Future Generation Computer Systems 114 (2021) 358–367

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer

Zhiyong Liu [a,1], Chuan Yang [b,1], Jun Huang [b], Shaopeng Liu [a], Yumin Zhuo [c], Xu Lu [a,*]

[a] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China
[b] Department of Ultrasonography, The First Affiliated Hospital of Jinan University, Guangzhou 510630, China
[c] Department of Urology, The First Affiliated Hospital of Jinan University, Guangzhou 510630, China

ARTICLE INFO

Article history:
Received 30 May 2020
Received in revised form 10 August 2020
Accepted 14 August 2020
Available online 20 August 2020

Keywords:
Prostate cancer
Ultrasound image
S-Mask R-CNN
Inception-v3
Instance segmentation
Image classification

ABSTRACT

The computer-aided diagnosis of prostate ultrasound images can aid in the detection and treatment of prostate cancer. However, the ultrasound images of the prostate sometimes come with serious speckle noise, low signal-to-noise ratio, and poor detection accuracy. To overcome this shortcoming, we proposed a deep learning model that integrates S-Mask R-CNN and Inception-v3 in the ultrasound image-aided diagnosis of prostate cancer in this paper. The improved S-Mask R-CNN was used to realize the accurate segmentation of prostate ultrasound images and generate candidate regions. The region of interest align algorithm was used to realize the pixel-level feature point positioning. The corresponding binary mask of prostate images was generated by the convolution network to segment the prostate region and the background. Then, the background information was shielded, and a data set of segmented ultrasound images of the prostate was constructed for the Inception-v3 network for lesion detection. A new network model was added to replace the original classification module, which is composed of forward and back propagation. Forward propagation mainly transfers the characteristics extracted from the convolution layer pooling layer below the pool_3 layer through the transfer learning strategy to the input layer and then calculates the loss value between the classified and label values to identify the ultrasound lesion of the prostate. The experimental results showed that the proposed method can accurately detect the ultrasound image of the prostate and segment prostate information at the pixel-level simultaneously. The proposed method has higher accuracy than that of the doctor's manual diagnosis and other detection methods. Our simple and effective approach will serve as a solid baseline and help ease future research in the computer-aided diagnosis of prostate ultrasound images. Furthermore, this work will promote the development of prostate cancer ultrasound diagnostic technology.

1. Introduction

The data that is extracted from GLOBOCAN 2018 shows that the incidence of prostate cancer is second to that of lung cancer [1], and this condition is a serious threat to man's physical and mental health. Diagnosis and comprehensive cure are the most effective means to prevent prostate cancer. Currently, ultrasound imaging technology has become an important method for the detection of prostate cancer, because it is cheap, cost-effective, and free of radiation and trauma. However, transrectal ultrasonography (TRUS) usually requires a doctor with clinical experience in diagnosis, and the doctor's diagnosis accuracy is affected by many factors. Therefore, an accurate and efficient computer-aided diagnosis (CAD) is essential to improve the efficiency of prostate-assisted therapy.

Segmentation is a crucial link in the CAD system of prostate ultrasound images. Considering that prostate ultrasound images have serious speckle noise and low signal-to-noise ratio, and most of the segmentation algorithms of prostate ultrasound images do not have pixel-level segmentation, the current prostate ultrasound image segmentation method is insufficient, and the positioning accuracy is not high. This limitation causes trouble and pressure to the doctor's judgment and work. Therefore, to obtain the final image containing only the information of the prostate region and reduce the interference of other factors, we adopted the improved S-Mask R-CNN network model to use the region of interest align (ROIAlign) algorithm [2] to segment the ultrasound image of the prostate.

Classification is another important link in the CAD system of the prostate ultrasound image. On the basis of segmentation, the

[*] Corresponding author.
E-mail address: bruda@126.com (X. Lu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.future.2020.08.015
0167-739X/© 2020 Elsevier B.V. All rights reserved.

prostate ultrasound image was classified, thus predicting whether the patient has prostate cancer through image classification. Traditional image classification methods mainly involve image preprocessing [3], image feature description and extraction [4], and classifier design. The performance difference of traditional image classification depends on the shape, texture, color, and underlying visual features of the feature extractor, classifier, and image. Various traditional classifiers can be used, including the K-nearest neighbor [5] and support vector machine [6]. However, the classification results are not satisfactory because of slight differences between the images of ultrasonic prostate cancer and serious noise interference. Accordingly, the convolutional neural network was introduced in deep learning [7–9], thus greatly improving the classification accuracy of prostate ultrasound images without requiring the manual feature description and extraction of target images. During transcrectal ultrasonography [10], the puncture place was marked and analyzed in the image. Subsequently, analysis of the images was performed with the Automated Urologic Diagnostic Expert(AUDEX) system, consisting of a personal computer connected to the ultrasound machine. Azizi et al. [11] used Temporal Enhanced Ultrasound(TeUS) to address the problem of prostate cancer in a clinical study of the biopsy cores [12]. The method involves capturing high-level latent features of TeUS with a deep learning approach followed by distribution learning to cluster. On the above research, the improved Inception-v3 [13] neural network lesion identification method was adopted in this paper, and the improved network depth enabled the convolutional neural network model to automatically extract the characteristic lesion images of prostate ultrasound images for classification. Then, the data set was constructed for training, and parameters were continuously optimized through batch gradient descent. Overfitting was eliminated using the dropout method.

Based on the above research, the present study introduced a deep learning model that integrates S-Mask R-CNN and Inception-v3 in the ultrasound image-aided diagnosis of prostate cancer. The results showed that the data set on building ultrasound images of prostate disease recognition rate increased.

A data set of prostate ultrasound images with segmentation labeling was constructed. Exactly 1200 images were randomly selected from the data of prostate ultrasound images, and a new data set was constructed for labeling. Among them, malignant and benignant images in simulated clinical medicine account for an appropriate proportion.

ROIAlign was used in the improved S-Mask R-CNN algorithm to retain the floating-point coordinates on the ultrasonic prostate images, thus boosting the image segmentation accuracy.

A deep learning model that integrates S-Mask R-CNN and Inception-v3 in the ultrasound image-aided diagnosis of prostate cancer was built, and multiple performance evaluation was carried out to identify prostate ultrasound image data sets. These processes are important for the study of prostate cancer identification.

2. Related work

Scholars have proposed the application of various ultrasonic image segmentation methods [14,15] in medicine, including level set [16], region growing [17], Markov random field [18], active control [19], and neural network [20]. Snake [21] proposed the contour algorithm composed of interactive marker control points. By matching the elastic deformation of the contour with the local features of the image to achieve balance, some energy functions of the contour were minimized to achieve the target region segmentation. However, the algorithm cannot deal with the topological change of the curve, and the target region cannot be segmented because of the complex topological structure. The uneven distribution of gray scale in the ultrasonic image renders the internal force of the noise image insufficient to make the contour converge to the global optimal value. Mclnerney [22] proposed a Snake model with adaptive topology. The algorithm uses the feature functions of triangular grids and grid points for the reference to determine boundary triangles and ensure that it can handle the contour of complex shapes. To provide an accurate prior information of the target region during the evolution of the active contour, Yuan [23] proposed an active contour model based on the local divergence similarity ratio to improve the robustness of the active contour to noise in the ultrasonic image. Considering the advantage of high quality of MRI and clear ultrasound images, a shape statistical model was constructed for the target area of MRI corresponding to the ultrasound image sequence to restrict the segmentation of the target area in the corresponding ultrasound image sequence [24]. The above research work has achieved good ultrasonic image segmentation, but these methods still encounter some problems.

The CNN segmentation method based on candidate regions is a popular research direction of target segmentation. CNN algorithms such as R-CNN [25], Fast R-CNN [26], Faster R-CNN [27], and Mask R-CNN are still being improved. These algorithms can not only achieve good segmentation effect for image segmentation, but also obtain satisfactory detection time.

In 2014, Girshick et al. [25] proposed the R-CNN algorithm, which was applied to VOC2007 data and stood out among various algorithms. Many scholars have focused on candidate region-based CNN [28,29]. However, R-CNN is limited by the double calculation in the feature extraction process, thus causing the algorithm to calculate slowly. Therefore, based on the research of R-CNN, Fast R-CNN first carries out feature extraction for the whole image. In addition, the ROIPooling [30,31] layer was introduced to unify the feature scale of the image. Then, softmax was used to replace the SVM to merge classification and border regression, thus greatly improving the detection accuracy and calculation time. Faster R-CNN improved the fast R-CNN algorithm [32–34], introduced the region proposal network (RPN) model, used nine types of anchors with different ratios of length and width to map on the feature map, and obtained the candidate regions. Candidate regions were applied to the deep network, which greatly improved the time and accuracy of the image detection.

In 2017, Kaiming et al. [35] proposed the Mask R-CNN algorithm based on Faster R-CNN. The Faster R-CNN algorithm adds an instance segmentation branch to obtain accurate pixel information through ROIAlign by using convolution network to generate the corresponding binary mask and complete the target detection and instance segmentation.

Szegedy et al. proposed Inception-v3 on the basis of Inception-v2 in 2016 [13]. Inception-v3 decomposes the convolution of any $n \times n$ into $1 \times n$ convolution followed by $n \times 1$ convolution, thus greatly reducing the number of parameters. This algorithm avoids overfitting by increasing the number of layers to enhance the nonlinear expression of the network. Different networks have different input size requirements, such as Inception-v3 for 299 × 299 pixels and ResNet for 224 × 224 pixels. Therefore, the ultrasonic images of prostate need to be down-sampled or up-sampled to the appropriate network input size.

3. Image segmentation

This paper adopts a method of prostate ultrasound image segmentation based on the improved Mask R-CNN, including the following steps: First, the improved S-Mask R-CNN network model was established. Then, the prostate ultrasound images to

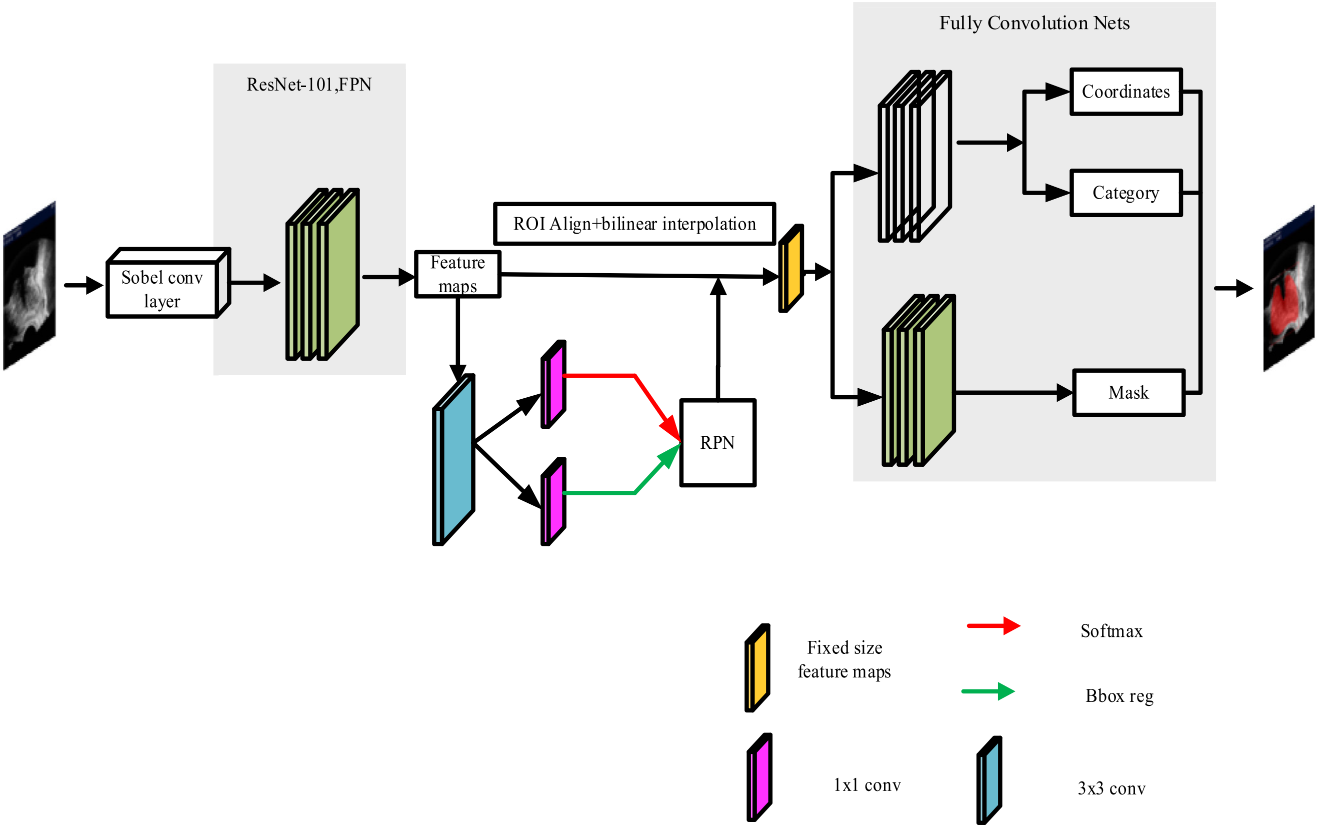


Fig. 1. Network structure of image segmentation.

be segmented was used as input into the network for segmentation. Finally, the ultrasound images of prostate after segmentation was used as output.

The image segmentation algorithm is mainly composed of three modules, namely, feature extraction, detection, and segmentation modules. The feature extraction module realizes the feature extraction of the image, the detection module realizes the positioning and classification of the objects in the image, and the segmentation module completes the instance segmentation by generating a binary mask through full convolution networks.

Fig. 1 shows the schematic diagram of the segmentation method in this paper. First, the prostate ultrasound image was used as input into the established model. Then, the RPN was used to quickly generate the candidate region box in the feature map, and the fixed size feature image was used as output through the ROIAlign. Then, the target box was classified and positioned in the detection module, the foreground and background of the prostate ultrasound image were predicted using the convolutional neural network in the segmentation module, the corresponding binary mask was painted to complete the case segmentation, and the prediction image of the prostate ultrasound image was used as output.

3.1. Sobel convolution

The Sobel operator [36] is commonly used in image sharpening. It is a first-order difference operator, which uses Gaussian smoothing and differential differentiation to detect the edge by calculating the difference between the gray values of adjacent pixels. As the first convolutional layer of the network, the pixels in the edge area of the prostate ultrasound images are enhanced. The algorithm first performs neighborhood weighted average and first-order differentiation to detect the edge. The algorithm uses two matrices to convolve the original image and thus calculate the estimated values of the gray difference partial derivatives in the X and Y direction. The two matrices of the Sober operator are as follows:

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{1}$$

where G_x and G_y are the gradients in the X and Y directions, respectively, after calculation by Sobel operator. The calculation of G_x and G_y are as follows:

$$\begin{aligned} Gx = {} & [f(x-1,y-1) + 2 \times f(x-1,y) + f(x-1,y+1)] \\ & - [f(x+1,y-1) + 2 \times f(x+1,y) + f(x+1,y+1)] \end{aligned} \tag{2}$$

$$\begin{aligned} Gy = {} & [f(x-1,y-1) + 2 \times f(x,y-1) + f(x+1,y-1)] \\ & - [f(x-1,y+1) + 2 \times f(x,y+1) + f(x+1,y+1)] \end{aligned} \tag{3}$$

Note that the Sobel operator is used to perform the convolution operation on the original figure for step size 1, followed by edge detection. High-pass filtering is enhanced, and the effect is shown in Fig. 2. After the processing of Sobel convolution kernel, the contrast between the target and the background becomes clearer. In the training phase, the Sobel convolutional layer is not trained, because its parameter variations will change the sharpening effect, resulting in uncontrollable results, considering the small size and variable shape of the prostate ultrasound images.

Considering that the surrounding environment is also strengthened, resulting in noise interference, threshold segmentation needs to be processed. This process is based on the difference in the gray scale between the target and the background to be extracted from the image, and it classifies the pixels into several categories according to the set threshold to separate the target and the background. By judging whether the feature

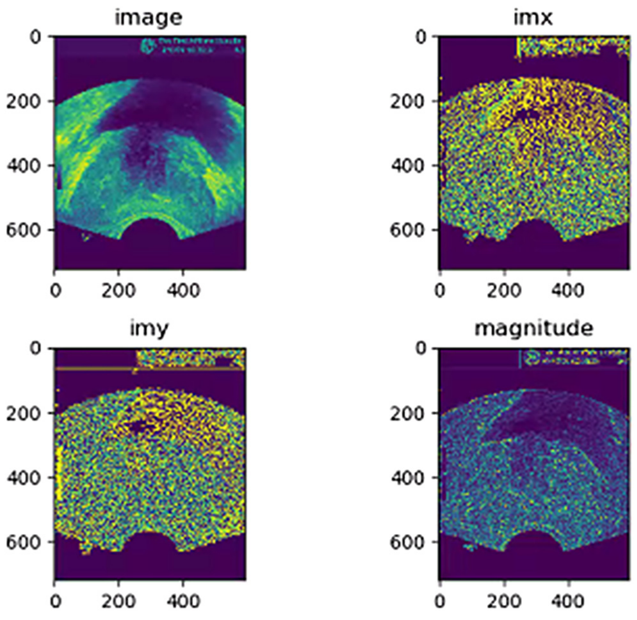


Fig. 2. Experimental result of Sobel.

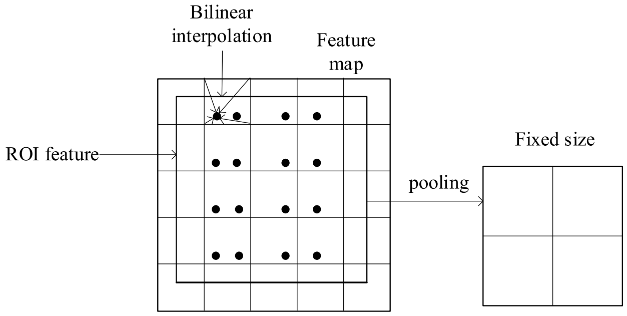


Fig. 3. Schematic diagram of RoIAlign algorithm.

attributes of each pixel in the image meet the requirements of the threshold value, it can determine whether the pixel in the image belongs to the target or background region to strengthen the target region and filter the noise.

3.2. RPN network

RPN can predict the image foreground and background through bounding boxes with different multiples and proportions of length and width, place the image box in a network, and then delimit the bounding box in the predicted feature image to quickly generate candidate regions. RPN uses the method of nine bounding boxes with three multiples of the reference bounding boxes to delimit the feature map. If the size of the reference is 32 pixels, the three bounding boxes represent three bounding boxes with length-to-width ratios of 1:1,1:2, and 2:1, and corresponding bounding boxes representing the size of 16 and 64 pixels are generated. Similarly, each of them has three anchors with a length to width ratio of 1:1, 1:2, and 2:1. RPN uses the above three multiples and three ratios of a total of nine scale anchors to perform sliding anchors on the feature map. An intersection over union (IOU) value greater than 0.5 indicates the foreground region. Meanwhile, it is subject to linear regression. The calculation formula for IOU is as follows:

$$IOU = \frac{G \cap D}{G \cup D} \tag{4}$$

When the candidate anchor generated by the RPN network and the correct target anchor in the training set. Where G represent the real region and D represents the segmentation network result; otherwise, it is denoted as a prediction error.

3.3. ROIAlign layer

For small target detection and instance segmentation of prostate ultrasound images, ROIpooling cannot meet the requirements of accurate feature point positioning. Therefore, the bilinear interpolation algorithm was adopted to replace the quantization process on the ROI feature graph generated on the ROIAlign layer, the floating point coordinates were retained, the quantization error was reduced, and the accurate mapping between the original and feature image pixels was realized. The formula for the bilinear interpolation algorithm is as follows:

First, for linear interpolation in the X direction,

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \tag{5}$$

where $R_1(x, y_1)$ and

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \tag{6}$$

where $R_2(x, y_2)$

Then, for the linear interpolation in the Y direction,

$$f(P) = f(x, y) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \tag{7}$$

where $f(x, y)$ is the pixel value of the point P to be solved, $f(Q_{11})$, $f(Q_{12})$, $f(Q_{21})$, and $f(Q_{22})$ are four known points, $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_1, y_2)$, and $Q_{22} = (x_2, y_2)$ are the pixel values, and $f\ (R_1)$ and $f(R_2)$ are the pixel values obtained by X interpolation.

The ROIAlign layer realizes the pooling operation for the generated candidate ROI, pools the feature maps of ultrasound images of prostate with different sizes through the ROIAlign layer, and maps them into fixed-size feature maps. Fig. 3 show the schematic diagram of the processing with the ROIAlign algorithm.

3.4. Loss function

In this paper, the S-Mask R-CNN model completes the detection and positioning of the prostate ultrasound image frame. It realizes the classification and segmentation of the prostate area and background. The loss function is composed of three parts and is defined as follows:

$$L_{loss} = L^*_{class} + L^*_{box} + L^*_{mask} \tag{8}$$

where L^*_{class} is the loss of classification, L^*_{box} is the loss of the positioning, and L^*_{mask} is the loss of segmentation.

$$L^*_{class} = L(\{p_i\}) \frac{1}{N_{class}} \sum L_{class}\left(p_i, p_i^*\right) \tag{9}$$

N_{class} is the number classified samples, i is the anchor, p_i is the probability, which is predicted to be the target for the anchor, $p_i^* = 1$ is the foreground, and $p_i^* = 0$ is the background.

The L^*_{box} formula is as follows:

$$L^*_{box} = L(\{t_i\}) = \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}\left(t_i, t_i^*\right) \tag{10}$$

where i is the corresponding to the anchor, N_{reg} is the number of anchors in the regression sample, t_i is the prediction position parameter of RPN, t_i^* is the real position parameter of RPN, L_{reg} is the regression loss of orientation, and $p_i^* L_{reg}$ means that when

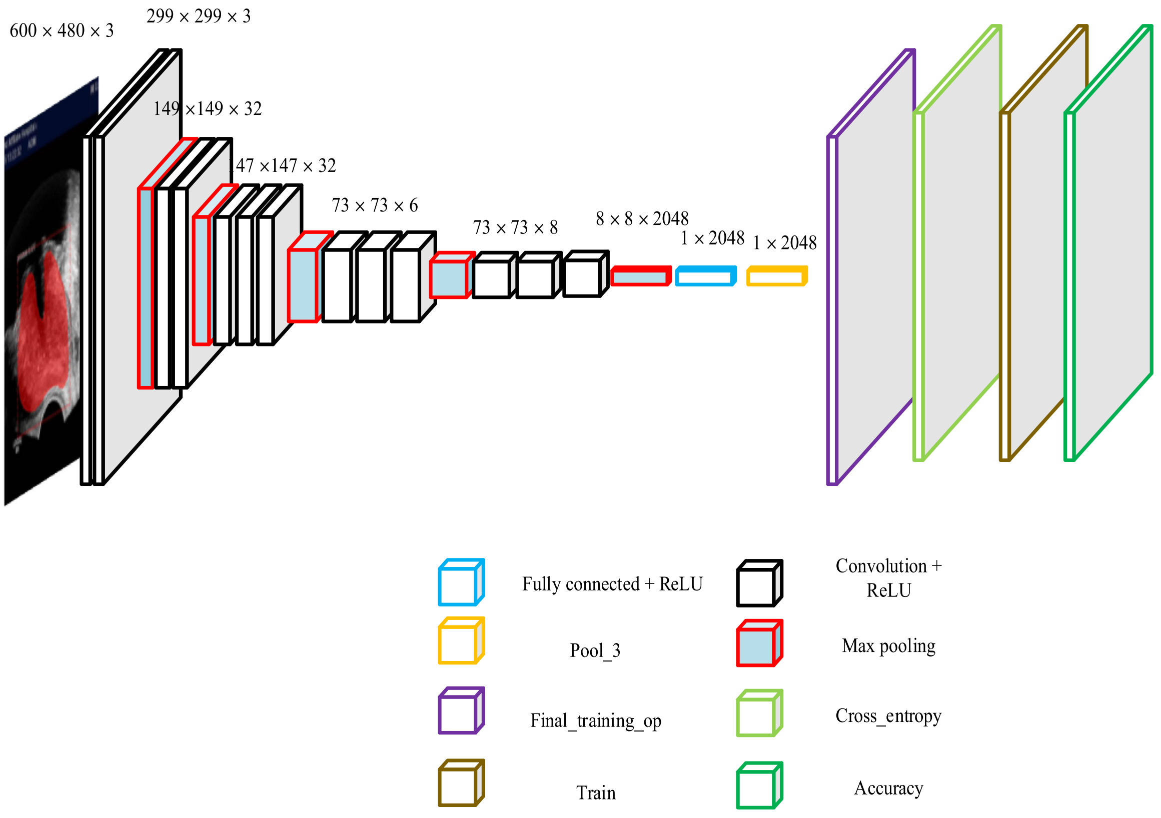


Fig. 4. Network structure of image classification.

the anchor is the foreground, the regression loss is calculated, the formula is:

$$L_{reg}\left(t_i, t_i^*\right) = smooth_{L_1}(t_i - t_i^*) \tag{11}$$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases} \tag{12}$$

The disadvantage of the L_1 loss function is that it has a fold point, which is not smooth and leading to instability. The purpose of constructing smooth function is to make loss function more robust to outliers, and the magnitude of gradient can be controlled during training.

The segmentation loss is the loss of a binary mask. When the candidate RPN detects one of the categories, the cross-entropy of this category is used as the error value for calculation. The loss of other categories are not included, and the formula is as follows:

$$L_{mask}^{*} = -\frac{1}{m^2}\sum_{1\leq i,j\leq m}\left[y_{ij}\log\hat{y}_{ij}^{k} + \left(1-y_{ij}\right)\log\left(1-\hat{y}_{ij}^{k}\right)\right] \tag{13}$$

where y_{ij} is the tag value of the coordinate point (i, j) in the $m \times m$ region, and $\hat{y}_{ij}^{k}$ is the predicted value of the kth class at this point. The formula k is the natural number less than or equal to 2.

3.5. Full convolutional network

Both full convolutional network (FCN) and traditional CNN contain convolutional and pooling layers. In this study, we used FCN to deconvolve the prostate ultrasound image of a convolutional layer at the end and carried out up-sampling to ensure that the size of the output image is consistent with the size of the input image. Finally, the line-by-line pixel was predicted using softmax classifier, and the category of each pixel was distinguished.

Table 1
Confusion matrix of the classification result.

Actual label	Predicted label	
	Malignant	Benignant
Malignant	TP	FN
Benignant	FP	TN

4. Image classification

4.1. Inception module

Inception-v3 belongs to a classification model of convolutional neural network, and it automatically converts the image size to 299 × 299. In comparison with the AlexNet in this network model, all connections from the AlexNet layer, instead of average for pooling, remarkably reduce the number of network model parameters, and asymmetric convolution kernels are used to increase the diversity at the same time.

4.2. Cross-entropy cost function

In the model training, the adjustment weight bias is required, and the cross-entropy cost function formula is as follows:

$$C = -\frac{1}{n} \times \sum_{x}[Ina + (1-y)In(1-a)] \tag{14}$$

where C is the cross-entropy cost function, x is the image sample, y is the actual value, and n is the output value.

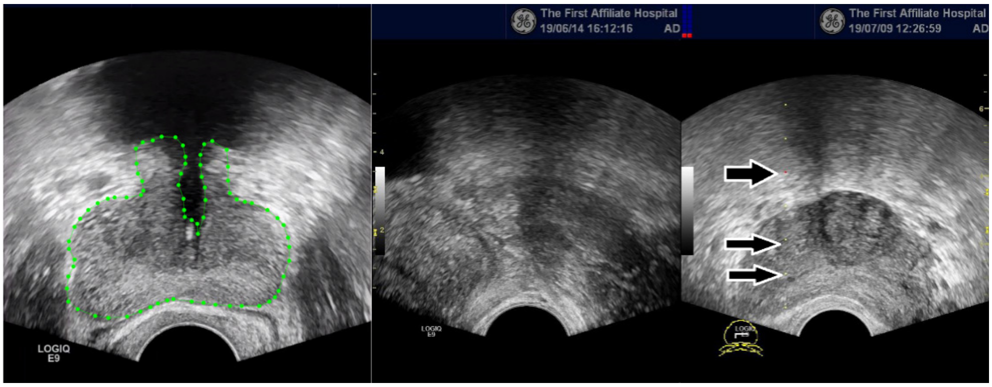


Fig. 5. Prostate ultrasound puncture and prostate area labeling.

The advantage of using cross-entropy is that the larger the error, the greater the magnitude of weight and bias adjustment; the smaller the error, the smaller the magnitude of weight and bias adjustment; In general, the smaller the cross-entropy value, the better performance of the constructed.

4.3. Dropout

Overfitting will affect the judgment of the model on the new data. Overfitting can be avoided using the method of Dropout to temporarily remove the nodes and to reduce the complexity of the model.

4.4. Network model training

In this paper, we used Inception-v3 to retain the matrix weight, bias, regular item coefficients, and other parameters of the pool-3 layer and its following layers and replaced the softmax layer of network classification. The following layer of pool-3 is the feature extraction part of the model. The original model uses a large amount of data training of ImageNet to greatly improve the feature extraction ability of the model. Considering the general characteristics of the underlying details of the image, the feature extraction ability was retained and was applied to the data set of the self-built prostate ultrasound image to obtain satisfactory results. The input and pool-3 layer play a very important role in image classification. After a series of feature extraction, the images in the data set were converted into TXT files, where abstract features were stored. The model structure after image replacement is shown in Fig. 4.

The image classification algorithm adopts the gradient descent method with a fast convergence rate in the experiment. The main purpose of the gradient descent algorithm is to enable the network to learn the optimal weight and bias, so as to make the cross-entropy cost function C as small as possible. Calculus was not used to solve the extreme value, because the CNN network has many variables, and the calculation is complex and difficult to achieve. Accordingly, the partial derivative of the gradient can be used.

The gradient descent vector is expressed in terms of ∇C, and the formula is as follows:

$$\nabla C = (\frac{\partial C}{\partial \omega_1}, \frac{\partial C}{\partial \omega_2}, \frac{\partial C}{\partial \omega_3}, \ldots, \frac{\partial C}{\partial \omega_n})^T \tag{15}$$

$$\nabla = \frac{df(\omega)}{d\omega}, \nabla C = \nabla C \cdot \nabla \omega \tag{16}$$

In the training of construction of the convolutional neural network of the data set, the learning rate is set to be constant at 0.01, and the updated formula of weight is as follows:

$$\omega_{i+1} = \omega_i - \eta \cdot \nabla f\ (\omega_i) \tag{17}$$

where ω_i is an immediate parameter, ω_{i+1} is the updated parameter, and η is the learning rate.

4.5. Classification result evaluation indicators

The confusion matrix of the classification result is as shown in Table 1. TP represents that the image classification model is judged to be malignant and the actual pathological results are malignant; FP represented that image classification model is judged to be malignant and the actual pathological results are benign; FN represents that the image classification model is judged to be benign and the actual pathological results are malignant. The precision and recall can be calculated as follows:

$$Precision = P = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = R = \frac{TP}{TP + FN} \tag{19}$$

F1 is based on the harmonic mean of precision and recall, and the formula is as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{20}$$

The researchers aimed to comprehensively investigate the precision, recall, and harmonic mean. The distribution calculates P and R for each category, and does arithmetic averaging to obtain macro-P and macro-R. Then, macro-F1 is obtained by calculating macro-P and macro-F1.

$$macro - P = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{21}$$

$$macro - R = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{22}$$

$$macro - F1 = \frac{2 \times macro - P \times macro - R}{macro - P + macro - R} \tag{23}$$

5. Data set and experimental environment construction

5.1. Data set

Prostate ultrasound images were collected from the First Affiliated Hospital of JiNan University and the Third Affiliated Hospital

Table 2
Experimental environment configuration.

OS	Ubuntu 16.04
CPU	Xeon E5-2690 V4 @2.20 GHz
GPU	NVIDIA GeForce GTX 1080Ti
Deep learning framework	Tensorflow 1.8.0
Programming language	Python 3.7.0

Table 3
Performance comparison of classical segmentation methods.

Method	mAP	DICE	IOU	AP	Time (s)
FCN	0.84	0.84	0.75	0.90	0.283
Our method	0.88	0.87	0.79	0.92	0.348

Table 4
Classification performance contrast.

Method	Category	Precision	Recall	F1-score	Support
Inceptionv3	Malignant	0.76	0.53	0.62	66
	Benignant	0.75	0.89	0.81	103
Xception	Malignant	0.83	0.30	0.44	66
	Benignant	0.68	0.96	0.80	103
ResNetV2	Malignant	0.64	0.35	0.45	66
	Benignant	0.68	0.87	0.76	103
ResNet-50	Malignant	0.65	0.61	0.62	66
	Benignant	0.76	0.79	0.77	103
Doctor	Malignant	0.32	0.42	0.36	66
	Benignant	0.70	0.62	0.66	103

of Sun Yat-sen University. Considering that the puncture guidance line of the ultrasound image of prostate puncture will affect the result of image classification, all images are matched in the folder named with the patient number of each patient, and the image before the puncture is backed up (Fig. 5). The random data set can be divided into three categories as follows: 60% of the training data set (including 422 images), 20% of the validation data set (including 140 images). The test data set was used to test the network model of the final effectiveness (the three dataset images shown in the prostate tissue pathology results consist of the appropriate proportion of non-cancer and cancer). Labelme software [37] was used by three professionals with more than 5 years of experience to delineate the boundaries of ultrasound image of prostate tissue. The primary and secondary pathological results of the puncture tissues were given Gleason [38] scores respectively. At last, the corresponding pathological result report was presented for each tissue submitted for examination. Patients with one or more malignant tissues were defined as malignant patients. Patients without any malignant tissue were defined as benignant.

5.2. Parameter settings

Experimental environment configuration is as shown in Table 2. In this paper, an improved S-Mask R-CNN network was used as the extractor of prostate region segmentation, and the CoCo pre-training model [39] was used as the generalization of parameters. An improved Inception-v3 network was used to classify positive and negative prostate images by using the Inception-v3 pre-training model. The improved S-Mask R-CNN+Inception-v3 model has certain feature extraction ability, thus reducing the training time. According to the training data sets with good models [35], the number of iterations was set to 100, the iteration number is 3000, the vector is 0.001, and the weight decay rate is 0.0001.

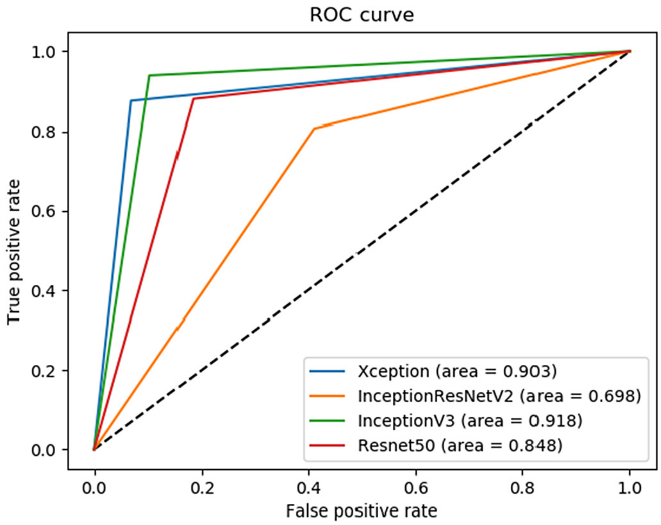


Fig. 6. Classification ROC curve.

6. Experimental results

6.1. Detection analysis

The time test experiment was verified on the prostate ultrasound data set. The 100 images were randomly selected from the above data for testing, and then the average time consumption was compared on FCN and S-Mask R-CNN segmentation algorithm respectively. The experimental results are shown in Table 3.

In comparison with the FCN algorithm, considering that the algorithm proposed in this paper has a high degree of complexity in the calculation of ROIAlign layer plus the addition of segmentation branches to the model, more time was consumed. This algorithm realizes the segmentation of prostate region information from the background and improves the detection and segmentation process.

To effectively test the performance of the classification model, we used the original data set provided by the hospital as the training set in the comparison test and then evaluated the performance of the model. The experiment compares the method in this paper with Xception [40] and InceptionResNetV2 on the original dataset, and the performance curve is shown in Fig. 6.

The X-coordinate represents the false positive rate, while the Y-coordinate represents the true positive rate. The closer the ROC [41] curve is to the point(0,1), the better the deviation is from the 45 degree diagonal, and the better the sensitivity and specificity. The area under the ROC curve is called AUC, AUC = [0.85,1] has good effect, AUC = [0.70,0,85] has average effect, and AUC = [0.50,0.70] has low effect. Fig. 6 shows that the number of false checks in this method is significantly less, and the accuracy is higher than that of other algorithms.

All the pathological results of the doctors were performed by three pathologists with more than five years of experience. Gleason score was obtained for the primary and secondary pathological results of the tissue to be examined, and then the total Gleason score was added up [38]. Finally, the corresponding pathological results were presented for each tissue to be examined. Table 4 shows the accuracy of prostate ultrasound image classification with the following order: Inception-v3 > Xception > ResNet-50 [42] > InceptionResNetV2 [43] >senior physicians.

Table 5
Classification results of different data set.

Method	Category	Precision	Recall	F1-score	Support
Original image+Inception-v3	Malignant	0.76	0.53	0.62	66
	Benignant	0.75	0.89	0.81	103
ROI+Inception-v3	Malignant	0.71	0.62	0.66	66
	Benignant	0.77	0.83	0.80	103
S-Mask R-CNN+Inception-v3	Malignant	0.80	0.55	0.65	66
	Benignant	0.76	0.91	0.83	103
S-Mask R-CNN+Xception	Malignant	0.59	0.62	0.61	66
	Benignant	0.75	0.73	0.74	103
S-Mask R-CNN+ResNetV2	Malignant	0.51	0.45	0.48	66
	Benignant	0.67	0.72	0.69	103
S-Mask R-CNN+ResNet-50	Malignant	0.68	0.64	0.66	66
	Benignant	0.78	0.81	0.79	103

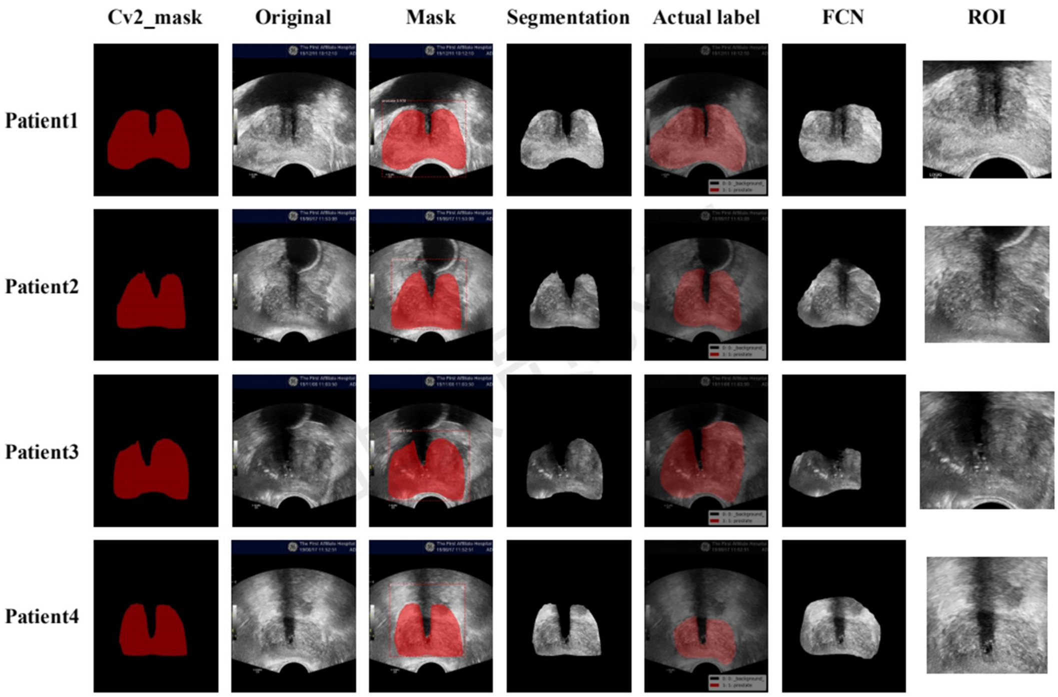


Fig. 7. Sample of test image segmentation results.

6.2. Building the effectiveness of the different data set analysis

In the classification, the performance comparison test results of different data sets show that the method used in this paper has a better classification effect than the original image+Inception-v3. The method was validated in the S-Mask R-CNN segmentation after constructing the validity of the data set.

Table 5 shows the accuracy of prostate ultrasound image classification with the following order: the network model of this paper > Bounding-box+Inception-v3 > original image+Inception-V3 > senior physicians

6.3. Results analysis

Based on the trained model, the data set of prostate ultrasound images was constructed for testing. The sample image segmentation results are shown in Fig. 7. The cv2_mask is the binary image after segmentation and the mask is shown in color, bounding box, category and confidences are also shown. In comparison with the FCN algorithm, for the algorithm in this paper, while completing the target positioning of prostate ultrasound images, the region of the prostate ultrasound images and the background are separated by the red binary mask to complete the detection and segmentation of the prostate region.

7. Conclusion

A deep learning model that integrates S-Mask R-CNN and Inception-v3 in ultrasound image-aided diagnosis of prostate cancer was proposed in this paper. A set of ultrasonic images of the prostate with segmentation and classification were constructed, and the model of this paper was trained on the set. This network model used the ROIAlign algorithm to achieve accurate positioning of feature points in prostate ultrasound images. The combined ResNet-101 and RPN networks improved the segmentation accuracy, and the corresponding binary mask of prostate ultrasound images was generated through FCN to achieve the segmentation effect of the prostate and background regions. The original model classification softmax layer was replaced with an improved Inception-v3 classification network model. Based on the comparison test of each classical network model, the method used in this paper can improve the detection without significantly increasing the complexity of calculation and model. The next step is to expand the current data set and use cubic linear interpolation algorithm instead of bilinear interpolation algorithm to improve the segmentation accuracy and thus improve the effect of model detection. Based on existing research, the model was optimized and applied to the 3D field.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (61802073); Scientific and Technological Planning Project of Guangzhou, China (201903010041); Medical Scientific Research Foundation of Guangdong Province, China (B2019023); Project Supported by Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2019).

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 68 (6) (2018) 394–424.

[2] K. He, G. Gkioxari, P. Dollar, et al., Mask R-CNN, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2961–2969.

[3] S. Bhattacharyya, A brief survey of color image preprocessing and segmentation techniques, J. Pattern Recognit. Res. 1 (1) (2011) 120–129.

[4] M.A. Vegarodriguez, Review feature extraction and image processing, Comput. J. 44 (2) (2004) 595–599.

[5] G. Amato, F. Falchi, Local feature based image similarity functions for KNN classification, in: Proceedings of the International Conference on Agents Ant Artificial Intelligence, Volume Artificial Intelligence, DBLP, Rome, Italy, 2011, pp. 157–166.

[6] T. Joachims, Making large-scale SVM learning practical, Adv. Kernel Methods Support Vector Learn. 8 (3) (2006) 499–526.

[7] L.O. Chua, T. Roska, CNN pardigm, IEEE Trans. Circuits Syst. Fundam. Theory Appl. 40 (3) (1993) 147–156.

[8] M. Matsugu, K. Mori, Y. Mitari, et al., Subject independent facial expression recognition with grobust face detection using a convolution network, Neural Netw. 16 (5–6) (2003) 555–559.

[9] P. Moeskops, M.A. Viergever, A.M. Mendrik, et al., Automatic segmentation of MR brain images with a convolutional neural network, IEEE Trans. Med. Imaging 35 (5) (2016) 1252–1261.

[10] J.J. Rosette, Computerized analysis of transcrectal ultrasongraphy images in the detection of prostate carcinoma, 1995.

[11] S. Azizi, S. Bayat, P. Yan, A. Tahmasebi, G. Nir, Detection and grading of prostate cancer using temporal enhanced ultrasound: combing deep neural networks and tissue mimicking simulations, Int. J. Comput. Assist. Radiol. Surg. (2017).

[12] S. Azizi, F. Imani, S. Ghavidel, A. Tahmasebi, B. Wood, P. Mousavi, P. Abolmaesumi, Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study, Int. J. Comput. Assist. Radiol. Surg. 11 (6) (2016) 947–956.

[13] S. Christian, Rethinking the inception architecture for computer vision(arXiv), 2015, 1512.00567v3 [cs.CV].

[14] J.A. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, IEEE Trans. Med. Imaging 25 (2006) 987–1010.

[15] K. Saini, M.L. Dewal, M. Rohit, Ultrasound imaging and image segmentation in the area of ultrasound: a review, Int. J. Adv. Sci. Technol. 11 (2010) 41–60.

[16] B. Liu, J. Huang, X. Tang, J. Liu, Y. Zhang, Combing global probability density difference and local gray level fitting for ultrasound image segmentation, Acta Automat. Sinica 36 (7) (2010) 951–959.

[17] K. Drukker, M.L. Giger, K. Horsch, M.A. Kupinski, C.J. Vyborny, E.B. Mendelson, Computerized lesion detection on breast ultrasound, Med. Phys. 29 (7) (2002) 1438–1446.

[18] D. Boukerroui, A. Baskurt, J.A. Noble, O. Basset, Segmentation of ultrasound images-multiresolution 2D and 3D algorithm based on global and local statistics, Pattern Recognit. Lett. 24 (4–5) (2003) 779–790.

[19] D.R. Chen, R.F. Chang, W.J. Wu, W.K. Moon, W.L. Wu, 3-D breast ultrasound segmentation using active contour model, Ultrasound Med. Biol. 29 (7) (2003) 1017–1026.

[20] D.R. Chen, R.F. Chang, W.J. Kuo, M.C. Chen, Y.L. Huang, Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks, Ultrasound Med. Biol. 28 (10) (2002) 1301–1310.

[21] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vis. 1 (1988) 321–331.

[22] T. McInerney, D. Terzopoulos, T-snakes: topology adaptive snakes, Med. Image Anal. 4 (2000) 73–91.

[23] J. Yuan, Active contour driven by local divergence energies for ultrasound image segmentation, IET Image Process. 7 (2013) 252–259.

[24] B. Ni, F.Z. He, Z.Y. Yuan, et al., Fibroid segmentation in ultrasonic image for constructing statistical deformation model from MRI, J. Comput.-Aid Comput. Graph 25 (2013) 817–822.

[25] R. Girshick, J. Donahue, T. Darrel, et al., Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587.

[26] R. Girshick, Fast R-CNN, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1440–1448.

[27] S. Ren, K. He, R. Girshick, et al., Faster R-CNN: Towards real-time object detection with region proposal networks, in: Advances in the Neural Information Processing Systems, 2015, pp. 91–99.

[28] J.R. Uijlings, K.E. Sande, T. Gevers, A.W. Semeulders, Selective search for object recognition, IJCV (2013) 2.

[29] J. Hosang, R. Benenson, P. Dollar, B. Schiele, What makes for effective detection proposals? IEEE Trans. Pattern Anal. Mach. Intell. (2015) 2.

[30] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, in: European on Conference on Computer Vision (ECCV), 2014, 1, 2.

[31] R. Girshick, Fast R-CNN, in: International Conference on Computer Vision, 2015, pp. 1–4, 6.

[32] A. Shrivastava, A. Gupta, R. Girshick, Training region based object detectors with online hard example mining, in: Computer Vision and Pattern Recognition (CVPR), 2016, 2, 5.

[33] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Computer Vision and Pattern Recognition (CVPR), 2017, 2, 4, 5, 7.

[34] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, Speed/accuracy trade-offs for modern convolutional object detectors, in: Computer Vision and Pattern Recognition (CVPR), 2017, 2, 3, 4, 6, 7.

[35] K. He, G. Gkioxari, P. Dollar, et al., Mask R-CNN, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2961–2969.

[36] Girish N. Chaple, Rd. Daruwala, Manoj S. Gofane, Sobel operator based edge detection methods for real time uses on FPGA, in: Proceedings of the 2015 International Conference on Technologies for Sustainable Development (ICTSD), 2015, pp. 1–4.

[37] B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, LabelMe: A datebase and web-based. Tool for image annotation, Int. J. Comput. Vis. 77 (1-3) (2008).

[38] Urological surgery society of Chinese medical association, China prostate cancer alliance. Chinese experts agree on prostate puncture, Chin. J. Urol. 37 (4) (2016) 241–244.

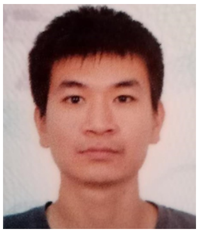
[39] T.Y. Lin, M. Maire, S. Belongie, et al., Microsoft coco: Common objects in context, in: Proceedings of the European Conference on Computer Vision, Springer, Cham, 2014, pp. 740–755.

[40] F. Chollet, Xcpetion: deep learning with depthwise separable convolutions, in: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Honolulu, HI, USA, 2017.

[41] CE. Metz, ROC methodology in radiologic imaging, Invest Radiol. 21 (9) (1986) 720–733.

[42] K.M. He, X.Y. Zhang, S.Q. Ren, et al., Deep residual learning for image recognition, in: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Las Vegas, NV, USA, 2016.

[43] C. Szegedy, S. loffe, V. Vanhoucke, et al., Inception-ResNet and the Impact of Residual Connections on Learning. arXiv preprint arXiv:1602.07261.



Zhiyong Liu is currently a postgraduate in the School of Computer Science, Guangdong Polytechnic Normal University, China. His research interests include medical image processing and deep learning.

Chuan Yang received her M.S. degree in Human Anatomy and Histoembryology from Sun yat-sen university, China in 2014. She is currently a attending physician in the Department of Ultrasound, the First Affiliated Hospital of Jinan University, China. Her research interests include ultrasonographic image and medical image analysis.

Jun Huang is currently a chief physician in the Department of Ultrasonography, the First Affiliated Hospital of Jinan University, China. Her research interests include ultrasonographic image and medical image analysis.

Shaopeng Liu received the Ph.D. degree in Computer Software and Theory from Sun Yat-Sen University, Guangdong, China, in 2013. Currently, he is a Lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, China. His research interests include machine learning and medical image analysis.

Yumin Zhuo is currently a chief physician in the Department of Urology, the First Affiliated Hospital of Jinan University, China. His research interests include urinary surgery and prostate disease.

Xu Lu received his M.S. and Ph.D. degrees in Control Theory and Control Engineering from Guangdong University of Technology, China in 2009 and 2015, respectively. He is currently an associate professor in the School of Computer Science, Guangdong Polytechnic Normal University, China. His research interests include medical image processing and artificial intelligence.