

Received April 2, 2020, accepted April 10, 2020, date of publication April 14, 2020, date of current version April 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2987925

Hybrid Automatic Lung Segmentation on Chest CT Scans

TAO PENG^{1,8}, (Member, IEEE), THOMAS CANHAO XU¹, (Member, IEEE), YIHUAI WANG¹,
HAILING ZHOU², (Member, IEEE), SEMA CANDEMIR³, (Member, IEEE),
WAN MIMI DIYANA WAN ZAKI⁴, (Member, IEEE)
SHANQ-JANG RUAN⁵, (Senior Member, IEEE),
JING WANG⁶, AND XINJIAN CHEN^{6,7}, (Senior Member, IEEE)

¹School of Computer Science and Technology, Soochow University, Suzhou 215006, China

²Institute for Intelligent Systems Research and Innovation, Deakin University, Geelong, VIC 3220, Australia

³The Ohio State University Wexner Medical Center, Columbus, OH 43210, USA

⁴INTEGRA, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, National University of Malaysia, Bangi 43600, Malaysia

⁵Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan

⁶School of Electrical and Information Engineering, Soochow University, Suzhou 215006, China

⁷State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China

⁸Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 2280, USA

Corresponding authors: Thomas Canhao Xu (thomasxu@suda.edu.cn) and Xinjian Chen (xjchen@suda.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Science Foundation of China under Grant 61672369 and Grant 61872257, in part by the National Natural Science Foundation of Jiangsu Province under Grant 17KJB520035, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

ABSTRACT Accurate lung segmentation in chest Computed Tomography (CT) scans is a challenging problem because of variations in lung volume shape, susceptibility to partial volume effects that affect thin antero-posterior junction lines, and lack of contrast between the lung and surrounding tissues. To address the need for a robust method for lung segmentation, we present a new method, called Pixel-based two-Scan Connected Component Labeling-Convex Hull-Closed Principal Curve method (PSCCL-CH-CPC), which automatically detects lung boundaries, and surpasses state-of-the-art performance. The proposed method has two main steps: 1) an image preprocessing step to extract coarse lung contours, and 2) a refinement step to refine the coarse segmentation result on the basis of the improved principal curve model and the machine learning model. Experimental results show that the proposed method has good performance, with a Dice Similarity Coefficient (DSC) as high as 98.21%. When compared with state-of-the-art methods, our proposed method achieved superior segmentation results, with an average DSC of 96.9%.

INDEX TERMS Automatic lung segmentation, chest CT scans, principal curve, closed principal curve method, machine learning.

I. INTRODUCTION

Various lung diseases cause millions of deaths and hospitalizations worldwide [1]. Early diagnosis, assisted by the wide use of lung scans to monitor and analyse organs, can lower the death rate [2]. Accurate lung segmentation is often performed as a necessary stage of quantitative lung image analysis because it is important for identifying and diagnosing lung diseases early [3]. Therefore, researchers have increasingly paid attention to lung segmentation and have proposed several methods for lung segmentation in recent years [4], [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval¹.

Obtaining accurate lung segmentation continues to require more attention because of the heterogeneity of the organ.

Because it contains a wealth of information and poses a low risk for patients, Computed Tomography (CT) scanning has become one of the most important imaging techniques. There are various anatomical challenges for lung volume segmentation methods (FIGURE 1). (1) lung shape and volume variation across patients due to differences in sex, age and physical condition, (2) susceptibility to partial volume effects that affect thin antero-posterior junction lines; and (3) intensity similarities between the lung and surrounding tissues such as airway and chest wall.

A number of strategies have been investigated for medical image segmentation techniques such as threshold

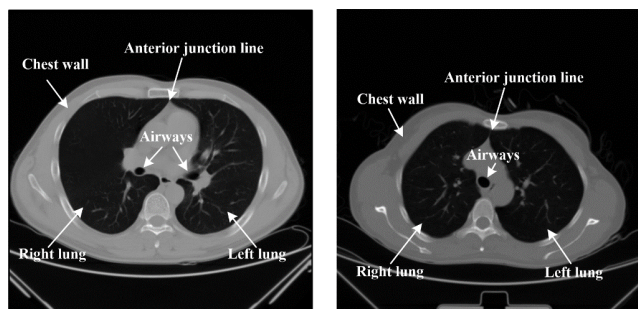


FIGURE 1. Anatomical feature in two chest CT images and their variations. Variations in lung shape, susceptibility to partial volume effects that affect thin antero-posterior junction lines, and lack of contrast between the lung and surrounding tissues make lung segmentation challenging on chest CT scans.

segmentation [6]–[8], region segmentation [9], [10], contour extraction approaches [11], [12], and atlas-based methods [13], [14]. Shen *et al.* [15] successfully minimize the over-segmentation of neighborhood regions by using a support vector machine (SVM) classifier to combine a two-way chain coding method. Although the average over-segmentation rate of this method can be as low as 0.3%, it always fails to re-include the juxta-pleural nodules that sit in consolidation areas. Khan and Kannan [16] have proposed an intelligent segmentation system based on fuzzy bitplane thresholding. The segmentation accuracy of the system is as high as 0.978, but the time complexity of the method is not discussed. Pratondo *et al.* [17] have integrated several machine learning models with region-based active contour models to segment medical images. Compared with the two previous approaches, the experimental results obtained by the contour extraction model provide a more realistic shape of the specific object, while reducing time complexity.

The critical goal of contour extraction approaches is to use a shape representation or curve approximation model to approximate the contour of the specific object. Oktay *et al.* [18] have presented an anatomically constrained neural networks model to enhance and segment cardiac images, but it requires prior knowledge of an anatomical shape. A unified segmentation model has been proposed based on the Fully Convolutional Neural Network (FCNN) and the Shape Representation Model (SRM) [19], but the training and testing datasets are limited. Keshani *et al.* [20] have explored a lung nodule segmentation system using an SVM classifier and active contour modeling. The method can achieve a detection rate of up to 89%, but it has not been validated on time complexity.

Recently, the principal curve model has attracted interests for segmenting abnormal organs from other neighboring normal structures, because of its strong abilities to effectively deal with noise input and obtain robust results [21]. You *et al.* [22] have proposed a semi-supervised system based on the principal curve method for lung CT image segmentation, and Peng *et al.* [23] have devised a principal curve model to detect lung contours; both of them detect tissue contours

with high accuracy. Furthermore, machine learning models can be used as classifiers for early diagnosis [24]. Therefore, combining machine learning with the principal curve model may be valuable for discriminant detection [25].

To segment lung tissues in chest CT scans accurately and efficiently, we propose a novel coarse-to-fine segmentation framework called Pixel-based two-Scan Connected Component Labeling-Convex Hull-Closed Principal Curve method (PSCCL-CH-CPC). Specifically, our pipeline consists of two cascaded stages. The first stage is coarse segmentation, where the PSCCL and CH are combined to extract the lung areas from the whole chest CT scans independently. The second stage refines the coarse segmentation result. The innovation of this stage is that it combines the improved CPC and the Backpropagation Neural Network method (BNN) to denote the smooth mathematical expression of lung contour. To the best of our knowledge, this is the first work to use an automatic lung segmentation method such as PSCCL-CH-CPC on chest CT scans.

The major contributions of this study are as follows:

(1) We propose a coarse-to-fine cascade segmenting framework that takes into account not only the overall lung but also the detail of the boundary between lung tissues.

(2) The lung and non-lung areas are separated automatically in the coarse segmentation stage, and the contour of the lung is also described independently.

(3) To the best of our knowledge, the improved CPC is the first attempt to describe the projection index of the closed dataset. We use a different initialization step than the Traditional Principal Curve method (TPC), and we add several stop conditions and constraint conditions.

(4) The parameters of the BNN are used to express the smooth and unified mathematical model, and the mathematical model is the proposed here for the first time. By combining the CPC with the BNN for training, the model error can be minimized to refine the coarse segmentation result.

The remainder of this paper is as follows. Section II presents the proposed method, and describes the detailed coarse-to-fine segmentation framework. Section III describes the database and evaluation metrics and presents the experimental results. Section IV draws conclusions.

II. METHODOLOGY

Appropriate pulmonary segmentation can identify normal and diseased tissues accurately. In this paper, the main steps of the proposed PSCCL-CH-CPC are the preprocessing step and the refinement step. Each step consists of several methods applied sequentially. Before we use the PSCCL-CH-CPC to segment a lung region, each image is enhanced by using Ying *et al.*'s method to improve the quality [26]. The flowchart of the proposed method is shown in FIGURE 2.

A. PREPROCESSING STEP

The preprocessing step completes the coarse segmentation and consists of the following key steps: region separation, lung region extraction, and outer contour extraction.

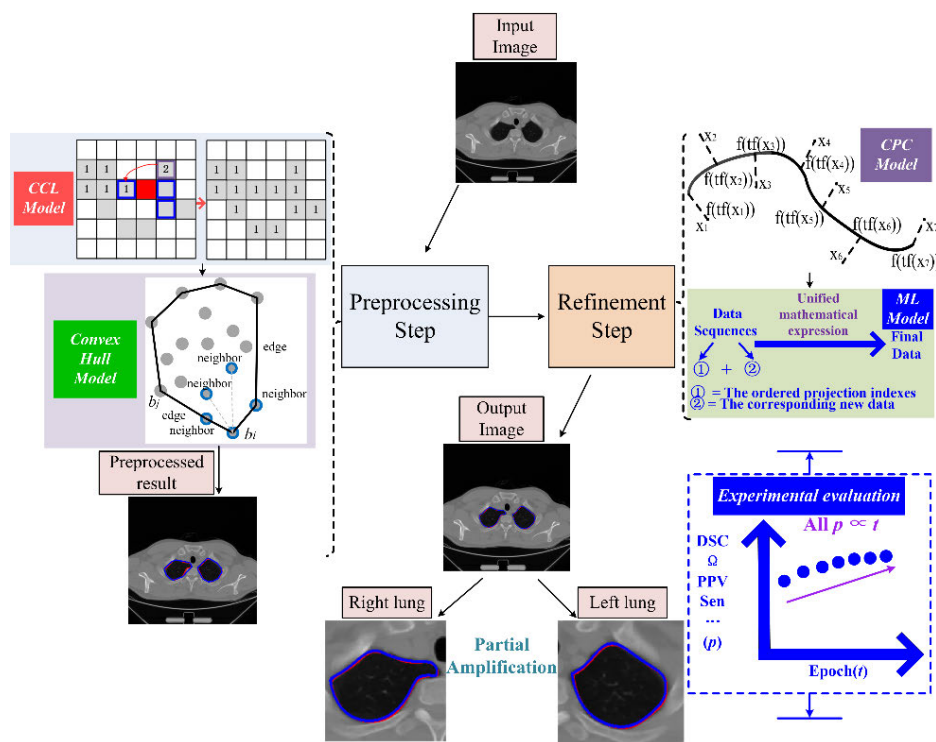


FIGURE 2. The flowchart of the proposed method. The system consists of the preprocessing and the refinement steps. The raw CT scans are regarded as input images. The preprocessing step mainly consists of the PSCCL and CH; this step is used to separate the lung and non-lung regions and complete the coarse segmentation. The refinement step combines the improved CPC and the ML model; this step is used to achieve high precision contour. The output can be quantitatively and qualitatively analyzed by the experimental evaluation part.

1) REGION ISOLATION

Because of the complexity of human organs, imaging devices cannot directly acquire the lung organs, so both lung and non-lung regions appear in the chest CT scans. To separate the lung and non-lung regions, we use the Otsu threshold algorithm (Otsu) [27], which selects the optimal threshold automatically. First, the histogram and probability for each level intensity are computed, and the weight ω_0 and class mean μ_0 are initialized. Second, an exhaustive search is performed, where the weight ω_i and the class mean μ_i are continuously updated. Then the variance of the classes σ_i^2 is calculated. Finally, the optimal threshold corresponding to the maximum inter-class variance is automatically obtained.

2) LUNG REGION EXTRACTION

The region isolation step effectively converts the raw CT scans into the binary images. In the lung region extraction step, we use the PSCCL to extract the separated lung regions independently. This study uses the PSCCL to record different organs by their own labels in the image so that each organ has a unique label. The first scan assigns temporary labels to all objects according to the principle of recording label equivalence [28], and the second scan replaces each temporary label according to the minimum label principle of equivalent classes [28].

For each binary CT lung image, the background and the objects are set to white and black, respectively. To be more intuitive, we used gray instead of black as the color of the objects to describe the flowchart of the PSCCL (See FIGURE 3). The details of the PSCCL are given in Algorithm 1. From FIGURE 3(a), First, we let P be the search direction and the row-by-row traverse is performed. When the object pixel $p(x, y)$ is found, then the point is labeled as No.1, which obeys the labeling rule. The principle of the labeling rule is to label the point based on the size of the number from small to large. At the same time, the neighbors of this point are obtained by the eight-neighbor algorithm [29], where the coordinates of the neighbors shown in FIGURE 3(b) are $(x + 1, y)$, $(x, y - 1)$, $(x - 1, y - 1)$, respectively. Then it will continue to traverse along the P-direction, as shown in FIGURE 3(c-d). If there are points that are not adjacent to the previous neighbors, then they are labeled No.2, and the rule of labeling is to label the point according to the size of the number from small to large. After traversing the second line, the next line is traversed. If the neighbor point $w(a, b)$ of the point $q(m, n)$ is labeled as both No.1 and No.2, the second scan is performed. The point $w(a, b)$ is re-labeled No.1 depending on the small label number, as shown in FIGURE 3(e-g). The loop search is always performed according to the above rules until all points have been traversed. FIGURE 4 shows the experimental result obtained by the Otsu and the PSCCL.

Algorithm 1 PSCCL

Input: the binary image I
Output: the coordinates of the data points of the lung
First scan:
for x in I'row
 for y in I'column
 if data[x][y] is not background
 Neighbors = connected elements with the value of the current element
 if neighbors is empty
 linked[NextLabel] = set containing NextLabel
 labels[x][y] = NextLabel
 NextLabel = NextLabel + 1
 else
 Find the smallest label
 L = neighbors labels
 labels[x][y] = min(L)
 for label in L
 linked[label] = union(linked[label], L)
Second scan:
for x in I'row
 for y in I'column
 if data[x][y] is not background
 labels[x][y] = find(labels[x][y])
 Store the coordinates of points which have the same label
 Extract the coordinates of the data points of the lung
return the coordinates of the data points of the lung

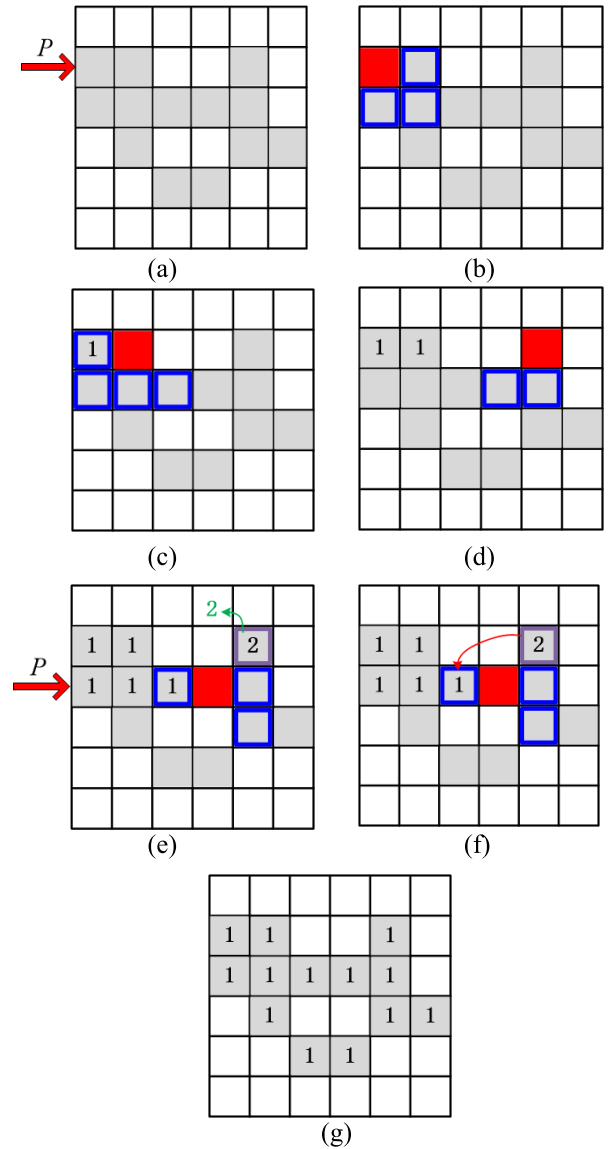


FIGURE 3. The flowchart of the PSCCL process. The search direction is determined in (a). In (b), we label the first point as No.1 which obeys the labeling rule, and the neighbors of this point are obtained. In (c) to (d), we label the point that is non-adjacent to the previous neighbors as No. 2. After scanning, we need to judge which label is suitable for the point with both No.1 and No.2 labels in (e) to (f). The result obtained is shown in (g).

3) OUTER CONTOUR EXTRACTION

Because there are many hollow parts in the extracted lung region, it is impossible to directly extract the outer contour of the lung by the contour extraction method [30], therefore the lung region obtained by the Otsu and PSCCL can be treated as a point set. We use the CH [31] to find the extreme points from the disordered point set. The principle is to calculate the direction of the intersection of the two vectors by Graham scan [32] and find all the vertexes along the boundary of the convex hull. The specific search process is shown in FIGURE 5.

The result obtained by the Otsu and PSCCL is the input of the CH. First, point p_0 , which is the point with the minimum y-coordinate, is selected; or when y-coordinates of several points are the same, the minimum x-coordinate point is selected. Second, if more than one point has the same polar angle taking p_0 as the vertex, all points but the one that is farthest from p_0 are removed. Third, the remaining points are numbered according to the principle of the polar angle changing from small to large. Finally, p_0, p_1, p_2 are pushed to stack. If there is no counterclockwise turn at a vertex, the vertex needs to be removed. When the following condition,

$$\overrightarrow{b_{j-1}b_j} \times \overrightarrow{b_{j-1}b_{j+1}} = (x_j - x_{j-1})(y_j - y_{j-1}) - (y_j - y_{j-1}) \times (x_{j+1} - x_{j-1}) < 0,$$

is satisfied, it is a counterclockwise turn. And $p_{j-1}(x_{j-1}, y_{j-1})$ and $p_j(x_j, y_j)$ denote the point at the top of S and the point next to the top of S, respectively, $p_i(x_i, y_i)$ is the point ready to stack.

B. REFINEMENT STEP

In order to obtain higher precision lung contours, the coarse segmentation results are fine-tuned by the refinement step, which combines the improved CPC with the BNN.

1) PRINCIPAL CURVE

The principal curve model is defined by Hastie as a smooth curve passing through the middle of the data [33]. Compared

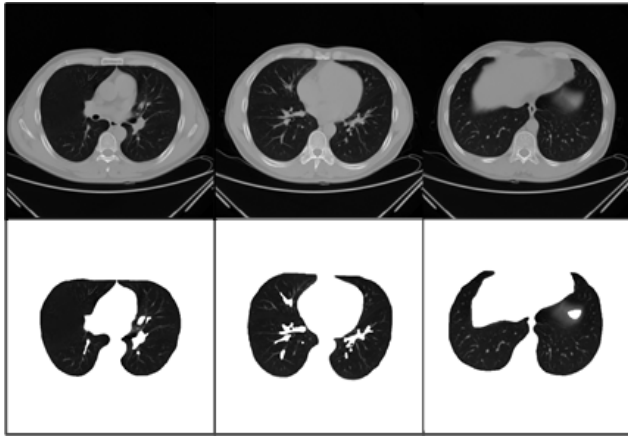


FIGURE 4. Results obtained by the Otsu and PSCCL. The first row represents the raw images randomly selected from the 2nd Hospital of Soochow University, and the second row shows the results obtained by the Otsu and PSCCL.

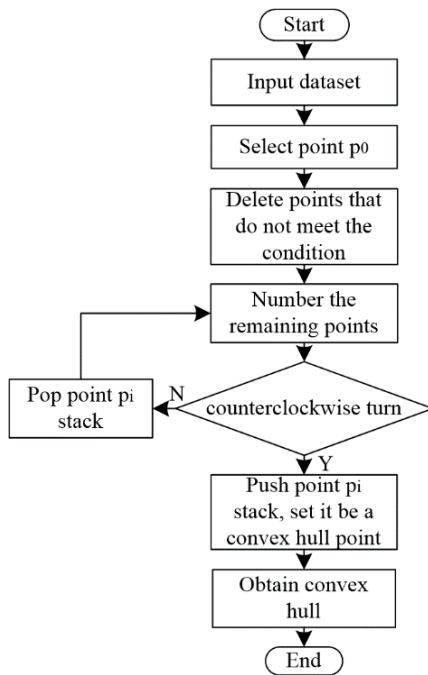


FIGURE 5. The process of obtaining the convex hull.

with the Traditional Principal Curve (TPC) model, we propose an improved CPC, while solves the issue of TPC to correctly describe the projection index of the dataset. The workflow of the CPC is shown in FIGURE 6.

a: TPC

Kegl *et al.*, proposed the TPC, whose key steps are projection and vertex optimization steps [34]. The detailed description is as follows: given a dataset $X_n = \{x_1, x_2, \dots, x_n\} \subseteq R^d$, then the shortest line segment of the first principal component line containing all projection data points is taken as the initial curve $f_{0,n}$ (initial step). After adding a new vertex, the position of each vertex is adjusted according to

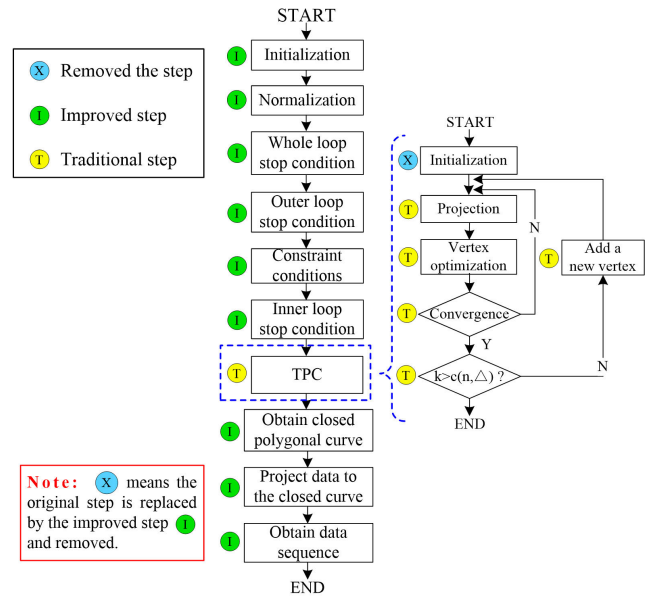


FIGURE 6. The flowchart of the CPC.

the principle of minimizing the penalty distance function (Eq. (5)). In each iteration, a new vertex obtained from the previous iteration is added to the curve $f_{k-1,n}$. the projection and vertex optimization steps are repeated for all previous vertexes until convergence to minimize the penalty distance function (Eq. (5)). Finally, the polygon curve can be obtained.

PROJECTION STEP

Let f be a polygon curve consisting of vertexes (v_1, \dots, v_{k+1}) and line segments (s_1, \dots, s_k) . The points of dataset X_n are projected to k edges and $k + 1$ vertexes of f , respectively. we set the nearest neighbor of any point x_i in X_n on f as v_i and assigned the v_i to a subset of V_i . when the point v_i is on the nearest neighbor segment s_i on f , the point v_i is also assigned to the subset S_i .

$$V_i = \{x \in X_n : \Delta(x, v_i) = \Delta(x, f), \Delta(x, v_i) < \Delta(x, v_m), m = 1, \dots, i - 1\} \tag{1}$$

$$S_i = \{x \in X_n : x \notin V, \Delta(x, s_i) = \Delta(x, f), \Delta(x, s_i) < \Delta(x, s_m), m = 1, \dots, i - 1\} \tag{2}$$

VERTEX OPTIMIZATION STEP

The goal of this step is to adjust the position of each vertex following the principle of the nearest distance from the data points to the principal curve [35], and the position of each line segment is also updated. Accordingly, we set $\pi(v_i) = r^2(1 + \cos \gamma_i)$, $\mu_+(v_i) = \|v_i - v_{i+1}\|^2$, $\mu_-(v_i) = \|v_i - v_{i-1}\|^2$,

the constraint $P(v_i)$ on the vertex v_i is as follows:

$$P(v_i) = \begin{cases} \mu_+(v_i) + \pi(v_{i+1}) & \text{if } i = 1 \\ \mu_-(v_i) + \pi(v_i) + \pi(v_{i+1}) & \text{if } i = 2 \\ \pi(v_{i-1}) + \pi(v_i) + \pi(v_{i+1}) & \text{if } 2 < i < k \\ \pi(v_{i-1}) + \pi(v_i) + \mu_+(v_i) & \text{if } i = k \\ \pi(v_{i-1}) + \mu_-(v_i) & \text{if } i = k + 1 \end{cases} \quad (3)$$

Let $\tau(v_i) = \sum_{x \in V_i} \Delta(x, v_i)$, $\sigma_+(v_i) = \sum_{x \in S_i} \Delta(x, s_i)$, $\sigma_-(v_i) = \sum_{x \in S_i} \Delta(x, s_{i-1})$. The average squared distance function $\Delta_n(v_i)$ is obtained as follows:

$$\Delta_n(v_i) = \begin{cases} v(v_i) + \sigma_+(v_i) & \text{if } i = 1 \\ \sigma_-(v_i) + v(v_i) + \sigma_+(v_i) & \text{if } 1 < i < k + 1 \\ \sigma_-(v_i) + v(v_i) & \text{if } i = k + 1 \end{cases} \quad (4)$$

The penalized distance function $G_n(v_i)$ can be expressed as

$$G_n(v_i) = \frac{1}{n} \Delta_n(v_i) + \lambda_p \frac{1}{k+1} P(v_i) \quad (5)$$

where the penalty factor $\lambda_p > 0$ and $\lambda_p = \lambda' \cdot \frac{k}{n^{1/3}} \cdot \frac{\sqrt{\Delta_n(f_{k,n})}}{r}$. λ' is the parameter of the method and is the constant value 0.13.

2) IMPROVEMENT

The CPC's main improvements over the TPC are initialization, normalization, stop conditions, and constraint conditions.

a: INITIALIZATION AND NORMALIZATION

The improved first principal component line begins with a closed square containing all the projection points, which is a big change for dealing with the closed dataset. The coordinates of the vertexes of the initial closed square are $\{(0.1, 0.1), (-0.1, 0.1), (-0.1, -0.1), (0.1, -0.1), (0.1, 0.1)\}$ along the counter-clockwise direction. The TPC's initialization has been replaced by the improved initialization step, shown in the FIGURE 6. To unify the dataset, the dataset $X = \{x_1, x_2 \dots x_n\}$ is normalized into the range of $\{(-1, -1) \sim (1, 1)\}$.

b: STOP CONDITIONS

When the number of line segments meet the condition $k > c(n, \Delta_n(f_{k,n})) = \beta n^{1/3} \Delta_n(f_{k,n})^{-1/2} r$ and the optimal parameter β is 0.3 [36], the whole loop will stop. The selection of the number of line segments depends on the average squared distance (Eq. (4)) to achieve robustness. When the variance of the noise is relatively small, we can use a relatively large number of segments to keep the approximation error low. When the variance of the noise is relatively large, a low approximation error does not improve the overall performance significantly, so at this time we choose to use a smaller number of segments k .

While in the inner loop, the value of the current distance function is compared with that of the last inner loop distance

function, when the reduced value is less than the maximum distance deviation $\Delta s = 0.002$, the inner loop will exit. While in the outer loop, the value of the current distance function is compared with that of the last outer loop distance function, and when the reduced value is less than the maximum distance deviation $\Delta s = 0.002$, the outer loop will exit. The distance deviation Δs is determined based on several trial runs.

c: CONSTRAINT CONDITIONS

The constraint conditions are mainly about the optimal selection of vertexes, line segments, and the shape of the contour curve. When inserting a new vertex, the whole dataset is selected to project to the segment lines or the determined vertexes, then the distance function from data points to the curve can be calculated. If the value of the distance function becomes smaller, then the position of each vertex will be adjusted; otherwise another new vertex is chosen to test. The longest line segment is selected to describe the approximate contour on the premise that the line segment contains the most projection points. The obtained contour curve keeps closed all the time, while the angle between the lines on the curve is $90^\circ < \theta < 180^\circ$.

3) A UNIFIED MATHEMATICAL MODEL FOR OBTAINING THE SMOOTH CONTOUR

The innovation of this step is to express the smooth and unified mathematical model by using the parameters of the machine learning model, and the mathematical model is the proposed here. Combining with the machine learning model for training, the model error can be minimized to refine the coarse segmentation result.

Because the feed-forward neural network with one hidden layer can be used to approximate any continuous function, we chose the BNN with only one hidden layer. The data sequence d is treated as the input of the model, and the output layer contains two units: x and y . x and y can be regarded as the continuous functions $x(t)$ and $y(t)$, respectively, on projection index t , where the projection index t is regarded as the independent variable, and the coordinates of the point $c(x,y)$ are the dependent variable. The data sequence d obtained by the CPC is expressed as $D = \{d_i = (t_i, c_i), i = 1, 2, \dots, n; t_1 < t_2 < \dots < t_n\}$, where the projection index t can be defined as follows:

$$t_f(x) = \sup \left\{ t : \|x - f(t)\| = \inf_{\tau} \|x - f(\tau)\| \right\} \quad (6)$$

where $\|\cdot\|$ is the associated Euclidean norm of R^d and τ is an auxiliary variable defined in R .

ReLU activation function $h_1(x) = \max(x, 0)$ and linear activation function $h_2(x) = x$ are used in the hidden layer and output layer, respectively. After training is complete, a unified and smooth neural network expression of lung contour is obtained as follows.

$$f(t) = (x(t_f(x)), y(t_f(y))) = \left(\sum_{i=1}^S \left(\frac{1}{1 + e^{-(t\omega_i - b_{1,i})}} \right) v_{i,1} \right)$$

TABLE 1. The parameters used in this model.

Type of the layer of BNN	Parameters of each layer	Parameters of whole model
Input Layer	#Input: data sequence (d) *Number of hidden neurons (S)	
Hidden Layer	#Number of hidden layers *Learning rate *Epoch *Momentum #Threshold (b _{1,i}) #Activation function: h ₁ (x)=max(x,0)	*Weight (w _i , v _{i,k})
Output Layer	#Output (x, y) #Threshold (b _{2,k}) #Activation function: h ₂ (x)=x	

$$- b_{2,1} \cdot \sum_{i=1}^S \left(\frac{1}{1 + e^{-(w_i - b_{1,i})}} \right) v_{i,2} - b_{2,2} \quad (7)$$

where S is the number of hidden neurons. w_i (i = 1, 2, . . . , S) and v_{i,k} (i = 1, 2, . . . , S; k = 1, 2) are the weights from the input layer to the i-th hidden neuron and from the i-th hidden neuron to the k-th output neuron, respectively. b_{1,i} (i = 1, 2, . . . , S) and b_{2,k} (k = 1, 2) are the thresholds of the i-th hidden neuron and the k-th output neuron, respectively. TABLE 1 shows the parameters used in this model, where # denotes the parameters validated in this paper, and * denotes the parameters discussed in our previous work [23].

III. EXPERIMENTS AND RESULTS

The dataset of chest CT scans used to evaluate the performance of the proposed model are provided by the 2nd Hospital of Soochow University. The dataset consists of 100 anonymous chest CT scans of 100 patients. For model training and quantitative evaluation, we used only one slice in the axial view from each patient to minimize physicians' manual labeling in this work. The format of the dataset is DICOM, and it contains 512*512*16-bit images. The chest CT images were obtained by a Philips CT Scanner (Philips Brilliance 40 Slice CT Scanner) using a high-resolution chest protocol [37], [38]. Each Ground Truth (GT) was marked and verified by five professional radiologists. Each radiologist independently checked their own marks along with the anonymous marks of the other radiologists, and the consensus GT was obtained by the majority voting of five experts' annotations. FIGURE 7 shows the workflow for extracting the GT.

In this section, we describe the series of experiments used to evaluate the proposed method. First, we describe the evaluation metrics (Section III-A). Second, we compare the experimental results achieved by the proposed method with those obtained by expert manual segmentation step by step, and discuss the optimal selection of parameters of the proposed method (Section III-B). Finally, we compare the proposed method with different state-of-the-art methods quantitatively and qualitatively (Section III-C and III-D). All experiments presented here were performed on a computer with Intel Core

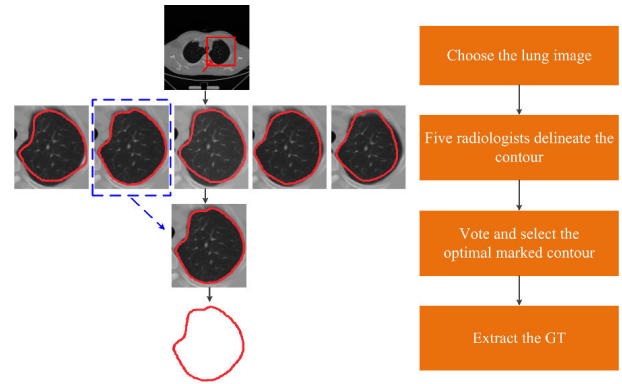


FIGURE 7. The workflow of extracting the GT.

TABLE 2. The definitions of TP, FP, FN and TN.

Test result label	Ground truth label	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

i5-4590 CPU @ 3.30GHz and GeForce GT 710 with 2GB memory.

A. EVALUATION METRICS

We evaluated the performance of lung segmentation methods with the following metrics: Dice Similarity Coefficient (DSC), Jaccard similarity coefficient (Ω), Sensitivity (Sen), Positive Predictive Value (PPV), Global Error (E), Average Symmetric Surface Distance (ASSD), Maximum Surface Distance (MSD), global Euclidean squared distance function (Δf), and execution time (t). Let TP, FP, FN and TN represents True Positive, False Positive, False Negative and True Negative, respectively, as shown in TABLE 2. The evaluation metrics can be expressed as below.

$$\Omega = \frac{TP}{FP + TP + FN}$$

$$DSC = \frac{2TP}{2TP + FP + FN}$$

$$Sen = \frac{TP}{TP + FN}$$

$$PPV = \frac{TP}{TP + FP}$$

It is worth noting that we used both Ω and DSC to evaluate the accuracy of the method. The transformation formula between these metrics is as follows:

$$\Omega = \frac{DSC}{2 - DSC} \quad (8)$$

We used Global Error (E) to evaluate the convergence of the machine learning model [39], and E_k is the mean square

error [40].

$$E = \sum_{i=1}^N E_k \tag{9}$$

Considering that DSC and Ω are both volume-based metrics, we used ASSD and MSD as another type of evaluation parameters: surface-based metrics [41].

We also used execution time and Δf to qualitatively evaluate the performance of the proposed algorithm; these metrics show the execution time of the whole system and the global Euclidean squared distance function, respectively [42]. The smaller the Δf , the closer the obtained contour f is to the real distribution of the dataset, where the accuracy of the obtained results is higher.

B. SELECTING OF THE BEST PERFORMANCE OF THE PROPOSED MODEL

In this section, we analyze the experimental results in detail from both the preprocessing and the refinement steps and select the optimal parameters of the proposed model.

1) PREPROCESSING

Many studies use the points manually delineated by professional radiologists as initial input; this process is expensive and the input is often difficult to obtain [43]. To solve this problem, our proposed preprocessing step automatically completes the coarse segmentation on the raw images. FIGURE 8 shows a randomly selected segmentation result obtained using the preprocessing step; the red line shows the GT, and the blue line shows the experimental results.

The output of the Otsu and PSCCL steps in the preprocessing step is used to extract lung regions. The CH is used only to obtain the coarse segmentation contour. When the lung has an irregular shape, the segmented contours obtained by CH only will deviate from the GT. To extract an accurate lung contour, we use the refinement step to fine-tune the coarse segmentation results.

2) REFINEMENT STEP

In this section, we use the refinement step consisting of the CPC and BNN to obtain the high-precision lung boundary. We used quantitative analysis to select the best performance of the proposed model and qualitative analysis to evaluate the optimal model. The quantitative results are the average value of all the results.

a: QUANTITATIVE ANALYSIS

FIGURE 9 through 13 show the quantitative experimental results of the DSC, Ω , Sen, PPV, and E under different numbers of hidden neurons after the refinement step. We randomly chose 50 cases from the hospital database to investigate the influence of different parameters of the proposed method. In FIGURE 9 through 12, the DSC, Ω , Sen, and PPV generally increase as the epochs increase. When the epochs are 1000, we obtained the highest DSC, Ω , Sen, and PPV

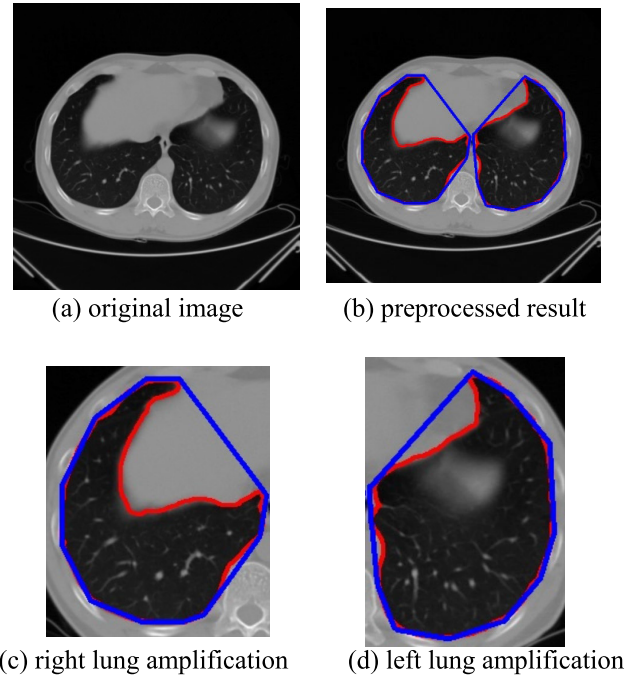


FIGURE 8. Illustration of lung boundaries obtained by the preprocessing step for one randomly selected case.

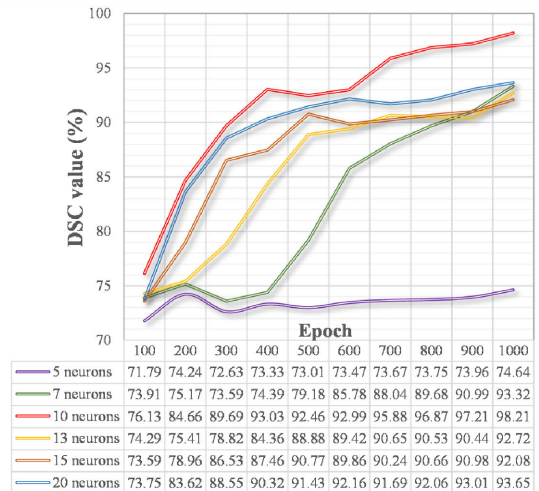


FIGURE 9. Changes in DSC at different numbers of epochs.

when the number of neurons was 10. We obtained the second comprehensive performance with 20 neurons, but the DSC, Ω , Sen, and PPV were 4.86%, 9.41%, 4.27%, and 4.02% lower, respectively. The DSC, Ω , Sen, and PPV were the smallest with 5 neurons. The trend of the evaluation parameters is similar with the other number of hidden neurons.

When either 13 or 15 hidden neurons are chosen, the DSC, Ω , Sen, and PPV decrease slightly and then rise again, as the epoch changes from 800 to 1000. This phenomenon is caused by the complicated model having too many parameters, which leads to overfitting. The evaluation parameters don't rise but decrease with the increasing epochs. In FIGURE 13,

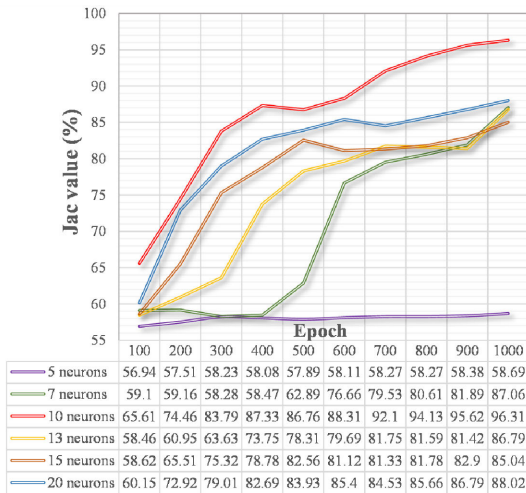


FIGURE 10. Changes in Ω at different numbers of epochs.

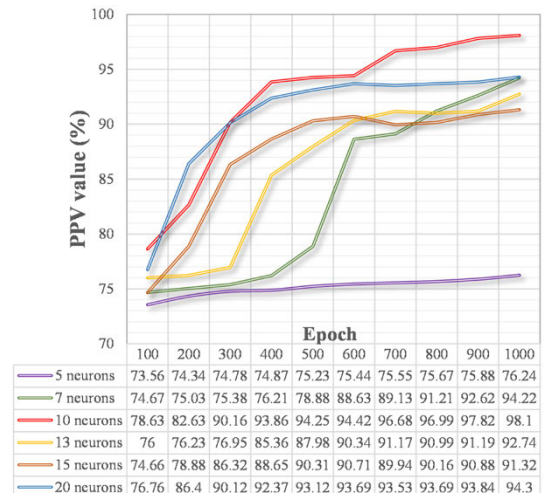


FIGURE 12. Changes in PPV at different numbers of epochs.

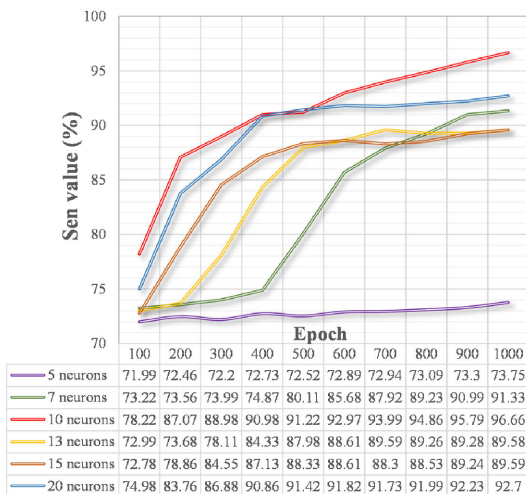


FIGURE 11. Changes in Sen at different numbers of epochs.

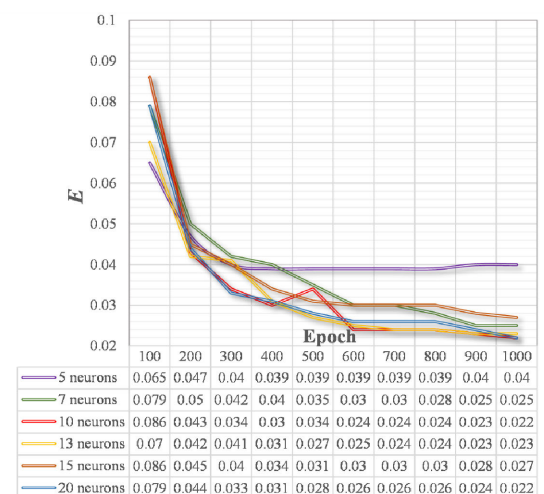


FIGURE 13. Changes in E at different numbers of epochs.

the E always decreases and eventually stabilizes. Except in the case of 5 neurons, all E values are lower than 0.03.

b: STEP-BY-STEP QUALITATIVE ANALYSIS OF THE PROPOSED PSCCL-CH-CPC

The preprocessing step of the proposed method consists of the Otsu, PSCCL, and CH, which performs the coarse segmentation to obtain the preprocess result. The improved CPC updates the positions of all vertexes according to the principle of minimizing the penalty distance function (Eq. (5)), while the position of each line segment is also adjusted, thus achieving the intermediate result. The final results show the lung contours after the refinement steps compared with the ground truth. FIGURE 14 illustrates step-by-step results of the proposed PSCCL-CH-CPC for one patient. We used the optimal model discussed in the previous section with 10 neurons and 1000 epochs for testing. For this test case only, GT contours were labelled on continuous 67 axial slices.

In FIGURE 14, the first row shows a representative slice on axial view, while the second row shows a coronal view.

As shown in FIGURE 14, the final results show good similarity with the ground truth.

C. COMPARISON WITH CONVENTIONAL METHODS

We compared the proposed method with the CPL-BNN and DBN-KNN semi-automatic models, which uses as little as 30% of the manually delineated points as inputs. The proposed PSCCL-CH-CPC is an automatic segmentation method, which uses the raw data as the input. We randomly chose three lung images from the clinical dataset as the research object.

Considering the optimal performance of the CPL-BNN, we set the hidden neurons to be 10 [23]. At the same time, considering the co-influence of time and accuracy on the model, we set DBN to two hidden layers, one with 25 neurons, and the other with 30 neurons. FIGURE 15 represents the

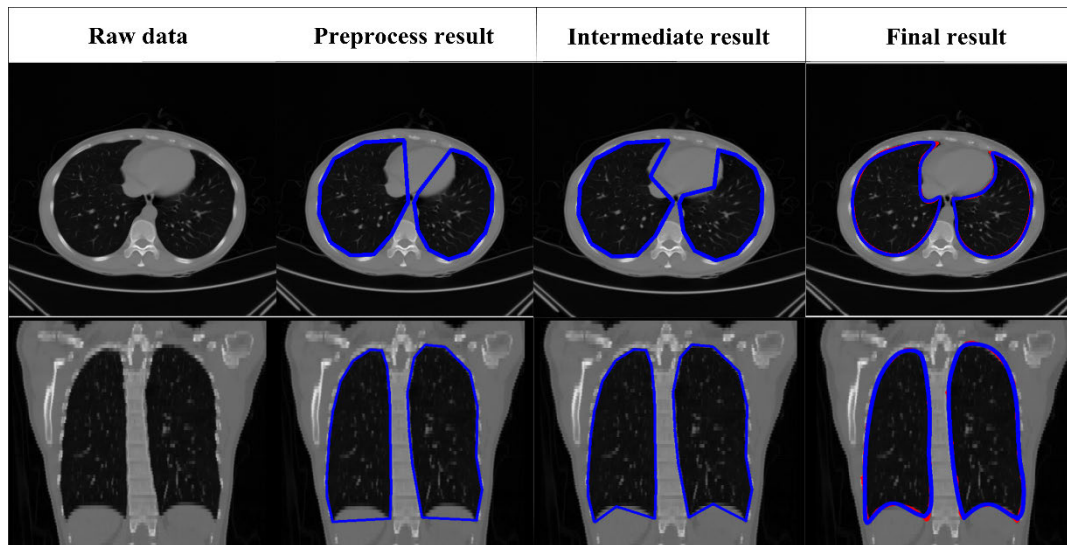


FIGURE 14. Illustration of lung contours after each step of the proposed PSCCL-CH-CPC method. The four columns represent the raw data, preprocess result, intermediate result, and final result, respectively. The first row shows an axial slice, and the second row shows the coronal view.

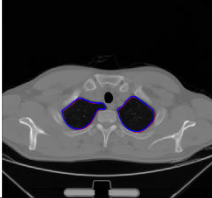
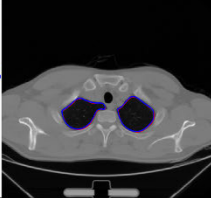
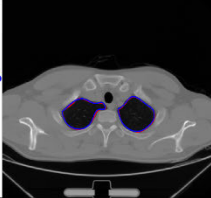
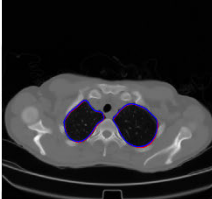
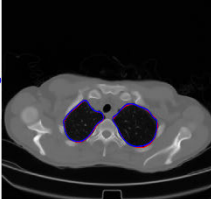
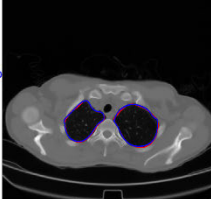
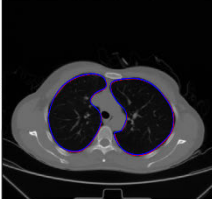
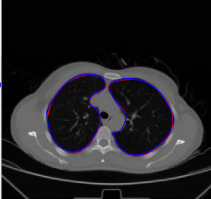

	PSCCL-CH-CPC	CPL-BNN	DBN-KNN
<i>Lung A</i>	 $DSC=98.2\%$ $\Delta f=0.024$	 $DSC=93.5\%$ $\Delta f=0.044$	 $DSC=88.4\%$ $\Delta f=0.045$
<i>Lung B</i>	 $DSC=97.9\%$ $\Delta f=0.034$	 $DSC=88.6\%$ $\Delta f=0.039$	 $DSC=83\%$ $\Delta f=0.043$
<i>Lung C</i>	 $DSC=97.2\%$ $\Delta f=0.06$	 $DSC=91.7\%$ $\Delta f=0.065$	 $DSC=86.2\%$ $\Delta f=0.067$

FIGURE 15. Global comparison among different algorithms on the clinical dataset.

global segmentation results obtained by different algorithms with 1000 epochs. FIGURE 16 and FIGURE 17 represent the partial magnification of the right and left lungs, respectively; the red lines and the blue lines denote the GT and the experimental results, respectively.

The DSC and Δf were used to evaluate qualitatively the performance of the three algorithms (FIGURE 15). Among the three models, the proposed PSCCL-CH-CPC had the best segmentation performance across different cases. The

PSCCL-CH-CPC's DSC was 10.4% higher (Lung B), and its Δf was 45.4% (Lung A) lower than the CPL-BNN, which had the second optimal performance. Among the three models, the DBN-KNN had the worst performance.

From FIGURE 16 and FIGURE 17, we can see that the three models have obvious deviations in the sudden and continuously turning areas, which are marked in frame 1 and frame 2. We used white and green frames to mark the most significant differences between the experimental results in the

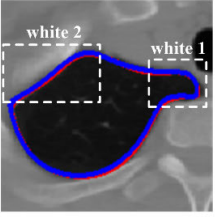
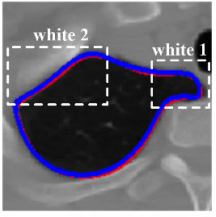
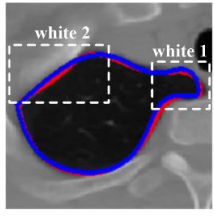
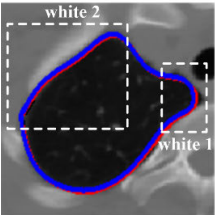
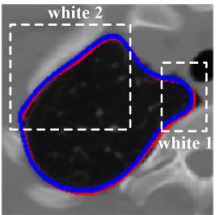
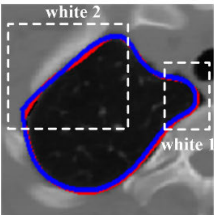
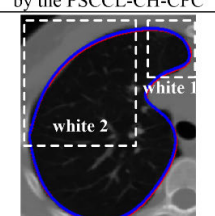
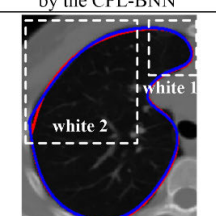
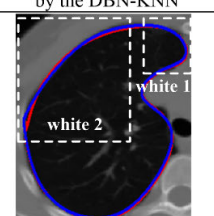
	PSCCL-CH-CPC	CPL-BNN	DBN-KNN
<i>Lung A</i>			
	The right <i>Lung A</i> dealt by the PSCCL-CH-CPC	The right <i>Lung A</i> dealt by the CPL-BNN	The right <i>Lung A</i> dealt by the DBN-KNN
<i>Lung B</i>			
	The right <i>Lung B</i> dealt by the PSCCL-CH-CPC	The right <i>Lung B</i> dealt by the CPL-BNN	The right <i>Lung B</i> dealt by the DBN-KNN
<i>Lung C</i>			
	The right <i>Lung C</i> dealt by the PSCCL-CH-CPC	The right <i>Lung C</i> dealt by the CPL-BNN	The right <i>Lung C</i> dealt by the DBN-KNN

FIGURE 16. Partial magnification graph of the right lung.

right and left lung, respectively. Considering how similar the situations of Lung A-Lung C are, we selected Lung B for detailed analysis.

Although the difference between the DSCs of the PSCCL-CH-CPC and the CPL-BNN was as high as 10.4% (FIGURE 15), both models detected the pulmonary boundary well at turning areas (FIGURE 16). The main reason for this phenomenon is that the CPL-BNN could not obtain the lung contour well at the continuous turning areas in left Lung B, which proves that using limited initial points leads to overfitting and seriously affects the model’s segmentation accuracy. The principal curve combined with machine learning model can fit the dataset better, but it consumes more training time. Sacrificing too much training time for high accuracy is not worthwhile, because the results are saturated. At the same time, the PSCCL-CH-CPC is a fully automatic method for contour extraction, which is better than the two semi-automatic methods.

The contour obtained by CPL-BNN is closer to the real distribution of the dataset than that obtained by the DBN-KNN, which proves that the CPL-BNN has a strong ability to fit the data accurately. This also shows that the proposed PSCCL-CH-CPC achieves the purpose of curve fitting through the refinement step so that it can continuously approach the center of the dataset and get the optimal results.

D. QUANTITATIVE EVALUATION

To evaluate the performance of our segmentation approach, we compared it with two state-of-the-art machine learning-based methods. In addition to DSC and Ω , two volume-based metrics, we used ASSD and MSD as surface-based metrics for evaluation. In order to compare performance of different methods under similar conditions, we used the Closed Polygonal Line and Backpropagation Neural Network Model (CPL-BNNM) [23] and Hull-Closed Polygonal Line Method (Hull-CPLM) [44] methods. We trained the CPL-BNNM and Hull-CPLM with the same 50 cases as the PSCCL-CH-CPC discussed in Section III-B. For all the three models, the remaining 50 cases were used for testing with 10 neurons and 1000 epochs. The DSC given by Eq. (8) was used to quantify the consistency between the segmented results and the GT provided manually by radiologists. TABLE 3 quantitatively compares our method and the state-of-the-art methods. All the execution time contains the training time. Our model showed better precision with higher average DSC, Ω , Sen, and PPV. Similarly, lower ASSD, MSD, and execution time mean better accuracy and efficiency. From TABLE 3, we can see that our method performed the best among the three models on all of the metrics, indicating our proposed method is effective and efficient.

TABLE 3. Comparison with other methods on 50 test cases.

Method	DSC	Ω	Sen	PPV	ASSD (mm)	MSD (mm)	Execution time (min)
CPL-BNNM [23]	93.9 %	91.3 %	94.1 %	93.6 %	3.271	26.18 6	35
Hull-CPLM [44]	95.9 %	93%	95.3 %	95.4 %	2.913	23.19 5	38
Ours	96.9 %	95.1 %	95.9 %	96.3 %	2.491	20.37 6	33

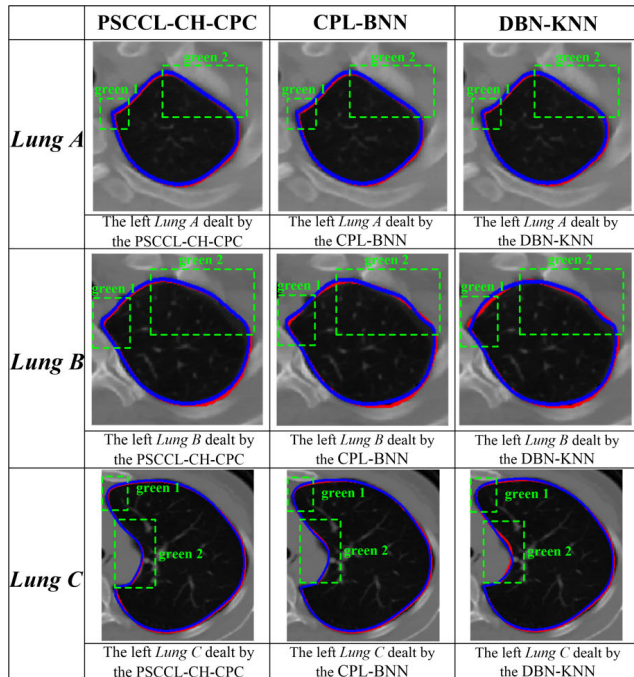


FIGURE 17. Partial magnification graph of the left lung.

IV. CONCLUSION

We have presented a novel model for fully automated lung segmentation called the PSCCL-CH-CPC. The innovation of the preprocessing step is to separate the lung and non-lung regions automatically. The effective combination of an improved CPC and a machine learning model to express a smooth mathematical expression of the lung contour is the innovation of the refinement stage of the proposed method, where the improved CPC and mathematical expression are proposed here for the first time. We tested the effectiveness and the robustness of our proposed model on a clinical dataset and evaluated its performance quantitatively and qualitatively by comparing it with state-of-the-art segmentation models using different evaluation metrics. Compared with the semi-automatic models, the proposed method saves the cost of manual annotation.

Thoracic dimensions and thoracoabdominal configuration differ between sexes. Females’ lung region, ribcage dimensions and diaphragm length are smaller than males’ [45]. In addition, males have a slightly larger cardiac diameter than females, and the average transverse cardiac diameter increases with age [46]. In the clinical decision-making process of CT, radiologists consider not only age and sex information but also visual clues. The proposed method does not

explicitly utilize age and sex information during model training and evaluation. Furthermore, lung segmentation is more challenging when the lung has moderate to high amounts of disease or abnormalities with a challenging shape or appearance [30]. In the future, we will systematically investigate whether our method is influenced by these factors. Our results have been validated by professional radiologists from the 2nd Hospital of the Soochow University. As segmentation is a critical initial step for any computer-aided detection and diagnosis system, the proposed hybrid technique has great potential to be integrated into a computer diagnosis system to assist the clinical analysis of chest diseases.

ACKNOWLEDGMENT

The authors would like to thank the 2nd Hospital of Soochow University for providing CT images, and Dr. Jonathan Feinberg for editing the article.

REFERENCES

- [1] A. El-Khuffash, A. T. James, J. D. Corcoran, P. Dicker, O. Franklin, Y. N. Elsayed, J. Y. Ting, A. Sehgal, A. Malikiwi, A. Harabor, A. S. Soraisham, and P. J. McNamara, “A patent ductus arteriosus severity score predicts chronic lung disease or death before discharge,” *J. Pediatrics*, vol. 167, no. 6, pp. 1354–1361, Dec. 2015.
- [2] F. Braidó, P. Santus, A. Corsico, F. Di Marco, G. Melioli, N. Scichilone, and P. Solidoro, “Chronic obstructive lung disease expert system validation of a predictive tool for assisting diagnosis,” *Int. J. Chronic Obstructive Pulmonary Disease*, vol. 13, pp. 1747–1753, May 2018.
- [3] P. Huang, S. Park, R. Yan, J. Lee, L. C. Chu, C. T. Lin, A. Hussien, J. Rathmell, B. Thomas, C. Chen, R. Hales, D. S. Ettinger, M. Brock, P. Hu, E. K. Fishman, E. Gabrielson, and S. Lam, “Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: A matched case-control study,” *Radiology*, vol. 286, no. 1, pp. 286–295, Jan. 2018.
- [4] S. Dai, K. Lu, J. Dong, Y. Zhang, and Y. Chen, “A novel approach of lung segmentation on chest CT images using graph cuts,” *Neurocomputing*, vol. 168, pp. 799–807, Nov. 2015.
- [5] J. Zhang, Y. Xia, H. Zeng, and Y. Zhang, “NODULE: Combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection,” *Neurocomputing*, vol. 317, pp. 159–167, Nov. 2018.
- [6] M. Wanet, J. A. Lee, B. Weynand, M. De Bast, A. Poncelet, V. Lacroix, E. Coche, V. Grégoire, and X. Geets, “Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: A comparison with threshold-based approaches. CT and surgical specimens,” *Radiotherapy Oncol.*, vol. 98, no. 1, pp. 117–125, Jan. 2011.
- [7] J. Lai and Q. Wei, “Automatic lung fields segmentation in CT scans using morphological operation and anatomical information,” *Bio-Med. Mater. Eng.*, vol. 24, no. 1, pp. 335–340, 2014.
- [8] A. O. de Carvalho Filho, W. B. de Sampaio, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, “Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index,” *Artif. Intell. Med.*, vol. 60, no. 3, pp. 165–177, Mar. 2014.
- [9] B. Ait Skourt, A. El Hassani, and A. Majda, “Lung CT image segmentation using deep neural networks,” *Procedia Comput. Sci.*, vol. 127, pp. 109–113, Jan. 2018.
- [10] H. Zhou, D. B. Goldgof, S. Hawkins, L. Wei, Y. Liu, D. Creighton, R. J. Gillies, L. O. Hall, and S. Nahavandi, “A robust approach for automated lung segmentation in thoracic CT,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Hong Kong, Oct. 2015, pp. 2267–2272.
- [11] S. Mukhopadhyay, “A segmentation framework of pulmonary nodules in lung CT images,” *J. Digit. Imag.*, vol. 29, no. 1, pp. 86–103, Feb. 2016.
- [12] A. A. Kiaei and H. Khotanlou, “Segmentation of medical images using mean value guided contour,” *Med. Image Anal.*, vol. 40, pp. 111–132, Aug. 2017.

- [13] A. Shimizu, K. Nakagomi, T. Narihira, H. Kobatake, S. Nawano, K. Shinozaki, K. Ishizu, and K. Togashi, "Automated segmentation of 3d ct images based on statistical atlas and graph cuts," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 214–223. [Online]. Available: <https://www.springerprofessional.de/en/medical-computer-vision-recognition-techniques-and-applications-/3593954>
- [14] L. Zhang, E. A. Hoffman, and J. M. Reinhardt, "Atlas-driven lung lobe segmentation in volumetric X-ray CT images," *IEEE Trans. Med. Imag.*, vol. 25, no. 1, pp. 1–16, Dec. 2005.
- [15] S. Shen, A. A. T. Bui, J. Cong, and W. Hsu, "An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy," *Comput. Biol. Med.*, vol. 57, pp. 139–149, Feb. 2015.
- [16] Z. F. Khan and A. Kannan, "Intelligent segmentation of medical images using fuzzy bitplane thresholding," *Meas. Sci. Rev.*, vol. 14, no. 2, pp. 94–101, Apr. 2014.
- [17] A. Pratondo, C.-K. Chui, and S.-H. Ong, "Integrating machine learning with region-based active contour models in medical image segmentation," *J. Vis. Commun. Image Represent.*, vol. 43, pp. 1–9, Feb. 2017.
- [18] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan, B. Kainz, B. Glocker, and D. Rueckert, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 384–395, Feb. 2018.
- [19] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, "Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks," *Med. Phys.*, vol. 45, no. 10, pp. 4558–4567, Oct. 2018.
- [20] M. Keshani, Z. Azimifard, F. Tajeripour, and R. Boostani, "Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system," *Comput. Biol. Med.*, vol. 43, no. 4, pp. 287–300, May 2013.
- [21] Z. Li, Y. Zhang, H. Gong, G. Liu, W. Li, and X. Tang, "An automatic and efficient coronary arteries extraction method in CT angiographies," *Biomed. Signal Process. Control*, vol. 36, pp. 221–233, Jul. 2017.
- [22] S. You, E. Bas, E. Ataer-Cansizoglu, J. Kalpathy-Cramer, and D. Erdogmus, "Principal curve based semi-automatic segmentation of organs in 3D-CT," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Boston, MA, USA, Aug. 2011, pp. 6220–6223.
- [23] T. Peng, Y. Wang, T. C. Xu, L. Shi, J. Jiang, and S. Zhu, "Detection of lung contour with closed principal curve and machine learning," *J. Digit. Imag.*, vol. 31, no. 4, pp. 520–533, Aug. 2018.
- [24] S. Li, P. Xu, B. Li, L. Chen, Z. Zhou, H. Hao, Y. Duan, M. Folkert, J. Ma, S. Huang, S. Jiang, and J. Wang, "Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features," *Phys. Med. Biol.*, vol. 64, no. 17, 2019, Art. no. 175012.
- [25] M. Bisele, M. Bencsik, M. G. C. Lewis, and C. T. Barnett, "Optimisation of a machine learning algorithm in human locomotion using principal component and discriminant function analyses," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0183990.
- [26] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *Proc. 17th Comput. Anal. Image. Pattern*, Ystad, Sweden, Aug. 2017, pp. 36–46.
- [27] Y. Feng, H. Zhao, X. Li, X. Zhang, and H. Li, "A multi-scale 3D otsu thresholding algorithm for medical image segmentation," *Digit. Signal Process.*, vol. 60, pp. 186–199, Jan. 2017.
- [28] L. He, Y. Chao, and K. Suzuki, "Two efficient label-equivalence-based connected-component labeling algorithms for 3-D binary images," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2122–2134, Aug. 2011.
- [29] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *J. Digit. Imag.*, vol. 28, no. 1, pp. 99–115, Feb. 2015.
- [30] A. Mansoor, U. Bagci, B. Foster, Z. Xu, G. Z. Papadakis, L. R. Folio, J. K. Udupa, and D. J. Mollura, "Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends," *RadioGraphics*, vol. 35, no. 4, pp. 1056–1076, Jul. 2015.
- [31] B. Hua and R. Baldick, "A convex primal formulation for convex hull pricing," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3814–3823, Sep. 2017.
- [32] K. R. Anderson, "A reevaluation of an efficient algorithm for determining the convex hull of a finite planar set," *Inf. Process. Lett.*, vol. 7, no. 1, pp. 53–55, Jan. 1978.
- [33] T. Hastie and W. Stuetzle, "Principal curves," *J. Amer. Statist. Assoc.*, vol. 84, no. 406, pp. 502–516, Mar. 1989.
- [34] B. Kegl and A. Krzyzak, "Piecewise linear skeletonization using principal curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 59–74, Aug. 2002.
- [35] E. Bas and D. Erdogmus, "Principal curves as skeletons of tubular objects," *Neuroinformatics*, vol. 9, nos. 2–3, pp. 181–191, Sep. 2011.
- [36] B. Kegl, A. Krzyzak, T. Linder, and K. Zeger, "Learning and design of principal curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 3, pp. 281–297, Mar. 2000.
- [37] A. Scott, R. Merrill, and X. Wu, "SU-FF-I-75: Verification of reported DLP for philips brilliance CT scanners," *Med. Phys.*, vol. 36, no. 6, pp. 2451–2452, Jun. 2009.
- [38] L. Vlachopoulos, A. Schweizer, M. Graf, L. Nagy, and P. Furnstahl, "Three-dimensional postoperative accuracy of extra-articular forearm osteotomies using CT-scan based patient-specific surgical guides," *BMC Musculoskeletal Disorders*, vol. 16, no. 1, p. 336, Dec. 2015.
- [39] R. Rodriguez, O. O. V. Villegas, V. G. C. Sanchez, J. Bila, and A. Mexicano, "Arrhythmia disease classification using a higher-order neural unit," in *Proc. 4th Int. Conf. Future Gener. Commun. Technol. (FGCT)*, Jul. 2015, pp. 1–6.
- [40] N. E. Krasowski, T. Beier, G. W. Knott, U. Kothe, F. A. Hamprecht, and A. Kreshuk, "Neuron segmentation with high-level biological priors," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 829–839, Apr. 2018.
- [41] R. M. H. Besseling, J. F. A. Jansen, G. M. Overvliet, M. J. Vaessen, H. M. H. Braakman, P. A. M. Hofman, A. P. Aldenkamp, and W. H. Backes, "Tract specific reproducibility of tractography based morphology and diffusion metrics," *PLoS ONE*, vol. 7, no. 4, 2012, Art. no. e34125.
- [42] L. Vincent, "Exact Euclidean distance function by chain propagations," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Maui, HI, USA, Jun. 1991, pp. 520–525.
- [43] E. Rios Velazquez, H. J. W. L. Aerts, Y. Gu, D. B. Goldgof, D. De Ruyscher, A. Dekker, R. Korn, R. J. Gillies, and P. Lambin, "A semiautomatic CT-based ensemble segmentation of lung tumors: Comparison with oncologists' delineations and with the surgical specimen," *Radiotherapy Oncol.*, vol. 105, no. 2, pp. 167–173, Nov. 2012.
- [44] T. Peng, Y. Wang, T. C. Xu, and X. Chen, "Segmentation of lung in chest radiographs using hull and closed polygonal line method," *IEEE Access*, vol. 7, pp. 137794–137810, 2019.
- [45] F. Bellemare, A. Jeanneret, and J. Couture, "Sex differences in thoracic dimensions and configuration," *Amer. J. Respiratory Crit. Care Med.*, vol. 168, no. 3, pp. 305–312, Aug. 2003.
- [46] Y. Mensah, K. Mensah, S. Asiamah, H. Gbadamosi, E. Idun, W. Brakohiapa, and A. Oddoye, "Establishing the cardiothoracic ratio using chest radiographs in an indigenous ghanaiian population: A simple tool for cardiomegaly screening," *Ghana Med. J.*, vol. 49, no. 3, pp. 159–164, Sep. 2015.



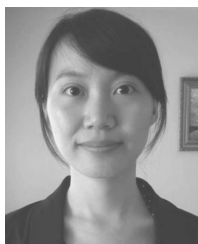
TAO PENG (Member, IEEE) is currently pursuing the Ph.D. degree with Soochow University, China. His main research interests include medical image processing and analysis, pattern recognition, machine learning, and their applications. He is a member of CCF.



THOMAS CANHAO XU (Member, IEEE) received the D.Sc. degree from the University of Turku, Finland, in 2012, and the M.Eng. degree from Zhejiang University, China, in 2007. He has been a System Engineer and an Analyst for several projects, since 2005. From 2008 to 2016, he was worked with the Software Engineering Laboratory and the Computer Systems Laboratory, University of Turku, as a Researcher, a Postdoctoral Researcher, and a Teaching Staff. He has been involved in several projects of Academy of Finland, EU, and China. He is currently an Associate Professor with Soochow University, China. His research interests include pattern recognition, machine learning, and their applications. He has authored or coauthored more than 50 international refereed articles. He has been a member of ACM, since 2007.



YIHUAI WANG received the Ph.D. degree from Hohai University, Nanjing, China. He is currently a Professor with Soochow University, Suzhou, China. He is also the Director of the Embedded System and the Internet of Things Technical Committee, Jiangsu Computer Federation. He is also the Director of the Embedded System and the Internet of Things Research Institute, Soochow University. His research mainly focuses on parallel systems, software engineering, embedded system and application, artificial intelligence, and pattern recognition. He is a Senior Member of CCF.



Her research interests include image segmentation, pattern recognition, and traffic modeling and simulation.



SEMA CANDEMIR (Member, IEEE) is currently a Research Scientist with The Ohio State University Wexner Medical Center. Before joining the OSU, she worked as a Research Scientist and a Postdoctoral Researcher with the National Library of Medicine, National Institutes of Health. Her research interest is medical image analysis.



Her research interests include intelligent software tools for medical image processing, analysis and diagnosis, study of retinal vascular in digital fundus image using image processing approach, and ocular diseases in ASPI screening systems.



SHANQ-JANG RUAN (Senior Member, IEEE) received the B.S. degree in computer science and information engineering from Tamkang University, Taiwan, in 1995, and the M.S. degree in computer science and information engineering and the Ph.D. degree in electrical engineering from National Taiwan University, Taiwan, in 1997 and 2002, respectively. He has served as an Electronic Officer with R.O.C. Air Force, from July 1997 to May 1999. From September 2001 to May 2002, he joined the Avant! Corporation, as a Software Engineer. From June 2002 to November 2003, he was with Synopsys, as a Senior Software Engineer. He is currently the Distinguished Professor with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology (NTUST), Taiwan, where he is also the Chairman of the Graduate Institute of Applied Science and Technology. His research interests include machine learning and image processing.



JING WANG received the B.S. degree in materials physics from the University of Science and Technology of China (USTC), in 2001, and the M.A. and Ph.D. degrees in physics from the State University of New York at Stony Brook, in 2003 and 2006, respectively. He finished his Postdoctoral Training with the Department of Radiation Oncology, Stanford University, in 2009. He joined UTSW in 2010, and he was promoted to a tenured Associate Professor, in 2016. He is currently an Associate Professor and the Director of the Data Analytics and Informatics, Department of Radiation Oncology, The University of Texas Southwestern Medical Center (UTSW). He has published more than 100 peer-reviewed journal articles. His research focuses on medical image reconstruction and processing, machine learning, and their applications in radiation therapy.



XINJIAN CHEN (Senior Member, IEEE) received the Ph.D. degree from the Center for Biometrics and Security Research, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2006. After completing the graduation, he joined Microsoft Research Asia, where he was involved in research on handwriting recognition. From 2008 to 2012, he conducted a Postdoctoral Research at several prestigious groups, such as the Medical Image Processing Group, Department of Radiology and Image Sciences, National Institutes of Health, University of Pennsylvania, and the Department of Electrical and Computer Engineering, University of Iowa. In 2012, he joined as a Full Professor with the School of Electrical and Information Engineering, Soochow University, where he is currently a Distinguished Professor, and also serves as the Director of the University Level Laboratory Medical Image Processing, Analysis and Visualization Laboratory. He has authored more than 70 high-quality international journal/conference articles. His research interests include medical image processing and analysis, pattern recognition, machine learning, and their applications.

...