# Coarse-to-Fine Approach: Automatic Delineation of Kidney Ultrasound Data

Tao Peng, Yiwen Ruan, Yidong Gu, Jiang Huang*, Caiyin Tang*, and Jing Cai*

**Abstract:** We present an automatic kidney segmentation method using ultrasound images. This method employs a coarse-to-fine approach to tackle the challenge of unclear and fuzzy boundaries. Four key innovations are introduced to enhance the segmentation process's accuracy and efficiency. First, an automatic deep fusion training network serves as a coarse segmentation strategy. Second, we propose an explainable mathematical mapping formula to better represent the kidney contour. Third, by utilizing the characteristics of the principal curve, a neural network automatically refines curve shapes, thus reducing model errors. Finally, we employ an intelligent searching polyline segment method for automatic kidney contour segmentation. The results show that our method achieves high accuracy and stability in segmenting kidney ultrasound images. This work's contributions include the deep fusion training network, intelligent searching polyline segment method, and explainable mathematical mapping formula, which are applicable to other medical image segmentation tasks. Additionally, this approach uses a mean-shift clustering model, supplanting standard projection and vertex optimization steps.

**Key words:** polyline segment technique; artificial neural network; explainable mathematical mapping formula; ultrasound kidney segmentation

## 1 Introduction

Ultrasound imaging is a highly valuable and non-invasive diagnostic tool, providing a cost-effective approach to kidney disease screening. Its capacity to deliver detailed images of the kidney's internal structures is essential for early detection and diagnosis. However, the manually delineating kidney contours in ultrasound images is laborious and time-consuming, often requiring significant expertise. This manual method burdens healthcare professionals and introduces potential variability and inconsistency in the results. Consequently, there is an increasing need for automated solutions that can efficiently and accurately perform kidney segmentation in ultrasound images,

• Tao Peng and Yiwen Ruan are with School of Future Science and Engineering, Soochow University, Suzhou 215222, China. E-mail: sdpengtao401@gmail.com; yiwenruan492@gmail.com.

• Yidong Gu is with Department of Medical Ultrasound, Suzhou Municipal Hospital, Suzhou 215006, China. E-mail: guyidong@njmu.edu.cn.

• Jiang Huang is with Department of Ophthalmology, The Second Affiliated Hospital of Soochow University, Suzhou 215004, China. E-mail: huangjiang2017@suda.edu.cn.

• Caiyin Tang is with Department of Radiology, Taizhou People's Hospital Affiliated to Nanjing Medical University, Taizhou 318020, China. E-mail: tangcaiyin123@outlook.com.

• Jing Cai is with Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong 999077, China. E-mail: jing.cai@polyu.edu.hk.

∗ To whom correspondence should be addressed.

  Manuscript received: 2023-10-13; revised: 2024-01-16; accepted: 2024-02-08

thus streamlining the diagnostic process and improving result reliability.

Nevertheless, this task is challenging due to several factors. The shape of the kidney varies among patients, and the intensity distribution within the kidney structure is non-uniform. Additionally, the kidney's boundary is often obscured by surrounding features like intestinal gas. This issue is compounded by the low contrast in ultrasound images, which complicates distinguishing the kidney's boundary from adjacent organs such as the spleen. These complexities underline the necessity for robust and accurate automated kidney segmentation methods in ultrasound imaging. Figure 1 shows ultrasound images depicting various complex kidney segmentation scenarios.

In our research, we propose a coarse-to-fine contour delineation approach for ultrasound kidney images, consisting of three key components. First, we use the Deep Fusion Training Network (DFTN) to establish an initial kidney contour. This involves combining the U-Net network[1] with Squeeze & Excitation (SE)[2] and Attention Gate (AG)[3] modules to enhance meaningful feature extraction. Next, these initial results are refined using an Intelligent Searching Polyline Segment (ISPS) method, paired with the Adaptive learning rate Back-propagation Neural Network (ABNN)[4]. Additionally, we introduce an interpretable mathematical model to describe the smooth kidney contour, based on parameters derived from ABNN.

Our method offers several advantages as follows:

(1) Incorporating DFTN in the coarse segmentation step improves feature extraction and suppresses less relevant information, surpassing traditional U-Net methods[1].
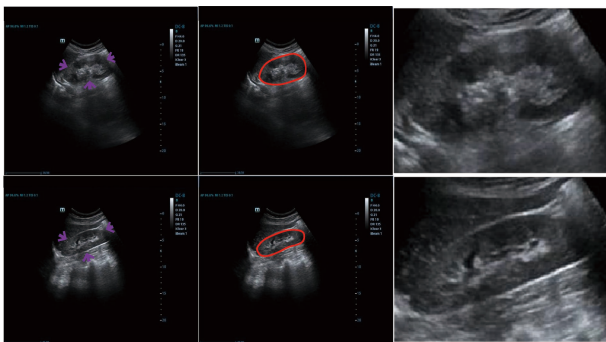


**Fig. 1   Ultrasound image shows unclear regions (marked by arrows). The first, second, and third columns are the original data, ground truth, and amplified results, respectively.**

(2) Previous studies employed the Closed Polyline Segment (CPS) method[4] to enhance the K-segments Polyline Segment (KPS) method[5] in handling closed data. However, these methods did not account for the influence of abnormal points. An improved CPS method[6] is proposed, which incorporated a vertex cleaning scheme to filter abnormal vertices and prevent mis-segmentation. Unlike KPS, CPS, and the improved CPS method, our work proposes the ISPS method. This method automatically determines vertices/clusters without prior information, combining a Neutrosophic Domain-inspired Mean-Shift Clustering (NDMSC) model[7] with an optimized vertex optimization module.

(3) We introduce an interpretable mathematical model for a smooth representation of the kidney contour.

(4) Our coarse-to-fine approach for ultrasound kidney segmentation effectively minimizes the influence of surrounding tissues.

## 2   Related Work

Recently, medical image segmentation techniques have been categorized into two main types: region-based and contour-based methods. Rahman and Uddin[8] developed a kidney segmentation method for ultrasound images, focusing on removing speckle noise, enhancing image quality, and extracting kidney regions for surgical procedures. This method involves restoration, speckle noise reduction, and Gabor filtering. While it effectively segments the kidney, it may require further refinement in cases with unclear boundaries. Jokar and Pourghassem[9] introduced a novel kidney segmentation algorithm for ultrasound images using the curvelet transform method with a shape prior. This algorithm achieves over 92% accuracy in extracting the outline of kidney tissue. The application of the curvelet transform method and a nonlinear function has shown promise in improving image quality and reducing noise, making it a robust method for organ segmentation in ultrasound images. This approach could significantly aid the clinical interpretation of kidney ultrasound images. Xie et al.[10] developed a technique for kidney segmentation using texture and shape priors to accurately delineate the kidney's boundary. Texture features are extracted using Gabor filters, and a texture model is created to assess similarities in internal and external regions. This

method has proven effective in experiments on natural and ultrasound images, outperforming other segmentation methods and manual segmentation. Chen et al.[11] introduced a multi-scale fusion network for precise ultrasound kidney segmentation. In this network, a multi-scale convolutional neural network architecture with multiple monitoring modules is used to enhance the model's flexibility. However, this method may yield incomplete or incorrect results in cases of unclear image boundaries. Contour-based segmentation, in contrast, offers a quicker and more detailed representation of anatomical structures compared to the region-based techniques, by utilizing shape representation and curve fitting.

Cerrolaza et al.[12] introduced an innovative active contour-based method for segmenting the renal collecting system, which simulates fluid propagation within the kidney. They combined a positive delta detector with a Gabor-based semi-automatic segmentation model, significantly improving segmentation accuracy. Weerasinghe et al.[13] implemented a multi-parametric fusion strategy in a deep convolutional neural network for fetal kidney segmentation in ultrasound images. However, a limitation of this strategy is the small dataset size, often resulting in overfitting during training. Ravishankar et al.[14] developed an automatic kidney morphology measurement algorithm for ultrasound data, using an edge analysis technique to accurately locate the kidney region. Experiments have shown the robust segmentation performance of this model.

## 3 Method

### 3.1 Problem formulation

Our method aims not only to obtain an accurate kidney contour, but also to eliminate any manual intervention. Utilizing the capability of deep learning to intelligently gain from image features (such as position, outline, and intensity of the region of interest), we initially apply a deep learning based model for coarse segmentation. Following this, we use the contour points set from the coarse segmentation, denoted as $P_n = \{p_1, p_2, ..., p_n\}$, where $p_i(x_i, y_i) \subseteq \mathbf{R}^d$, and use it as the input for our PS-related method. This input includes the order of vertex distribution, represented as $D = \{d_1, d_2, ..., d_n\} \in R^d = \{(t_i, (x, y)), i = 1, 2, ..., n, 0 \leq t_1 < t_i < t_n \leq 1\}$. Here, $t_i$ represents the order of vertex distribution, while $t_i$ and $y$ correspond to the $x$-axis and $y$-axis

coordinates of the vertex $v_i$ of the resulting boundary. Notably, the PS-related method generates a contour comprising multiple segments. To address this, we have developed a mathematical mapping function to produce a smooth kidney contour with optimized vertices, ensuring close alignment with the actual contour, as shown in the following:

$$(\text{Outcome}_o(x(t)), \text{Outcome}_o(y(t))) =$$
$$\left( \sum_{j=1}^{K} \frac{1}{1 + e^{\sum_{i=1}^{H} -(t \times w1_i - a_i)}} w2_{j,1} - b_{j,1}, \right.$$
$$\left. \sum_{j=1}^{K} \frac{1}{1 + e^{\sum_{i=1}^{H} -(t \times w1_i - a_i)}} w2_{j,2} - b_{j,2} \right) \quad (1)$$

$$f(t) = (x(t), y(t)) = \left( \frac{1}{1 + e^{-\text{Outcome}_o(x)}}, \frac{1}{1 + e^{-\text{Outcome}_o(y)}} \right) \quad (2)$$

where Outcome ($\cdot$) is the outcome of the output units; $K$ and $H$ represent the amount of output and hidden neurons, respectively; $b_j (j = 1, 2)$ indicates the output-layer threshold under the $j$-th neuron; $a_i$ represents the model threshold in the hidden layer under the $i$-th hidden neuron; and w1 and w2 represent the hidden-layer and output-layer weights, respectively.

Several challenging issues are considered in our project: (1) The deep neural network has gained significant interest in automated boundary delineation works. Nevertheless, due to the low contrast between the kidney and neighboring organs, finding an optimal deep neural network for ultrasound kidney segmentation remains difficult. (2) The primary steps in current Polyline Segment (PS) related methods[5] are projection and vertex optimization, where the number of vertices is typically decided through artificial intervention. A significant challenge in this study is determining the vertices automatically to describe the segmentation contour. (3) After applying the PS-related method, the contour consists of several line segments. Developing a mathematical map function to represent the smooth kidney contour poses a substantial challenge.

### 3.2 Our method

This coarse-to-fine approach comprises two primary parts: (1) the coarse segmentation step and (2) the refinement step. The workflow of this method is illustrated in Fig. 2. Stage 1 is the deep fusion training-related coarse segmentation step, and Stage 2 is the refinement step.
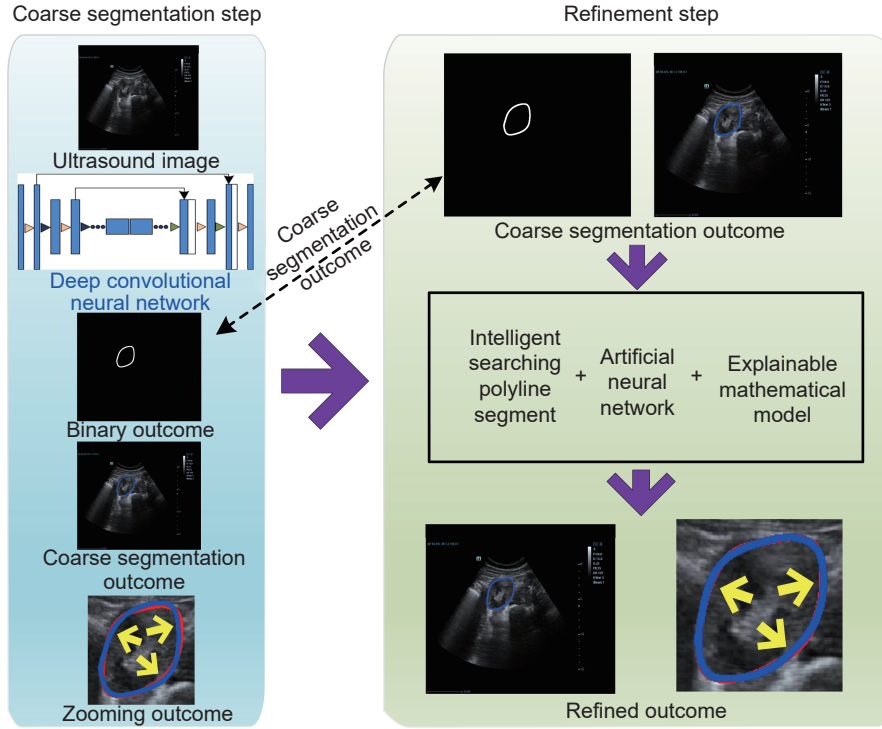
**Fig. 2　Workflow of the coarse-to-fine-related approach. In the coarse segmentation step, the DFTN neural network is utilized for the automatic coarse segmentation of the Region Of Interest (ROI). The refinement step primarily includes the ISPS method and ABNN. The ISPS method is applied to determine the vertex sequence and the ROI boundary, which consists of segments. This vertex sequence is used as input for ABNN, helping to reduce the global model error during training. Post-training, ABNN's parameters are employed to interpret the smooth ROI boundary, as depicted in Eqs. (1) and (2). The blue outline represents the segmentation results, while the red outline indicates the ground truth contour.**

### 3.3　Coarse segmentation step

To effectively extract and localize key features for kidney segmentation, we utilize the AG module's capacity to emphasize salient features and the SE module's ability to enhance significant features while suppressing less relevant ones[2]. We integrate the SE and AG modules into the U-Net architecture[1] for coarse segmentation, employing both serial and parallel variants, as depicted in Fig. 3.

Let $H_s$ and $W_s$ be the spatial height and width, respectively, and $C'$ and $C$ are the number of input and output channels, respectively. We suppose that an input feature map in the real number set, $X \in \mathbf{R}^{C' \times H_s \times W_s}$, it passes through several encoder or decoder blocks $F(\cdot)$, to produce the output feature map $U \in \mathbf{R}^{C \times H_s \times W_s}$, $F$: $X \rightarrow U$. $U$ is produced by merging the information of $X$ through a series of convolutional layers and non-linearity transformation functions (i.e., Rectified Linear Unit (ReLU) and Sigmoid[15]). To achieve a calibrated $U'$ from $U$, we adopt serial (Figs. 3a and 3b) and parallel architectures (Fig. 3c) containing the SE

module $F_{SE}(\cdot)$, or the attention gate module, $F_{AG}(\cdot)$, blocks in different orders.

The DFTN model incorporates a U-Net architecture with encoder and decoder blocks. Encoder blocks consist of two $3 \times 3$ convolutions with instance normalization and Leaky ReLU for feature extraction. Decoder blocks feature skip connections and convolutional layers for upsampling and local feature integration. The SE module calibrates the model through global max-pooling (squeeze) and $1 \times 1$ convolutions with ReLU and Sigmoid activations (excitation). The AG module combines outputs from downsampling and upsampling layers, followed by ReLU, a $1 \times 1$ convolution, sigmoid activation, and multiplication. This enhances the precision of segmentation by adjusting predictions based on pertinent information. These modules are strategically placed within the U-Net structure to improve the accuracy of kidney segmentation in ultrasound images.

### 3.4　Obtain vertices sequence

We propose the ISPS method for acquiring the vertices

(a) Serial AG block followed by SE block

(c) Parallel architectures

- 🟢 SE block
- 🔴 AG block
- ⚪ Concatenation
- 🔵 Encoder or decoder blocks
- ↗ Feature from encoding path as SE's input
- ↗ Feature from decoding path as SE's input
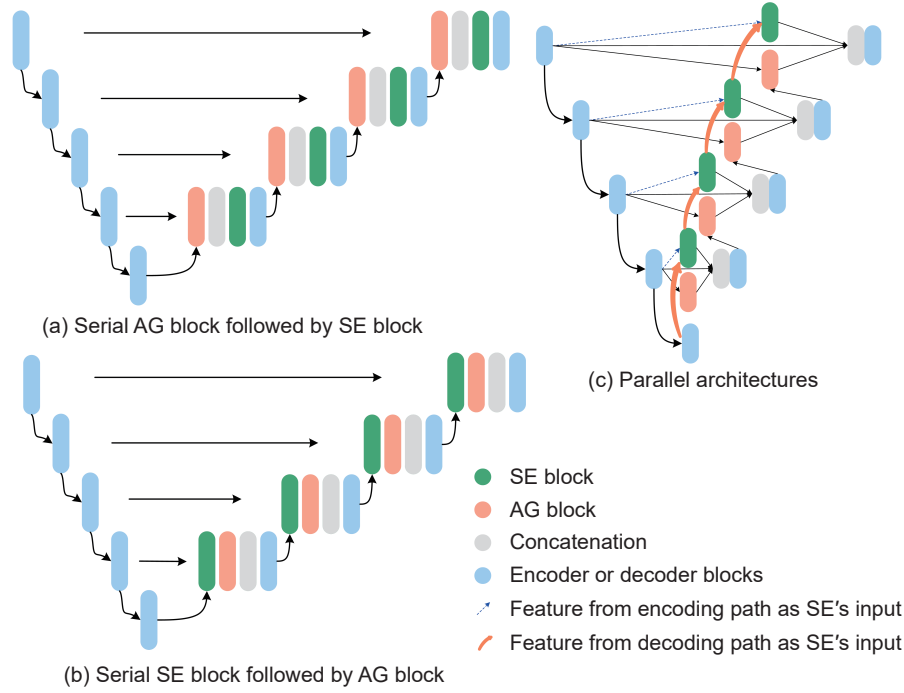
(b) Serial SE block followed by AG block

**Fig. 3    Illustration of variants structure. (a) Serial structure of AG followed by SE (Serial Architecture Variant 1, SAV-1), (b) serial structure of SE followed by AG (Serial Architecture Variant 2, SAV-2), and (c) parallel structure employing features from decoding (Parallel Architecture Variant 1, PAV-1) or encoding (Parallel Architecture Variant 2, PAV-2) path as SE's input.**

sequence, which includes the sequence number of vertices $t$ and vertices' coordinates $v_i$ $(x_i, y_i)$. Unlike conventional PS-related methods[5], our method automatically determines the number of vertices without prior information. Figure 4 illustrates the difference between the conventional PS-related method and our ISPS method.

### 3.4.1   NDMSC method

Mean Shift Clustering (MSC), initially proposed by Cheng[16], identifies data points belonging to the same
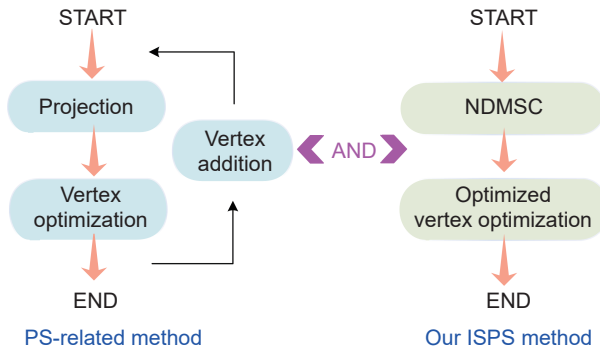


**Fig. 4    Comparison between the conventional PS-related method and our ISPS method. The PS-related approach involves projection, vertex optimization, and vertex addition steps. In contrast, our ISPS method includes an optimized vertex optimization step and the NDMSC model.**

cluster and automatically determines the number of vertices or clusters. We enhance this approach with the NDMSC model[7], integrating a neutrosophic domain-inspired filter into the MSC method. Additionally, to improve data consistency, we preprocess the data with min-max normalization before applying the NDMSC model.

### 3.4.2   Vertex optimization module

#### 3.4.2.1   Vertex optimization step of standard PS-related methods

Standard PS-related methods primarily utilize the vertex optimization step to minimize the penalty distance function $G_n(v_i)$, thereby updating the position of each vertex $v$ or segment $s$[5]. The penalized distance function $G_n(v_i)$ is defined as follows:

$$G_n(v_i) = \frac{1}{n}\Delta_n(v_i) + \lambda\frac{1}{b+1}P(v_i) \qquad (3)$$

where $\Delta_n(v_i)$ denotes the local squared distance from data points $p$ to vertex $v$[5] and $P(v_i)$ denotes the triangle-based curvature penalty at vertex $v_i$[17]. The penalty factor $\lambda$ is defined as $\lambda = \lambda' \cdot \dfrac{b}{n^{1/3}} \cdot \dfrac{\sqrt{\Delta_n(f_{b,n})}}{r}$, where $n$ and $b$ are the numbers of data points and vertices, respectively. $\lambda'$ is an adjustable parameter, set

to a constant 0.13 following several experimental trials[18].

### 3.4.2.2 Vertex optimization step of standard PS-related methods

The previous PS-related works[5, 18] employed the curvature penalty $P(v_i)$, computed using numerous triangle functions and piecewise functions, necessitating time-intensive computational operations. Consequently, our work redesigns the penalized distance function $P(v_i)$, replacing it with an addition and average-based penalized distance function $D(v_i)$. The revised penalized distance function $G_n'(v_i)$ is

$$G_n'(v_i) = \frac{1}{n}\Delta_n(v_i) + \lambda\frac{1}{b+1}D(v_i) \tag{4}$$

where $D(v_i)$ represents the penalty imposed on the total curvature of the Principal Curve (PC) shown as follows:

$$D(v_i) = \frac{1}{b}(\sum_{i=1}^{b}\Delta(p_i, v_i)) \tag{5}$$

### 3.5 Explainable mathematical formula of kidney boundary

After the ISPS method generates the vertex sequence, the next critical step is creating a mathematical model for a smooth kidney contour. This model incorporates an ABNN, which optimizes based on input parameters, like $t$ and vertex coordinates, aiming to minimize mean square error[4]. We employ Sigmoid and ReLU activation functions in forward propagation to fine-tune the model for optimal outcomes[15]. Finally, ABNN's parameters are used to interpret the smoothing equations in Eqs. (1) and (2), significantly improving segmentation accuracy and interpretability.

## 4 Experiment

To assess our proposed method both qualitatively and quantitatively, we conduct various experiments on a clinical kidney dataset. This dataset includes ultrasound images from 100 patients, captured through trans-abdominal ultrasound scanning using the Mindray DC-8 System with a 1.3–5.7 MHz low-resolution linear transducer. The mechanical index is maintained at 1.3, with a probe detection depth of 200 mm, a frequency of 4 MHz, and an amplifier gain ranging from 3 to 33 dB. Patient ages varied from 13 to 76, and all images are in the digital imaging and communications in medicine file format, with an original resolution of 1200 pixel ×

900 pixel. Three expert radiologists manually delineate the images, with the majority opinion serving as the Ground Truth (GT). Additionally, we employed two standard metrics, the Dice Similarity Coefficient (DSC) and the Jaccard similarity coefficient ($\Omega$)[4], for quantitative analysis,

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{6}$$

$$\Omega = \frac{TP}{FP + TP + FN} \tag{7}$$

where TP, FP, FN, and TN represent true positive, false-positive, false-negative, and true negative, respectively, as shown in Table 1.

### 4.1 Settings

Our kidney dataset, comprising images from 100 patients (1200 images), is randomly divided into three groups: 70 patients (840 images) for training, 10 patients (120 images) for validation, and 20 patients (240 images) for testing. The distributions of left and right kidney images are equal in the training, validation, and testing sets. All kidney slices are resized to 600 pixel × 450 pixel. To develop the optimal deep learning model, training data augmentation is conducted by rotating all training slices within a range of [−15°, 15°] to reach a total of 2000 training slices. In the DFTN training, the Dice loss function calculates loss, and the Adam optimizer[13] updates the learning rate. We begin with a learning rate of 0.001, reducing it on a plateau with a patience of 30 and implementing early stopping at a maximum of 300 epochs. Moreover, the ABNN method, based on machine learning, is optimized using stochastic gradient descent, with an initial learning rate of 0.4, a momentum value of 0.9, and a cap of 1000 epochs. All experiments are executed on a computer with an Intel Core i7-8750H CPU and a Geforce GTX 1070 GPU with 8 GB of memory.

### 4.2 Evaluation under multiple variants

Table 2 presents the testing results of four variant models (SAV-1, SAV-2, PAV-1, and PAV-2) in terms of the DSC value. Additionally, Fig. 5 displays the qualitative outcomes of these four variant methods.

**Table 1    Definitions of TP, FP, FN, and TN.**

| Test result label | Positive | Negative |
|---|---|---|
| Positive | TP | FP |
| Negative | FN | TN |

**Table 2** **Performance among multiple variants. SAV-1 and SAV-2 employ serial structures with AG followed by SE, and SE followed by AG, respectively. PAV-1 and PAV-2 represent parallel structures using features from the decoding and encoding paths, respectively, as inputs for the SE module (SD means standard deviation).**

(%)

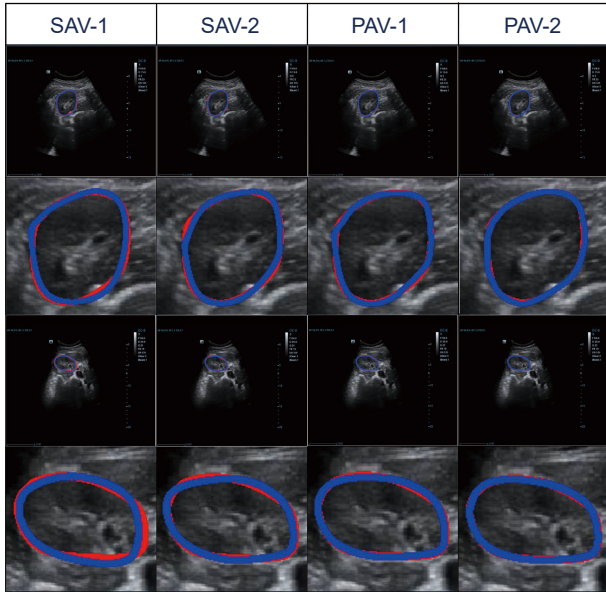| Method | DSC + SD | $\Omega$ + SD |
|--------|----------|---------------|
| SAV-1 | 87.1 + 12.9 | 78.6 + 14.1 |
| SAV-2 | 87.7 + 13.1 | 79.9 + 13.7 |
| PAV-1 | 88.7 + 12.6 | 80.5 + 13.2 |
| PAV-2 | 89.2 + 12.4 | 81.1 + 13.2 |



**Fig. 5  Display outcomes of four variant methods. These images, representing two randomly chosen slices, showcase multiple structures. The first two rows feature images from the left kidney in the axial view, while the remaining rows show the right kidney in the same view. Blue and red curves indicate the test result and GT, respectively. The first and third rows depict the overlay between the test outcome and GT, with the other rows providing zoomed-in visualizations.**

parallel architecture variants (PAV-1 and PAV-2) demonstrate superior performance compared to Serial Architecture Variants (SAV-1 and SAV-2), with DSC and $\Omega$ values of 89.2 + 12.4% and 81.1 + 13.2%, and 88.7 + 12.6% and 80.5 + 13.2%, respectively. Notably, compared to PAV-1, PAV-2 shows a mean improvement of 0.5% in DSC and 0.7% in $\Omega$. Thus, we select PAV-2 as the coarse segmentation stage of our model.

### 4.3  Ablation study

To thoroughly evaluate our proposed method's effectiveness, we conduct an ablation study, systematically analyzing each component's impact. A summary of all the ablation experimental models is provided in Table 2, with the model abbreviations detailed in Appendix A.

According to Table 2, employing PAV-2 for coarse segmentation, our method achieves the initial segmentation outcome (Model 1). The refinement stage is used to optimize this outcome, where MSC + Vertex Optimization Step (VOS) + Back-propagation Neural Network (BNN) serves as the baseline model for refinement. This stage utilizes the advantages of the principal curve and mean-shift clustering for intelligent data fitting and automatic determination of vertices/clusters' number, respectively, while utilizing neural network learning to reduce modeling error (Model 2). To improve Model 2's handling of indeterminate information, particularly image noise, we develop the neutrosophic domain-inspired NDMSC (Model 3). To reduce time loss during computing curvature penalty, the addition and average-based Optimized Vertex Optimization Step (OVOS) module is adopted instead of the triangle-based VOS module (Model 4). Combining the merits of neutrosophic domain-inspired NDMSC, a new method is presented (Model 5). Differing from Model 5, our model (Model 6) uses the ABNN model to enhance the stability and adaptability of the neural network model.

The results of this study are detailed in Table 3 for quantitative assessment and Fig. 6 for qualitative comparison of different models. Table 3 shows that models with a refinement step consistently outperform Model 1, which lacks this step. These refined models consistently achieve DSC values above 91.7%, signifying a marked improvement in accuracy. The DSC and $\Omega$ values improved by up to 5.26% and 8.63%, respectively, compared to Model 1. Notably, the standard deviations of these metrics in the refined models are significantly lower, indicating enhanced stability and reliability in their performance.

### 4.4  Comparison with other methods

To assess the precision and robustness of our method,

**Table 3　Comparison of model ablation. All models, except Model 1, use coarse-to-refinement architectures. Model 1: PAV-2; Model 2: PAV-2+MSC+VOS+BNN; Model 3: PAV-2+NDMSC+VOS+BNN; Model 4: PAV-2+MSC+OVOS+BNN; Model 5: PAV-2+NDMSC+OVOS+BNN; Model 6: PAV-2+NDMSC+OVOS+ABNN.**

(%)

| Model | DSC + SD | $\Omega$ + SD |
|---|---|---|
| Model 1 | 89.2 + 12.4 | 81.1 + 13.2 |
| Model 2 | 91.7 + 9.6 | 83.9 + 10.2 |
| Model 3 | 92.3 + 8.1 | 85.4 + 9.4 |
| Model 4 | 92.7 + 7.3 | 86.6 + 7.4 |
| Model 5 | 93.2 + 6.7 | 87.4 + 6.9 |
| Model 6 | 93.9 + 6.1 | 88.1 + 6.6 |

we compare it with current leading methods, including hybrid and deep-learning-based approaches. Initially, we compare our method with our previous work, H-SegNet, designed for lung segmentation in chest radiographs[19]. Additionally, we compare it with deep learning methods, like the Mask Region-based Convolutional Neural Networks (Mask-RCNN) for object detection[20] and nested U-Net architecture (U-Net++) for medical image segmentation[21].

As indicated in Table 4, our method outperforms deep learning methods, with hybrid approaches achieving DSC and $\Omega$ values up to 7.93% and 9.57% higher, respectively. This improvement is largely attributed to the refinement step's effectiveness in fine-tuning coarse segmentation results. Overall, our
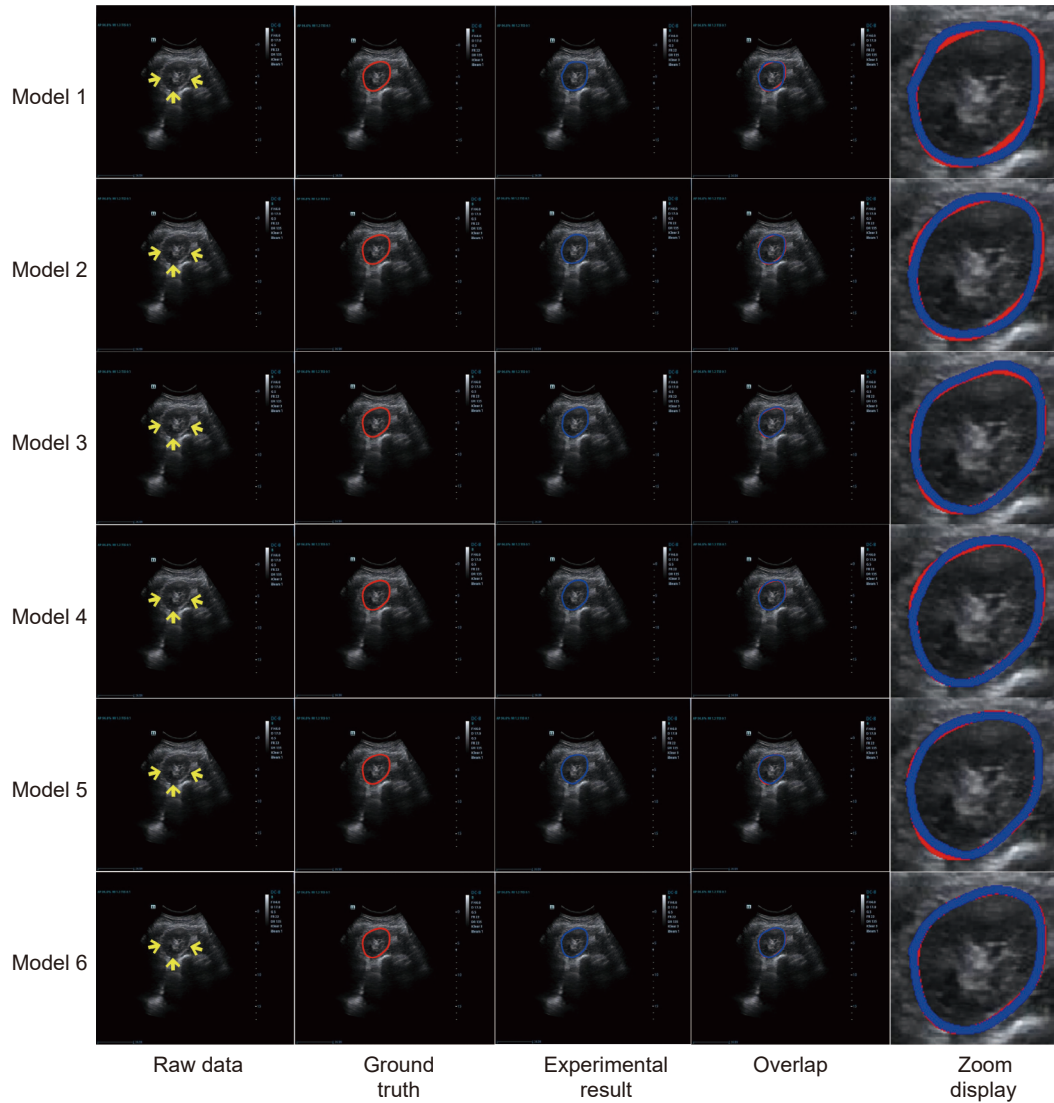


**Fig. 6　Analyzing the impact of each component in our approach. The arrows highlight areas with unclear or fuzzy boundaries. Red curve represents the ground truth, while the blue curve depicts the segmentation result.**

**Table 4    Comparison with state-of-the-art methods (DSC and $\Omega$ in mean + SD).**

| Reference | Method | Model | DSC + SD (%) | $\Omega$ + SD (%) |
|---|---|---|---|---|
| Ref. [19] | H-SegNet | Hybrid | 92.9 + 6.5 | 87.4 + 6.9 |
| Ref. [20] | Mask-RCNN | Deep learning | 87.0 + 12.7 | 80.4 + 14.6 |
| Ref. [21] | U-Net++ | Deep learning | 88.7 + 11.1 | 80.5 + 11.6 |
| Current | Our method | Hybrid | 93.9 + 6.1 | 88.1 + 6.6 |

proposed method demonstrates superior performance across all metrics.

# 5    Conclusion

To address the challenges of accurate kidney segmentation in ultrasound images, we propose a coarse-to-refinement architecture. Key contributions of our method include: (1) an automatic deep fusion training network for coarse segmentation; (2) a novel intelligent searching polyline segment approach; and (3) an explainable mathematical mapping function of the kidney contour. The algorithm's kidney contour is compared with manual segmentation by experts, and quantitative measurements are conducted. The efficacy of our proposed method is demonstrated through ablation studies, comparisons with leading methods, and experiments involving noise introduction. Inspired by Refs. [22–24], future work will explore multi-modal data across different tasks.

# Appendix

## A    Abbreviation of mentioned model/metrics

Here is the abbreviation of our model and evaluation metrics, shown in Table A1.

## B    Illustration of formulas of smooth the explainable kidney boundary

(1) Illustration of Formula (1)

Our work employs the ABNN for training, where the output layer's outcome is obtained during the forward-propagation stage, and model arguments are optimized in the backpropagation stage. When the forward-propagation stage is executed, the Sigmoid activation function $h_1 = \frac{1}{1+e^{-x}}$ is utilized at the input-to-hidden layer, and ReLU activation function $h_2 = \frac{e^{-x}-1}{2}$ is utilized at the hidden-to-output layer. We will mainly illustrate the obtaining procedure of Eq. (1).

First, let $H$ represent the amount of hidden neurons, the input of the hidden layer $\text{Hid}_\text{I}$ is defined,

**Table A1    Abbreviation of used model and metrics.**

| Classe | Abbreviation | Full name |
|---|---|---|
| General | GT | Ground Truth |
| | TP | True Positive |
| | FP | False Positive |
| | FN | False Negative |
| | TN | True Negative |
| Model | DFTN | Deep Fusion Training Network |
| | ISPS | Intelligent Searching Polyline Segment method |
| | MSC | Mean Shift Clustering method |
| | NDMSC | Neutrosophic Domain-inspired Mean-Shift Clustering |
| | PS | Polyline Segment method |
| | CPS | Closed Polyline Segment method |
| | BNN | Back-propagation Neural Network |
| | ABNN | Adaptive learning rate Back-propagation Neural Network |
| Module | SE | Squeeze & Excitation |
| | AG | Attention Gate |
| | OVOS | Optimized Vertex Optimization Step |
| | VOS | Vertex Optimization Step |
| | SAV-1 | Serial Architecture Variant 1 |
| | SAV-2 | Serial Architecture Variant 2 |
| | PAV-1 | Parallel Architecture Variant 1 |
| | PAV-2 | Parallel Architecture Variant 2 |
| Metric | DSC | Dice Similarity Coefficient |
| | $\Omega$ | Jaccard Similarity Coefficient |
| | SD | Standard Deviation |

$$\text{Hid}_\text{I} = \sum_{i=1}^{H} t \times \text{w}1_i - a_i \tag{A1}$$

Then, with $\text{Hid}_\text{I}$, the outcome of the hidden layer $\text{Hid}_\text{O}$ is acquired,

$$\text{Hid}_\text{O} = \frac{1}{1+e^{-\text{Hid}_\text{I}}} = \frac{1}{1+e^{\sum_{i=1}^{H} -(t \times \text{w}1_i - a_i)}} \tag{A2}$$

Given the coordinate of vertex with two dimensions (i.e., $x$- and $y$- axis), we select the number of output neurons $K = 2$, and then the input of the output layer $\text{Outcome}_\text{I}$ is calculated,

$$\text{Outcome}_\text{I} = \sum_{j=1}^{K} \text{Hid}_\text{o} \text{w2}_j - b_j =$$

$$\sum_{j=1}^{K} \frac{1}{1 + e^{\sum\limits_{i=1}^{H} -(t \times \text{w1}_i - a_i)}} \text{w2}_j - b_j \quad (A3)$$

With $\text{Outcome}_\text{I}$, the outcome of the output layer $\text{Outcome}_\text{O}$ is computed,

$$\text{Outcome}_\text{O} = \text{Outcome}_\text{I} = \sum_{j=1}^{K} \text{Hid}_\text{o} \text{w2}_j - b_j =$$

$$\sum_{j=1}^{K} \frac{1}{1 + e^{\sum\limits_{i=1}^{H} -(t \times \text{w1}_i - a_i)}} \text{w2}_j - b_j \quad (A4)$$

In the end, the outcome of two output-layer neurons $\text{Outcome}_\text{o}(\cdot)$ are calculated as follows:

$$(\text{Outcome}_\text{o}(x(t)), \text{Outcome}_\text{o}(y(t))) =$$

$$\Big( \sum_{j=1}^{K} \frac{1}{1 + e^{\sum_{i=1}^{H} -(t \times \text{w1}_i - a_i)}} \text{w2}_{j,1} - b_{j,1},$$

$$\sum_{j=1}^{K} \frac{1}{1 + e^{\sum_{i=1}^{H} -(t \times \text{w1}_i - a_i)}} \text{w2}_{j,2} - b_{j,2} \Big) \quad (A5)$$

(2) Illustration of Formula (2)

From Ref. [25], an arithmetic function for the organ ROI is developed, employing the Sigmoid activation function on both input-to-hidden and hidden-to-output layers,

$$g(t) = \big(x(t)_g, y(t)_g\big) = \left( \frac{1}{1 + e^{-\text{Outcome}_\text{o}(x)}}, \frac{1}{1 + e^{-\text{Outcome}_\text{o}(y)}} \right) \quad (A6)$$

To tackle the issue of gradient vanishing, this work substitutes the ReLU activation function for the Sigmoid function at the hidden-to-output layer[26], where $x(t)$ and $y(t)$ are

$$x(t)_f = x(t)_g = \frac{1}{1 + e^{-\text{Outcome}_\text{o}(x)}} \quad (A7)$$

$$y(t)_f = y(t)_g = \frac{1}{1 + e^{-\text{Outcome}_\text{o}(y)}} \quad (A8)$$

Following Eqs. (A6−A8), the arithmetic equation of the ROI boundary is shown as

$$f(t) = \big(x(t)_f, y(t)_f\big) = \left( \frac{1}{1 + e^{-\text{Outcome}_\text{o}(x)}}, \frac{1}{1 + e^{-\text{Outcome}_\text{o}(y)}} \right) \quad (A9)$$

## Acknowledgment

## References

[1] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015, pp.234−241.

[2] A. G. Roy, N. Navab, and C. Wachinger, Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks, *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 540–549, 2019.

[3] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, *Med. Image Anal.*, vol. 53, pp. 197–207, 2019.

[4] 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. Imaging*, vol. 31, no. 4, pp. 520–533, 2018.

[5] B. Kegl and A. Krzyzak, Piecewise linear skeletonization using principal curves, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 59–74, 2002.

[6] T. Peng, C. Tang, Y. Wu, and J. Cai, H-SegMed: A hybrid method for prostate segmentation in TRUS images via improved closed principal curve and improved enhanced machine learning, *Int. J. Comput. Vis.*, vol. 130, no. 8, pp. 1896–1919, 2022.

[7] Y. Guo, A. Sengür, Y. Akbulut, and A. Shipley, An effective color image segmentation approach using neutrosophic adaptive mean shift clustering, *Measurement*, vol. 119, pp. 28–40, 2018.

[8] T. Rahman and M. S. Uddin, Speckle noise reduction and segmentation of kidney regions from ultrasound image, in *Proc. Int. Conf. Informatics, Electronics and Vision (ICIEV)*, Dhaka, Bangladesh, 2013, pp. 1–5.

[9] E. Jokar and H. Pourghassem, Kidney segmentation in ultrasound images using curvelet transform and shape prior, in *Proc. Int. Conf. Communication Systems and Network Technologies*, Gwalior, India, 2013, pp. 180–185.

[10] J. Xie, Y. Jiang, and H.-T. Tsui, Segmentation of kidney from ultrasound images based on texture and shape priors, *IEEE Trans. Med. Imaging*, vol. 24, no. 1, pp. 45–57, 2005.

[11] G. Chen, J. Yin, Y. Dai, J. Zhang, X. Yin, and L. Cui, A novel convolutional neural network for kidney ultrasound images segmentation, *Comput. Methods Programs Biomed.*, vol. 218, p. 106712, 2022.

[12] J. J. Cerrolaza, C. Meyer, J. Jago, C. Peters, and M. G. Linguraru, Positive delta detection for alpha shape segmentation of 3D ultrasound images of pathologic kidneys, in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*,

Munich, Germany, 2015, pp.711−718.

[13] N. H. Weerasinghe, N. H. Lovell, A. W. Welsh, and G. N. Stevenson, Multi-parametric fusion of 3D power Doppler ultrasound for fetal kidney segmentation using fully convolutional neural networks, *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 2050–2057, 2021.

[14] H. Ravishankar, P. Annangi, M. Washburn, and J. Lanning, Automated kidney morphology measurements from ultrasound images using texture and edge analysis, SPIE medical imaging, in *Proc. SPIE 9790, Medical Imaging 2016*: *Ultrasonic Imaging and Tomography*, San Diego, CA, USA, 2016, pp. 359–365,

[15] A. Apicella, F. Donnarumma, F. Isgrò, and R. Prevete, A survey on modern trainable activation functions, *Neural Netw.*, vol. 138, pp. 14–32, 2021.

[16] Y. Cheng, Mean shift, mode seeking, and clustering, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, 1995.

[17] T. Peng, C. Wang, Y. Zhang, and J. Wang, H-SegNet: Hybrid segmentation network for lung segmentation in chest radiographs using mask region-based convolutional neural network and adaptive closed polyline searching method, *Phys. Med. Biol.*, vol. 67, no. 7, p. 075006, 2022.

[18] 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, 2000.

[19] T. Peng, Y. Wu, J. Qin, Q. J. Wu, and J. Cai, H-ProSeg: Hybrid ultrasound prostate segmentation based on explainability-guided mathematical model, *Comput.*

*Methods Programs Biomed.*, vol. 219, p. 106752, 2022.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, Mask R-CNN, in *Proc. IEEE Int. Conf. Computer Vision* (*ICCV*), Venice, Italy, 2017, pp. 2980–2988.

[21] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, UNet: Redesigning skip connections to exploit multiscale features in image segmentation, *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, 2020.

[22] J. Liu, G. Tan, W. Lan, and J. Wang, Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks, *BMC Bioinformatics*, vol. 21, no. Suppl6, p. 123, 2020.

[23] W. Lan, Y. Dong, Q. Chen, R. Zheng, J. Liu, Y. Pan, and Y. P. Chen, KGANCDA: Predicting circRNA-disease associations based on knowledge graph attention network, *Brief. Bioinform.*, vol. 23, no. 1, p. bbab494, 2022.

[24] W. Lan, T. Yang, Q. Chen, S. Zhang, Y. Dong, H. Zhou, Y. Pan, Multiview subspace clustering via low-rank symmetric affinity graph, *IEEE Trans. Neural Netw. Learning Syst.*, doi: 10.1109/TNNLS.2023.3260258.

[25] Y. H. Wang, Y. Guo, Y. C. Fu, and Z. Y. Shen, An algorithm for learning principal curves with principal component analysis and back-propagation network, in *Proc. Seventh Int. Conf. Intelligent Systems Design and Applications*, Rio de Janeiro, Brazil, 2007, p. 447–453.

[26] K. Hara, D. Saito, and H. Shouno, Analysis of function of rectified linear unit used in deep learning, in *Proc. Int. Joint Conf. Neural Networks* (*IJCNN*), Killarney, Ireland, 2015, pp. 1–8.
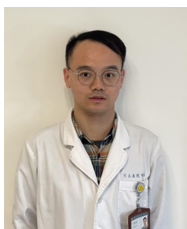
**Tao Peng** received the PhD degree in computer secience from Soochow University, China in 2019. From 2020 to 2022, he was a postdoctoral researcher at Department of Health Technology and Informatics, The Hong Kong Polytechnic University, and Department of Radiation Oncology, University of Texas Southwestern Medical Center, USA, successively. He is currently an associate professor at School of Future Science and Engineering, Soochow University, China. His main research interests include medical image processing, pattern recognition, machine learning, and their applications.

**Yidong Gu** received the Bachelor of Medical and Master of Medicine from Xuzhou Medical University, China in 2009, and Soochow University, China in 2019, respectively. He is currently a sonographer at Department of Medical Ultrasound, Suzhou Municipal Hospital, China. His main research interests are the diagnosis of male infertility and prostate disease.

**Jiang Huang** received the Bachelor of Medicine from Xi'an Jiaotong University, China in 2004, the Master of Medicine in ophthalmology and optometry from Wenzhou Medical University, China in 2007, and the Doctor of Medicine from Soochow University, China in 2017. He is an ophthalmologist chief physician at Department of Ophthalmology, The Second Affiliated Hospital of Soochow University, China. He specializes at vitreous and retina disease, and is proficient in surgery of retinal disease, such as retinal detachment, diabetic retinopathy, and retinal vein occlusion.
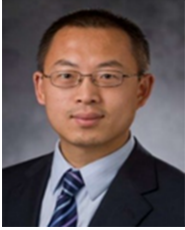
**Yiwen Ruan** currently is an undergraduate student at Soochow University. Her main research interest is AI-based medical image segmentation.

**Jing Cai** received the PhD degree in engineering physics from University of Virginia, USA in 2006. He entered the ranks of academia as assistant professor at Duke University, USA in 2009, and was promoted to associate professor in 2014. He joined The Hong Kong Polytechnic University, China in 2017, where he is currently a full professor and the funding programmer leader of medical physics at Department of Health Technology and Informatics. His main research interests contain medical image processing, pattern recognition, and AI.

**Caiyin Tang** received the Bachelor of Medicine and the Master of Medicine from Xuzhou Medical University, China in 2008, and Southeast University, China in 2017, respectively. He is a radiologist at Department of Radiology, Taizhou People's Hospital Affiliated to Nanjing Medical University. His main research interest is the diagnosis of angiography, lung, kidney, and breast disease.