractional order in the range of [0, 1]; Γ shows the Gamma function; objective function sum($\cdot$) is the sum function (Chen et al., 2020) and adjustment parameter $ap = \min\{w_1(r), w_2(r)\}$.

2.2.3 An explainability-guided mathematical model for obtaining the smooth contour

The contribution of this step is to develop a smooth mathematical model of the lung, which is denoted by the parameters of the FOBL (shown in Eq. (13) and Eq. (14)). By using the FOBL for training to minimize the total error E, this step fine-tunes the coarse segmentation results. After completing training, the newly proposed mathematical expression of the lung contour can be obtained as follows:

$$f(t) = \left(X(c(x(t))), Y(c(y(t))) \right) = \left(\frac{c(x(t)) + 1 - \sqrt{1 - (c(x(t)))^2}}{2 \times c(x(t))}, \frac{c(y(t)) + 1 - \sqrt{1 - (c(y(t)))^2}}{2 \times c(y(t))} \right) \qquad (13)$$

where $X(\cdot)$ and $Y(\cdot)$ show the x-axis and y-axis coordinates of the points of the generated lung contour, respectively. $c(\cdot)$ shows the output of the FOBL's output units, corresponding to $c(x(t))$ and $c(y(t))$, shown as below,

$$c(x, y) = \left(c(x(t)), c(y(t)) \right) = \left(\frac{e^{\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,1} \cdot b_1} - e^{-(\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,1} \cdot b_1)}}{e^{\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,1} \cdot b_1} + e^{-(\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,1} \cdot b_1)}}, \frac{e^{\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,2} \cdot b_2} - e^{-(\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,2} \cdot b_2)}}{e^{\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,2} \cdot b_2} + e^{-(\sum_{j=1}^{h} \frac{1}{1+e^{-(tw1j-a_j)}} w2_{j,2} \cdot b_2)}} \right) \qquad (14)$$

where $c(x(t))$ and $c(y(t))$ are the two output units. $x(t)$ and $y(t)$ can be regarded as the continuous functions, respectively, on projection index t, where the projection index t is regarded as the independent variable, and the x-axis coordinates $x(t)$ and y-axis coordinates $y(t)$ of the data points are the dependent variables.

3 Multi-site datasets and evaluation metrics

In this study, we evaluated our proposed method by using three different CXR datasets: 1) Shenzhen hospital chest x-ray dataset (SZCX) (Jaeger et al., 2014), 2) Japanese Society of Radiological Technology dataset (JSRT) (Shiraishi et al., 2000), and 3) Montgomery County chest x-ray dataset (MC) (Jaeger et al., 2014).

- SZCX: This dataset is collected by Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China. SZCX dataset contains 662 CXRs, where 326 are normal cases and the remaining 336 are abnormal cases with tuberculosis manifestations.
- JSRT: The JSRT dataset is one of the most widely used public lung CXR datasets, which has been available more than two decades ago. All the CXRs are represented by the 12-bit gray-level image with the size of 2048 × 2048 pixels.
- MC: The MC dataset is collected in collaboration with the Department of Health and Human Services of Montgomery County, Maryland, USA. This dataset consists of 138 posterior-anterior CXRs, where 80 CXRs are normal and the other 58 CXRs are abnormal with different degrees of tuberculosis manifestations.

We evaluated the model's performance on segmentation accuracy by using several parameters, including Dice Similarity Coefficient (DSC)/F1-score[2], Jaccard Similarity Coefficient (Ω)/Intersection-Over-Union (IoU)[3], and Accuracy (ACC). These parameters are defined as follows:

$$DSC\ (F1-score) = \frac{2TP}{2TP+FP+FN} \tag{15}$$

$$\Omega\ (IoU) = \frac{TP}{FP+TP+FN} \tag{16}$$

$$ACC = \frac{TP+TN}{TP+FN+FP+TN} \tag{17}$$

where TP, FP, FN, and TN represent True Positive, False Positive, False Negative, and True Negative, respectively.

4. Experimental and results

We first used the SZCX dataset to determine the hyperparameter (i.e., hidden neurons and epochs) of proposed method during both training and validation steps (Section 4.1), where we used 400 of them for training, 162 for validation, and the other 100 for testing. Then, we used all 100 testing cases, comprising 50 normal and 50 abnormal cases, for qualitative and quantitative evaluation (Section 4.2). In Section 4.3, we used the other two datasets (JSRT and MC) to test the stability and generalizability of the model obtained in Section 4.2. Finally, we compared our proposed method with other state-of-the-art methods (Section 4.4).

All the ground truths of lung contours have been marked and verified by five board-certified radiologists. Each radiologist independently checked their own marks along with the anonymous marks of the other radiologists, and the consensus ground truths were obtained by the majority voting of five experts' annotations. All experiments have been implemented on a computer with an Intel Core i5-9300H CPU and a Geforce GTX 1650 GPU with 4GB memory.

4.1. Determining the hyperparameter of proposed method on SZCX

In this section, we evaluated the impact of the hyperparameter (i.e., hidden neurons and epochs) on the performance of refinement step of the proposed method. Due to the image resolution of each dataset is different, each testing case was down-sampled to 256×256. After coarse segmentation stage, the DSC, Ω, and ACC of the testing results of Mask-RCNN are 83.6%, 82.1%, and 83.1%, respectively, where the detail of configuration of Mask-RCNN is shown in Section 2.1. Then, we use the refinement step to fine-tune the coarse segmentation results.

During training and validation stages, Fig. 5 and Fig. 6 show the results obtained with different hidden neurons and epochs, respectively. Brown, blue, and green curves show the change trends of the DSC, Ω, and ACC, respectively. Each dot (i.e., brown rhombic, blue triangle, and green square) shows the average result at different hidden neurons or epochs, respectively. Overall, the trend of training results is more stable, while the curve of the validation results is more oscillated. To the best of our knowledge, the validation set is mainly used to determine the optimal model (Yu et al., 2018), and we will mainly discuss the validation results here.

To investigate the impact of hidden neurons, we set the constant epochs of 1000. From Fig. 5 (b), when the hidden neuron is 1, all the metrics (i.e., DSC, Ω, and ACC) of the validation result were only near 86%. The main reason is that due to the limited hidden neurons, FOBL may not handle the complex issue successfully. After increasing the hidden neurons, the validation results increase as well. When the number of hidden neurons rises to 10, we can obtain the optimal performance of the model, where the DSC, Ω, and ACC of the validation result are 96.8%, 95.3%, and 96.4%, respectively. After that, the hidden neurons continue increasing, the validation

---

[2] https://en.wikipedia.org/wiki/F-score
[3] https://en.wikipedia.org/wiki/Jaccard_index

accuracy starts to decrease, where overfitting occurs.
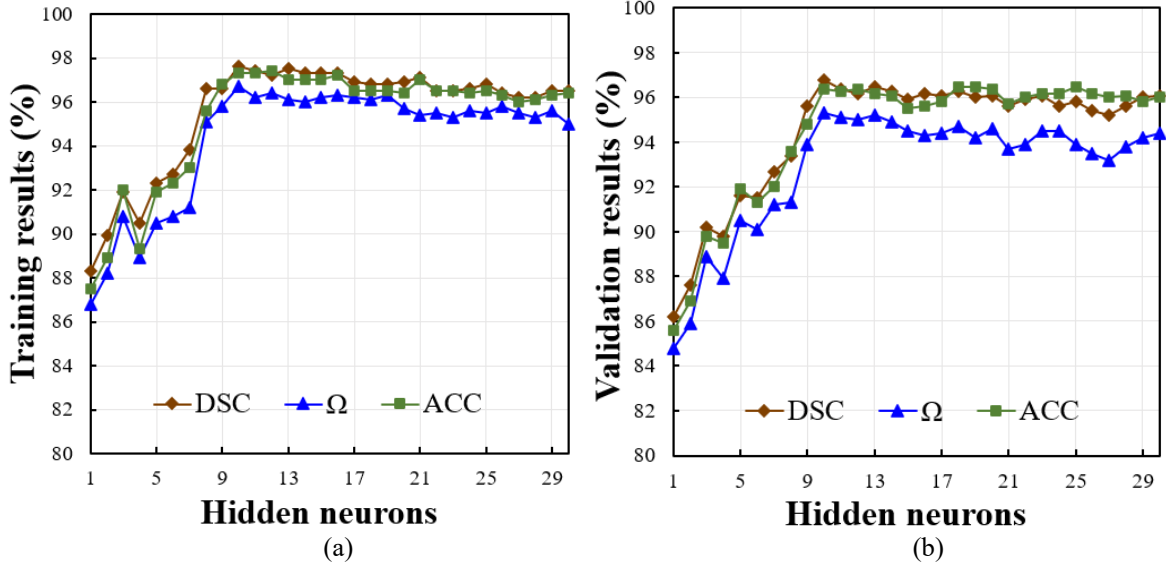

(a)   (b)

Fig. 5. Training and validation results at different hidden neurons.

According to the previous results, we used the optimal selection of 10 neurons in the following experiments, and investigated the impact of epochs for the performance of the proposed method, shown in Fig. 6. From Fig. 6 (b), the metrics increase at the beginning stage and remain stable after reaching the peak. When it reaches 1000 epochs, the model gets the best model.
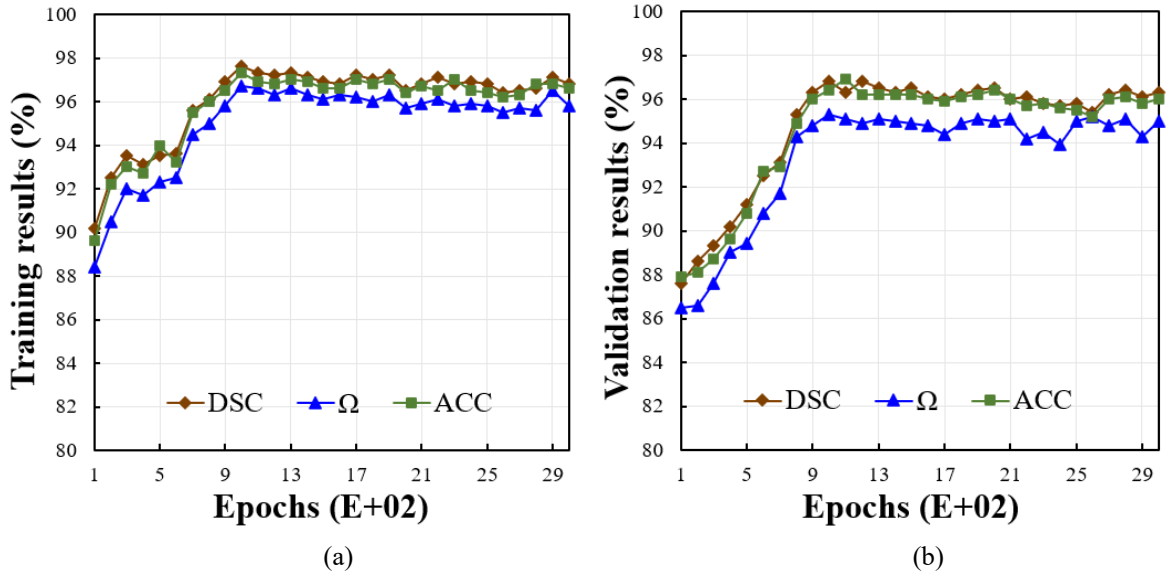

(a)   (b)

Fig. 6. Training and validation results at different epochs.

4.2. Segmentation performance on SZCX

Based on the previous results, we set 10 neurons and 1000 epochs for the following experiments. We first quantitatively evaluated the performance of the proposed method on different resolutions, where each testing case was downsampled to 256×256, 512×512, and 1024×1024. Then, we randomly selected one slice for step-by-step comparison and chose another eight results—five normal cases and three nodule-containing cases—from the 100 testing cases at 256×256 resolution for the global and local qualitative display. We carried out the global qualitative comparison step by step and illustrated the Apical Regions (AR) and CostoPhrenic angle Regions (CPR) of the global results in Section 4.2.1. Finally, we used several evaluation metrics to test the stability of the proposed method, with each image at a resolution of 256×256 (Section 4.2.2).

4.2.1. Qualitative evaluation

A. Step by step comparison

The qualitative testing experimental results are denoted step by step, as shown in Fig. 7. The input of the proposed method is the raw data and GT shows the ground truth marked and verified by five board-certified radiologists (first column). The preprocessing step of the proposed method is Mask-RCNN (second column). The refinement step combines ACPS with FOBL, where ACPS is used obtain the data sequences and closed curve consisting of segments, and FOBL is used to train the data sequences and denote a mathematical model of smooth contour using its parameters. The proposed ACPS is used to update the positions of all the vertices according to the principle of minimizing the penalty distance function (Kégl et al., 2000), while the position of each line segment is also adjusted. Therefore, the intermediate result can be achieved, which is obtained by ACPS using the results of Mask-RCNN as the input (third column). The final result shows the experimental result compared with the ground truth (fourth column). From the second row (results of different steps) of Fig. 7, we can find that the result of Mask-RCNN is the worst, while the results between ACPS and proposed method are similar. From the third row (results of ROI) of Fig. 7, we can find that the result of proposed method is more accurate and smoother.
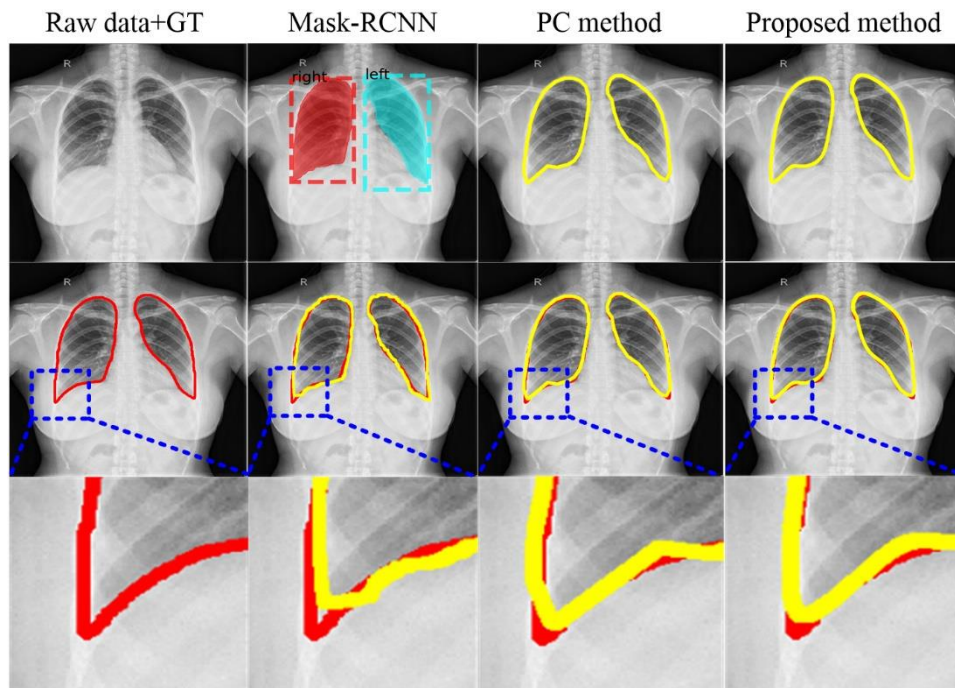

Fig. 7. Qualitative discussion of experimental results step by step. The first column shows the raw data, Ground Truth (GT), and the ROI of GT. The second, third, and fourth row show the results of Mask-RCNN, principal curve method, and our proposed method, respectively. Meanwhile, the second, third, and fourth row show the segmented result, compared result, and ROI of the compared result, respectively. Yellow curve shows the experimental result and red curve shows the GT.

B. Global comparison

Fig. 8 shows eight CXR results (Lung A-Lung H) randomly selected from the 100 testing results for qualitative evaluation. The first five cases (Lung A-Lung E) are normal, and the last three cases (Lung F-Lung H) are from patients with lung disease. Each row shows the flow through which the segmentation results were obtained. "Preprocessed Result-A" denotes the results obtained by the Mask-RCNN, which consist of the mask, annotated name, box, and ROI contour. "Preprocessed Result-B" shows only the contour. The final result, called

"Compared with GT," shows the experimental result obtained after the refinement step, and it is also compared with the ground truth. As shown in Fig. 8, the experimental results have good similarities with the ground truths.
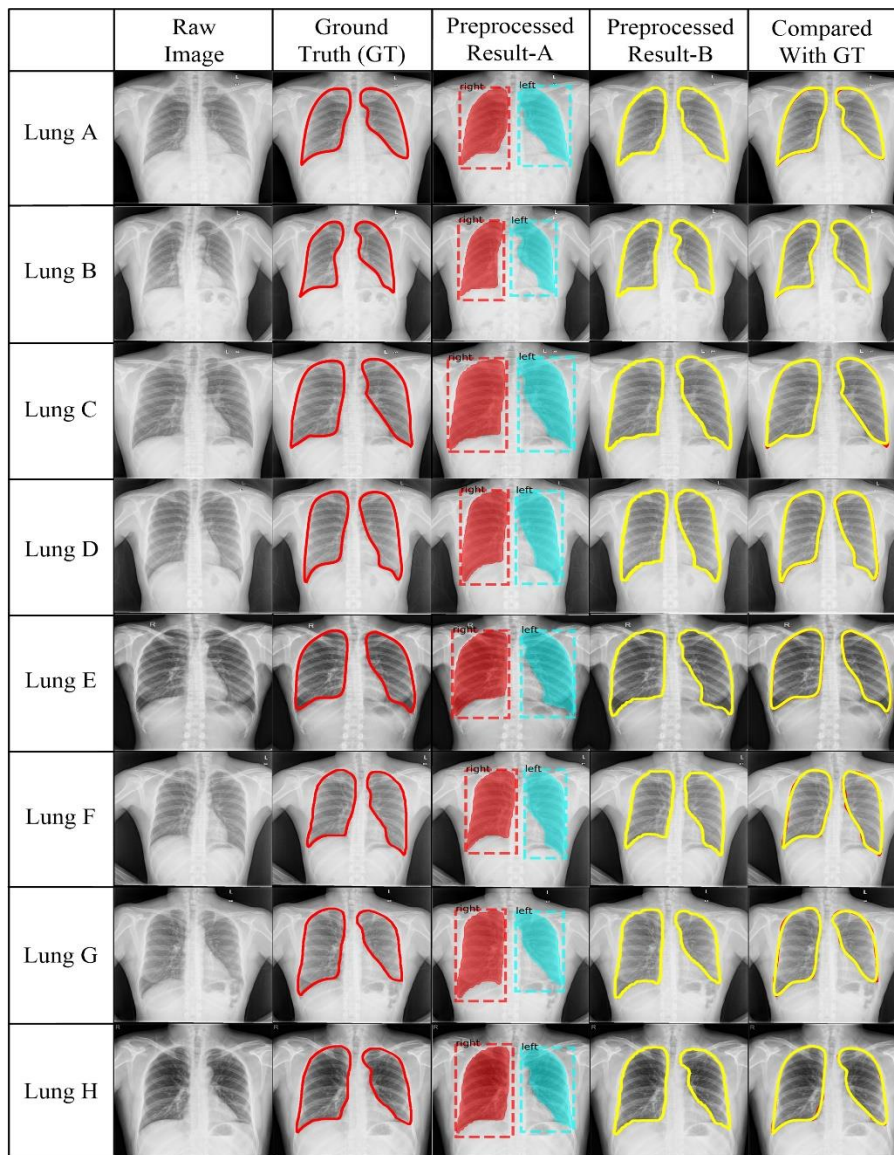

Fig. 8. Several raw images are randomly selected for assisting qualitative analysis. "Preprocessed Result-A" shows the preprocessed result with the mask, box, contour, and annotated name. "Preprocessed Result-B" denotes the preprocessed result consisting only of the contour. "Compare with GT" shows the results after the refinement step and compares with the ground truth.

C. AR and CPR of lung

Accurate segmentation of AR and CPR is important for subsequent diagnosis of diseases such as tuberculosis or lung cancer, which often start at the AR. A blunt CostoPhrenic (CP) angle is often considered to show abnormal pulmonary (Leong et al., 2020; Peng, Xu, Wang, & Li, 2020). To demonstrate the proposed method's performance in these regions, we tested our method at the top 25% and the bottom 25% of the lung. Fig. 9 and Fig. 10 denote the AR and CPR, respectively, of the segmentation results, which are zoomed images from Fig. 8 ("Compared with GT" column). Yellow and red contours show the automatic segmentation and the ground truth, respectively. Sharp corner areas (too small CP angle) and big turning areas (AR) make the segmentation more challenging. Nevertheless, the principal curve method shows a strong ability to fit the data accurately. Overall, the proposed
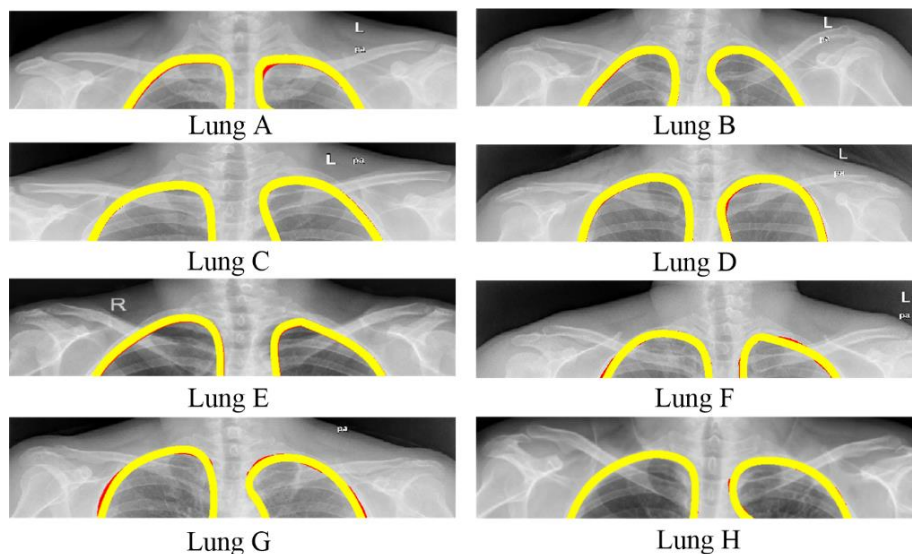
method can accurately extract these lung regions.

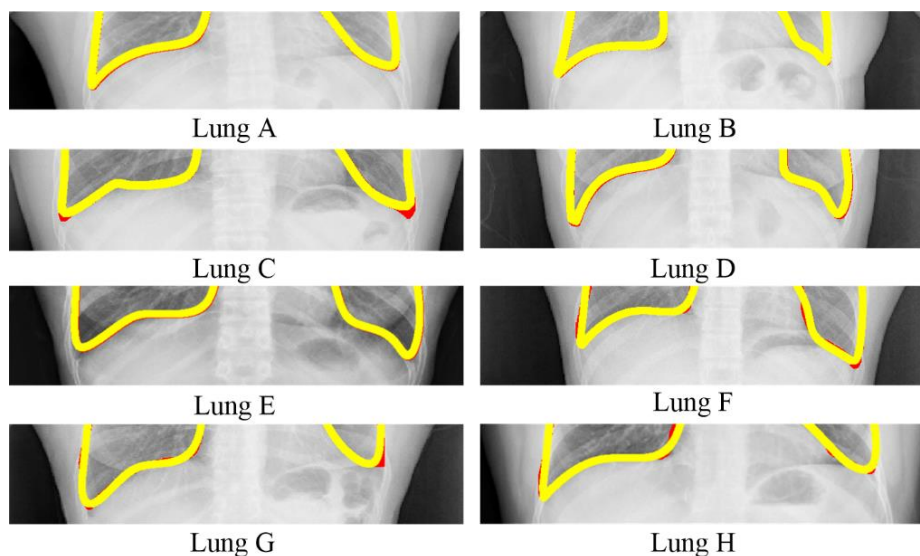

Fig. 9. Segmentation results for AR.

Fig. 10. Segmentation results for CPR.

4.2.2. Quantitative evaluation

We first used DSC, Ω, and ACC as metrics to evaluate the performance of our proposed A-LugSeg at 256×256, 512×512, and 1024×1024 resolutions. All (All), negative (-), and positive (+) cases represent all 100 testing cases, normal cases (50), and abnormal radiographs (50), respectively.

For all CXRs in the test dataset, DSC values obtained by the proposed method were all above 0.955, and the average DSC obtained by using "ALL" results on different resolutions was about 0.965 (Fig. 11). All the Ω and ACC values are higher than 0.93 and 0.95, respectively. Higher resolution leads to lower DSC, Ω, and ACC (Candemir et al., 2014), as the performance was the best on CXRs at 256×256 resolution. However, the DSC, Ω, and ACC deviated only slightly at different resolutions (ranging from 0.958 to 0.971, from 0.939 to 0.955, and from 0.952 to 0.97 for "ALL"), which shows that our method's performance is generally stable across different image resolutions. The optimal DSC, Ω, and ACC is obtained at 256×256 resolution, which are 0.971, 0.955, and 0.97, respectively.
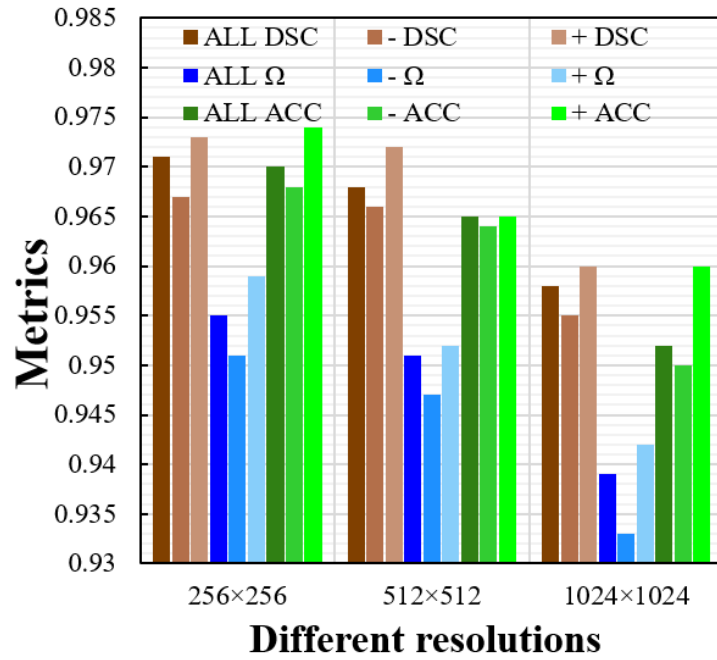
15

Fig. 11. The segmentation performance as measured by DSC, Ω, and ACC on SZCX at different resolutions. DSC, Ω, and ACC denote the brown-related, blue-related, and green-related histogram, respectively. All (All), negative (-), and positive (+) cases represent all 100 testing cases, normal cases (50), and abnormal radiographs (50), respectively.

We further evaluated the accuracy and stability of the proposed method by using the boxplot of different evaluation metrics on the 100 testing cases (Fig. 12). Key components of the boxplot include median value, mean value, and outliers (Li et al., 2019; Pereira et al., 2016). From Fig. 12, the median and mean value of DSC, Ω, and ACC are 0.977 and 0.971, 0.961 and 0.955, and 0.976 and 0.97, respectively. Furthermore, the DSC was above the mean value of DSC (0.971) for 58% of the images and above 0.98 for 38% of the images. The ACC for 68% of the images was higher than the mean value of ACC (0.97) and above 0.98 for 23% of the images. The Ω was higher than the mean value of Ω (0.955) for 64% of the images. Overall, due to the impact of the outlier points, the mean value is lower than the median value. There are seven radiographs with scores of Ω below 0.9, with the lowest score of Ω being 0.876.
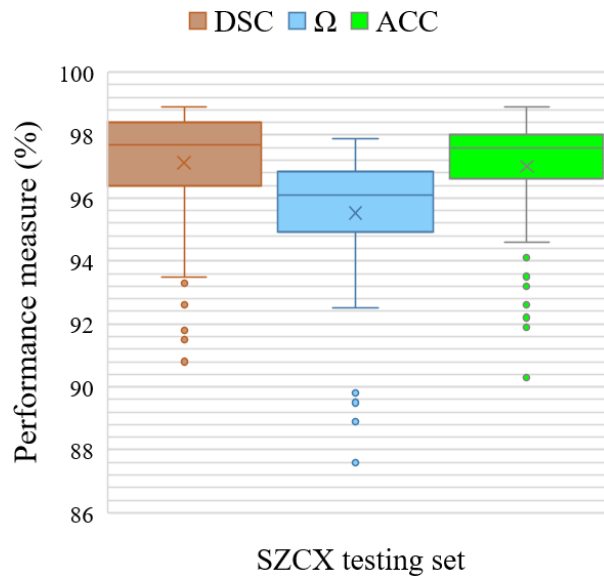
4.3. Performance evaluation on the other public datasets

We used two additional public datasets (JSRT and MC) to test the generalizability of the model trained on the SZCX dataset. We described these datasets in Section 3. Because the CXRs have different resolutions, we resized all the CXRs to 256×256 in this experiment. Fig. 13 shows the boxplot of each quantitative result, as measured by DSC, Ω, and ACC, of all the JSRT and MC datasets. The average DSC, Ω, and ACC for JSRT dataset were 0.973, 0.958, and 0.972, respectively. Meanwhile, the average DSC, Ω, and ACC for MC dataset were 0.971, 0.955, and 0.97, respectively. These values of both datasets are comparable to that of the SZCX test cases (0.971, 0.955, and 0.97), thus demonstrating our model's generalizability. From Fig. 12 and Fig. 13, compared with all the metrics of both SZCX and MC datasets, the outlier points of JSRT dataset are less. The main reason is that the majority of the CXR images in both SZCX and MC datasets have more tuberculosis manifestations inside the lung regions, which can lead to severe shape deformations.
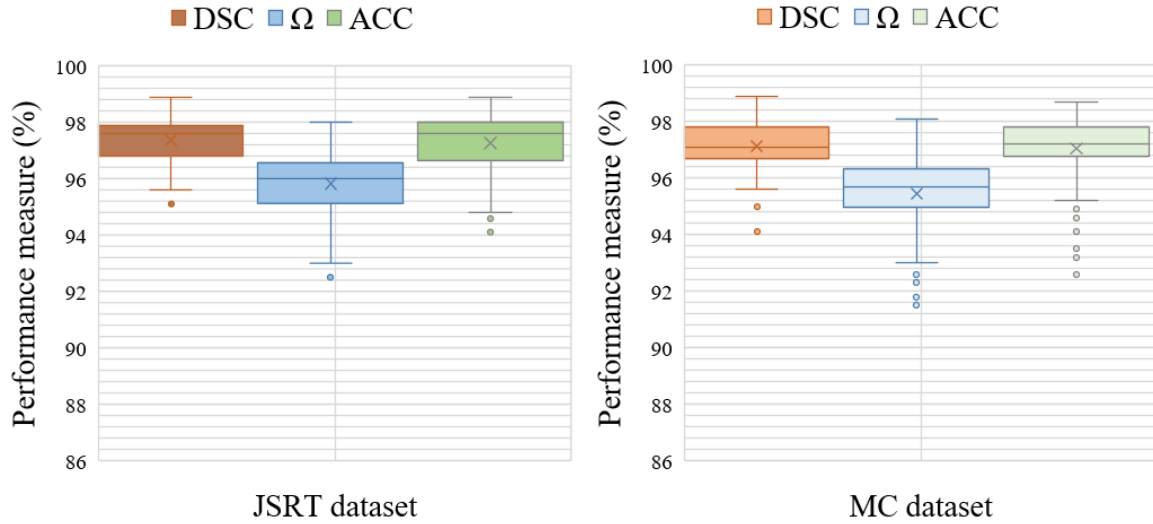

Fig. 13. DSC of each CXR image in the JSRT and MC datasets. Both JSRT and MC datasets are used, where JSRT dataset contains 247 CXRs and MC dataset contains 138 CXRs.

4.4. Comparison with state-of-the-art methods on public datasets

We have demonstrated via the previous experiments that the proposed method is both effective and efficient for lung segmentation. To further illustrate the effectiveness of our method, we compared our method with other existing state-of-the-art lung contour extraction methods on the two popular public datasets, JSRT and MC. We used each of these datasets in its entirety for testing. To evaluate the performance comprehensively, we have added another metrics for evaluation, including precision/positive predictive value (PPV), sensitivity/recall, and Specificity.

$$Precision \ (PPV) = \frac{TP}{TP + FP} \tag{18}$$

$$Sensitivity \ (Recall) = \frac{TP}{TP + FN} \tag{19}$$

$$Specificity = \frac{TN}{FP + TN} \tag{20}$$

We compared with state-of-the-art methods by using the average values of all the used metrics, shown in

Table 2. We also investigated the statistical significance of the differences for DSCs between our method and different models (Peng et al., 2018; Peng et al., 2019) using paired *t*-test with a significance level of 0.05 (Khiyali et al., 2017) on different testing datasets (JSRT and MC), shown in Table 3. We considered the following two categories of methods:

- Machine learning (ML) models (Novikov et al., 2018; Souza et al., 2019; Rashid et al., 2018; Mansoor et al., 2020; Nishio et al., 2021): Novikov et al., (2018) presented the InvertedNet with Exponential Linear Units activation functions (ELU) for multi-class segmentation (i.e., lung, clavicle, and heart) in chest radiographs. Souza et al., (2019) proposed an automatic method that uses a ResNet-based Deep Convolutional Neural Network (RDCNN) to segment lung fields in chest radiographs. Rashid et al., (2018) proposed the Fully Convolutional Neural Network (FCNN) for lung segmentation from CXRs, which was validated on MC dataset. Mansoor et al., (2020) presented a generic approach to lung field segmentation that uses Deep Space and Shape Learning (DSSL). Furthermore, Nishio et al., (2021) designed a Modified U-net model (MU-net) combining the U-net with the Bayesian optimization for lung CXR segmentation, where Bayesian optimization is used for optimizing the U-net's parameters.

- Hybrid models (T. Peng et al., 2019; Peng et al., 2018; Reamaroon et al., 2020; Afzali et al., 2021; Zou et al., 2021): Peng et al., (2018) combined Closed Polyline Searching algorithm (CPS) with Back Propagation algorithm (BP) to detect the lung contour. Furthermore, Peng et al., (2019) proposed a hybrid method called Hull-CPS to detect the boundaries of the lung Region of Interest (ROI), and Reamaroon et al., (2020) proposed the Total Variation-based Active Contour algorithm (TVAC) for lung segmentation. Both methods have been tested on JSRT and MC datasets, respectively, with good performance. Afzali et al., (2021) proposed the Active Shape Model (ASM) using Center Of Mass (COM)-based model for lung CXR segmentation. The model has been tested on the hybrid dataset (JSRT and MC). Zou et al., (2021) presented a recursive cascaded unsupervised registration network (RCINet) for lung segmentation in CXRs.

According to Table 2, the proposed method obtains a higher result among all the metrics than other existing state-of-the-art methods to prove the better performance, except for the works (Rashid et al., 2018; Mansoor et al., 2020; Nishio et al., 2021). Compared with the proposed method (A-LugSeg), the method (Rashid et al., 2018) has better performance on MC dataset using ACC, the method (Mansoor et al., 2020) has better Ω. Furthermore, there are higher DSC and Specificity values in the method (Nishio et al., 2021). One reason is that we used a more challenging evaluation strategy where our model was trained on one dataset (SZCX) while tested on the other two datasets (JSRT/MC). However, Rashid et al., (2018) used each dataset individually for training/validation/testing. Furthermore, both works (Rashid et al., 2018; Mansoor et al., 2020) have combined all the datasets and then randomly selected CXRs as training/validation/testing set.

Overall, the proposed method has an excellent performance, especially for precision (PPV) and sensitivity (Recall). Compared with other methods, the proposed method obtained the best precision (PPV), no matter we used the individual dataset (JSRT/MC) or hybrid datasets (JSRT+MC) for testing. Furthermore, the proposed method obtains the best Sensitivity (Recall), which proves that the A-LugSeg can obtain more samples correctly classified from the whole dataset.

Table 2

Quantitative comparison with the state-of-the-art methods. "JSRT/MC" shows the result of each dataset, and "JSRT+MC" shows the result of the hybrid datasets. All the papers are sorted according to the year of publication.

						Performance evaluation			
Paper	**Method**	**Model**	**Data**	**DSC/ F1-score**	**Ω/ IoU**	**Accuracy (ACC)**	**Precision/ PPV**	**Sensitivity /Recall**	**Specificity**

(Novikov et al., 2018)	Inverted Net with ELU	ML	JSRT	0.974	0.949	-	-	-	-
(Peng et al., 2018)	CPS-BP	Hybrid	JSRT/ MC	0.913/ 0.895	0.881/ 0.87	0.91/ 0.887	0.946/ 0.93	0.936/ 0.935	0.956/ 0.945
(Rashid et al., 2018)	FCN (U-Net)	ML	JSRT/ MC	0.951/ 0.954	-	0.971/ **0.977**	-	0.951/ 0.954	0.98/ 0.985
(Peng et al., 2019)	Hull-CPS	Hybrid	JSRT/ MC	0.965/ 0.961	0.932/ 0.93	0.97/ 0.97	0.973/ 0.964	0.946/ 0.938	0.97/0.965
(Souza et al., 2019)	RDCNN	ML	MC	0.94	0.88	0.969	-	0.975	0.967
(Mansoor et al., 2020)	DSSL	ML	JSRT	0.969	**0.961**	-	-	-	-
(Reamaroon et al., 2020)	TVAC	Hybrid	JSRT/ MC	0.95/ 0.956	-	-	-	-	-
(Afzali et al., 2021)	ASM+COM	Hybrid	JSRT+ MC	0.966	0.934	-	-	-	-
(Zou et al., 2021)	RICNet	Hybrid	JSRT/ MC	0.947/ 0.953	-	-	-	-	-
(Nishio et al., 2021)	MU-net	ML	JSRT/ MC	**0.976**/ **0.973**	0.954/ 0.949	-	-	0.987/ 0.967	0.985/ **0.992**
Our model	**A-LugSeg**	**Hybrid**	**JSRT/ MC**	**0.973**/ **0.971**	**0.958**/ **0.955**	**0.972**/ **0.97**	**0.989**/ **0.984**	**0.995**/ **0.991**	**0.993**/ **0.989**
			JSRT +MC	**0.972**	**0.956**	**0.97**	**0.982**	**0.992**	**0.99**

Table 3

Statistical analysis using paired *t*-test for DSCs on different testing datasets between different models and our method.

	JSRT	MC
(Peng et al., 2018)	<0.001	<0.001
(Peng et al., 2019)	<0.001	<0.001

5. Discussion and Conclusion

In this paper, we have proposed a new hybrid model for accurate and robust lung segmentation from CXRs. The innovations of our proposed method include: (1) an automatic and hybrid segmentation strategy; (2) an improved principal curve model; and (3) a smooth mathematical expression of the ROI contour. To demonstrate the applicability of our proposed segmentation method to lungs with various shapes, we used three clinical CXR

datasets to validate the performance of the proposed method. Both qualitative and quantitative experimental results show that: (1) the proposed A-LugSeg can obtain consistently accurate results regardless of dataset (SZCX, JSRT, or MC), evaluation metric (Ω, DSC, or ACC), or different kinds of CXRs (normal or abnormal); and (2) the proposed A-LugSeg has good accuracy in challenging regions, such as AR and CPR regions. While we focus on lung contour identification in this work, the method proposed in this work can be readily applied to segment any object with a smooth boundary in an image.

Although the proposed method achieves overall accurate segmentation, it still yields a few unsatisfactory segmentation results. As shown in Fig. 12, one case had a low Ω score of 0.876, where the DSC and ACC of the corresponding case in SZCX are 0.919 and 0.911, respectively. The inferior performance of the proposed method for this case may be owing to the extraordinary condition of the left lung, as shown in Fig. 14. This patient has bilateral pulmonary tuberculosis (PTB), where the apex of the right lung presents a similar intensity to the background due to the disease. Lung diseases such as tuberculosis can severely affect the lung's shape. In the extreme case of a collapsed lung or an effusion, the lung shape will differ dramatically from the shape of a healthy lung (Candemir et al., 2014). These abnormal lung shapes can cause problems for auto-segmentation methods. For these extraordinary cases, manual segmentation may be needed to obtain accurate lung contours.
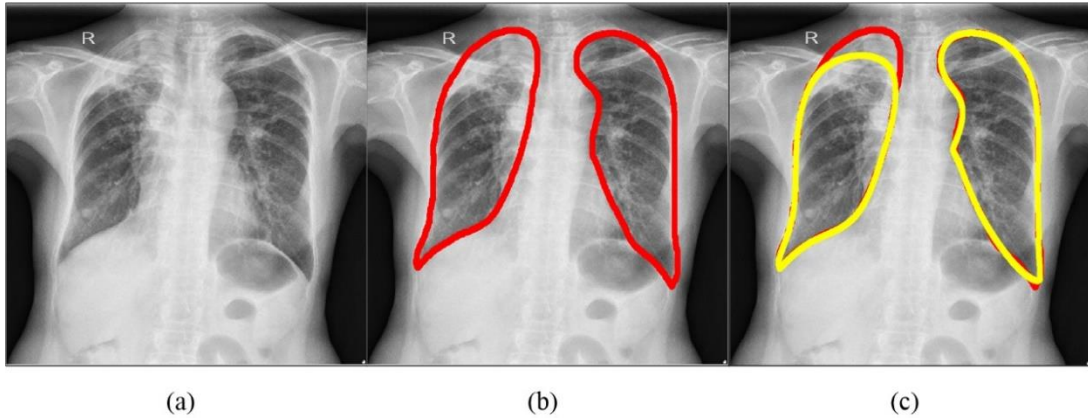

Fig. 14. Unsatisfactory segmentation result of SZCX, which is the lowest score Ω of 0.876 (Fig. 12). (a) denotes the raw data, (b) denotes the ground truth, and (c) shows the result compared to the ground truth, where the yellow contour is the segmented result, and the red contour is the ground truth.

The preprocessing step of the proposed method is based on a deep convolutional neural network architecture (Mask-RCNN) for automatic segmentation. We only use 400 and 162 CXRs from SZCX for training and validation, respectively. Compared with other works (Flores et al., 2019) that use more than 1000 images for training, the 400 training CXRs that we used for our method seem to be limited. This could explain why using Mask-RCNN to initially detect the lung contour did not obtain very accurate results in the coarse segmentation of the lung. The Mask-RCNN's performance for initial coarse segmentation could be improved by data augmentations, such as image transformation (Peng, Estrada, Pedersoli, & Desrosiers, 2020) and Generative Adversarial Networks (GAN)-based methods (Mahapatra, & Ge, 2020). We will investigate whether the data augmentation step would affect the initial segmentation results and whether improving the initial coarse segmentation would further improve the final segmentation results after the refinement step.

In this paper, we use the refinement step to fine-tune the coarse segmentation results, where the parameters of the FOBL are used to represent an explainability-guided mathematical model of the lung contour (Eq. (13) and Eq. (14)). In the FOBL, several parameters, such as connection weights and thresholds, must be optimized to achieve accurate segmentation results during training. In our implementation, we randomly initialized the connection weights and thresholds of the FOBL. However, the trained network depends on the initial connection weights and thresholds, and the solutions could fall into a local minimum rather than into the global optimum. To

overcome this issue, we may adopt different methods, such as differential evolution algorithms, to select the optimal initial connection weights and thresholds of the FOBL (Leema et al., 2016; Vivekanandan, & Sriman Narayana Iyengar, 2017).

In the future, we will further consider the below aspects to improve the performance of our method and validate its applicability to other organs and imaging modalities. *First*, the proposed network architecture contains two subnetworks, which increases the memory burden for segmentation. Therefore, model compression is desired for real-time clinical applications. *Second*, we plan to validate our method on other organs in the future, such as prostate, kidney, and bladder. Furthermore, we plan to test on multi-organs, instead of one organ. Considering that the contours of the prostate are usually much smoother than lungs (which contain small CP angles), our proposed A-LugSeg may be suitable for segmenting these smooth organs as well. *Finally*, for multi-modality imaging, the performance of our proposed method may be improved by fusing segmentation results from individual modalities. For example, Positron Emission Tomography (PET) and Computed Tomography (CT) are two commonly used imaging technologies for cancer diagnosis and staging. We can test our method on both of them for co-segmentation that take advantage of the superior tumor contrast from PET images and the anatomical information from CT images.

Acknowledgment

References

Afzali, A., Babapour Mofrad, F., & Pouladian, M. (2018). Inter-Patient Modelling of 2D Lung Variations from Chest X-Ray Imaging via Fourier Descriptors. *Journal of Medical Systems*, *42*(11), 233.

Afzali, A., Babapour Mofrad, F., & Pouladian, M. (2021). 2D Statistical Lung Shape Analysis Using Chest Radiographs: Modelling and Segmentation. *Journal of Digital Imaging*, *34*(3), 523–540.

Ataer-Cansizoglu, E., Bas, E., Kalpathy-Cramer, J., Sharp, G. C., & Erdogmus, D. (2013). Contour-based shape representation using principal curves. *Pattern Recognition*, *46*(4), 1140–1150.

Bagheri, F., Tarokh, M. J., & Ziaratban, M. (2021). Skin lesion segmentation based on mask RCNN, Multi Atrous Full-CNN, and a geodesic method. *International Journal of Imaging Systems and Technology*, *31*(3), 1609–1624.

Cai, L., Long, T., Dai, Y., & Huang, Y. (2020). Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis. *IEEE Access*, *8*, 44400–44409.

Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., Singh, R. K., Zhiyun Xue, Karargyris, A., Antani, S., Thoma, G., & McDonald, C. J. (2014). Lung Segmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration. *IEEE Transactions on Medical Imaging*, *33*(2), 577–590.

Chakraborty, S., & Mali, K. (2021). SuFMoFPA: A superpixel and meta-heuristic based fuzzy image segmentation approach to explicate COVID-19 radiological images. *Expert Systems with Applications*, *167*, 114142.

Chandra, T. B., Verma, K., Singh, B. K., Jain, D., & Netam, S. S. (2020). Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. *Expert Systems with Applications*, *158*, 113514.

Chen, M.-R., Chen, B.-P., Zeng, G.-Q., Lu, K.-D., & Chu, P. (2020). An adaptive fractional-order BP neural network based on extremal optimization for handwritten digits recognition. *Neurocomputing*, *391*, 260–272.

Chondro, P., Yao, C.-Y., Ruan, S.-J., & Chien, L.-C. (2018). Low order adaptive region growing for lung segmentation on plain chest radiographs. *Neurocomputing*, *275*, 1002–1011.

Farhangi, M. M., Frigui, H., Seow, A., & Amini, A. A. (2017). 3-D Active Contour Segmentation Based on Sparse Linear Combination of Training Shapes (SCoTS). *IEEE Transactions on Medical Imaging*, *36*(11), 2239–2249.

Flores, C. F., Gonzalez-Garcia, A., van de Weijer, J., & Raducanu, B. (2019). Saliency for fine-grained object recognition in domains with scarce training data. *Pattern Recognition*, *94*, 62–73.

Gamage, C., Wijesinghe, I., & Perera, I. (2019). Instance-Based Segmentation for Boundary Detection of Neuropathic Ulcers Through Mask-RCNN. In *2019 International Conference on Artificial Neural Networks (ICANN)* (pp. 511–522). Springer.

Geetha, S., Narayanamoorthy, S., Manirathinam, T., & Kang, D. (2021). Fuzzy Case-Based Reasoning approach for finding COVID-19 patients priority in hospitals at source shortage period. *Expert Systems with Applications*, *178*, 114997.

Hastie, T., & Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association*, *84*(406), 502–516.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2961–2969). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 770–778). IEEE.

Jaeger, S., Candemir, S., Antani, S., Wáng, Y.-X. J., Lu, P.-X., & Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative Imaging in Medicine and Surgery*, *4*(6), 475–477.

Kégl, B., & Krzyzak, A. (2002). Piecewise Linear Skeletonization Using Principal Curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 59–74.

Kégl, B., Linder, T., & Zeger, K. (2000). Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 281–297.

Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A., & Segovia, F. (2015). Early diagnosis of Alzheimer′s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, *151*, 139–150.

Khiyali, Z., Manoochri, M., Jeihooni, A. K., Heydarabadi, A. B., & Mobasheri, F. (2017). *Educational Intervention on Preventive Behaviors on Gestational Diabetes in Pregnant Women: Application of Health Belief Model*. 11.

Leema, N., Nehemiah, H. K., & Kannan, A. (2016). Neural network classifier optimization using Differential Evolution with Global Information and Back Propagation algorithm for clinical datasets. *Applied Soft Computing*, *49*, 834–844.

Leong, P., Ruane, L. E., Phyland, D., Koh, J., MacDonald, M. I., Baxter, M., Lau, K. K., Hamza, K., & Bardin, P. G. (2020). Inspiratory vocal cord closure in COPD. *European Respiratory Journal*, *55*(5), 1901466.

Li, H., Li, A., & Wang, M. (2019). A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Computers in Biology and Medicine*, *108*, 150–160.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *2014 International Conference on European Conference on Computer Vision (ECCV)* (pp. 740–755). Springer.

Mahapatra, D., & Ge, Z. (2020). Training data independent image registration using generative adversarial networks and domain adaptation. *Pattern Recognition*, *100*, 107109.

Mansoor, A., Cerrolaza, J. J., Perez, G., Biggs, E., Okada, K., Nino, G., & Linguraru, M. G. (2020). A Generic Approach to Lung Field Segmentation from Chest Radiographs using Deep Space and Shape Learning. *IEEE Transactions on Bio-Medical Engineering*, *67*(4), 1206–1220.

Nishio, M., Fujimoto, K., & Togashi, K. (2021). Lung segmentation on chest X-ray images in patients with severe abnormal findings using deep learning. *International Journal of Imaging Systems and Technology*, *31*(2), 1002–1008.

Novikov, A. A., Lenis, D., Major, D., Hladuvka, J., Wimmer, M., & Buhler, K. (2018). Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs. *IEEE Transactions on Medical Imaging*, *37*(8), 1865–1876.

Peng, J., Estrada, G., Pedersoli, M., & Desrosiers, C. (2020). Deep co-training for semi-supervised image segmentation. *Pattern Recognition*, *107*, 107269.

Peng, T., Wang, Y., Xu, T. C., & Chen, X. (2019). Segmentation of Lung in Chest Radiographs Using Hull and Closed Polygonal Line Method. *IEEE Access*, *7*, 137794–137810.

Peng, T., Wang, Y., Xu, T. C., Shi, L., Jiang, J., & Zhu, S. (2018). Detection of Lung Contour with Closed Principal Curve and Machine Learning. *Journal of Digital Imaging*, *31*(4), 520–533.

Peng, T., Xu, T. C., Wang, Y., & Li, F. (2020). Deep Belief Network and Closed Polygonal Line for Lung Segmentation in Chest Radiographs. *The Computer Journal*.

Peng, T., Xu, T. C., Wang, Y., Zhou, H., Candemir, S., Zaki, W. M. D. W., Ruan, S.-J., Wang, J., & Chen, X. (2020). Hybrid Automatic Lung Segmentation on Chest CT Scans. *IEEE Access*, *8*, 73293–73306.

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*, *35*(5), 1240–1251.

Privalov, M., Beisemann, N., Barbari, J. E., Mandelka, E., Müller, M., Syrek, H., Grützner, P. A., & Vetter, S. Y. (2021). Software-Based Method for Automated Segmentation and Measurement of Wounds on Photographs Using Mask R-CNN: A Validation Study. *Journal of Digital Imaging*.

Rashid, R., Akram, M. U., & Hassan, T. (2018). Fully Convolutional Neural Network for Lungs Segmentation from Chest X-Rays. In *2018 International Conference on Image Analysis and Recognition (ICIAR)* (pp. 71–80). Springer.

Reamaroon, N., Sjoding, M. W., Derksen, H., Sabeti, E., Gryak, J., Barbaro, R. P., Athey, B. D., & Najarian, K. (2020). Robust segmentation of lung in chest x-ray: Applications in analysis of acute respiratory distress syndrome. *BMC Medical Imaging*, *20*(1), 116.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149.

Shen, S., Bui, A. A. T., Cong, J., & Hsu, W. (2015). An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Computers in Biology and Medicine*, *57*, 139–149.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., & Doi, K. (2000). Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule: Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules. *American Journal of Roentgenology*, *174*(1), 71–74.

Shu, Y., Zhang, J., Xiao, B., & Li, W. (2021). Medical image segmentation based on active fusion-transduction of multi-stream features. *Knowledge-Based Systems*, *220*, 106950.

Souza, J. C., Bandeira Diniz, J. O., Ferreira, J. L., França da Silva, G. L., Corrêa Silva, A., & de Paiva, A. C. (2019). An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Computer Methods and Programs in Biomedicine*, *177*, 285–296.

Tu, L., Styner, M., Vicory, J., Elhabian, S., Wang, R., Hong, J., Paniagua, B., Prieto, J. C., Yang, D., Whitaker, R.,

& Pizer, S. M. (2018). Skeletal Shape Correspondence Through Entropy. *IEEE Transactions on Medical Imaging*, *37*(1), 1–11.

Vidal, P. L., de Moura, J., Novo, J., & Ortega, M. (2021). Multi-stage transfer learning for lung segmentation using portable X-ray devices for patients with COVID-19. *Expert Systems with Applications*, *173*, 114677.

Vivekanandan, T., & Sriman Narayana Iyengar, N. C. (2017). Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. *Computers in Biology and Medicine*, *90*, 125–136.

Vuola, A. O., Akram, S. U., & Kannala, J. (2019). Mask-RCNN and U-net Ensembled for Nuclei Segmentation. In *2019 IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 208–212). IEEE.

Wan Ahmad, W. S. H. M., W Zaki, W. M. D., & Ahmad Fauzi, M. F. (2015). Lung segmentation on standard and mobile chest radiographs using oriented Gaussian derivatives filter. *BioMedical Engineering OnLine*, *14*(1), 20.

Wang, J., Wen, Y., Gou, Y., Ye, Z., & Chen, H. (2017). Fractional-order gradient descent learning of BP neural networks with Caputo derivative. *Neural Networks*, *89*, 19–30.

Wang, X.-F., Min, H., & Zhang, Y.-G. (2015). Multi-scale local region based level set method for image segmentation in the presence of intensity inhomogeneity. *Neurocomputing*, *151*, 1086–1098.

Yang, S., Chen, D., Zeng, X., & Pudney, P. (2014). A Greedy Algorithm for Constraint Principal Curves. *Journal of Computers*, *9*(5), 1125–1130.

Yang, W., Liu, Y., Lin, L., Yun, Z., Lu, Z., Feng, Q., & Chen, W. (2018). Lung Field Segmentation in Chest Radiographs From Boundary Maps by a Structured Edge Detector. *IEEE Journal of Biomedical and Health Informatics*, *22*(3), 842–851.

Yu, W., Tang, C., Hobbs, B. P., Li, X., Koay, E. J., Wistuba, I. I., Sepesi, B., Behrens, C., Rodriguez Canales, J., Parra Cuentas, E. R., Erasmus, J. J., Court, L. E., & Chang, J. Y. (2018). Development and Validation of a Predictive Radiomics Model for Clinical Outcomes in Stage I Non-small Cell Lung Cancer. *International Journal of Radiation Oncology*Biology*Physics*, *102*(4), 1090–1097.

Zeng, N., Li, H., & Peng, Y. (2021). A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease. *Neural Computing and Applications*.

Zeng, N., Li, H., Wang, Z., Liu, W., Liu, S., Alsaadi, F. E., & Liu, X. (2021). Deep-reinforcement-learning-based images segmentation for quantitative analysis of gold immunochromatographic strip. *Neurocomputing*, *425*, 173–180.

Zeng, N., Wang, Z., Zhang, H., Kim, K.-E., Li, Y., & Liu, X. (2019). An Improved Particle Filter With a Novel Hybrid Proposal Distribution for Quantitative Analysis of Gold Immunochromatographic Strips. *IEEE Transactions on Nanotechnology*, *18*, 819–829.

Zhang, H., Pedrycz, W., Miao, D., & Zhong, C. (2013). A global structure-based algorithm for detecting the principal graph from complex data. *Pattern Recognition*, *46*(6), 1638–1647.

Zou, W., Luo, Y., Cao, W., He, Z., & He, Z. (2021). A cascaded registration network RCINet with segmentation mask. *Neural Computing and Applications*, 1–17.