Boundary delineation in transrectal ultrasound images for region of interest of prostate

Tao Peng[a, b, c, 1, *], Yan Dong[d, 1], Gongye DI[e, 1], Jing Zhao[f, 1], Tian Li[b, 1], Ge Ren[b, 1], Lei Zhang[g], Jing Cai[b, *]

[a] School of Future Science and Engineering, Soochow University, Suzhou, China

[b] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, China

[c] Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

[d] Department of Ultrasonography, The First Affiliated Hospital of Soochow University, Suzhou, China

[e] Department of Ultrasonic, Taizhou People's Hospital Affiliated to Nanjing Medical University, Taizhou, China

[f] Department of Ultrasound, Tsinghua University Affiliated Beijing Tsinghua Changgung Hospital, Beijing, China

[g] Medical Physics Graduate Program and Data Science Research Center, Duke Kunshan University, Kunshan, Jiangsu, China

Abstract

Accurate and robust prostate segmentation in transrectal ultrasound (TRUS) images is of great interest for ultrasound-guided brachytherapy for prostate cancer. However, the current practice of manual segmentation is difficult, time-consuming, and prone to errors. To overcome these challenges, we developed an accurate prostate segmentation framework (A-ProSeg) for TRUS images. The proposed segmentation method includes three innovation steps: (1) acquiring the sequence of vertices by using an improved polygonal segment-based method with a small number of radiologist-defined seed points as prior points; (2) establishing an optimal machine learning-based method by using the improved evolutionary neural network; and (3) obtaining smooth contours of the prostate region of interest using the optimized machine learning-based method. The proposed method was evaluated on 266 patients who underwent prostate cancer brachytherapy. The proposed method achieved a high performance against ground truth with a Dice similarity coefficient (DSC) of 96.2% ± 2.4%, a Jaccard similarity coefficient (Ω) of 94.4% ± 3.3%, and an accuracy (ACC) of 95.7% ± 2.7%, which are all higher than the state-of-the-art methods. Sensitivity evaluation on different noise levels demonstrated that our method achieved high robustness against changes in image quality. Meanwhile, an ablation study was performed, and the significance of all the key components of the proposed method was demonstrated.

Keywords: Prostate segmentation, transrectal ultrasound, global closed polygonal segment, distributed-based memory differential evolution, neural network, explainability-guided mathematical model.

1 Introduction

Prostate cancer is one of the leading causes of death among men worldwide. Transrectal ultrasound (TRUS) is routinely utilized in the diagnosis and brachytherapy of prostate disease [1]. Prostate gland segmentation is an essential step of TRUS-guided brachytherapy because it defines the target volume to be irradiated [2]. Precise segmentation is a prerequisite for successful treatment because segmentation errors may lead to an under-dose to the target and/or an over-dose to the surrounding normal tissues. The current clinical practice involves manual segmentation of the prostate, which is a tedious and time-consuming process; moreover, its accuracy often depends on the clinician's experience [3]. Despite many advances in medical image segmentation, TRUS prostate

---

[1] Authors contributed to this manuscript equally.
* Corresponding author:
E-mail addresses: sdpengtao401@gmail.com (Tao Peng), jing.cai@polyu.edu.hk (Jing Cai)

segmentation is accompanied by unique challenges owing to poor image quality and intensity heterogeneity, as shown in Fig. 1.
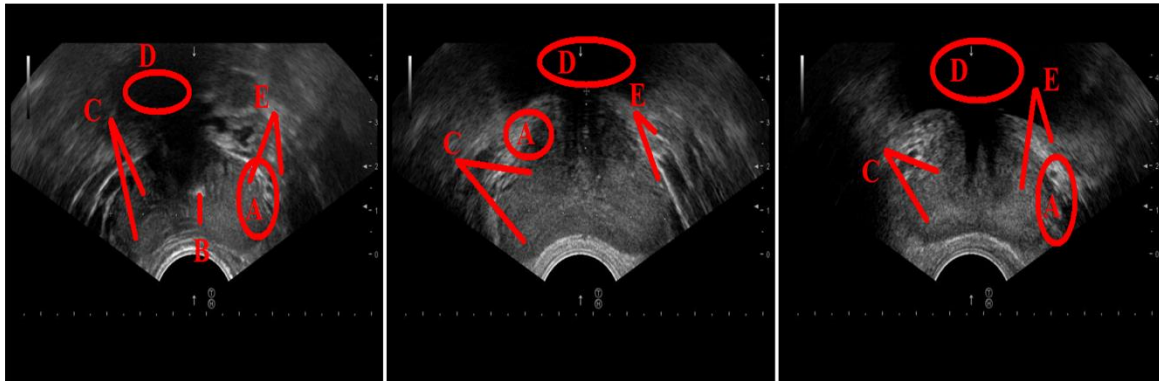

Fig. 1 Challenges in prostate segmentation caused by different factors. A) low signal-to-noise ratio (SNR), B) micro-calcification, C) intensity heterogeneity within the prostate, D) shadow artifacts, and E) speckle noise.

1.1 Related studies

Recently, four common imaging modalities have been used for disease diagnosis, including X-rays, magnetic resonance imaging (MRI), computed tomography (CT), and TRUS [4]. Unlike the other three modalities, TRUS has a higher dependence on prior information regarding shape features for prostate segmentation [5]. This is mainly because the outline cue of intrinsic ultrasound images is inadequate for accurate prostate outline detection [6]. In the field of medical image segmentation, the currently used methods are generally categorized into two types—fully automatic methods and semi-automatic methods—as briefly summarized below.

(i) *Fully automatic methods*: Currently, automatic prostate segmentation in TRUS images is a significant topic of interest [7]. A polar transformer network [8], which is used to segment the prostate in the polar coordinate system instead of the image coordinate system, was developed for ultrasound prostate segmentation. From the perspective of the polar coordinate system, the prostate surface is well parameterized through a radius map on the object surface regarding the coordinate of the centroid point. This assists the proposed network in improving the segmentation accuracy. However, the selection of the location of the centroid point may affect the location of the prostate. In addition, the performance of this model is severely influenced when the shape of the prostate is not convex-based. A novel feature pyramid network [9] equipped with attention modules to generate deep attentive features for prostate segmentation in TRUS images has been proposed. This method generates accurate results because the attention module emphasizes the salient features well. However, a total of forty patients' data was used for training and testing, which potentially caused the overfitting owing to limited training samples. Hence, the author mentioned that they would concentrate on assessing the generalizability of their method on a larger dataset. In addition, no validation data was used for hunting for the optimal hyper-parameters of the neural network. A unity squeeze-and-excitation (SE) network model [10] has been reported, which combines the SE modules with Unity Network (U-Net) for prostate zonal segmentation. Owing to the advantage of the SE module in boosting meaningful features and suppressing the less useful ones, the Dice similarity coefficient (DSC) of the proposed method is as high as 0.91. However, the dataset was only divided into the training and testing groups, without a validation set for optimal model selection. A deep-learning model has been used to determine the effects of different parameters of the dataset, including dataset size, image quality, and image type, on the ultrasound prostate segmentation task [11]. When using 100 to 4,023 images for training, the corresponding DSC of the proposed method ranged from 0.72 to 0.91, in a dataset that included clinical prostate brachytherapy images. In contrast, the DSC of the proposed method was only 0.85 when using 1,000 images for training. However, owing to the difficulties in acquiring ultrasound prostate images, obtaining more than 4,000 images is a challenge for resource-

limited hospitals.

(ii) ***Semi-automatic methods***: When using semi-automatic methods for prostate segmentation, the intervention of an expert radiologist is always sought to initialize the implementation procedure or refine the final result, which yields accurate and robust results [12]. A semi-automatic segmentation method [13] has been proposed and evaluated in prostate TRUS images. This method uses the inherent characteristics of the original images, with information on the shape, before fine-tuning the final segmentation results. However, the model is susceptible to noise and speckles. In addition, a multi-pipeline framework, including feature extraction, classification, and level set segmentation [14], has been developed. The segmentation step was implemented after classifying the obtained region of interest (ROI) features and background features, whereas a shape prior was used to correct the final segmentation results. However, 47 images were used for training and the other 85 images were used for testing, which may potentially cause overfitting at the training stage. A multi-label joint learning-based method [15] has also been described for ultrasound prostate segmentation. This method potentially incorporates prior knowledge of the dataset, such as space location information about the TRUS volume, the clinical target, and the planning target. However, the precision of labels can potentially influence the performance of the proposed method. In addition, the accuracy of the method is severely affected by the prostate's variable shape and size. A semi-supervised learning method was designed [16] for prostate segmentation using prior knowledge of the shadow artifacts. During training, the model can learn how to distinguish shadow artifacts from the prostate ROI, which improves the segmentation accuracy.

1.2 Contributions/innovations of this work

In our study, we introduce the technical contributions of our boundary delineation method in the ultrasound image processing area, as indicated below:

1. During TRUS-guided brachytherapy, segmenting the prostate gland is significant, which can accurately determine the target volume that is to be irradiated. Meanwhile, it can ensure the dosage of the target is appropriate. The purpose of our study is to achieve a high-accuracy contour that can make sure successful treatment during TRUS-guided brachytherapy.

2. Considering that the precise detection of prostate contour from ultrasound images is challenging work, the DSCs of recent fully-automatic methods are nearly 0.9 [16] [17]. Our study develops a semi-automatic contour detection approach that adopts radiologist-defined data seed points as prior, resulting in a DSC of 0.962.

3. Owing to their excellent capability to deal with noisy input, many researchers have used the principal curve (PC) technology to separate unusual tissues from the neighboring normal tissues [18]. Nevertheless, the potential problem with PC methods is that the number of vertices must be pre-decided by the users, and our developed technique herein focuses on this issue.

Compared with current methods, there are four primary novelties of the segmentation method proposed in this study:

1. We present a hybrid segmentation architecture that combines an improved, closed-principal curve-based method with an improved machine-learning method. This method uses the principal curve-based method for automatically approaching the center of the dataset and machine learning during the training stage to reduce model error.

2. Our previous studies [18] [19] [20] were our first attempts to utilize a closed *K*-polygonal segment (CKPS) method for medical image segmentation with promising performance. Based on these previous studies, in this study, we propose a novel global closed polygonal segment (GCPS) method, with improvements in 1) vertex cleaning, 2) the vertex-merge model, and 3) constraint conditions.

3. A distributed-based memory differential evolution (DMDE) method has been newly proposed for the

3

optimal initialization of the Caputo fractional-order backpropagation learning (CFBL) method [21]. Unlike the originally distributed differential evolution framework proposed in Ref. [22], we used a new memory-based differential evolution (DE) model to replace the original DE when processing the slave node.

4. Through careful analysis and experimental verification, an intrinsic problem was found with the principal curve-based method: the method lacks a mechanism to obtain a smooth contour. To solve this problem, a map function (represented by parameters of the CFBL) was designed, which can smooth the prostate contour while matching the ground truth (GT) contour.

The initial results of this study were published as a conference paper at the 2021 IEEE International Conference on Bioinformatics and Biomedicine conference [23]. This manuscript extends the contents of the conference paper in several aspects.

1. A detailed review of the literature has been included to place our work within the context of existing techniques. We have presented our work in more detail and have included comprehensive results, which were constrained in our conference paper owing to the limited number of pages allowed.

2. We have added more details of each experiment as follows. *First*, we have added qualitative and quantitative experiments to discuss the impact of patient age. *Second*, we used higher noise with a lower signal-to-noise ratio to corrupt the testing data, which was then used to test the accuracy and robustness of our method. *Third*, we adjusted the number of validation and testing datasets. We used more testing data to evaluate the prediction performance adequately. *Fourth*, we performed more ablation experiments to investigate the influence of each component of our method. *Fifth*, we added the standard deviation of each metric (i.e., DSC, Jaccard similarity coefficient (Ω), accuracy (ACC), and Hausdorff distance (HD)) to evaluate the robustness of our method further.

3. A comprehensive comparison of our methods with those reported in the literature is presented. In the initial paper, we compared our deep learning architecture (i.e., Mask RCNN [24]) with hybrid architecture (i.e., Hull-CPS [19] and a deep-belief network (DBN)-CPS [20]). Our model was also compared with a more recent hybrid segmentation architecture known as the H-SegMod method, which was proposed in our previous study [25]. Furthermore, our A-ProSeg was compared with a well-known transformer-based architecture known as the UTNet method [26], and good segmentation architecture named as U-Net [27] and U-Net++ [28]. Meanwhile, the statistical significance of all the methods was also investigated.

2. Methods

2.1 Problem statement

Let the initial seed points set be $P = \{p_i, p_2, ..., p_n\}$, which represent the limited manually selected seed points. We expect to acquire a smooth outline of the prostate ROI, for which the seed point set P is known as a prior. Let us assume that the seed point P is perfect without any abnormal point p_j, and we expect to find a suitable polygonal segment-based method [18] for training, which is a variant of the principal curve-based methods [29]. After training, the optimal vertex sequence D, which consists of the vertex sequence t and vertex coordinates $v(x_i, y_i)$ [30], was adaptively found. At this time, the results represented the polygon-based prostate contour [18], which was not smooth. Then, using t as the input of the CFBL method and the vertex coordinates (v_x, v_y) for minimizing the mean square error (MSE) [9], CFBL was used as the training model. When the training was completed, CFBL had the best performance, and the MSE and DSC values were infinitely close to 0 and 1, respectively. Finally, the coordinates of the optimized vertices were used to represent the prostate contour, as shown in Eq. (1) and Eq. (2). This mathematical expression has been denoted in our previous conference paper [23]. For ease of understanding,

both of those equations are presented again:

$$f(t) = \left(x(t), y(t) \right) = \left(\frac{c(x(t)) + 1 - \sqrt{1 - (c(x(t)))^2}}{2 * c(x(t))}, \frac{c(y(t)) + 1 - \sqrt{1 - (c(y(t)))^2}}{2 * c(y(t))} \right)$$ (1)

In Eq. (1), the output layer's neuron $c(\bullet)$ consisting of $c(x)$ and $c(y)$ is obtained, which are regarded as the continuous functions $c(x(t))$ and $c(y(t))$, respectively, on the vertex sequence t, as shown below:

$$\left(c(x(t)), c(y(t)) \right) = \left(\frac{e^{\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{1,i})}} w2_{i,1} - b_{2,1}} - e^{-(\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{1,i})}} w2_{i,1} - b_{2,1})}}{e^{\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{1,i})}} w2_{i,1} - b_{2,1}} + e^{-(\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{1,i})}} w2_{i,1} - b_{2,1})}}, \frac{e^{\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{2,i})}} w2_{i,2} - b_{2,2}} - e^{-(\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{2,i})}} w2_{i,2} - b_{2,2})}}{e^{\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{2,i})}} w2_{i,2} - b_{2,2}} + e^{-(\sum_{i=1}^{h} \frac{1}{1+e^{-(tw1_i - b_{1,i})}} w2_{i,2} - b_{2,2})}} \right)$$ (2)

where w and b denote the weight and threshold of the CFBL model, respectively, and h represents the number of hidden neurons. The details of both equations can be found in our conference paper [23].

However, two challenging issues may appear in this framework:

1. The quality of dataset P would be affected when there are abnormal points p (i.e., uneven distribution or incorrect locations). Hence, we need to find the appropriate polygonal segment-based methods to remove the unavailable points and obtain the optimal vertex sequence D.

2. After training, the model parameters of the optimal CFBL model would be adopted to represent the coordinates of the optimized vertices used to update the prostate contour, as shown in Eq. (1) and Eq. (2). However, the option of the CFBL model's parameters would affect the performance of CFBL, causing the selection of the CFBL with optimal model parameters to be a challenging issue.

2.2 Overall architecture

In this section, we present the use of our hybrid A-ProSeg architecture for obtaining a smooth contour of the arbitrary ROI using a small number of manually selected seed points P as a prior (<8% of points as a prior). The main steps of our model are as follows:

1. Using the GCPS, a polygon-based curve f comprising segments $S = \{s_1, s_2, .. s_{is}\}$ was obtained.
2. The DMDE was used to achieve the best initialization of the parameters (i.e., weight and threshold) of the CFBL model.
3. The precision of the prostate contour was improved by constraining the global deviation E, shown as $E = \sum_{k=1}^{m} E_k$, during CFBL training, in which E_k represents the MSE calculated by the actual and expected outputs.
4. The smooth mathematical-model-based coordinates of the optimized vertices (shown in Eq. (1) and Eq. (2)) were denoted to represent the prostate contour.

Fig. 2 illustrates the overall A-ProSeg architecture. The details of our A-ProSeg architecture are shown in the following subsections.
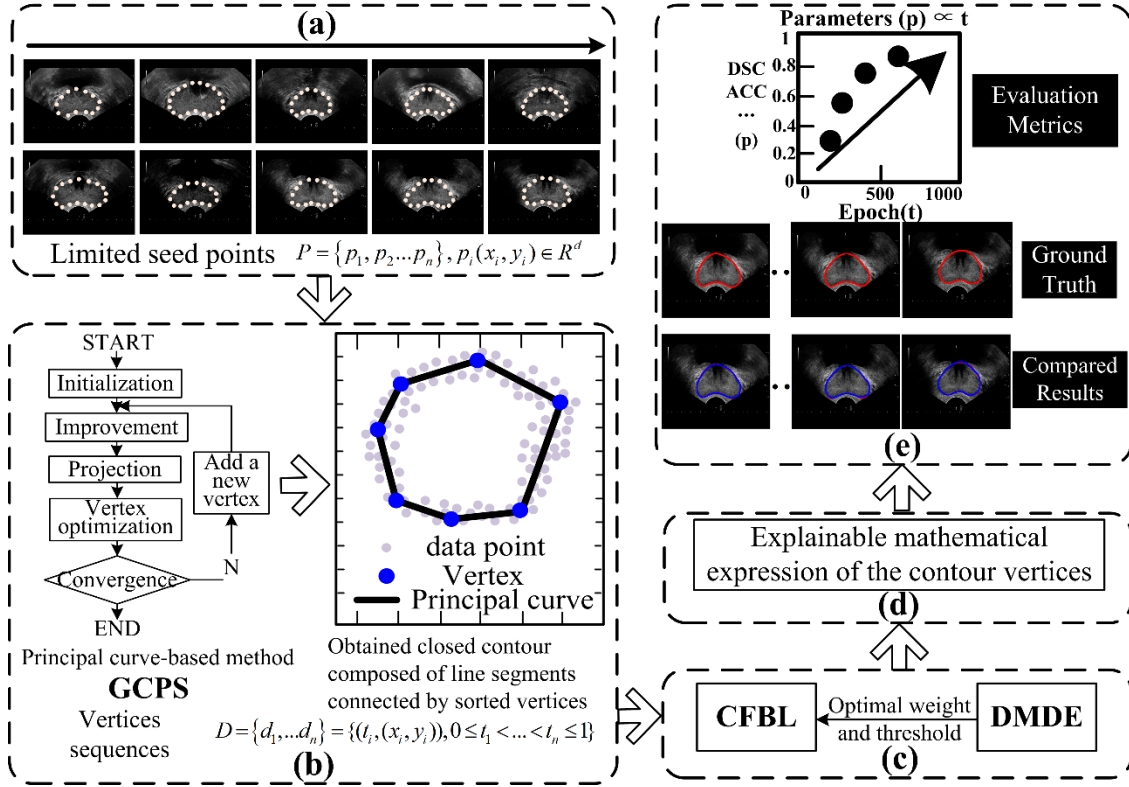
Fig. 2 The A-ProSeg architecture.

2.3 GCPS

Because the traditional K-segments polygonal segment (KPS) model [31] cannot fit the closed dataset, our previous model, termed the CKPS model, was first developed by adding a new initial step and several judging conditions [18] [19] [20]. However, both of these methods improved the model's usefulness, by dealing with scattering, uneven distributions, and abnormal data [32]. In this study, inspired by the work of [32] and Zhang et al. [33], the GCPS was devised to handle the aforementioned issues through several modified steps. Furthermore, several constraint conditions were added. The improvements in the GCPS model mainly comprise three parts: 1) vertex cleaning, 2) the vertex-merge model, and 3) constraint conditions.

2.3.1 Traditional KPS

A traditional KPS [31] model was previously proposed, with the main steps including the initial way (T0), a projection step (T1), a vertex optimization step (T2), converge conditions (T3), and a new vertex addition (T4). The details of these steps have been previously described [31].

2.3.2 Our previous model, termed the CKPS model

Compared with the KPS model, the CKPS model[18] [19] [20] includes several improvements, i.e., a new initial step (T5) and judging conditions (T6).

In the improved initialization step (T5), a closed square was used as the starting curve, replacing the first principal component line [31] used in the traditional KPS model. Furthermore, the judging conditions (T6) described by [18] mainly comprise three aspects:1) the whole loop exits if the segment number $k > c(n, \Delta(f)) = \lambda^* n^{1/3} \Delta(f)^{-1/2} r$; 2) both the inner and outer loops should meet the distance constraint conditions; and 3) there are several additionally added conditions, including optimal selections of vertices, segments, and the contour curve's shape.

2.3.3 Our newly proposed GCPS model

To decrease the possibility of generating a distorted principal curve caused by abnormal vertices, based on the CKPS model, we added several improvements, including vertex cleaning (I0), vertex merging (I1), and several constraint conditions (I2). All of the improvements to the GCPS model (i.e., I0, I1, and I2) were implemented after the vertex optimization step (T2), as shown in Fig. 3.
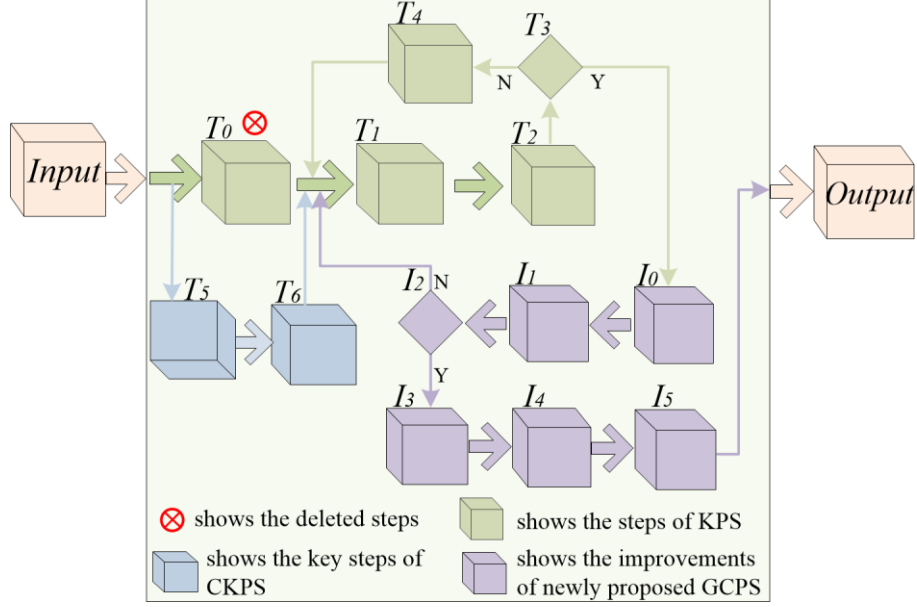

Fig. 3 The workflow of the GCPS model. T0–T4 show the main steps of the traditional KPS model, and T5–T6 show the main steps of the CKPS model. I0–I5 show the key improvements in our newly proposed GCPS model.

1. Vertex cleaning (I0)

We used $flag(v_i)$ to represent the vertex cleaning label, the initial value of which is 1. When the distance of the i-th segment $l_{si} >$ data radius r, $flag(v_i)$ is set to 0, and the vertex v_i is removed. In addition, l_{si} needs to satisfy the rule that $l_{si} = \|v_{i+1} - v_i\| \quad 1 \le i \le m$, in which m represents the number of vertices of the polygonal curve.

The data radius r decides the dataset scale and needs to satisfy $r = \max_{x \in P} \left\| x - \frac{1}{n} \sum_{y \in P} y \right\|$, in which x and y represent the coordinates of the x- and y-axis of p, respectively.

2. Vertex merging (I1)

Owing to the presence of abnormal vertices that may cause distortion of the principal curve f, a vertex merging model was developed to handle this issue, as shown in Algorithm 1.

Algorithm 1 The vertices merging model

Input: $f_0 = (v, s)$, the initial principal curve of vertices to be merged

Output: f_1, reduction of f_0

1:    **for** each vertex $v_i \in f_0$ **do**

2:       Set v_l and v_r be the left and right neighboring vertex of v_i, respectively

4:       Obtain the Euclidean distance $d(v_l, v_i)$ and $d(v_i, v_r)$ based on the work of [34]

5:       Obtain the average distance threshold $d_t = \frac{1}{m} \sum_{i=1}^{m} d(v_{i-1}, v_i)$

6:	**if** $d(v_l, v_i) > d(v_i, v_r) / 3$ or $d(v_l, v_i) < 2 \times d_t$ **then**
7:	Remove v_i, and connect v_l and v_r
8:	**end if**
9:	**end for**
10:	**for** each vertex $v_i \in f_0$ **do**
11:	Obtain the directional angle θ_{vlvi} between v_l and v_i
12:	$\theta_{vlvi} = arctan[(v_i \times y - v_l \times y) / (v_i \times x - v_l \times x)]$
13:	Obtain the directional angle between v_i and v_r
14:	$\theta_{vivr} = arctan[(v_r \times y - v_i \times y) / (v_r \times x - v_i \times x)]$
15:	Obtain $\beta = |\theta_{vlvi} - \theta_{vivr}|$
16:	**if** $\beta > \pi$ **then**
17:	$\beta = 2\pi - \beta$
18:	**else**
19:	Remove v_i, and connect v_l and v_r
20:	**end if**
21:	**end for**

3. Constraint conditions (I2)

Two constraint conditions were proposed to decide whether the freshly inserted vertex v_i should be deleted. If it satisfies either of the following constraint condition, the vertex v_i must be removed: 1) the vertex v_i exceeds the range of the data radius r or 2) very few points p_i projecting to the edge s_i or vertex v_i exist.

2.4 DMDE

When the initial value of the model parameters of neural networks (NNs) is selected stochastically, it easily gets trapped in the local minima of the error surface in the NN's training stage [35]. The DE network is good at global searching and has been used widely. For example, Leema N et al. [36] used DE networks to search the optimal weights of NNs. However, traditional DE networks do not consider improving their performance from certain aspects, such as using a suitable migration technique, memorizing the value of optimal parameters from the last loop, and improving population diversity [37]. To solve these problems, a new method termed the DMDE was designed by combining distributed- and memory-based techniques and multi-mutation operators.

2.4.1 Distributed-based technique

The distributed-based mechanism [22] was adopted to migrate the optimal individual in each parallel subpopulation to the latter subpopulation in the ring topology. However, a new memory-based DE was developed to replace the original DE when processing the slave node [22].

2.4.2 Memory-based technique

A memory-based DE was proposed to save the best mean mutation factor (uF) and mean crossover rate (uCR) from the last loop and then use these two parameters in the next loop. The primary steps of the memory-based DE were as follows.

(i) The initial values of the uF and uCR were set in the range of [0, 1]. In addition, the recent iteration number $G <$ the maximum iteration number, $GMax$, where the initial value of G is 1.

(ii) The initial values of F and CR were randomly selected within the range of [uF, 1] and [uCR, 1], respectively.

(iii) The mutant individual v_a^{G+1} was produced through the mutation step, as shown in Section 2.4.3.

(iv) The experimental individual u_a^{G+1} was obtained in the crossover step, as follows:

$$u_a^{G+1} = \begin{cases} vec_a^{G+1}, & if \ rand[0,1] \le CR \\ x_a^G, & otherwise \end{cases} \tag{3}$$

(v) The next candidate was obtained in the selection step, as follows:

$$x_a^{G+1} = \begin{cases} u_a^{G+1}, & if \ f(u_a^{G+1}) < f(x_a^G) \\ x_a^G, & otherwise \end{cases} \tag{4}$$

(vi) S_{CR} and S_F were the successful crossover and mutation probabilities, respectively, where both the parameters were required to meet Lehmer's function ($mean_L(\bullet)$) [38] as follows:

$$mean_L(S_{CR}) = \frac{\sum_{CR \in S_{CR}} CR^2}{\sum_{CR \in S_{CR}} CR} \tag{5}$$

$$mean_L(S_F) = \frac{\sum_{F \in S_F} F^2}{\sum_{F \in S_F} F} \tag{6}$$

(vii) uF and uCR were modified as follows:

$$uF = (1 - ap) \times uF + ap \times mean_L(S_F) \tag{7}$$

$$uCR = (1 - ap) \times uCR + ap \times mean_L(S_{CR}) \tag{8}$$

where the adjusted parameter ap was randomly chosen within the range of [0, 1].

If $G < Gmax$, the model proceeded to step (ii), in which the values of both the uF and uCR in this loop were used for the next loop, while $G = G + 1$. On the contrary, if $G \ge GMax$, it exited the loop and proceeded to the next step.

(viii) The optimal individual was determined.

2.4.3 Multi-mutation operators

To enhance the diversity of the population, two-mutation operators were utilized to generate the mutant vector individual vec_a^{G+1} as follows:

$$vec_a^{G+1} = \begin{cases} x_{a_1}^G + rand_1 \times (x_{a_2}^G - x_{a_3}^G), & if \ rand[0,1] < p_G \\ x_{a_1}^G + rand_2 \times (x_{a_2}^G - x_{a_3}^G) + rand_3 \times (x_{a_4}^G - x_{a_5}^G), & otherwise \end{cases} \tag{9}$$

where $rand$ denotes a random value within the range of [0, 1]. A_k ($k \in [1, 5]$) is an integer and randomly selected within [1, N_p], where N_p is the number of solutions. P_G is the probability when using the mutation operator, denoted as follows:

$$p_G = p_{min} + \frac{G \times (p_{max} - p_{min})}{GMax} \tag{10}$$

where p_{min} and p_{max} are the minimal and maximal probabilities, respectively, when using the mutation operator.

2.5 CFBL

Because the Caputo-type fractional gradient descent (CFGD) module [39] has good heredity and memory, we used a CFGD-based CFBL in this study [21]. Unlike backpropagation neural networks (BPNNs) [36], the key advantage of CFBL is that the CFGD is used to replace the gradient descent method [40] to optimize the weight vector at the forward propagation stage. In addition, CFBL has input, hidden, and output layers, which can explain the arbitrary continuous function and achieve the arbitrarily expected precision [41].

At the forward propagation stage of the CFBL, the sigmoid activation function $h_1 = 1 / (1+e^{-x})$ was adopted from the input to the hidden layers, whereas the tanh activation function $h_2 = (e^x - e^{-x}) / (e^x + e^{-x})$ was utilized from the hidden to the output layers. Meanwhile, the backpropagation stage was used to update the weight of the CFBL. After training, the coordinates of the optimized vertices were used to denote the prostate contour, which was

expressed by an explainability-guided mathematical model, as shown in Eq. (1) and Eq. (2). For better understanding, the details of CFBL are illustrated in Appendix 1 and 2.

3. Experimental settings

3.1 Materials

The Tsinghua Changgung Hospital (TCH) prostate dataset used in this study [23] contained TRUS images of 266 patients undergoing brachytherapy (945 slices), which was obtained by a Hitachi HI VISION Avius® ultrasound diagnostic equipment and an integrated high-resolution linear transducer whose frequency is within [4MHz, 8MHz]. Each patient had both axial and sagittal view slices. The original data were in Digital Imaging and Communications in Medicine (DICOM) format, and the matrix size of the dataset was 1,024 × 768 pixels. All data images were rescaled to 600 × 450 pixels. The GT contours of the prostate were manually delineated by three professional radiologists, and the final GT was then determined by a majority vote based on the three experts' annotations. If the mark of each radiologist obtains the same number of votes, they will discuss and determine the consensus ground truth. Fig. 4 shows the process of obtaining the GT contour. The evaluation metrics, i.e., DSC, Ω, ACC, and HD, were used, where ACC and HD are the boundary-based metrics, and DSC and Ω are the region-based metrics [19] [42].

Region-based metrics:

$$\Omega = \frac{TP}{FP+TP+FN} \tag{11}$$

$$DSC = \frac{2TP}{2TP+FP+FN} \tag{12}$$

Boundary-based metrics:

$$ACC = \frac{TP+TN}{TP+FN+FP+TN} \tag{13}$$

$$HD(X,Y) = \max(\max_{x \in X} \min_{y \in Y} \|x-y\|_2, \max_{y \in Y} \min_{x \in X} \|x-y\|_2) \tag{14}$$

where TP, FP, FN, and TN indicate true positive, false positive, false negative, and true negative, respectively. *X* and *Y* represent the experimental result and ground truth, respectively.
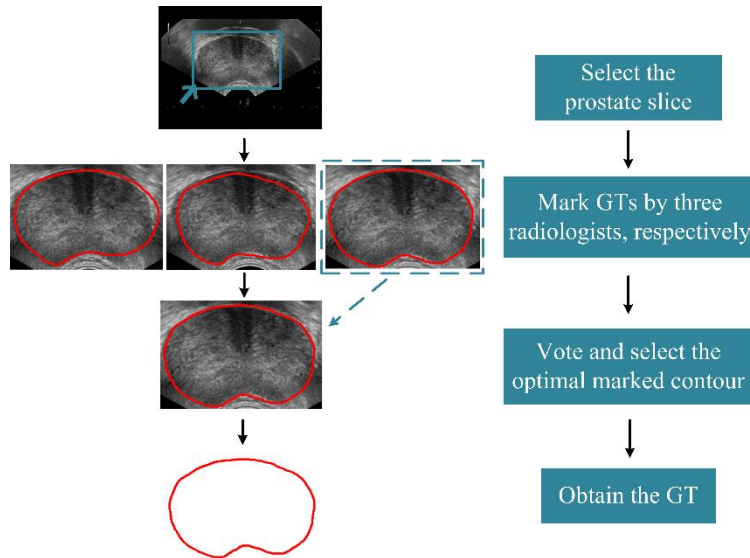

Fig. 4 Extraction of the prostrate ground truth contour.

3.2 Parameter settings

One hundred and ninety patients (675 slices) were randomly chosen for training, with 20 patients (66 slices) for validation and 56 (204 slices) for testing. Each patient had several axial slices, but only one sagittal slice. The neural network-based CFBL model was optimized using the stochastic gradient descent scheme, adopting an initial learning rate, momentum, and maximum epoch of 0.4, 0.9, and 1,000, respectively. Each experiment was conducted on a personal computer with an Intel® Core™ i7-8750H processor and a Geforce GTX 1070 Max-Q graphics processing unit with 8 GB of memory. To assist the readers' understanding, the parameter settings for our initial model are provided in Appendix 3.

4. Results

4.1 Step-by-step visualization

For ease of understanding, Fig. 5 shows the progressively improved qualitative experimental outcomes. The original data and the GT, manually determined by three physicians (first column), were the inputs of our A-ProSeg model. After adopting the GCPS model to achieve a polygon comprising the segments, the approximate shape of the contour could be determined (second column). The advantage of the GCPS model is that it allows the optimization of the location of each vertex based on the rule of minimizing the penalty distance [31], in which the location of each segment is also updated. After executing the DMDE-CFBL, the final result was obtained (third column). The DMDE was used to achieve the best initialization of the parameters (i.e., weight and threshold) of the CFBL model. After CFBL training, the model parameters were used to explain the smooth mathematical expression of the prostate contour to ensure the experimental outcomes were more precise and smoother. To make the model more intuitive, the third row was introduced to denote the zooming visualization of the compared outcomes (second row), in which the red and blue lines show the GT and experimental outcomes, respectively. As shown in the third row (visualization of the ROI), the outcome of our A-ProSeg model was the best among all steps.
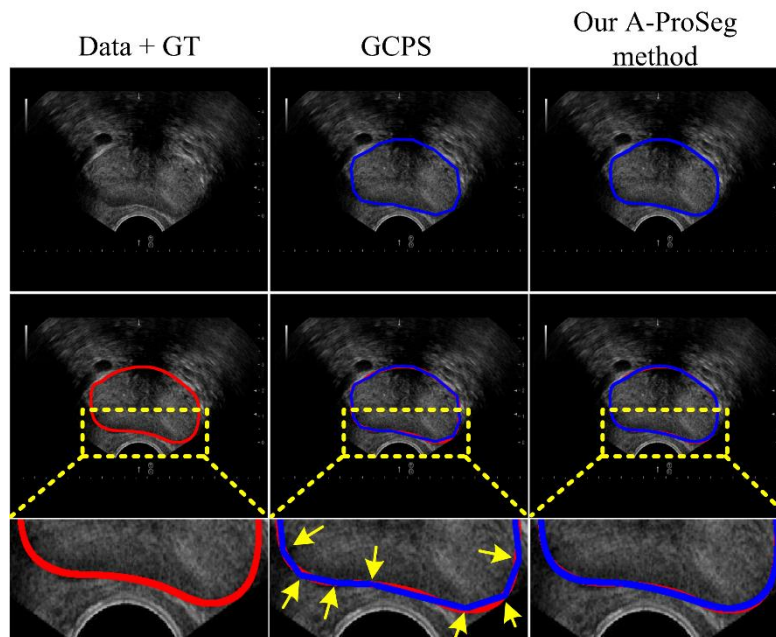

Fig. 5 Qualitative illustration of the experimental results step by step.

4.2 Robustness to noise

Given that different levels of salt and pepper noise may result in different levels of influence on the

segmentation results, many recent studies have chosen the optimal signal-to-noise ratio (SNR) through repeated trials. A patch-based fuzzy local similarity c-means method [43] was used for image segmentation, and the noise densities were set in the range of [0.05, 0.2], corresponding to an SNR range of [0.8, 0.95]. An active contour model [44] has previously been proposed for medical image segmentation, and good performance has been demonstrated with the salt and pepper noise with different noise densities set in the range of [0.05, 0.2], which corresponds to an SNR range of [0.8, 0.95]. A lower SNR value results in poorer image quality [43]. To better validate the robustness of our proposed A-ProSeg model, the SNR was set with lower values, including 0.9, 0.85, and 0.8. In addition, another overlap evaluation criterion [45] was used for evaluation, which is adopted to evaluate the influence of different levels of noise on raw data, calculated as follows:

$$overlap = \frac{\left|redR \cap blueR\right|}{redR} \tag{15}$$

where the overlap rate (overlap) illustrates the scale of overlap between the red (redR) and blue (blueR) areas.

Table 1 presents the robustness of our method tested with variable degrees of salt and pepper noise. The values indicate that when the SNR decreases, the mean values of all the metrics (i.e., DSC, Ω, and ACC) and the overlap also decrease. However, the mean values of all the metrics were greater than 92%, indicating that no matter whether the raw testing set was corrupted by different levels of noise, the performance of our model has satisfactory performance.

Furthermore, two slices were randomly selected for visualization validation, as shown in Fig. 6. The first three rows show the visualization results of one slice, and the last three rows show the results of the other slice. The visualization results of each slice contain three parts, including the original image, the histogram of the initial images, and the segmentation-compared results. According to the histogram row (second and fifth row), the red line shows the number of pixels at the different gray values of the raw images, whereas the blue line shows the number of pixels at the different gray values of the corrupted images. From the histogram row, the level of damage to the images at different SNRs can be seen. The blue line shows the segmentation results of the proposed method, whereas the red line shows the GT.
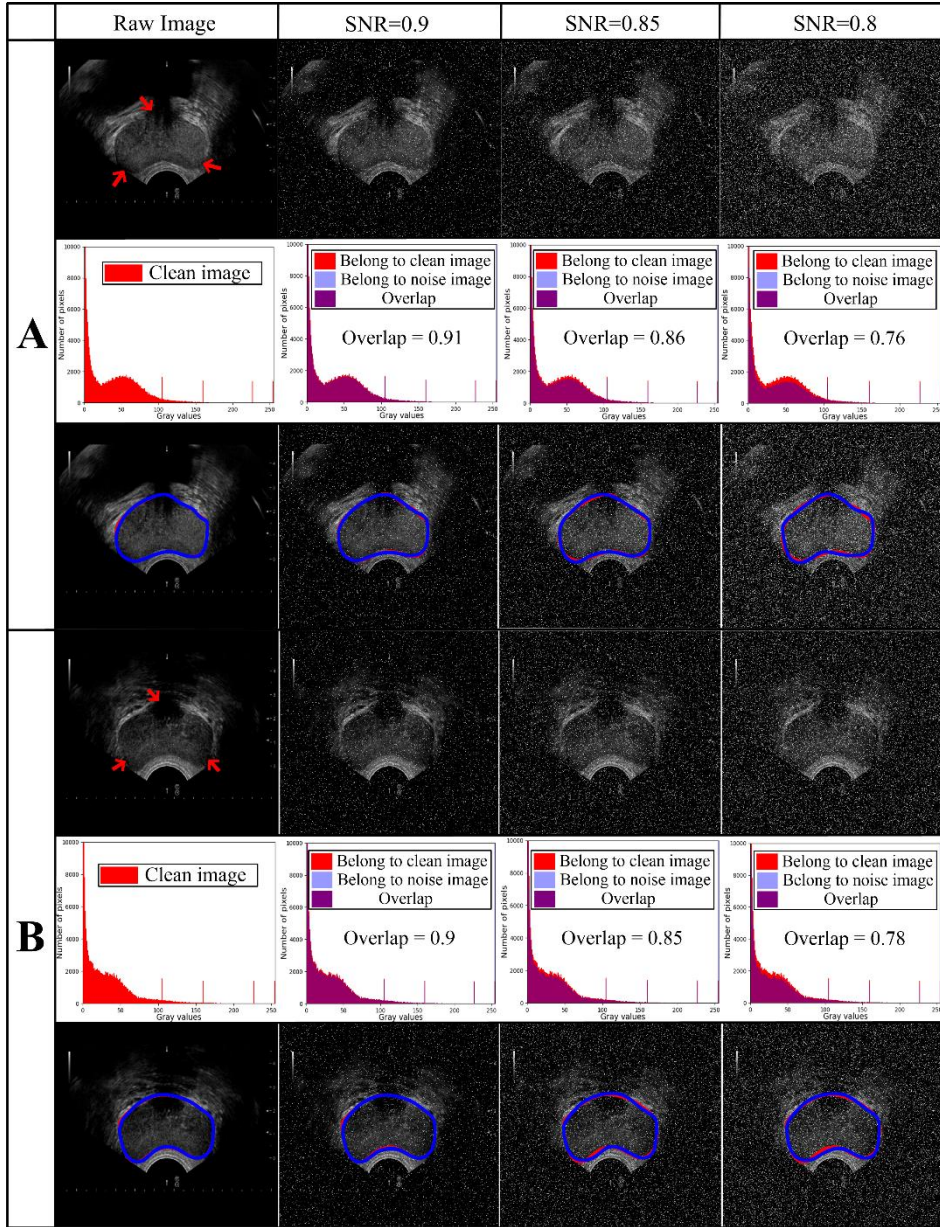
Fig. 6 Visualization results of the proposed method at different levels of salt and pepper noise. The red arrow indicates the missing and blurred boundaries of the prostate. In the second and fifth rows, the *overlap* metric is addopted for evaluating the effect of various degrees of noise on original data, and details are shown below: first, the number of pixels for the gray value within [0, 255] in the ultrasound image is calculated. Secondly, the overlap region is shown, where the red and light blue represent the areas that only belong to clean images and noisy images, respectively, and the purple area represents the pixel overlap area between clean and noisy images. Finally, the *overlap* metric is calculated for evaluation.

Table 1 Performance of the proposed method under varying degrees of salt and pepper noise.

Different degrees of noise	DSC (%)	Ω (%)	ACC (%)	HD (mm)
Original testing set	96.2 ± 2.4	94.4 ± 3.3	95.7 ± 2.7	1.9 ± 0.9
SNR = 0.9	94.9 ± 2.9	93.3 ± 3.5	94.6 ± 3.2	2.7 ± 1.6
SNR = 0.85	94.2 ± 3.8	92.8 ± 4.3	93.9 ± 3.4	2.9 ± 1.7
SNR = 0.8	93.5 ± 4	92.4 ± 4.7	93.3 ± 3.7	3.3 ± 1.9

4.3 Ablation study (AS)

In this section, an AS is discussed to demonstrate the impact of each component of our A-ProSeg model. Three evaluation metrics, i.e., DSC, Ω, and ACC, were used. All of the methods used the same number of slices for data training, validation, and testing. The results of the AS are shown in Table 2. We used the CKPS-BPNN as the baseline model, where it inherits the characteristic of the principal curve to fit the data automatically, while using a neural network for training to decrease the model error (AS1). To improve the performance of AS1, and solve the issue that the neural network is always trapped in local minimum during training, the memory-based DE model is adopted (AS2). Further enhancing the capability of AS2, the distributed-based scheme is added to well hunt for the optimal individual (AS3). Based on the AS3 model, the global-based closed polygonal segment (GCPS) is developed, whose improvements are the vertex cleaning scheme, the vertex-merge model, and constraint conditions (AS4). Different from AS4, our A-ProSeg model uses the CFBL model to improve the heredity and memory performance of the neural network model.

The baseline model (CKPS-BPNN) had the lowest DSC, Ω, and ACC values of 90.3%, 89%, and 90%, respectively. After making some improvements (i.e., MDE, DMDE, GCPS, and CFBL) on the baseline model (CKPS-BPNN), the DSC, Ω, and ACC values increased by approximately 3.43%–6.53%, 3.37%–6.16%, and 3.44%–6.33%, respectively. Overall, our A-ProSeg model (GCPS + DMDE + CFBL) achieved the best results. Three patients were randomly chosen for qualitative demonstration, as shown in Fig. 7. Table 3 indicates the *P*-values from the paired Student's *t*-test on the DSCs between our method (AS5) and other different ASs.

Table 2 The quantitative performance of the model demonstrated using an ablation study.

AS	Models	DSC (%)	Ω (%)	ACC (%)	HD (mm)
AS1	Baseline	90.3 ± 4.5	89 ± 5.7	90 ± 4.6	4.3 ± 2.2
AS2	CKPS-MDE-BPNN	93.4 ± 3.6	92 ± 4.4	93.1 ± 3.9	3.3 ± 1.8
AS3	CKPS-DMDE-BPNN	93.8 ± 3.6	92.4 ± 4.1	93.4 ± 3.5	3.1 ± 1.9
AS4	GCPS-DMDE-BPNN	95.8 ± 2.5	94 ± 3.7	95.3 ± 3.4	1.9 ± 1.0
AS5	Our A-ProSeg	96.2 ± 2.4	94.4 ± 3.3	95.7 ± 2.7	1.9 ± 0.9

CKPS: closed K-polygonal segment model   GCPS: global closed polygonal segment model

MDE: memory-based differential evolution model   DMDE: distributed-based memory differential evolution

BPNN: backpropagation neural network   CFBL: Caputo fractional-order backpropagation learning model

Table 3 *P*-values from paired Student's *t*-tests comparing the dynamic susceptibility contrast values between our method (AS5) and the other ASs.

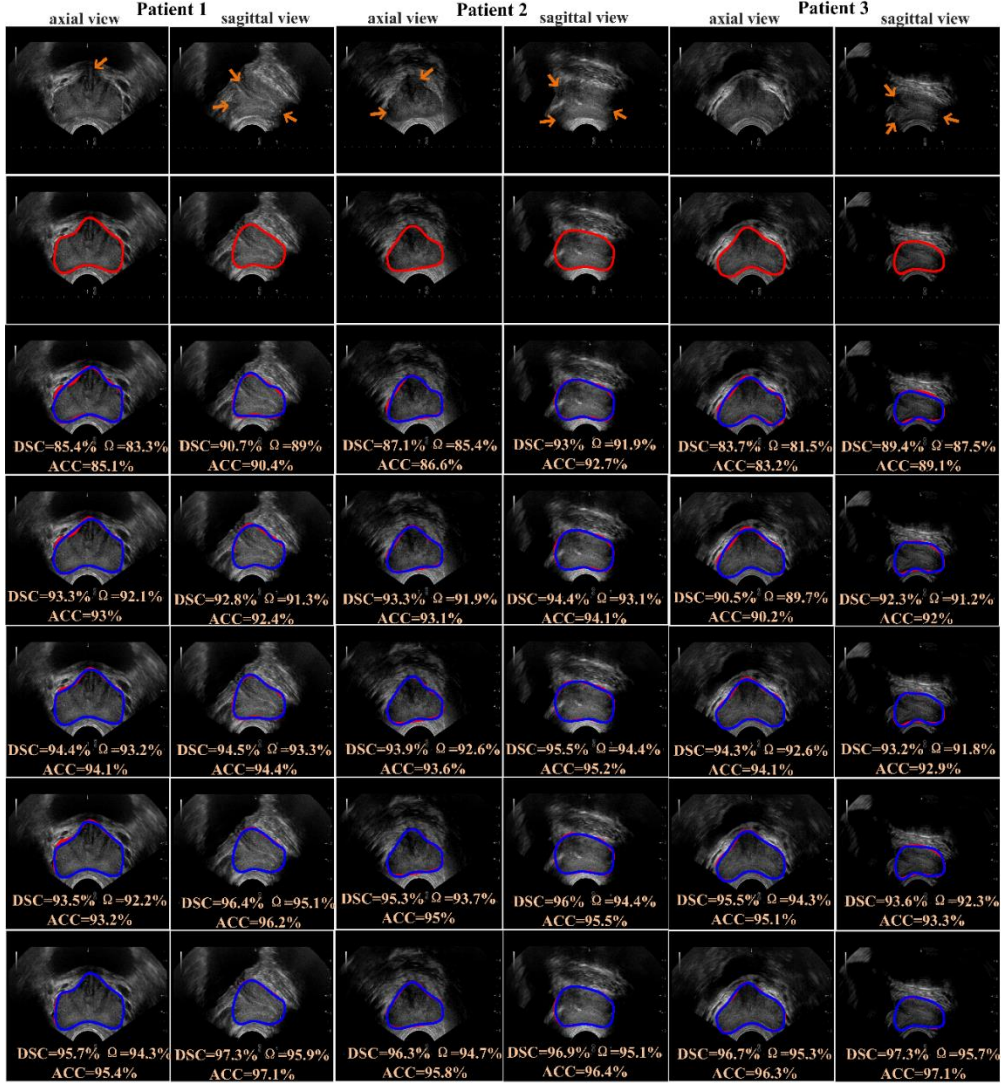|	AS1	AS2	AS3	AS4
p-value to our method (AS5)	<0.001	<0.001	<0.01	<0.1

Fig. 7 Qualitative demonstration of ablation study of three patient cases. The first row represents the original data. The second row represents the GT manually contoured by professional radiologists. Rows 3 to 7 represent the results of five comparison techniques (AS1, AS2, AS3, AS4, and AS5). The red and blue curves show the GT and experimental results, respectively. The orange arrow indicates the missing or blurred outlines of the prostate.

4.4 Comparison with state-of-the-art (SOTA) methods

Our hybrid A-ProSeg model was compared with SOTA methods, including the hybrid prostate segmentation method (H-SegMod) [25], U-Net [27], U-Net++ [28], and the U-shape hybrid transformer network (UTNet) [26]. H-SegMod is the method developed in our previous study. It is a hybrid, semi-automatic method that uses limited prior points to assist the model in locating the prostate contour cue in TRUS images. UTNet is a fully automatic, self-attention-based convolutional NN and is widely used in the medical image segmentation field. It integrates the features of the self-attention module to capture long-range associative features. Moreover, a Jiangsu Province Hospital of Chinese Medicine (JPHCM) dataset [5] was included for external evaluation to assess the generalizability of our A-ProSeg model. The JPHCM dataset contained 55 brachytherapy patients (total of 393 slices) gathered by the Nanjing University of Chinese Medicine, Nanjing, China. The ultrasound slices were acquired via a Vinno 70 Lab digital ultrasound diagnostic system (Vinno, Suzhou, China) and an integrated ultrasound probe whose frequency is within the range of 4-8 MHz.

Owing to the influence of the limited training data on deep learning methods, image rotation was used to augment the number of training data sizes, where each raw image was rotated within [−15°, 15°] until a total of 1,500 training images were achieved. Nevertheless, both H-SegMod and our method used only original data without data augmentation.

4.4.1 Internal evaluation on the TCH dataset

1. Deep learning methods without using prior points

The segmentation performance of all of the methods is shown in Table 4. Table 5 also presents the *P*-values from the paired Student's *t*-test on the DSCs between our method and the comparison methods. In the hybrid models (i.e., H-SegMod and our A-ProSeg model), the DSC, Ω, and ACC values increased by approximately 4.27%–5.48%, 5.04%–5.82%, and 4.64%–5.86%, respectively, compared with the corresponding values for the deep-learning-based model (UTNet). Overall, our A-ProSeg model achieved the best performance, with DSC, Ω, and ACC of $96.2 \pm 2.4\%$, $94.4 \pm 3.3\%$, and $95.7 \pm 2.7\%$, respectively.

Table 4 Quantitative comparison of the hybrid models with existing state-of-the-art models on the TCH dataset.

Publication	Model	DSC (%)	Ω (%)	ACC (%)	HD (mm)	End-to-end time (h)	Testing time (s)
U-Net	Deep learning	90.4 ± 7.2	88.6 ± 7.9	90.4 ± 8.4	4.2 ± 1.8	~6	3
U-Net++	Deep learning	91.4 ± 7.7	89.5 ± 8.4	90.6 ± 7.9	4.6 ± 1.9	~22	5
UTNet	Deep learning	91.2 ± 8.1	89.2 ± 9.6	90.4 ± 8.4	4.1 ± 2.2	~20	4
H-SegMod	Hybrid	95.1 ± 2.9	93.7 ± 4.4	94.6 ± 3.5	2.1 ± 1.1	~2.1	3
Our A-ProSeg	Hybrid	96.2 ± 2.4	94.4 ± 3.3	95.7 ± 2.7	1.9 ± 0.9	~2	3

Table 5 *P*-values from paired Student's *t*-tests comparing the dynamic susceptibility contrast values between our A-ProSeg model and the other state-of-the-art models on the TCH dataset.

	H-SegMod	UTNet	U-Net++	U-Net
p-value to our A-ProSeg	<0.001	<0.1	<0.1	<0.1

2. Deep learning methods using 8% of points as a prior

Table 6 represents the capability of all the models on both accuracy and efficiency views, and we also use the *p*-value from the paired student's *t*-test on the DSC values between our model and other state-of-the-art models for evaluation, shown in Table 7. From Table 6, the testing time of all the methods is close. However, compared with deep learning models, the hybrid methods perform better. The potential reasons for this phenomenon are that 1) the hybrid methods inherit the characteristic of a principal curve to fit the dataset automatically so that to achieve higher accuracy, and 2) the hybrid methods adopt the multilayer perceptron neural network for training, whose architecture is more straightforward than deep learning neural network and needs less time for training. Overall, our model obtained the most satisfactory capability, with DSC, Ω, ACC, and HD of $96.2 \pm 2.4\%$, $94.4 \pm 3.3\%$, $95.7 \pm 2.7\%$, and 1.9 ± 0.9 (mm), respectively.

Table 6 Quantitative comparison of the hybrid models with existing state-of-the-art models on the Tsinghua Changgung Hospital dataset.

Publication	Model	DSC (%)	Ω (%)	ACC (%)	HD (mm)	End-to-end time (h)	Testing time (s)
U-Net	Deep learning	94.1 ± 3.4	93.2 ± 4.7	93.8 ± 4.1	2.9 ± 1.4	~3	3
U-Net++	Deep learning	94.7 ± 3.7	93 ± 4.1	94.2 ± 3.8	2.6 ± 1.2	~8	4
UTNet	Deep learning	94.4 ± 3.8	93.3 ± 3.9	94.1 ± 4.2	2.7 ± 1.2	~7	3
H-SegMod	Hybrid	95.1 ± 2.9	93.7 ± 4.4	94.6 ± 3.5	2.4 ± 1.1	~2.1	3
Our A-ProSeg	Hybrid	96.2 ± 2.4	94.4 ± 3.3	95.7 ± 2.7	1.9 ± 0.9	~2	3

Table 7 *P*-values from paired Student's *t*-tests comparing the dynamic susceptibility contrast values between our A-ProSeg model and the other state-of-the-art models on the TCH dataset.

	H-SegMod	UTNet	U-Net++	U-Net
p-value to our A-ProSeg	<0.001	<0.01	<0.01	<0.01

4.4.2 External evaluation on the JPHCM dataset

1. Deep learning methods without using prior points

The JPHCM dataset contains 393 slices from 95 patients. All of the slices were resized to 600 × 450 pixels and used for testing the capability of all the methods, including our A-ProSeg model. Table 8 shows the capability of all of the SOTA models on the JPHCM dataset. The DSC, Ω, and ACC values of our A-ProSeg model increased by 1.59%–5.99%, 0.64%–7.55%, and 1.38%–6.48%, respectively, compared with the corresponding values for the other SOTA models (Table 8). Compared with our A-ProSeg model, the *P*-values for comparisons with the H-SegMod and UTNet models when using the DSC results were < 0.01 and < 0.1, respectively. More details on the statistical analysis are provided in Table 9.

Table 8 Quantitative comparison of the hybrid models with the existing state-of-the-art models on the JPHCM dataset.

Publication	Model	DSC (%)	Ω (%)	ACC (%)	HD (mm)	Testing time (s)
U-Net	Deep learning	89.2 ± 9.2	86.1 ± 10.4	88.6 ± 9.4	4.5 ± 2.2	3
U-Net++	Deep learning	90.7 ± 9.2	87.9 ± 9.8	89.6 ± 8.6	4.3 ± 2.1	4
UTNet	Deep learning	90.4 ± 8.9	87.4 ± 9.7	89.4 ± 9	4.5 ± 2.2	4
H-SegMod	Hybrid	94.3 ± 3.1	93.4 ± 4.5	93.9 ± 3.8	3.1 ± 1.6	3
Our A-ProSeg	Hybrid	95.8 ± 2.5	94 ± 3.5	95.2 ± 2.9	2.1 ± 1.2	3

Table 9 *P*-values from paired Student's *t*-tests comparing dynamic susceptibility contrast values between different methods on the JPHCM dataset.

	H-SegMod	UTNet	U-Net++	U-Net
p-value to our A-ProSeg	<0.01	<0.1	<0.1	<0.1

2. Deep learning methods using 8% of points as a prior

   To better compare the optimal performance of all the methods on the same conditions, we used the same number of prior points to assist deep learning models for prediction. The quantitative evaluation of all the methods is shown in Table 10, and Table 11 represents the statistical analysis. Due to the guidance of prior points, all the methods achieve satisfactory performance, where the maximum deviations of DSC, Ω, and ACC between our A-ProSeg method and deep learning model (U-Net) are only 2.2%, 1.5%, and 2.1%, respectively.

Table 10 Quantitative comparison of the hybrid models with existing state-of-the-art models on the JPHCM dataset.

Publication	Model	DSC (%)	Ω (%)	ACC (%)	HD (mm)	Testing time (s)
U-Net	Deep learning	93.7 $\pm$ 3.7	92.6 $\pm$ 4.9	93.2 $\pm$ 4.4	3.1 $\pm$ 1.5	3
U-Net++	Deep learning	94.5 $\pm$ 3.9	92.6 $\pm$ 4.5	94.1 $\pm$ 3.9	2.7 $\pm$ 1.4	3
UTNet	Deep learning	94.2 $\pm$ 4.3	93.2 $\pm$ 5.3	93.7 $\pm$ 4.9	2.9 $\pm$ 1.4	3
H-SegMod	Hybrid	94.3 $\pm$ 3.1	93.4 $\pm$ 4.5	93.9 $\pm$ 3.8	3.1 $\pm$ 1.6	3
Our A-ProSeg	Hybrid	95.8 $\pm$ 2.5	94 $\pm$ 3.5	95.2 $\pm$ 2.9	2.1 $\pm$ 1.2	3

Table 11 *P*-values from paired Student's *t*-tests comparing dynamic susceptibility contrast values between different methods on the JPHCM dataset.

	H-SegMod	UTNet	U-Net++	U-Net
p-value to our A-ProSeg	<0.001	<0.01	<0.01	<0.01

5. Discussion

   Accurate prostate segmentation in TRUS is difficult because the prostate outline is often ambiguous or may disappear. In this study, we developed a novel A-ProSeg segmentation method that uses AI-based and polygonal segment-based methods. Compared with SOTA segmentation methods, our method makes three key contributions. First, semi-automatic and fully automatic methods are currently used for ultrasound segmentation tasks. The DSCs of most existing fully automatic models [16] [17] are approximately 0.9. Using limited radiologist-defined points as a prior, our study achieved satisfactory outcomes, and the DSC of our study increased by approximately 6.88% compared with the fully automatic models [16] [17]. Second, compared with deep-learning-based segmentation methods that use more than 4,000 images for training with a DSC of 0.92 [46], our approach used 675 training images with a DSC of 0.96. The main reason for this performance is that our method uses prior information while inheriting the advantage of the principal curve to be close to the dataset's center automatically. Third, unlike traditional principal curve techniques, our method contains a smooth mathematical mapping function, achieved by CFBL, to yield a smooth prostate boundary. Our experimental outcomes demonstrate that our approach achieved satisfactory capability. The remainder of this section comprehensively summarizes our findings.

   Discussion of the A-ProSeg model's innovations: Our main findings are that our proposed model (a) had clear

step-by-step visualization (Section 4.1); (b) produced robust results across noise levels (Section 4.2); (c) was more accurate than the current SOTA prostate segmentation methods on internal and external evaluations, and using with and without prior points (Section 4.4); (d) was able to handle most of the image artifacts effectively, such as missing and blurred boundaries, based on principal curve-based and machine-learning-based methods (Overall); and (e) was able to yield a smooth contour through an explainability-guided mathematical model (Overall). Moreover, we included an ablation study to show the importance of each component of the proposed method (Section 4.3).

Influence of different noise levels: We used a histogram to indicate the relationship between the gray values (x-axis) and the number of pixels (y-axis) (Fig. 6 in Section 4.2). As per our findings, grayscale-based ultrasound images are mostly composed of black pixels, and the gray value equals zero [47]. Hence, in this study, a limited number of pixels (y-axis), in the range of [0, 10,000], was selected for display. After the damage caused by different levels of noise (SNR decreases), the *overlap* decreases, illustrating that the image was damaged by higher levels of noise. In addition, based on the quantitative performance of our A-ProSeg model, all of the evaluation metrics (i.e., DSC, Ω, and ACC) were maintained at a level greater than 92%, indicating that our model has good robustness, even when dealing with severely damaged images (Table 1 and Fig. 6). Fig. 8, which shows the relevant zoomed-in images from Fig. 6, shows that the degree of image damage increased with decreasing SNRs.
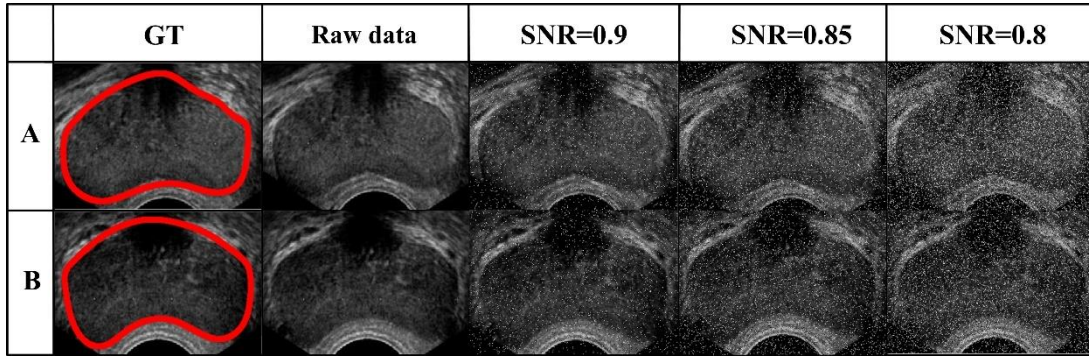

Fig. 8 Zoomed-in images of Fig. 6 corrupted by various degrees of salt and pepper noise.

Difficulty in acquiring data: As shown in Fig. 7 in Section 4.3, only two image views (axial and sagittal) were used for assessments. Several axial-view and only one sagittal view slices were used per patient. The TRUS slice acquisition procedure is shown below. After manually inserting an ultrasound probe into the patient's rectum, the physician adjusted the probe to achieve a regular cross-sectional prostate slice. Nevertheless, it is nearly impossible for a physician to rotate the probe in the rectum to acquire a coronal-view slice owing to the patient's pain tolerance. In addition, owing to limitations in the probe's rotation angle, acquiring several sagittal view slices is challenging. Thus, we acquired only one sagittal view slice (i.e., the slice with the maximum cross-section that was the clearest) per patient.

Influence of each improvement of our A-ProSeg model: As indicated by the results for models AS1 to AS5 (our model), presented in Table 2 in Section 4.3, the performance of the model increased sequentially. Based on the AS1 model, both AS2 and AS3 introduced a DE-based scheme to determine the optimal initialization of NNs to resist the NN getting trapped in the local minima during training. Upon combination with the GCPS scheme, model AS4 exhibited better performance than AS3 owing to the ability of GCPS to handle abnormal vertices. Compared with model AS4, our A-ProSeg model adopted the CFBL scheme with good heredity and memory and exhibited the most satisfactory performance.

Comparison of our method using different numbers and locations of prior points: This work used 8% of points of contours manually delineated by radiologists (closer to the prostate contour) as the prior points. To better assess the impact of the different numbers and locations of prior points, the seed points closer to the shadowed

areas were re-collected, where DSC and testing time were adopted as the evaluation metrics. The evaluation performance of our method is shown in Table 12. From Table 12, when using the prior points around the prostate outline, the performance of our approach was better than using the prior points closer to shadowed areas. Our method has the best accuracy when using <50% of prior points, but the prior information used is too strong. When using fewer prior points with less delineation time, our method needs more time for training (<1% of points situation). When using more seed points, our approach becomes stable faster, while using too much time for delineation (<50% of points situation). All in all, our method achieves the best performance using <8% of points as prior.

Table 12 Performance of our method using different numbers and locations of prior points

Method	Metrics	<1% of points	<5% of points	<8% of points	<10% of points	<15% of points	<30% of points	<50% of points
Closer to shadowed areas	DSC (%)	81.2 ± 8.4	84.7 ± 6.3	86.9 ± 4.4	87.5 ± 3.5	90.6 ± 2.7	94.4 ± 1.4	97.1 ± 0.9
	End-to-end time (h)	2.1	2	2	2.2	2.3	2.3	2.4
	Testing time (s)	5	4	4	4	3	2	2
Closer to the prostate contour	DSC (%)	92.6 ± 4.5	94.1 ± 3.4	**96.2 ± 2.4**	96.4 ± 2.4	96.9 ± 1.7	97.3 ± 1.2	98.5 ± 0.7
	End-to-end time (h)	1.9	1.9	**2**	2.3	2.3	2.4	2.4
	Testing time (s)	4	3	**3**	3	3	2	2

Comparison between our method and delineation of professional radiologists: In this work, the GT contours of the prostate manually delineated by each professional radiologist were used for comparison. From Table 13, we can find that compared with GTs, GT3 spends the least end-to-end time for delineation, while the DSC is the lowest. In addition, GT1 achieves the best DSC while spending the longest end-to-end time. Overall, our method achieves the highest DSC, and the total spending time is reasonable.

Table 13 Evaluation between our method and delineation of professional radiologists.
GTs 1, 2, 3 were manually delineated by three professional radiologists, respectively.

Method	DSC	End-to-end time (h)	Testing time (s)
GT1	93.3 ± 4.1	2.2	5
GT2	92.6 ± 4.5	1.9	4
GT3	91.7 ± 5.1	1.5	3
Our method	**96.2 ± 2.4**	**2**	**3**

Comparison with the SOTA models on two datasets: Tables 4–11 in Section 4.4 show comparisons of two types of models: deep-learning-based and hybrid-based models. Compared with the deep-learning-based models, the hybrid models produce more accurate segmentation results with fewer training slices, indicating that the hybrid models comprising the principal curve-based and neural network-based techniques achieve improved performance for data fitting. Overall, our A-ProSeg model achieved the best performance.

Comparison between inter-institutional datasets: According to the data presented in Tables 4–11 in Section 4.4, all methods performed better on the JPHCM dataset than the TCH dataset. The main reasons underlying this

observation can be as follows. First, the datasets were acquired using different equipment and detection parameters, and the quality of the two datasets was different. The prostate was more challenging to segment in the TCH dataset than the JPHCM dataset, owing to the high gain of the images, which made the prostate boundary blurry. Second, the total JPHCM dataset was used for prediction, which is more challenging for all of the models. In addition, one randomly selected slice in the JPHCM dataset was used for qualitative external evaluation, as shown in Fig. 9. Overall, our A-ProSeg model had the most satisfactory performance of the models tested.
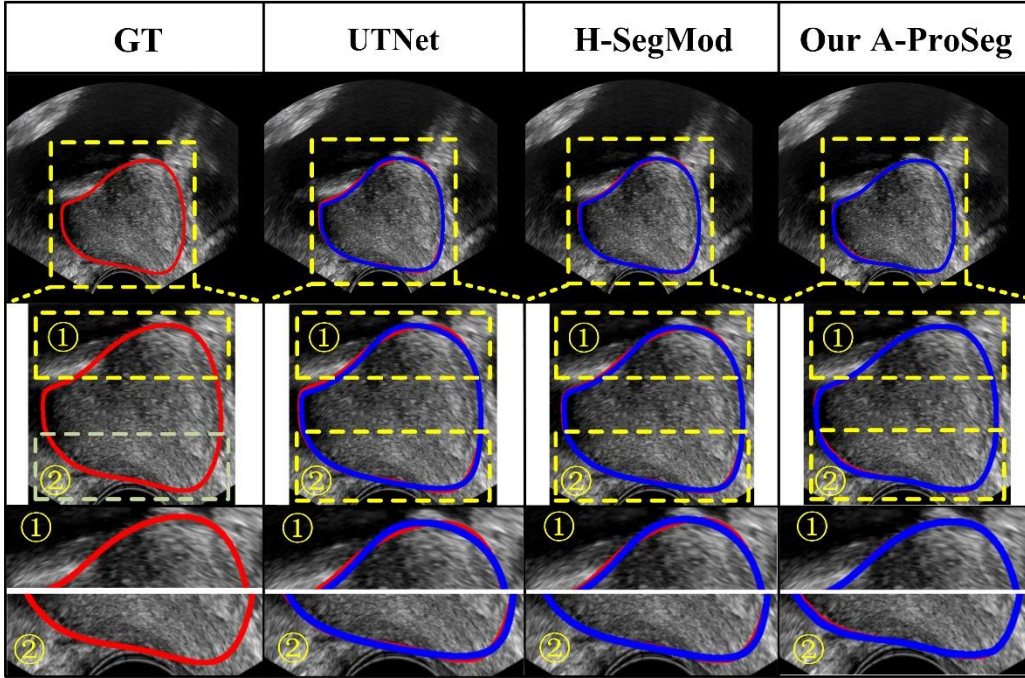

Fig. 9 Qualitative external evaluation on the JPHCM dataset.

Evaluation of our method on different subregions using 8% of points as a prior: To assess the influence of data on different subregions (i.e., apex, mid-gland, and base subregions), we re-collected data from 25 JPHCM patients for testing, containing 100 apex slices and 100 base slices. To achieve optimal performance, we used 8% prior points for all the methods. The quantitative experimental outcomes of our method are indicated in Table 14, and Fig. 10 shows the qualitative effects. Table 14 shows that our method achieves the best performance using various metrics (i.e., DSC, Ω, ACC, and HD) on various subregions (i.e., apex, mid-gland, and base areas).

Table 14 Quantitative evaluation of state-of-the-art models using various subregions.

Method	Apex slice		Mid-gland slice		Base slice	
U-Net	DSC (%)	89.9 ± 8.4	DSC (%)	89.2 ± 9.2	DSC (%)	88.1 ± 9.5
	Ω (%)	87.3 ± 9.8	Ω (%)	86.1 ± 10.4	Ω (%)	84.4 ± 10.5
	ACC (%)	89.6 ± 8.6	ACC (%)	88.6 ± 9.4	ACC (%)	86.9 ± 9.9
	HD (mm)	4.5 ± 2.2	HD (mm)	4.5 ± 2.2	HD (mm)	4.9 ± 2.4
U-Net++	DSC (%)	91.5 ± 8.2	DSC (%)	90.7 ± 9.2	DSC (%)	90.2 ± 8.4
	Ω (%)	89.2 ± 8.7	Ω (%)	87.9 ± 9.8	Ω (%)	86.7 ± 9.3
	ACC (%)	91.1 ± 7.6	ACC (%)	89.6 ± 8.6	ACC (%)	89 ± 8.1
	HD (mm)	4.2 ± 2	HD (mm)	4.3 ± 2.1	HD (mm)	4.5 ± 2.1
UTNet	DSC (%)	91.1 ± 7.5	DSC (%)	90.4 ± 8.9	DSC (%)	89.3 ± 9.1
	Ω (%)	88.7 ± 9.1	Ω (%)	87.4 ± 9.7	Ω (%)	87.9 ± 9.4
	ACC (%)	90.4 ± 8.8	ACC (%)	89.4 ± 9	ACC (%)	88.6 ± 8.8
	HD (mm)	4.5 ± 2.2	HD (mm)	4.5 ± 2.2	HD (mm)	4.7 ± 2.2

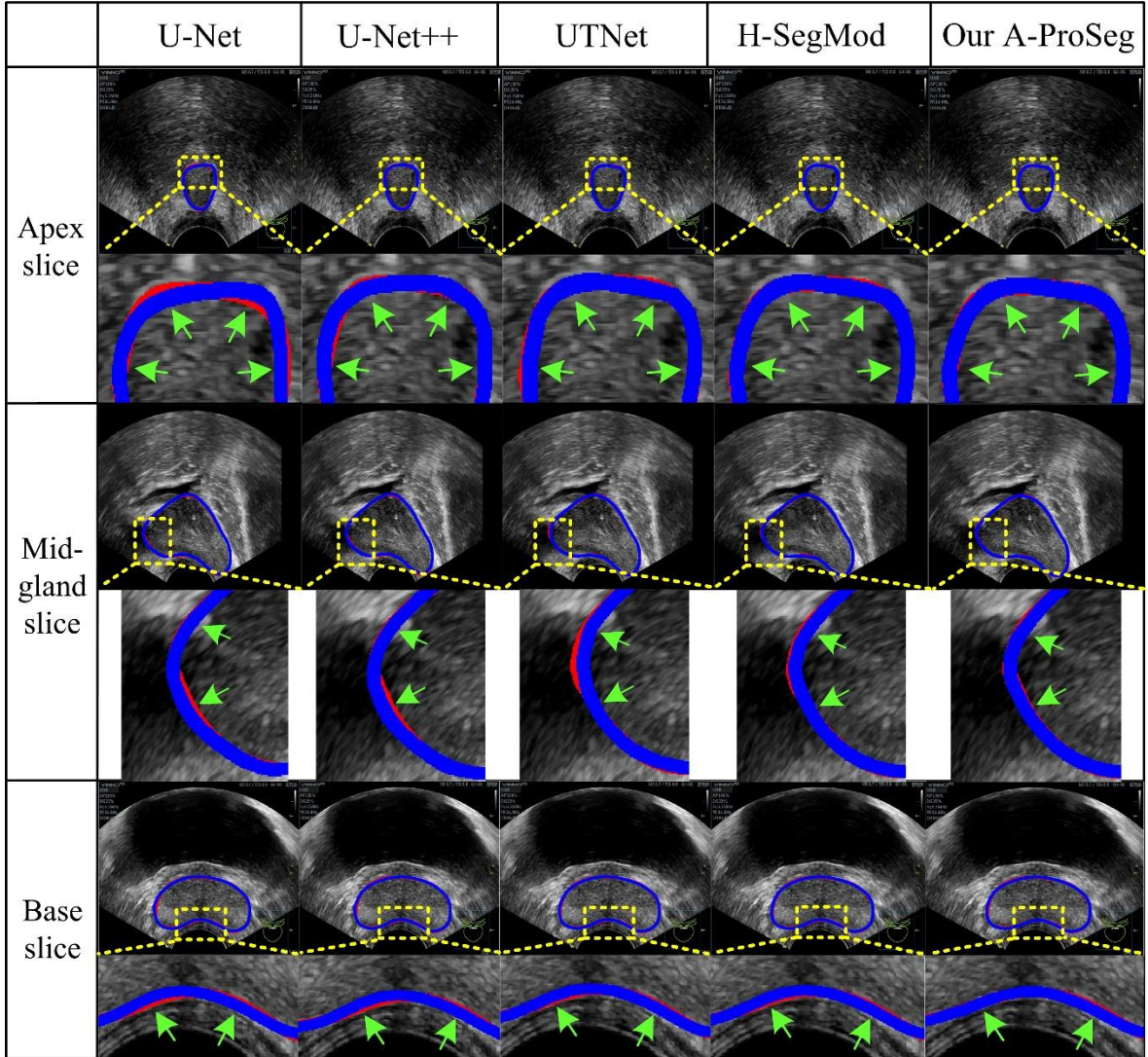|						
H-SegMod	DSC (%)	94.8 ± 2.8	DSC (%)	94.3 ± 3.1	DSC (%)	93.8 ± 3.3
	Ω (%)	93.9 ± 4.3	Ω (%)	93.4 ± 4.5	Ω (%)	93.1 ± 4.6
	ACC (%)	94.3 ± 3.4	ACC (%)	93.9 ± 3.8	ACC (%)	93.5 ± 3.9
	HD (mm)	2.8 ± 1.5	HD (mm)	3.1 ± 1.6	HD (mm)	3.5 ± 1.6
Our method	DSC (%)	96.1 ± 2.3	DSC (%)	95.8 ± 2.5	DSC (%)	95.4 ± 2.6
	Ω (%)	94.1 ± 3.4	Ω (%)	94 ± 3.5	Ω (%)	93.6 ± 3.7
	ACC (%)	95.6 ± 2.7	ACC (%)	95.2 ± 2.9	ACC (%)	94.8 ± 3.2
	HD (mm)	2 ± 1.1	HD (mm)	2.1 ± 1.2	HD (mm)	2.3 ± 1.4

Fig. 10 Qualitative visualization of the randomly selected slices at different subregions, including mid-gland, apex, and base regions. Every two rows represent the slice and its corresponding zooming display of the region of interest. The blue and red curves represent the experimental results and ground truth, respectively. In addition, we use the green arrow, pointing to the areas with obvious contrast in experimental results.

Improvements in our A-ProSeg model, compared with the previous model, H-SegMod: According to the data presented in Tables 4–11 in Section 4.4, our method exhibited better performance than our previous model, termed H-SegMod [25]. To achieve this goal, several improvements were included.

1. Based on the H-SegMod model, our method further used a vertex-merge model to handle abnormal

vertices to avoid generating a distorted principal curve *f*.

2. Unlike the H-SegMod model, our method utilized a memory-based evolution network to reuse the optimal parameters in every loop.

3. We used the CFGD module to replace the gradient descent method [40] used in our previous H-SegMod model to ensure good heredity and memory of the NN.

Future work: Our proposed method achieved some promising results; however, the performance of our method could be further enhanced in the future. *First*, the use of our proposed method can be further investigated under different modalities (i.e., CT and MRI) or for different organs (i.e., kidney, bladder, and gallbladder). *Second*, model compression is very important to our cascaded framework, especially for resource-constrained systems, and is also necessary for real-time disease diagnosis. *Third*, we wish to transfer the semi-automatic method to a fully automatic method, which would have more significant clinical applications. *Fourth*, corrupted testing data, in which the SNRs were selected at 0.9, 0.85, and 0.8, were used to evaluate the robustness of our A-ProSeg model. In the future, we wish to test the limits of the robustness of our model using lower SNRs, and add a constraint condition to use evaluation metrics (i.e., DSC, Ω, and ACC) greater than 0.9. *Fifth*, owing to the lack of several primary properties, including model robustness, safety, transparency, accountability, non-discrimination and fairness, data privacy, and governance, both traditional and explainable artificial intelligence (AI) cannot be defined as trustworthy AI [48] [49]. With in-depth interactions between humans and machines, trustworthy AI methods will become a research hotspot. In the future, we aim to develop novel techniques to complete the transformation of our A-ProSeg model.

6. Conclusions

We developed a novel A-ProSeg method for segmentation in TRUS prostate images. Our method contains several innovations, including (1) an enhanced principal curve-based method, (2) an enhanced evolution NN, and (3) an explainability-guided mathematical function of the prostate ROI contour. To demonstrate the accuracy and robustness of our method, quantitative and qualitative evaluations were conducted by comparing our method with SOTA segmentation models. Owing to its accurate prostate segmentation, our method holds great promise for application in TRUS-based brachytherapy for prostate cancer.

Acknowledgement

1. Ethical Statement

This study involves a retrospective use of patients' standard of care images, where the clinicians have obtained patients' agreement before the transrectal prostate ultrasound examination, which is an item covered by the medical insurance program.

Appendix

1. Traditional BPNN

Because a three-layer network can express arbitrary nonlinear functions and achieve arbitrarily expected precision [50], we used a three-layer architecture involving the input, hidden, and output layers comprised of the neurons $\{I_1,...,I_l\}$, $\{H_1,..., H_h\}$, and $\{O_1,..., O_k\}$, respectively. The forward and backward propagation stages were the primary stages of the BPNN; the forward propagation stage was used to calculate the outcome of each layer; and the backward propagation stage was used to update the model parameters (i.e., weight).

During the forward propagation stage, the total error E can be calculated as follows:

$$E = \sum_{i=1}^{n} E_i = \frac{1}{2} \sum_{i=1}^{n} (o_i - eo_i)^2 \tag{16}$$

where E_i represents the square error a and o and eo represent the experimental and expected outputs of the output layer, respectively, where eo represents the GT.

During the back-propagation stage, the weights between the layers were updated based on the gradient descent (GD) method [51] as follows:

$$w_{2hk}(G+1) = w_{2hk}(G) - LR_1 \times \frac{\partial E(w)}{\partial w_{2hk}(G)} \tag{17}$$

$$w_{1ih}(G+1) = w_{1ih}(G) - LR_2 \times \frac{\partial E(w)}{\partial w_{1ih}(G)} \tag{18}$$

where G indicates the current iteration number, and LR_1 and LR_2 indicate the learning rates from the input- to the hidden- layers and the hidden- to the output- layers, respectively.

2. Our CFBL

Recently, research groups have commonly adopted the gradient descent module to minimize the model error of the BPNN model error [36]. Owing to the satisfactory memory and heredity of the CFGD technique [39] and the favorable performance of L_2 regularization to resist overfitting without revising the network structure, and enlightened by the work of Chen et al. [52], CFBL was developed in this study.

2.1 Fractional-order backpropagation learning (FBL)

The primary contribution of FBL is its use of the CFGD method instead of the GD method [40] to optimize the model weight of the backpropagation step.

According to the theory rule, in the backpropagation stage, the weight vector is optimized, as shown in Eq. (19) and Eq. (20):

$$w_{2hk}(G+1) = w_{2hk}(G) - LR_1 \times Caputo(w_{2ij}(G)) \times \frac{\partial E(w)}{\partial w_{2hk}(G)} \tag{19}$$

$$w_{1ih}(G+1) = w_{1ih}(G) - LR_2 \times Caputo(w_{1ih}(G)) \times \frac{\partial E(w)}{\partial w_{1ih}(G)} \tag{20}$$

where LR_1 and LR_2 are within [0, 1], and $Caputo$ (•) indicates the Caputo derivative operator [21] defined as follows:

$$Caputo(w_{1ih}(G)) = \frac{1}{(1-\alpha) \times \Gamma(1-\alpha)} \times w_{2ho}(G) \times F'(w_{1ih}(G) \times t_h) \times t_h \times (w_{1ih}(G) - cp)^{1-\alpha} \tag{21}$$

$$Caputo(w_{2hk}(G)) = \frac{1}{(1-\alpha) \times \Gamma(1-\alpha)} \times g(w_{1ih}(G) \times t_h) \times (w_{2hk}(G) - cp)^{1-\alpha} \tag{22}$$

where t represents the projection index, α indicates the fractional parameter in the range of [0, 1], Γ represents the gamma formula, $F(•)$ indicates the sum formula [52], and $cp = \min\{w_{1ih}(G), w_{2hk}(G)\}$ is the calibration parameter.

2.2 L_2 regularization

Owing to the limited training set, FBL is always influenced by the overfitting phenomenon [53], and L_2 regularization is a common technique to resist overfitting without revising the network architecture [50]. Therefore,

L_2 regularization was introduced into FBL to generate the CFBL. After combing with L_2 regularization, the model error function E_{L2} was updated as follows:

$$E_{L_2} = E + \frac{\lambda}{2} \times \|w\|^2 \tag{23}$$

where $\|w\|^2$ indicates the sum of the squares of total weight vectors, and $\lambda \geq 0$ represents the regularization parameter.

According to Eq. (23), Eq. (19) and Eq. (20) can be updated as follows:

$$w_{2hk}(G+1) = w_{2hk}(G) - LR_1 \times Caputo(w_{2ij}(G)) \times \frac{\partial E_{L_2}(w)}{\partial w_{2hk}(G)} \tag{24}$$

$$w_{1ih}(G+1) = w_{1ih}(G) - LR_2 \times Caputo(w_{1ih}(G)) \times \frac{\partial E_{L_2}(w)}{\partial w_{1ih}(G)} \tag{25}$$

3. Initial setting of parameters of our A-ProSeg

Here is the setting of the initial parameters of our A-ProSeg, shown in Table 15.

Table 15 Setting of initial parameters of our method.

	Parameters	Initial value
Global setting	Total number of TCH / JPHCM data	945 slices / 95 patients (393 slices)
	Number of training set	190 patients (675 slices)
	Number of validation slices	20 patients (66 slices)
	Number of testing slices	56 patients (204 slices)
	Number of testing slices	70 patients (300 slices)
	Scale of each slice	600×450
GCPS	Vertices cleaning label $flag(v_i)$	1
	Data radius (r)	1
	Optimal weight parameter (λ)	0.13
DMDE	Number of solutions (Np)	150
	Mutation factor (F)	0.8
	Crossover factor (CR)	0.6
	Average mutation factor (uF)	1
	Average crossover rate (uCR)	0.9
	Probability of using the mutation operator (p_G)	1
	Current iteration number (G)	1
	Max iteration number ($GMax$)	500
CFBL	Number of input neurons (I)	1
	Number of hidden neurons (h)	10
	Number of output neurons (k)	2
	Weight from input to hidden, and hidden to output layer (w_1 / w_2)	1
	Threshold at the hidden / output neuron (b_1 / b_2)	1
	Learning rate (LR)	0.4
	Momentum (m)	0.9
	Current iteration number (G)	1
	Max epoch (G_{max})	1,000

Reference

1. Wang Y, Zheng Q, Heng PA (2018) Online robust projective dictionary learning: shape modeling for MR-TRUS registration. IEEE Trans Med Imaging 37:1067–1078

2. Lim S, Jun C, Chang D, et al (2019) Robotic Transrectal Ultrasound-Guided Prostate Biopsy. IEEE Trans Biomed Eng 66:2527–2537

3. Shaaer A, Davidson M, Semple M, et al (2019) Clinical evaluation of an MRI-to-ultrasound deformable image registration algorithm for prostate brachytherapy. Brachytherapy 18:95–102.

4. Rajpurkar P, Chen E, Banerjee O, Topol EJ (2022) AI in health and medicine. Nat Med 28:31–38.

5. Peng T, Wu Y, Qin J, et al (2022) H-ProSeg: Hybrid ultrasound prostate segmentation based on explainability-guided mathematical model. Comput Methods Programs Biomed 219:106752.

6. Liu Z, Yang C, Huang J, et al (2021) Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. Future Gener Comput Syst 114:358–367.

7. Jiang J, Guo Y, Bi Z, et al (2022) Segmentation of prostate ultrasound images: the state of the art and the future directions of segmentation algorithms. Artif Intell Rev.

8. Xu X, Sanford T, Turkbey B, et al (2022) Polar transform network for prostate ultrasound segmentation with uncertainty estimation. Med Image Anal 78:102418.

9. Wang Y, Dou H, Hu X, et al (2019) Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound. IEEE Trans Med Imaging 38:2768–2778

10. Rundo L, Han C, Nagano Y, et al (2019) USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. Neurocomputing 365:31–43.

11. Orlando N, Gyacskov I, Gillies DJ, et al (2022) Effect of dataset size, image quality, and image type on deep learning-based automatic prostate segmentation in 3D ultrasound. Phys Med Biol 67:074002.

12. Singh RP, Gupta S, Acharya UR (2017) Segmentation of prostate contours for automated diagnosis using ultrasound images: A survey. J Comput Sci 21:223–231.

13. Wu P, Liu Y, Li Y, Shi Y (2013) TRUS image segmentation with non-parametric kernel density estimation shape prior. Biomed Signal Process Control 8:764–771.

14. Wu P, Liu Y, Li Y, Liu B (2015) Robust Prostate Segmentation Using Intrinsic Properties of TRUS Images. IEEE Trans Med Imaging 34:1321–1335.

15. Nouranian S, Ramezani M, Spadinger I, et al (2016) Learning-Based Multi-Label Segmentation of Transrectal Ultrasound Images for Prostate Brachytherapy. IEEE Trans Med Imaging 35:921–932.

16. Xu X, Sanford T, Turkbey B, et al (2022) Shadow-consistent semi-supervised learning for prostate ultrasound segmentation. IEEE Trans Med Imaging 41:1331–1345.

17. Orlando N, Gillies DJ, Gyacskov I, et al (2020) Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. Med Phys 47:2413–2426.

18. Peng T, Wang Y, Xu TC, et al (2018) Detection of Lung Contour with Closed Principal Curve and Machine Learning. J Digit Imaging 31:520–533

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

20. Peng T, Xu TC, Wang Y, Li F (2020) Deep Belief Network and Closed Polygonal Line for Lung Segmentation in Chest Radiographs. Comput J.

21. Wang J, Wen Y, Gou Y, et al (2017) Fractional-order gradient descent learning of BP neural networks with Caputo derivative. Neural Netw 89:19–30

22. Ge Y, Yu W, Lin Y, et al (2018) Distributed differential evolution based on adaptive mergence and split for large-scale optimization. IEEE Trans Cybern 48:2166–2180.

23. Peng T, Zhao J, Wang J (2021) Interpretable mathematical model-guided ultrasound prostate contour extraction using data mining techniques. In: IEEE 15th International Conference on Bioinformatics and

Biomedicine (BIBM). pp 1037–1044

24. He K, Gkioxari G, Dollar P, Girshick R (2017) Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, pp 2961–2969

25. Peng T, Tang C, Wu Y, Cai J (2022) Semi-automatic prostate segmentation from ultrasound images using machine learning and principal curve based on interpretable mathematical model expression. Front Oncol 12:

26. Gao Y, Zhou M, Metaxas D (2021) UTNet: A hybrid transformer architecture for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp 61–71

27. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp 234–241

28. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2020) Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 39:1856–1867

29. Hastie T, Stuetzle W (1989) Principal curves. J Am Stat Assoc 84:502–516.

30. Kégl B, Krzyzak A (2002) Piecewise Linear Skeletonization Using Principal Curves. IEEE Trans Pattern Anal Mach Intell 24:59–74

31. Kégl B, Linder T, Zeger K (2000) Learning and design of principal curves. IEEE Trans Pattern Anal Mach Intell 22:281–297

32. Zhang J, Chen D, Kruger U (2008) Adaptive Constraint K-Segment Principal Curves for Intelligent Transportation Systems. IEEE Trans Intell Transport Syst 9:666–677.

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

34. Liu Q, Du S, van Wyk BJ, Sun Y (2020) Niching particle swarm optimization based on Euclidean distance and hierarchical clustering for multimodal optimization. Nonlinear Dyn 99:2459–2477.

35. Zeng Y-R, Zeng Y, Choi B, Wang L (2017) Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. Energy 127:381–396

36. Leema N, Nehemiah HK, Kannan A (2016) Neural network classifier optimization using Differential Evolution with Global Information and Back Propagation algorithm for clinical datasets. Appl Soft Comput 49:834–844

37. Su Y, Chi R (2017) Multi-objective particle swarm-differential evolution algorithm. Neural Comput & Applic 28:407–418.

38. Zhang J, Sanderson AC (2009) JADE: adaptive differential evolution with optional external archive. IEEE Trans Evol Computat 13:945–958

39. Xiao M, Zheng WX, Jiang G, Cao J (2015) Undamped oscillations generated by Hopf bifurcations in fractional-order recurrent neural networks with Caputo derivative. IEEE Trans Neural Netw Learning Syst 26:3201–3214

40. Qian N (1999) On the momentum term in gradient descent learning algorithms. Neural Netw 12:145–151

41. Sun Y, Zhang Q (2018) Optimization design and reality of the virtual cutting process for the boring bar based on PSO-BP neural networks. Neural Comput & Applic 29:1357–1367.

42. Karimi D, Salcudean SE (2019) Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks. IEEE Trans Med Imaging.

43. Tang Y, Ren F, Pedrycz W (2020) Fuzzy C-Means clustering through SSIM and patch for image segmentation. Appl Soft Comput 87:105928.

44. Badshah N, Atta H, Shah SIA, et al (2020) New Local region based model for the segmentation of medical images. IEEE Access 8:175035–175053.

45. Ali S, Madabhushi A (2012) An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. IEEE Trans Med Imaging 31:1448–1460.

46. Lei Y, Tian S, He X, et al (2019) Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. Med Phys 46:3194–3206.

47. Steffel CN, Brown R, Korcarz CE, et al (2019) Influence of ultrasound system and gain on grayscale median values: Influence of ultrasound system and gain on grayscale median values. J Ultrasound Med 38:307–319.

48. Floridi L (2019) Establishing the rules for building trustworthy AI. Nat Mach Intell 1:261–262.

49. Liang W, Tadesse GA, Ho D, et al (2022) Advances, challenges and opportunities in creating data for trustworthy AI. Nat Mach Intell 4:669–677.

50. Rice L, Wong E, Kolter JZ (2020) Overfitting in adversarially robust deep learning. International Conference on Machine Learning. pp 8093–8104

51. Hecht-Nielsen R (1992) Theory of the backpropagation neural network. Neural networks for perception 65–93

52. Chen M-R, Chen B-P, Zeng G-Q, et al (2020) An adaptive fractional-order BP neural network based on extremal optimization for handwritten digits recognition. Neurocomputing 391:260–272

53. Bao C, Pu Y, Zhang Y (2018) Fractional-order deep backpropagation neural network. Comput Intell Neurosci 2018:1–10