Utah State University

DigitalCommons@USU

8-2020

Visual Saliency Estimation and Its Applications

Fei Xu
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Computer Sciences Commons

Visual Saliency Estimation and Its Applications

by

Fei Xu

A dissertation proposal submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Computer Science

Approved:

_____   _____
Heng-Da Cheng, Ph.D.   Lie Zhu, Ph.D.
Major Professor   Committee Member

_____   _____
Haitao Wang, Ph.D.   Curtis Dyreson, Ph.D.
Committee Member   Committee Member

_____   _____
Vicki Allan, Ph.D.   Richard S. Inouye, Ph.D.
Committee Member   Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, UT

2020

ABSTRACT

Visual Saliency Estimation and Its Applications

by

Fei Xu, Doctor of Philosophy

Utah State University, 2020

Major Professor: Heng-Da Cheng,
Ph.D. Department: Computer Science

Visual Saliency Estimation (VSE) aims to imitate the human visual system to estimate the degree of human attention attracted by different image regions and locate the salient object. The current VSE approaches on natural images model generic visual stimuli based on lower-level image features, e.g., central-bias spatial distance, local/global contrast, and feature correlation, and still suffered from some drawbacks. First, these methods formulated the center-bias constraint using the image center, which fail when the objects are near the image borders. Second, many region-based methods were based on the assumption that every small region is homogeneous. Such limitation makes these methods cannot achieve good results when the images have complicated backgrounds.

This research focuses on solving these challenges by proposing a new framework with more robust task-related priors, and extend the framework into the other real problems.

The new framework formulated VSE on natural images as a quadratic program (QP) problem. It proposes an adaptive center-based bias hypothesis to replace the most common image center-based center-bias, which is much more robust when the objects are far away from the image center. Second, it models the smoothness term on saliency statistics of each

color and forces the pixels with similar colors to have similar saliency statistics. The new smoothness term is more robust than that based on region dissimilarity when the image had a complicated background or low contrast. The new approach achieves the best performance among 11 latest methods on three public natural image datasets.

Three approaches based on the framework are utilized to imitate the radiologists' attention to detect breast tumors from breast ultrasound images. The first approach proposes a novel hybrid framework for tumor saliency estimation (TSE), which is the first optimization framework to detect tumor saliency by integrating both high-level domain-knowledge and robust low-level saliency assumptions. The second approach estimates the tumor saliency via breast anatomy modeling based on Neutro-Connectedness. It proposes a new objective function to handle the images without tumors and achieved more accurate detection results. The strategy of modeling image structure based on Neutro-Connectedness theory can be applied to model tissue relationships in other medical image tasks such as thyroid tumor detections, liver tumor detections, lung tumor detections, etc. The third approach utilizes a deep neural network to generate semantic breast anatomy. A new background map generation method weighted by the semantic probability and spatial distance is proposed to improve the performance. The experimental results demonstrate the new approach obtains the best performance comparing with the latest methods on two breast ultrasound image datasets.

(105 pages)

PUBLIC ABSTRACT

Visual Saliency Estimation and Its Applications

Fei Xu

The human visual system can automatically emphasize some parts of the image and ignore the other parts when seeing an image or a scene. Visual Saliency Estimation (VSE) aims to imitate this functionality of the human visual system to estimate the degree of human attention attracted by different image regions and locate the salient object. The study of VSE will help us explore the way human visual systems extract objects from an image. It has wide applications, such as robot navigation, video surveillance, object tracking, self-driving, etc.

The current VSE approaches on natural images models generic visual stimuli based on lower-level image features, e.g., locations, local/global contrast, and feature correlation. However, existing models still suffered from some drawbacks. First, these methods fail in the cases when the objects are near the image borders. Second, due to imperfect model assumptions, many methods cannot achieve good results when the images have complicated backgrounds. In this work, I focuses on solving these challenges on the natural images by proposing a new framework with more robust task-related priors, and I apply the framework to low-quality biomedical images.

The new framework formulates VSE on natural images as a quadratic program (QP) problem. It proposes an adaptive center-based bias hypothesis to replace the most common image center-based center-bias, which is much more robust even when the objects are far away from the image center. Second, it models a new smoothness term to force similar

color having similar saliency statistics, which is more robust than that based on region dissimilarity when the image has a complicated background or low contrast. The new approach achieves the best performance among 11 latest methods on three public datasets. Three approaches based on the framework by integrating both high-level domain-knowledge and robust low-level saliency assumptions are utilized to imitate the radiologists' attention to detect breast tumors from breast ultrasound images.

ACKNOWLEDGMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Backgrounds

Visual saliency estimation (VSE) aims to model human visual mechanisms and plays an important role in automatic object detection and image segmentation [1-3]. It measures the degrees of human attention attracted by different image regions. The image center and local & global contrasts are three typical clues modeled in VSE approaches. Generally, VSE approaches can be classified into two categories based on saliency generation strategies. First, the direct mapping methods [4-9] transform image features into saliency values by using predefined maps; second, the optimization models [10-17] focus on modeling different hypotheses into one unified framework, and the saliency values are generated by optimization techniques. The approaches in the former category are simple and faster; however, they have poor performance using images with low contrast and big objects. While approaches in the latter category can achieve better performance by automatically adapting and balancing different components of the models.

In recent years, VSE for natural images analysis is one of the most popular research topics. One of the earliest saliency models proposed by Itti et al. [4] was based on the biological model proposed by Koch and Ullman [5]. It generated an early bottom-up computational framework for saliency detection. Hou et al. [6] proposed the spectral residual method to detect salient objects based on Fourier Transform. The method

calculates saliency values fast but was only sensitive to the boundary pixels of the salient objects. The phase spectrum based saliency detection was another frequency domain-related method proposed by Guo et al. [7]. To address the problems of poor borders and low-resolution map, Achanta et al. [8] proposed an efficient frequency-turned approach to generate a full resolution saliency map. However, this method still could not highlight full salient region.

In [9], Cheng et al. showed that the region-based global contrast method could achieve quite a good saliency detection. They also demonstrated that the color space smoothing on the saliency map was an effective approach to reduce saliency noise. The method not only could highlight object boundaries but also could highlight object regions. However, it failed when the objects had low contrast.

All the methods mentioned above are directly mapping models: calculating saliency values based on a predefined map: pixel or region color difference, spectral residual, phase information, and gradient. The methods used the unified models to process various images, and they failed in many cases, such as big smooth objects, low image contrast, complicated background and not centered objects. Therefore, optimization models based visual saliency estimation method have attracted more and more attention recently.

Chang et al. [10] modeled the saliency detection as a quadratic program (QP) optimization problem. They modeled three constraints in the objective function: saliency estimation term, objectness estimation term and the interaction term. The constraints were

heavily dependent on saliency and objectness priors, and the smooth term was only defined based on adjacent superpixels. Chen et al. [11] proposed an optimization method based on the low-rank matrix, which modeled the center-bias prior, semantic prior and color prior in the objective function. The method treated the low-rank problem as an approximation convex problem with equality constraints, which could only obtain local optimal. Jiang et al. [12] proposed a supervised method with regional contrast and background feature constraints. It used random forest regression to learn the saliency estimator based on richer descriptors. Kim et al. [13] proposed a supervised method based on random forest regression. They extracted high-dimensional features in color space and modeled the saliency detection as a linear program (LP) optimization problem. Li et al. [14] modeled the saliency detection as a QP problem. They modeled the region's rarity, center-bias, and regions' correlation hypotheses. The method used the image center to model the center-bias hypotheses, which failed if the salient object is close to image borders, and the method generated an inaccurate saliency map when the image has a complicated background. In [15] and [16], the saliency detection was formulated as a random walk problem in the absorbing Markov chain with boundary prior constraint.

Existing optimization-based saliency estimation methods still suffered from some drawbacks. First, these methods formulated the center-bias constraint using the image center, which made them have low performance when the objects are near the image borders. Second, many region-based methods assumed that every small region was homogeneous, and they modeled the smoothness term based on region features. Due to the

limitations, these methods cannot achieve good results when the images have complicated backgrounds.

1.2 Outline

The rest of the dissertation is organized as follows. In Chapter 2, we propose the generalized bottom-up framework for salient object detection on natural images. In Chapter 3, we apply the framework into breast ultrasound images and propose a novel hybrid tumor saliency estimation approach by integrating breast ultrasound image knowledge and low-level saliency assumptions for breast tumor detection. In Chapter 4, we estimate the tumor saliency via breast anatomy modeling based on Neutro-Connectedness. It proposes a new objective function to handle the images without tumors and achieve more accurate detection results. In Chapter 5, a new approach is discussed, which utilize a deep neural network to generate semantic breast anatomy. A new background map generation method weighted by the semantic probability and spatial distance is proposed to improve the performance. The future work is discussed in Chapter 6.

CHAPTER 2

A ROBUST VSE FRAMEWORK FOR SALIENT OBJECT DETECTION

2.1 Basic Idea

We formulate VSE as a quadratic program (QP) problem with robust constraints. We do not need to assume that the closer the image region to the image center is, the higher saliency value the region may have. Alternatively, we use the newly proposed adaptive center to model the center-bias constraint. The adaptive center is calculated based on image local contrast and can locate the object automatically. In addition, we do not need to assume that the small regions generated by the superpixels method or other segmentation methods are homogeneous. Therefore, the smoothness term is not defined on the region features but the saliency statistics (sum) of each color.

2.2 Problem formulation

We model the VSE problem as a convex optimization problem, and the optimization is to assign optimal saliency values for a set of image regions $\{R_i\}_{i=1}^{N}$. To facilitate the discussion, we define $S = (s_1, s_2, \cdots, s_n)^T$ as a vector of saliency values for N image regions, where $s_i \in [0, 1]$ denotes the saliency value of the ith image region.

The problem is formulated as

$$
\begin{aligned}
\text{minimize} \quad & E(S) = E^{data}(S) + \beta E^{smooth}(S) \\
\text{subject to} \quad & 0 \le s_i \le 1, i = 1, 2, \cdots, N \\
& \sum_{i=1}^{N} s_i = 1 \\
& B^T S = 0, B = (b_1, b_2, \cdots b_N), b_i = \{0,1\}
\end{aligned}
\tag{1}
$$

where the data term E^{data} models the region-based global contrast and the adaptive center-bias, E^{smooth} models the color-based smoothing, and β is defined to balance the influence

of the two terms; b_i is assigned to *1* if the *i*th region is adjacent to the image border, and *0* otherwise.

We utilize the method in [18] to segment the image into *N* regions. Similar to the method in [9], we extract regions' histograms in lab color space as the region features and only keep *n* bins for each histogram. In the data term, the newly proposed adaptive center-bias hypothesis and the global contrast-based hypothesis are modeled. It forces the image region with high global contrast value and/or short distance to the adaptive center to have high saliency value. In the smoothness term, we model the robust correlation hypothesis based on color saliency.

2.2.1 Data term

$$E^{data}(S) = S^T(D + \alpha C) \tag{2}$$

In Eq. (2), $E^{data}(S)$ is a linear function of *S;* the terms $D = (d_1, d_2, \cdots, d_N)^T$ and $C = (c_1, c_2, \cdots, c_N)^T$ are two N × 1 vectors denoted the global contrast and adaptive center-bias of the image regions, respectively. Large value of d_i indicates the low global contrast between the *i*th region and other image regions, and small d_i indicates the high global contrast (Eqs. (3) - (5)). The term $S^T D$ defines the cost on the global contrast, and $S^T C$ defines the cost based on the newly proposed adaptive center-bias. The relative importance between the two terms is specified by α which will be discussed in section 2.4.2.

$D = (d_1, d_2, \cdots, d_N)^T$ is the global contrast vector defined by

$$d_i = \sum_{j=1}^{N} w_{ij} \tag{3}$$

$$w_{ij} = \exp\left(-ds(H_i, H_j)\right) \tag{4}$$

where w_{ij} is the color contrast between the *i*th region and the *j*th regions, and H_i and H_j are

the color histograms of region i and j, respectively. $ds(H_i, H_j)$ defines the dissimilarity between two regions using the Bhattacharyya distance as

$$ds(H_i, H_j) = \sqrt{1 - \sum_k \sqrt{H_i(k)H_j(k)}} \tag{5}$$

where $H_i(k)$ and $H_j(k)$ are the values of the kth bins of histograms H_i and H_j, respectively.

The term C models the proposed adaptive center-bias. Most methods used the image center to calculate center-bias values and failed when the objects are far from the image center. Here, we propose a new method based on an adaptive center to compute the center-bias value. C is defined by

$$C = \exp\left(\|RC_i - AC\|_2 / d_D\right) \tag{6}$$

$$d_i = \sum_{i=1}^{M} p_i w^l(p_i) / \sum_{i=1}^{M} w^l(pi) \tag{7}$$

where RC_i is the coordinate of the ith region center, AC is the adaptive center, $\|\cdot\|_2$ is the l_2 norm, d_D is the diagonal distance of the image, $p_i = (x_i, y_i)$ is the coordinates of the ith pixel, M is the number of pixels, and $w^l(p_i)$ is the local contrast of the ith pixel. The local contrast map is calculated by using the local range filter. We first transform the color image into gray image, then the local range output $w^l(p_i)$ is computed by using $\max\left(\{g(p)|p \in NH_{3\times3}^{p_i}\}\right) - \min\left(\{g(p)|p \in NH_{3\times3}^{p_i}\}\right)$, where $g(p)$ is the gray value of the pth pixel, $NH_{3\times3}^{p_i}$ is a set that contains all the pixels (including p_i) of the 3-by-3 neighborhood around p_i.

As shown in Eq. (7), the adaptive center AC is the weighted center of the local contrast map, e.g., if the energy of the local contrast map concentrated on the left side of the image, the adaptive center will move to the left. To overcome the effect of the local contrast noise, we use the mean value of the local contrast values as a threshold (discussed in section 4.1), and the local contrast values less than the threshold will be set as 0.

2.2.2 Smoothness term

Most region-based methods [10, 11, 14] model the smoothness term based on the region-based correlation hypothesis: two similar regions should have similar saliency values if they are close to each other. The similarity between two regions is calculated by using the mean of the regions' color values. These methods can achieve good results for smooth images; but if the image regions are inhomogeneous, the mean values cannot describe the region features precisely and the performance will drop dramatically. In [9], Cheng et. cl demonstrated that replacing saliency value of each color using the weighted average saliency values of similar colors can reduce large amount of noise. Inspired by their work, we model the smoothness term on color saliency, which makes similar colors have similar saliency statistics (sum).

$$E^{smooth}(S) = (S - A \times S)^T(S - A \times S) \tag{8}$$

In Eq. (8), $E^{smooth}(S)$ is a quadratic function of S, and A is an $N \times N$ transition matrix which is proposed to smooth the saliency map.

$$A = H \times V \times H^T$$
$$H = \begin{pmatrix} H_1^T \\ H_2^T \\ \vdots \\ H_N^T \end{pmatrix}, H_i = (h_1, h_2, \cdots h_n)^T \tag{9}$$

In Eq. (9), H is an $N \times n$ matrix containing histograms of N regions; and V is an $n \times n$ matrix saving the L_2 distance between each color pairs (Lab color); only top 1/4 nearest color distances are kept, and the others are set to zero; H_i is column vector saves the values of the ith region's color histogram; the maximum number of histogram bins n is 1728 ($12 \times 12 \times 12$), only valid bins ($h_i \neq 0$) are employed in computation.

In order to smooth region saliency values based on color saliency, we introduce

the transition matrix A. $A \times S$ will output the smoothed saliency values of each region based on color saliency.

The histogram matrix H^T realizes the mapping from region saliency values (S) to color saliency values ($H^T \times S$). The value of the ith element in $H^T \times S$ is the sum of saliency values of all pixels with color value i.

The color distance matrix V works as a weighted mean filter. It replaces the saliency values of each color by the average saliency values of its m-nearest neighbors. The value of the ith element in $V \times H^T \times S$ is the smoothed saliency value of color i. $H \times V \times H^T \times S$ maps the smoothed saliency value of each color to region saliency value.

The proposed smoothness term is a QP problem with linear equality and inequality constraints. The original problem can be rewritten as follows:

$$
\begin{aligned}
\text{minimize } f_0(S) = \ & S^T(D + \alpha C) + \\
& \beta(S - A \times S)^T(S - A \times S) \\
\text{subject to } \ & 0 \leq s_i \leq 1, i = 1,2, \cdots, N \\
& \sum_{i=1}^{N} s_i = 1 \\
& B^T S = 0, B = (b_1, b_2, \cdots b_N), b_i = \{0,1\}
\end{aligned}
\tag{10}
$$

2.3 A primal-dual interior-point method for QP optimization

We formulate the visual saliency estimation as a QP problem with linear equality and inequality constraints, and the primal-dual method can be applied to optimize the problem globally. The inequality constraints can be rewritten as a set of functions:

$$
\begin{aligned}
f_k(S) = -S_k \leq 0, k = 1,2, \cdots, N \\
f_k(S) = S_{k-N} - 1 \leq 0, k = N + 1, N + 2, \cdots, 2N
\end{aligned}
\tag{11}
$$

In Eq. (11), N is the number of image regions and S_k is the saliency value of the kth region. We write all inequality constraints in a matrix (Eq. (12)).

$$f(S) = \begin{bmatrix} f_1(S) \\ f_2(S) \\ \vdots \\ f_{2N}(S) \end{bmatrix} = \begin{bmatrix} -S \\ S-1 \end{bmatrix}_{2N \times 1} \tag{12}$$

And the derivative matrix of Eq. (12) is

$$Df(S) = \begin{bmatrix} \nabla f_1(S)^T \\ \nabla f_2(S)^T \\ \vdots \\ \nabla f_{2N}(S)^T \end{bmatrix} = \begin{bmatrix} -I \\ I \end{bmatrix}_{2N \times N} \tag{13}$$

where I is the identity matrix.

In the primal-dual interior method, we update the dual residual (r_d), primal residual (r_p), and the centrality residual (r_c) in each iteration, and stop the optimization processing when the sum of the L_2 norms is less than 10^{-6}.

The dual residual is

$$\begin{aligned} r_d &= \nabla f_0(S)^T + Df(S)^T \lambda + v(1)O + v(2)B \\ &= D + \alpha C + 2\beta(I - A)^T(S - AS) + \\ &\quad \begin{bmatrix} -I \\ I \end{bmatrix}^T \lambda + v(1)O + v(2)B \end{aligned} \tag{14}$$

where vectors $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_{2N})^T$ and $v = (v_1, v_2)^T$ are the dual feasible parameters, and O is a $2N \times 1$ vector and all the values are 1s.

The primal residual is

$$r_p = \begin{bmatrix} O^T S - 1 \\ B^T S \end{bmatrix} \tag{15}$$

The centrality residual is

$$r_c = -diag(\lambda)f(S) - (1/t)O \tag{16}$$

where t is the step is size and initialized as 1.

The partial derivatives of r_d, r_p and r_c with respect to variables S, λ and v are:

$$\frac{\partial r_d}{\partial S} = 2\beta(I - A)^T(I - A),$$

$$\frac{\partial r_c}{\partial S} = diag(\lambda) \times \begin{bmatrix} -I \\ I \end{bmatrix}, \frac{\partial r_p}{\partial S} = \begin{bmatrix} O^T \\ B^T \end{bmatrix} \tag{17}$$

$$\frac{\partial r_d}{\partial \lambda} = \begin{bmatrix} -I \\ I \end{bmatrix}^T, \frac{\partial r_c}{\partial \lambda} = diag(f(s)), \frac{\partial r_p}{\partial \lambda} = 0_{2 \times 2N} \tag{18}$$

$$\frac{\partial r_d}{\partial v} = \begin{bmatrix} O & B \end{bmatrix}, \frac{\partial r_c}{\partial v} = 0_{2 \times 2N}, \frac{\partial r_p}{\partial v} = 0_{2 \times 2} \tag{19}$$

In each iteration, we get the Newton step $(\Delta S, \Delta\lambda, \Delta v)$ by solving Eq. (20) using the partial derivatives in Eqs. (17) - (19).

$$\begin{bmatrix} \frac{\partial r_d}{\partial S} & \frac{\partial r_d}{\partial \lambda} & \frac{\partial r_d}{\partial v} \\ \frac{\partial r_c}{\partial S} & \frac{\partial r_c}{\partial \lambda} & \frac{\partial r_c}{\partial v} \\ \frac{\partial r_d}{\partial v} & \frac{\partial r_p}{\partial \lambda} & \frac{\partial r_p}{\partial v} \end{bmatrix} \begin{bmatrix} \Delta S \\ \Delta\lambda \\ \Delta v \end{bmatrix} = - \begin{bmatrix} r_d \\ r_c \\ r_p \end{bmatrix} \tag{20}$$

The variables S, λ and v are updated using the following equations.

$$S^{k+1} = S^k + t^k \times \Delta S, \lambda' = \lambda + t^k \times \Delta\lambda,$$
$$v' = v + t^k \times \Delta v \tag{21}$$

In Eq. (21), t^k is the step size and updated by using the line search method in each iteration; t^0 and S0 are initialized as 1 and $(1/n)(1,1,\cdots,1)^T$, respectively.

In order to deal with the salient regions connecting to image border, the saliency value of each boundary region is replaced by the average saliency value of its three most similar regions (Eq. (5)).

2.4 Experimental results

2.4.1 Evaluation metrics

All experiments are performed by using Matlab (R2014a, MathWorks Inc., MA) on a Windows-based PC equipped with a dual-core (3.6 GHz) processor and 8 GB memory.

The precision-recall (P-R) curve, F-measure and mean absolute error (MAE) are employed to evaluate the overall performance of saliency detection method. The precision and recall ratios are defined as following:

$$Precison = \frac{|SM \cap GT|}{|SM|}, Recall = \frac{|SM \cap GT|}{|GT|}$$

where SM denotes the binary saliency map, while GT is the ground truth binary map, and $|SM|$ denotes the white pixel number of the saliency map. The P-R curve shows the mean precision and recall rate of all saliency maps on a dataset. For each method, the P-R curve is calculated by segmenting the saliency map with threshold range from 0 to 255, and computing the precision and recall rates by comparing the thresholding result with the ground truth. To obtain the average precision and recall rates, it uses an adaptive thresholding method [8], which chooses two times the mean saliency value as the threshold. The F-measure [9] and MAE [19] are defined as

$$F_\gamma = \frac{(1 + \gamma^2)Precison \cdot Recall}{\gamma^2 \cdot Precison + Recall}$$

$$MAE = \sum_{i=1}^{M} |S(p_i) - G(p_i)|$$

where γ^2 is set to 0.3 as suggested in published saliency detection methods, p_i is the coordinate of the ith image pixel, $S(p_i)$ is the saliency value of the ith pixel, and G is the binary ground truth.

In this section, we validate the performance of the newly proposed method on three public datasets: ASD [8], SED1 [20], ECSSD [21]. The ASD dataset contains 1000 images selected from MSRA [37] dataset, and the manually labeled boundaries of salient objects are used for the ground truth; the SED1 dataset has 100 images and pixel-wise masks; and

each image in SED1 only has one object; the ECSSD dataset has 1000 images and their corresponding users labeled ground truths.

We compare the proposed method with the most recent 10 state-of-the-art methods, including LC [22], spectrum residual (SR) [6], phase spectrum (PS) [7], frequency-tuned (FT) [8], histogram-based contrast (HC) [9], region-based contrast (RC) [9], fusing generic objectness and visual saliency (SVO) [10], context-aware (CA) [23], low rank (LR) [11], estimating visual saliency (EST) [14]. We use the corresponding original implementations for all methods. We implemented the method PS [7] because we could not obtain the original implementation.

In addition, we validate the effectiveness of the newly proposed adaptive center-based constraint and the color saliency statistics-based smooth term using the datasets.

2.4.2 Parameter tuning

On SED1 dataset, we evaluate the performance of the proposed method with five thresholds: 0.3, 0.5, mean value, median value, and the tenth maximum value. The
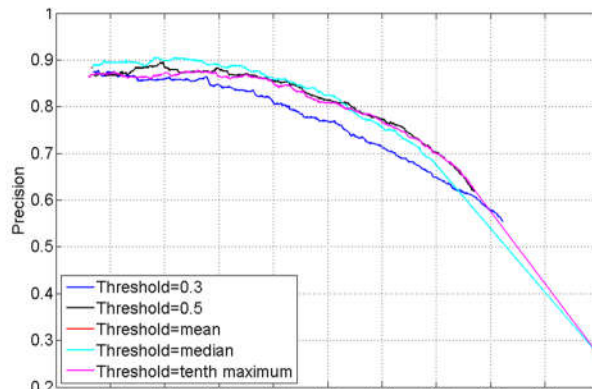
Fig. 2.1. Local contrast threshold tuning: α and β are set to $2\sqrt{N}$ and $2N$, respectively.
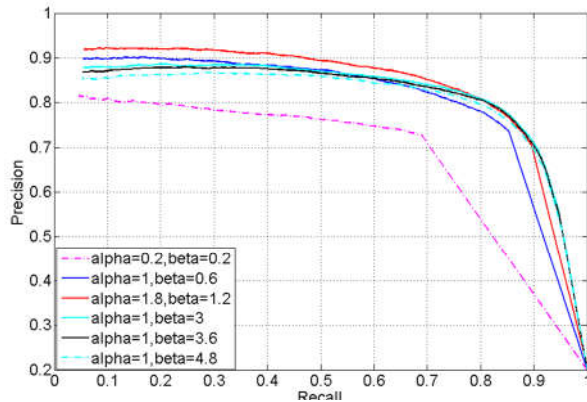
Fig. 2.2. Parameters tuning with different α and β.

proposed method can generate better results if we use the mean or median value as the threshold (see Fig. 2.1). Therefore, we set the mean value of the local contrast map as the threshold in all our experiments. Parameter α is used to balance the influence of the adaptive center-based term $(S^T D)$ and the global contrast based term$(S^T C)$; parameter β is applied to balance the influence of data term and the smoothness term. On ASD dataset, we evaluate the performance of the proposed method with α and β range from 0 to 10N. The step size is $0.2\sqrt{N}$, and the step size of β is 0.4N. We obtain better P-R curve when α is equal to $1.8\sqrt{N}$ and β is equal to 1.2N (see Fig. 2.2), and we will choose the two values in all experiments.

2.4.3 Example results

We compare the newly proposed method with the latest five state-of-the-art methods (RC, SVO, CA, LR, and EST) on several sample images from the dataset, and the overall performance comparisons are shown in the section 2.4.6.

| Original | RC[9] | SVO[10] | LR[11] | CA[23] | EST[14] | Ours | GT |
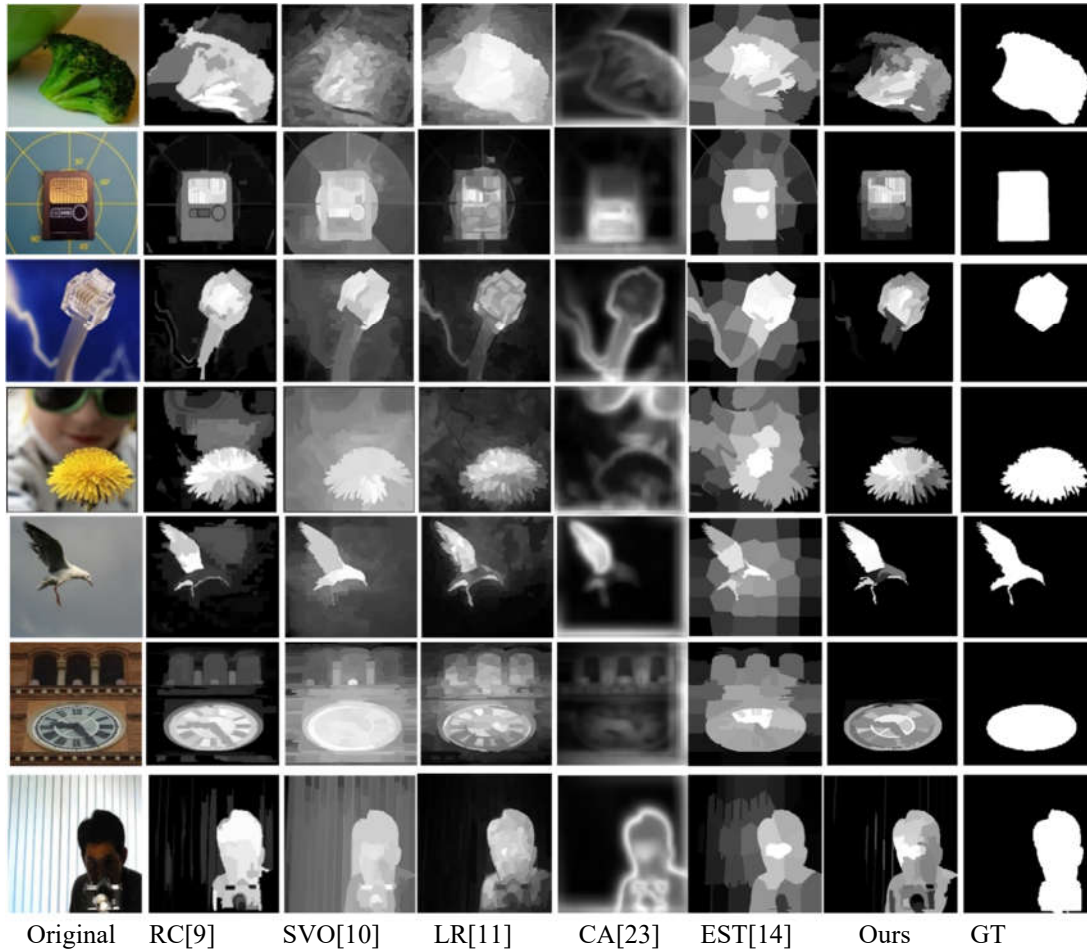
Fig.2.3. Example results of six different methods.

As shown in Fig. 2.3, the regions around the salient objects will get high saliency values in RC method; though the SVO method can highlight the entire salient objects, it also generates high saliency values for the background; the CA method gives high saliency values to object boundaries, but it fails to highlight the entire objects; the EST method obtains good results when a salient object is uniform and has high contrast with the background, but it will fail when the object is not uniform and the image has low contrast. The saliency estimation results in Fig. 2.3 demonstrate that the proposed method can highlight the entire salient object and generate low saliency values to the background, in

spite of not uniform objects or low image contrast.

2.4.4 The effectiveness of adaptive center constraint

Similar to the procedures in section 2.2, we compare the methods using image center-based constraints with our method using adaptive center constraint. In our methods, α and β are set to $2\sqrt{N}$ and $2N$, respectively. As shown in Fig.2.4, the proposed method using adaptive center-based data term is more robust than other center-bias methods.
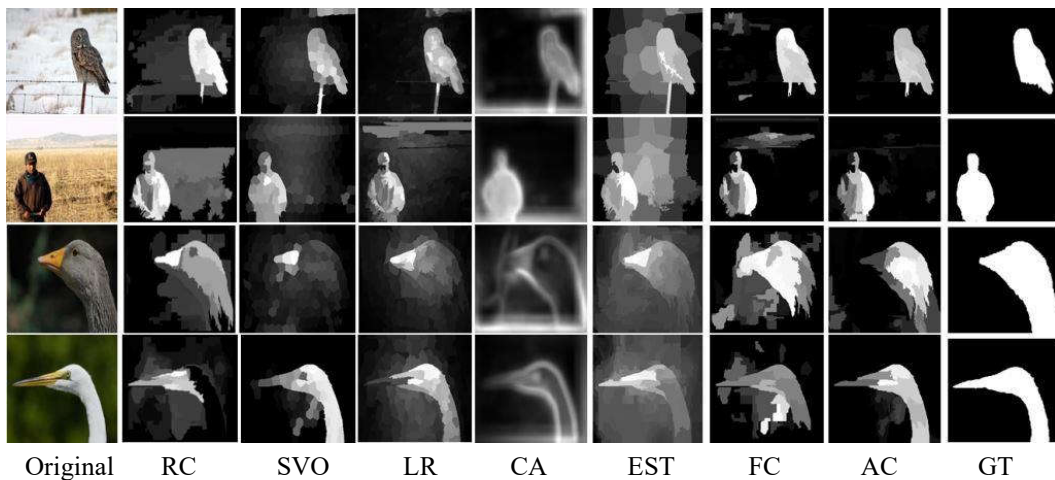
Fig.2.4. Examples of adaptive center effectiveness. FC is fixed center. AC is adaptive center. From the last three columns we can see that using the adaptive center can make the high saliency more concentrated on the objects than using the fixed center.

2.4.5 The effectiveness of the newly proposed smoothness term

We compare the saliency estimation results of the method using the newly proposed smooth term with that of the method using region correlation-based smooth term [14]. In this experiment, we first test our method with the newly proposed smooth term on the two datasets; and then test our method again by replacing the smooth term with the regions
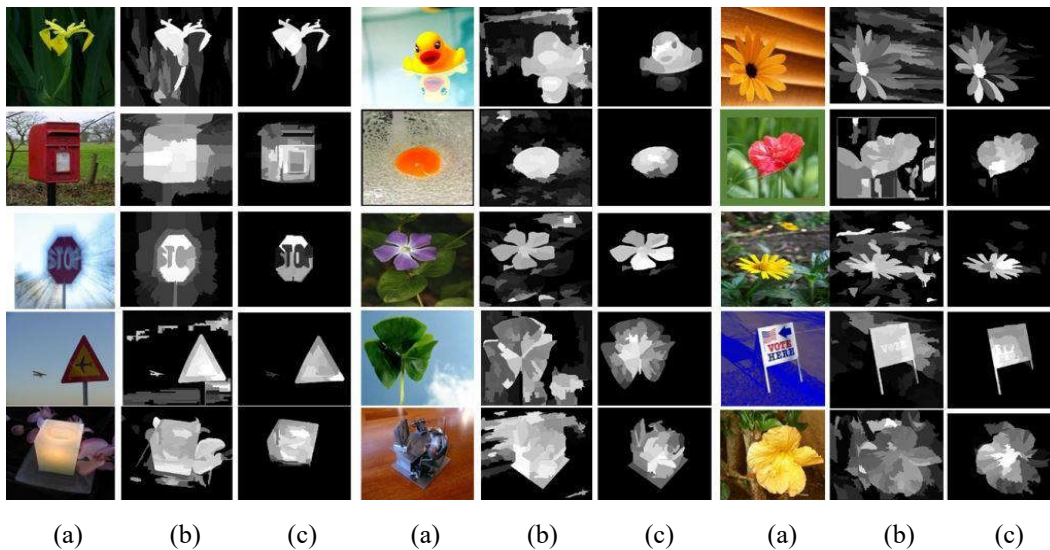
Fig.2.5. The effectiveness of the new smoothness term. (a) the original images (b) the results of using regions' correlation term (c) the results of using our smoothness term.

correlation-based one. The parameters are the same with those in [14]. The results of 15 images are shown in Fig. 2.5.

The results in Fig. 2.5 demonstrate that the newly proposed color saliency-based smooth term is more robust than the region correlation-based smooth term: our method with the newly proposed smooth term can generate more accurate saliency map; however, our method with the region-correlation smooth term generates high background saliency values especially when the background is complicated.

2.4.6 Comparison with state-of-the-art

In this section, we compare the overall performance of the proposed approach with 10 state-of-the art methods (SR [6], PS [7], FT [8], HC [9], RC [9], CA [23], SVO[10], LRS [11], EST [14], and LC [22]), on ECSSD, ASD and SED1 datasets. As shown in Figs. 2.6-2.8, the proposed method outperforms 10 methods (SR, PS, FT, HC, RC, CA, SVO,
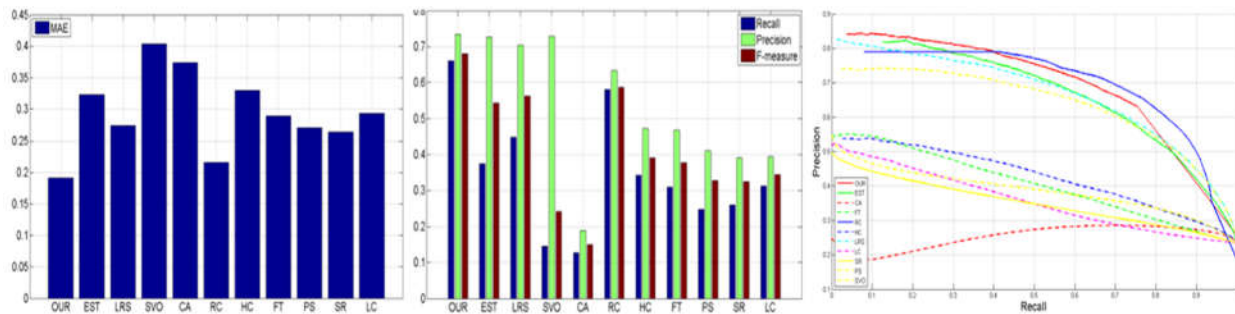
Fig. 2.6: Evaluation results of different methods on the ECSSD dataset.
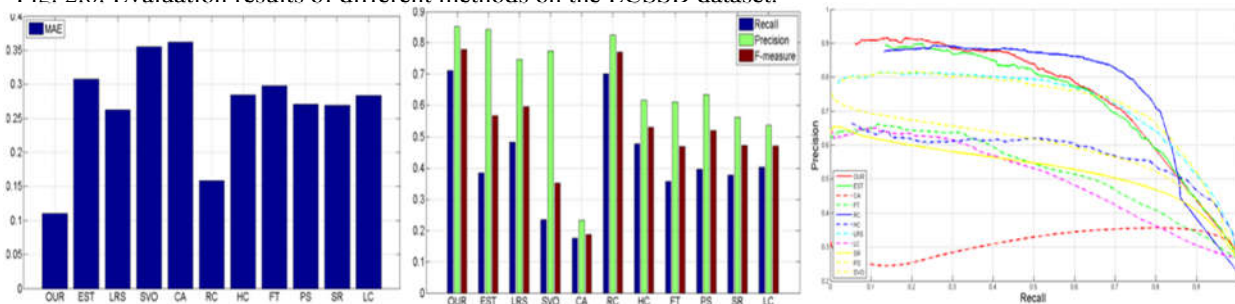
Fig. 2.7: Evaluation results of different methods on the SED1 dataset.
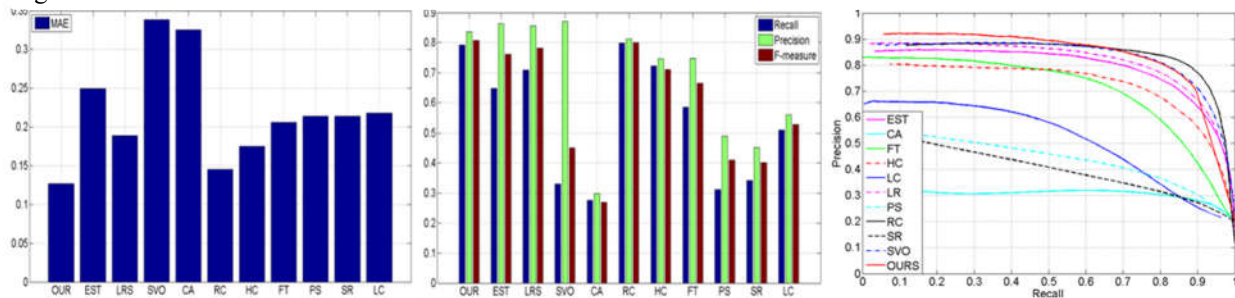
Fig. 2.8: Evaluation results of different methods on the ASD dataset.

LRS, EST and LC) on the three datasets; our MAE values, the F-measure values and P-R curve are better than all 10 methods on ECSSD and SED1 datasets.

CHAPTER 3

A HYBRID VSE FRAMEWORK FOR TUMOR DETECTION

In this chapter, a new VSE approach is applied to imitate the radiologists' attention to detect the tumor using breast ultrasound (BUS) images.

3.1 VSE in breast ultrasound images

Breast cancer is the most frequently diagnosed cancer and account for about 29% of all new female cancer cases [24]. Automatic BUS segmentation is a key component in computer-aided diagnosis (CAD) systems and has the advantages of operator-independence and high reproducibility [25, 26]. However, developing automatic segmentation approaches for BUS images is challenging, due to the speckle noise, low contrast, weak boundary, and artifacts; furthermore, strong priors to object features such as tumor size, shape and echo strength vary considerably across patients and machine settings cannot work well on images from multiple sources [27].

Many automatic BUS segmentation approaches have been proposed in the last decade [25-34]. The major strategy of most automatic approaches is to locate tumors automatically by modeling domain-related priors. However, some strong constraints such as the number of tumors, tumor size, and predefined tumor locations, were utilized in the approaches, which result in dramatic performance degradation in clinical practice where BUS images could be collected under different settings or situations such as low image contrast, more artifacts, containing no tumor/more than one tumors per image, etc.

Therefore, it is crucial to develop automatic BUS segmentation techniques that are invariant and robust to images settings.
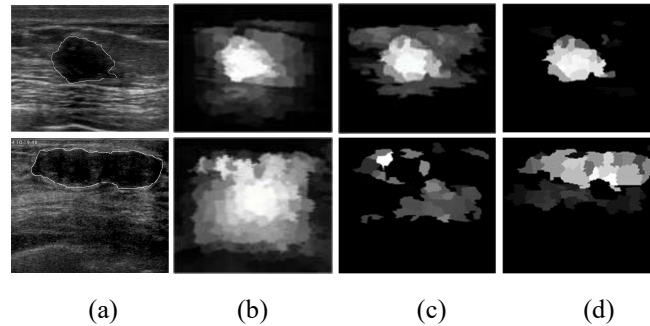
Fig. 3.1. Tumor saliency estimation examples. (a)Two BUS images with the ground truths (white boundaries); (b) results of the method in [17]; (c) results of the method in [27]; (d) results of the proposed method.

Tumor saliency estimation aims to model the visual clues of tumors in BUS images that attract radiologists' attention during the tumor detection process. The saliency estimation outputs the saliency value of each BUS image pixel regarding the pixel's possibility of belonging to a tumor. In [2], Shao et al. proposed a computational saliency estimation model for fully automatic tumor segmentation. The model combined tumor prior knowledge and saliency estimation hypothesis and achieved very good performance using their BUS image dataset. However, it had two main drawbacks: 1) it always outputs a salient region and could not deal with images without tumor; 2) the computational model failed to handle the images with large tumors, shadows, and low contrast (Fig. 3.1). Xie et al. [1] proposed to model intensity, blackness ratio, and superpixel contrast; and the final saliency value of each pixel was the average of values of the three components. It shared the same drawbacks with [1] due to the nature of direct mapping and the strategy of "winner-take-all."
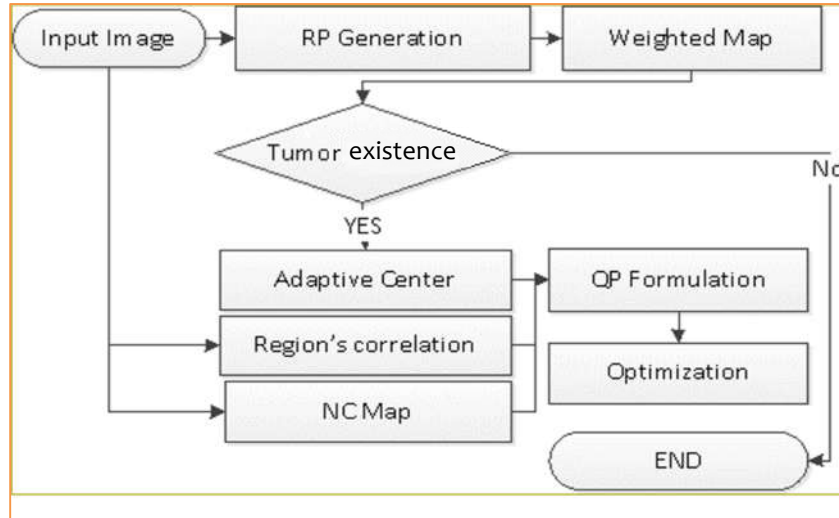
Fig. 3.2. Flowchart of the proposed method.

Xu et al. proposed a general bottom-up saliency estimation model [3] that integrated many robust hypotheses: the global contrast, adaptive center-bias, boundary constraint and the smoothness term based on the color statistic. The model was flexible, and the global optimum could be reached by using the primal-dual interior-point method. However, the model always outputs a salient region and could not handle BUS images without tumors due to the equality constraint used.

To solve the above drawbacks, we propose a novel hybrid framework for tumor saliency estimation, which follows a two-step strategy. The first step determines if a BUS image has tumor(s) utilizing the adaptive reference point (RP) generation [20]. Weighted maps are generated iteratively based on the relative locations of two consecutive RPs. The final RP indicates the possible location of tumor; and the final weighted map gives the possibility of local regions in a tumor. In the second step, it formulates the tumor-saliency estimation problem as a quadratic programming (QP) problem which integrates both high-

level domain-knowledge and robust low-level saliency assumptions. In this framework, it incorporates Neutro-Connectedness (NC) [25] to generate more robust and accurate boundary connectedness and to measure the corresponding degree of confidence simultaneously. The adaptive center-bias and regions' correlation hypothesis are also integrated in the framework. Fig. 3.2 shows the flowchart of the proposed framework on BUS images.

3.2 Tumor existence determination

Existing tumor saliency detection or segmentation methods assume that there exists a tumor in each BUS image and cannot handle the image without tumor; however, as an automatic tumor detection or segmentation system, it is important to identify whether there is a tumor or not. Besides, the convex optimization frameworks cannot deal with the image without the salient object. The equality constraint will force that there must be at least one salient object in the saliency map.

In [26], Xian et al. proposed an algorithm to automatically generate the adaptive reference point (RP) based on the breast anatomy. The RP was generated accurately and fast and could detect the darker regions (candidates of the tumors). The weighted map was constructed based on the RPs and the intensity map. The region is nearer the RP, the intensity value of the region in the weighted map is higher, vice versa. As shown in Fig.3.3, the weighted map enhanced the low -intensity pixels/regions near RP and decayed the high-intensity pixels/regions far away from RP.
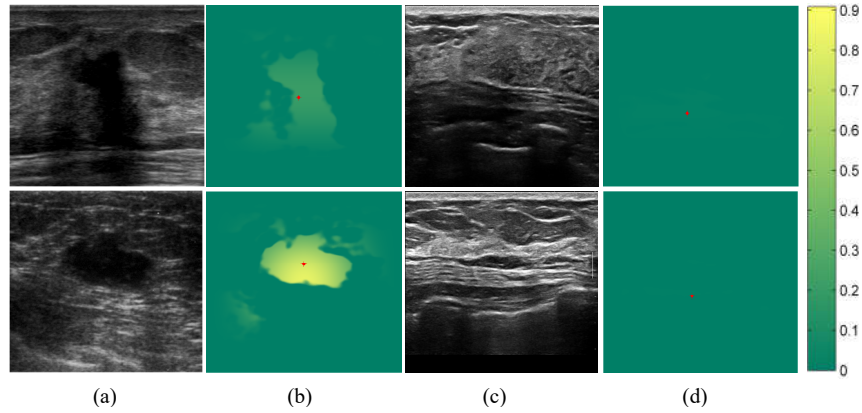
Fig. 3.3. Weighted maps. (a) BUS images with tumor; (b) and (d) weighted
maps with RPs (marked with red color); (c) BUS images without tumor.

Based on the observation, the weighted map of the BUS image without tumor is
smoother than that of the BUS images with tumor. In Fig. 3.3, the four max intensities of
the weighted maps are 0.043, 0.0152,0.9086 and 0.0035, listing from left to right, top to
bottom, respectively. It chooses the local maximum, mean, and standard deviation of the
weighted map as the feature vector and applies threshold or Decision Tree to classify. The
tumor existence determination result and discussion are in Section 3.4.

3.3 Tumor Saliency Estimation

Researchers have applied several saliency hypotheses to construct mathematics
models for visual saliency estimation, such as rarity hypothesis, center-bias hypothesis,
correlation hypothesis, etc. In this work, it utilizes the adaptive center-bias, regions'
correlation hypotheses, the boundary NC map, and weighted map to model the tumor-
saliency estimation problem as a convex optimization problem.

Firstly, it used a quick shift algorithm in [30] to over-segment the image into N
superpixels, noted as $\{R_i\}_{i=1}^{i=N}$. Similar to the method in [27], it extracts regions' average

intensities as the region features. To facilitate the discussion, it defines $S = (s_1, s_2, \cdots, s_n)^T$ as a vector of saliency values for N image regions, where s_i denotes the saliency value of the ith image region and $s_i \in [0, 1]$. The optimization of the model is to assign the optimal saliency values for a set of image regions.

3.3.1 Problem formulation

The problem is formulated as

$$\begin{aligned} \text{minimize} \quad & E(S) = S^T(T + \alpha C + \gamma W) + \\ & + \beta \sum_{i=1}^{N} \sum_{j=1}^{N} (s_i - s_j)^2 w_{ij} D_{ij} \\ \text{subject to} \quad & 0 \le s_i \le 1, i = 1, 2, \cdots, N; \\ & \sum_{i=1}^{N} s_i = 1 \end{aligned} \tag{22}$$

In Eq. (22), the term $T = (t_1, t_2, \cdots, t_N)^T$ denotes the NC map, and t_i defines the NC between the ith region and the boundary; the term $C = (c_1, c_2, \cdots, c_N)^T$ is the distance map, and c_i defines the distance between the ith region and the adaptive-center; the term $W = (e_1, e_2, \cdots, e_N)^T$ is the weighted map, and e_i is the value of the ith region; the terms w_{ij} and D_{ij} define the similarity and the spatial distance between the ith and the jth regions, respectively. The term $S^T T$ defines the cost using the NC map, the term $S^T W$ defines the cost on the weighted map, and the term $S^T C$ defines the cost based on the adaptive center-bias. The last term is the smoothness that forces the regions with similar features to have similar saliency values.

The formulated problem is a typical QP problem with linear equality and inequality constraints. The original problem can be rewritten as follows:

$$minimize\ f_0(S) = \left(\sum_{i=1}^{N} s_i t_i + \alpha \sum_{i=1}^{N} s_i c_i + \gamma \sum_{i=1}^{N} s_i e_i\right)$$
$$+ \beta \sum_{i=1}^{N} \sum_{j=1}^{N} (s_i - s_j)^2 w_{ij} D_{ij}$$
$$subject\ to\ \ 0 \le s_i \le 1, i = 1,2,\cdots,N;$$
$$\sum_{i=1}^{N} s_i = 1$$

(23)

3.3.2 NC map generation

Boundary connectivity is an effective prior utilized in many visual saliency

estimation models [3, 19, 35-38]. Most models define the boundary connectivity by using

the shortest path between the local regions and the boundary. However, such connectivity

cannot handle noisy data well. The Neutro-Connectedness (NC) theory [29, 30] introduced

a new domain, the degree of confidence, to measure the confidence of the connectedness.

The new domain is very useful to avoid the fake connectedness caused by uncertainty, such

as noise.

In [30], the NC of two regions contains three parts: the degree of truth, the degree

of confidence, and the degree of false, $NC(i,j) = [T(i,j), I(i,j), 1 - T(i,j)]$ where i and

j indicate the ith and jth pixels or regions, respectively.

Here, NC map is defined on the image region $\{R_i\}_{i=1}^{i=N}$. To calculate the NC triplet

between all the regions with the boundary set, it applies the definitions of NC and

computation algorithm in [29,30]. For more details of NC theory, refer [29,30]. The three

basic ideas of NC are summarized as follows:

a) NC of two adjacent regions i and j

$$\mu_T(i,j) = exp\left(-|Gray_i - Gray_j|/\sigma^2\right)$$

(24)

$$\mu_I(i,j) = 1 - max\left(h(i), h(j)\right)$$

(25)

where $Gray_i$ is the regions average gray level of the ith region, $\sigma^2 = 0.5$, and $h(i)$ is the inhomogeneity of the ith region [29,30].

b) NC of a path

The degree connectedness of a path is defined as the minimum value of μ_T along the path, and the confidence is the maximum μ_I value along the path.

c) NC of any two regions

The degree of connectedness is defined by the strongest path of all paths connecting the two regions. It uses the confidence of the strongest path as the degree of confidence of the two regions. For more information about how to deal with ties, refer [29,30].

As the particular characteristic that no tumor is touching the border, it sets the border regions as the background seeds to generate the NC map by using the algorithm in [29]. Fig. 3.4 shows some samples, T maps and I maps.

(a) (b) (c)
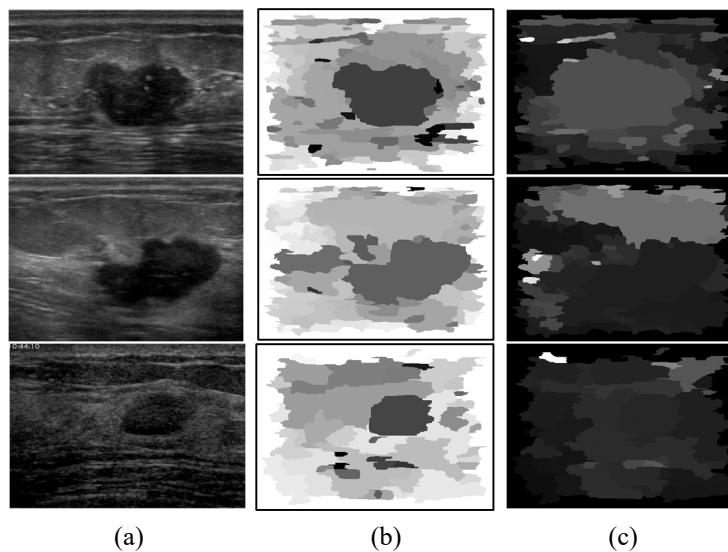
Fig. 3.4. T map and I map samples. (a) BUS images with tumor; (b) T maps; (c) I maps.

T and I maps of image regions employed in the framework are defined as $T = (t_1, t_2, \cdots, t_N)^T$ and $I = (I_1, I_2, \cdots, I_N)^T$, and they are $N \times 1$ vectors.

3.3.3 Adaptive center bias and weighted map generation

Traditional saliency estimation models usually use the image centers as important visual clues to estimate the saliency maps. However, they failed when objects are far away from the centers. The approach in [3] solved this problem on natural images by estimating the adaptive center using the weighted local contrast map. However, the local contrast map was sensitive to noise and could not achieve good performance on BUS images. Instead of detecting the top and bottom lines of mammary layer [2], a new tumor detection approach was proposed by utilizing the RP and weighted map [29].

The weighted map vector in the framework is defined as $W = (e_1, e_2, \cdots, e_N)^T$

$$e_i = exp\left(-m_i/\sigma^2\right) \tag{26}$$

where m_i is the mean value of the ith region in the weighted map, and $\sigma^2 = 0.5$.

The reference point is used as the adaptive center in the saliency detection model. It is defined as $C = (c_1, c_2, \cdots, c_N)^T$

$$c_i = exp\left(\|SC_i - RP\|_2/d_D\right) \tag{27}$$

where SC_i is the coordinate of the ith region's center and the value is in $[0,1]$. RP is the reference point position. $\|\cdot\|_2$ is the l_2 norm. d_D is equal to $\sqrt{2}/2$.

3.3.4 Regions' correlation

It uses the region correlation hypothesis to force the closer similar regions to have similar saliency value.

It defines w_{ij} as the similarity, and D_{ij} as the spatial distance between the ith and the jth regions.

$$w_{ij} = \exp\left(-\left|Gray_i - Gray_j\right|/\sigma^2\right) \tag{28}$$

$$D_{ij} = \exp\left(-\left\|SC_i - SC_j\right\|_2/d_D\right) \tag{29}$$

where *Gray* is the regions average gray level vector, In Eq. (29), SC_i is the coordinate of the ith region's center and the value is in $[0,1]$. $\sigma^2 = 0.5$. d_D is equal to $\sqrt{2}/2$. $\|\cdot\|_2$ is the l_2 norm.

3.3.5 Optimization

It uses the primal-dual method to optimize the QP with linear inequality and equality constraints [3]. It can obtain the global optimal value quickly. There are three important steps to apply the primal-dual interior-point method: (1) modify the KKT conditions and obtain the dual, prime and centrality residuals; (2) obtain the primal-dual search direction; and (3) update S and the dual variables.

In the primal-dual interior method, t^0 and S^0 are initialized as 1 and $(1/N)E^T$, respectively; and the dual residual, primal residual, and the centrality residual are updated in each iteration, and the optimization processing stops when the sum of the l_2 norms is less than 10^{-6}.

3.4 Experimental results

3.4.1 Datasets, metrics and setting

In this section, it validates the performance of the newly proposed method using a

BUS image dataset containing 706 ultrasound images, in which 96 images have no tumors,

and 610 images have tumors [27].

Metrics of VSE is the same as section 2.4.1.

Metrics of tumor existence determination: two metrics, precision ratio, recall ratio

are utilized:

$$PR = \frac{|TC|}{|TS|}, RR = \frac{|TC|}{|Tmset|}$$

where $|TC|$ is the number of correct detected images with tumors, and $|TS|$ is the total

number of images detected with tumors; $|Tmset|$ is the total number of images with

tumors in the dataset.

Parameter setting: all the experiments are based on the parameters: $\alpha = 10, \beta = 2$,

and $\gamma = 80$.

3.4.2 Tumor existence determination

Based on the observation, the maximum value of the weighted map is very useful

for tumor detection. Simple thresholds are applied to the maximum value, and the result

is shown in Table 1. It uses the Decision Tree classifier with 10-fold cross-validation. The

mean accuracy is 100%.

TABLE 3.1. Results of thresholding

Thresholds	PR	RR
0.02	98.07%	99.84%
0.03	99.84%	99.51%
0.04	99.02%	99.67%
0.05	99.84%	99.84%
0.057	100%	99.34%

3.4.3 Parameter tuning for the TSE framework

α, β and γ. As presented in previous section, the detection framework has 4 major parts. The NC map or weighted map cannot always provide the correct information to generate the saliency map (see Figs 3.4). It is very important to balance the effect of each part.

The values of α, β and γ are used to balance the influence of the adaptive center-bias term, weighted map and smooth term, respectively. It evaluates the performance of the proposed method with α ranging from 0 to 200, β ranging from 0 to 200, and γ ranging from 0 to 200 on the randomly selected subset containing 20 images. There are three stages to choose the parameters. In the first stage, it makes the step size of three parameters be 40 and roughly obtain the range of each parameter, which can achieve better P-R curve performance and MAE value if the P-R curve is similar. In the second stage, the step size is 10. The step size is 2 in the third stage. As shown in Fig. 3.5, we obtain a better P-R curve when α is close to 10, β is close to 2 and γ is close to 80, respectively.
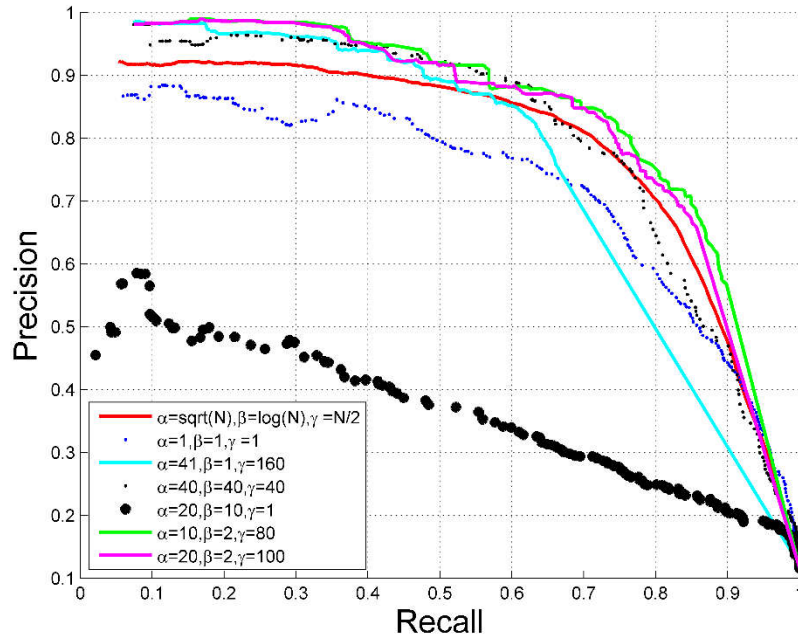
Fig. 3.5. The P-R curve of different parameter values.

3.4.4 The effectiveness of NC term in the TSE framework

The NC map with the boundary connectivity based on the graph shortest path is computed. In this experiment, it used the algorithm of [37] and defined the edge weight for each pair of adjacent nodes as $\left| Gray_i - Gray_j \right|$ to obtain the background map, where $Gray_i$ is the average gray level of the ith region. The examples of 5 images are shown in Fig. 3.6.

The results in Fig.3.6 demonstrate that the two methods can achieve better results on the smooth BUS images (the 3rd and 4th rows). The method based on graph shortest path (GS) fails to handle the BUS images with too small or too large tumors, or poor quality with noise. And the maps generated by NC method are much smoother than that of GS method.

Original images   GS map   NC map
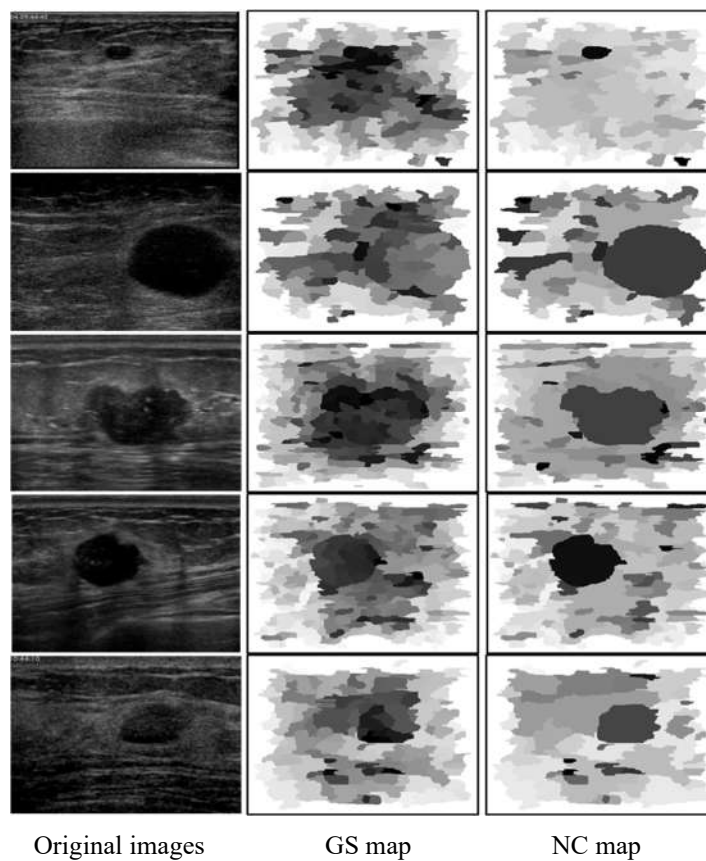
Fig. 3.6. The GS map and NC map samples.

3.4.5 Overall performance of the TSE framework

The proposed method is compared with most recently published methods SMTD [2], OMRC [3], RBD [35] and RRWR [17]. The saliency results of ten BUS images from the dataset are shown in Fig. 3.7. The proposed method can make the high saliency value concentrate on the tumor and the background areas have low saliency value while the methods SMTD, RBD and RRWR made the background regions around the tumor have higher salient values. This situation will cause the recall ratio of the methods higher but the precision ratio lower. And the result maps are more accurate than that of other methods. Especially, SMTD and OMRC cannot produce good saliency maps, even miss the big

| BUS Images | SMTD | OMRC | RRBD | RRWR | Ours | Ground truth |
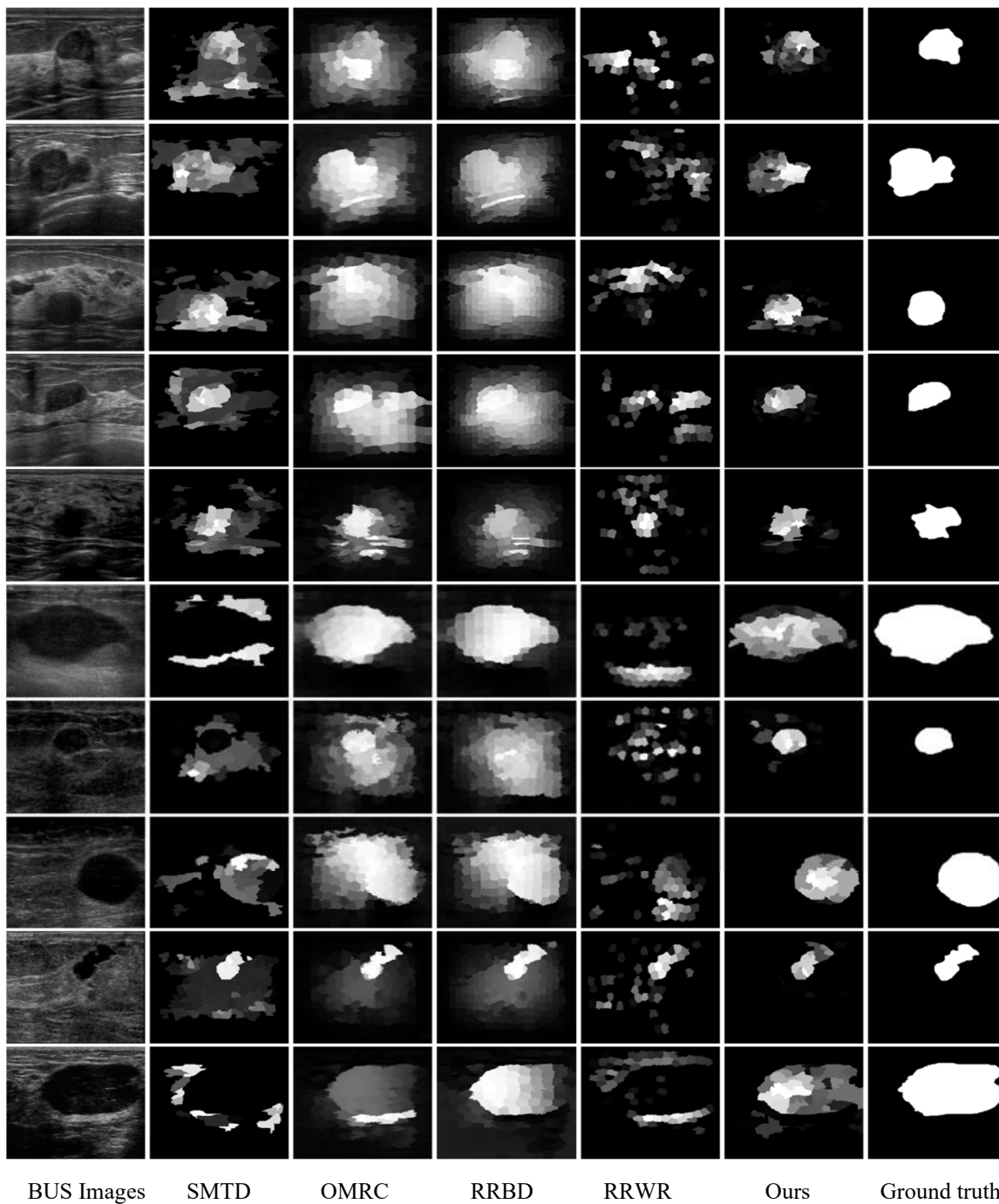
Fig. 3.7. The extensive results.

tumors (the 6th and 10th rows of Fig. 3.7).

The performance of the proposed method is evaluated using the metrics and the dataset: MAE values, the F-measure values, and P-R curves. As shown in Fig. 3.8, the

proposed method is better than other methods. The methods, SMTD, RBD and RRWR, can obtain relatively high average recall ratios, but the precision ratios and F-measures are low. It is because the saliency maps generated by those methods make the tumors, as well as the background around the tumors have high saliency value.
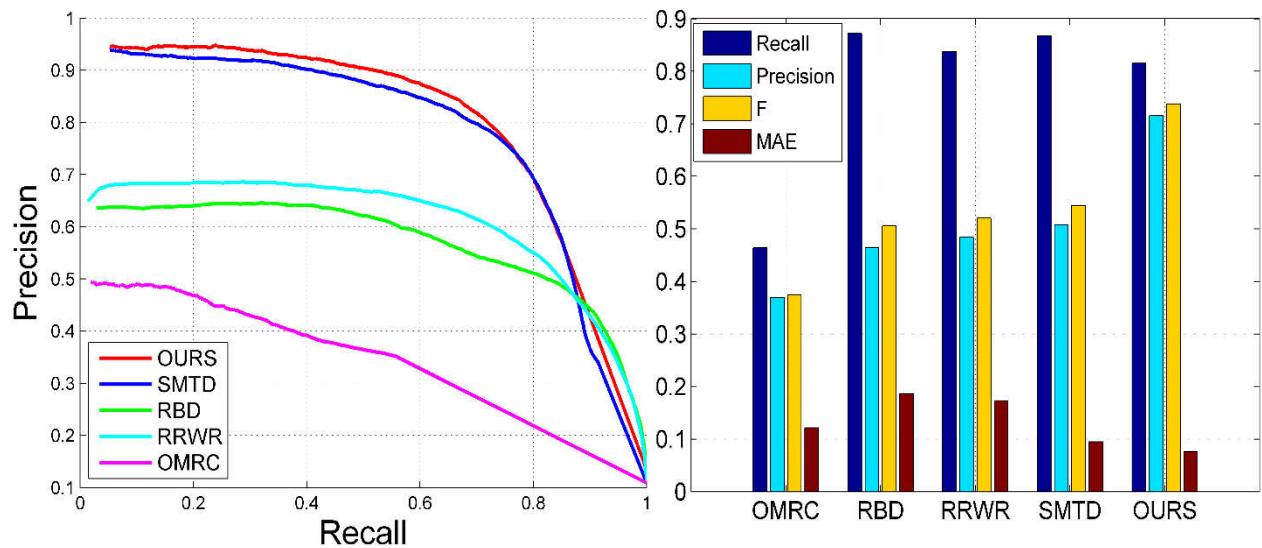
Fig. 3.8. The PR curve and the recall, precision, F-measure and MAE of five methods.

CHAPTER 4

TUMOR SALIENCY ESTIMATION VIA BREAST ANATOMY MODELING

In this chapter, a new VSE framework is applied to imitate the radiologists' attention to detect the breast tumor using breast ultrasound (BUS) images via breast anatomy modeling.

4.1 VSE in BUS images via breast anatomy modeling

In BUS images, tumors typically attract the attention of radiologists even under very different imaging conditions. Examples of applying VSE to BUS images are shown in Fig. 4.1. Many approaches [1, 2, 44] were proposed to model the visual cues attracting radiologists' attention. In [2], Shao *et al.* proposed a model based on saliency estimation for fully automatic tumor detection. The model combined tumor prior knowledge and human visual saliency estimation hypothesis and achieved very good performance using

(a)   (b)   (c)   (d)   (e)   (f)
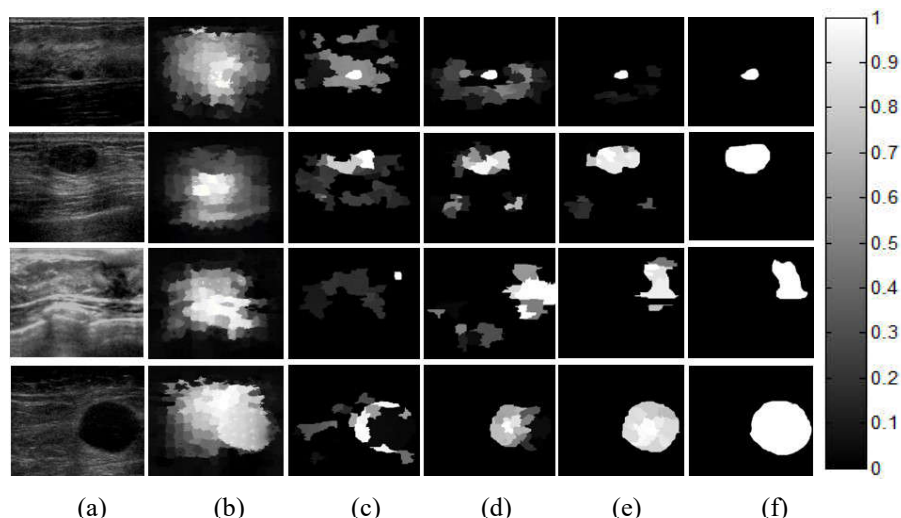
Fig. 4.1. Visual saliency estimation for BUS images. (a) Four original BUS images; (b-d) results of [17], [2], and [44], respectively; (e) results of the proposed method; and (f) the ground truth (GT). The region with higher intensity indicates the region has higher probability belonging to a tumor.
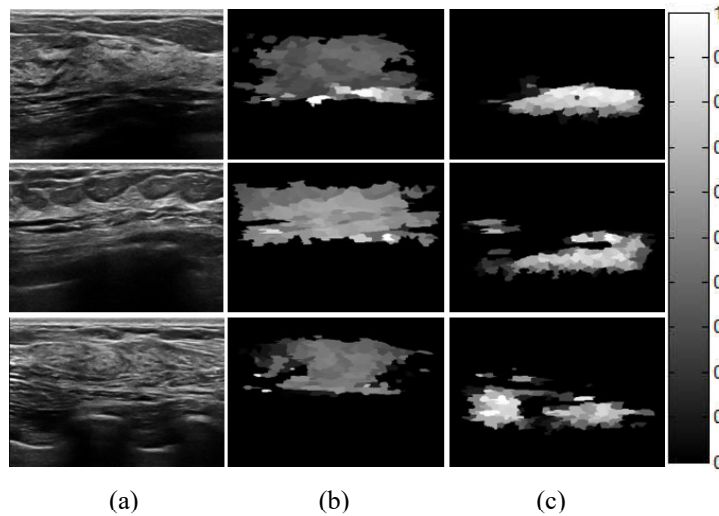
Fig. 4.2. The methods [2] and [44] always generated salient regions in the images without tumors. (a) Original BUS images without tumors, (b) and (c) the saliency maps generated by using method [2] and [44], respectively. The region with higher intensity indicates that the region belongs to a tumor with higher probability.

their own BUS image dataset. The model could locate the tumor accurately most of the cases; however, it had two main drawbacks: 1) always outputs a salient region even there was no tumor in the image (Fig. 4.2 (b)); and 2) could not deal with some special cases well, such as missing some parts of the objects when the images had large tumors (the 2nd and 4th rows of Fig, 4.1 (c)), and low contrast (the 3rd row of Fig. 4.1 (c)). Xie *et al*. [1] computed tumor saliency by comprising intensity, blackness ratio, and superpixel contrast separately; and the average of the values of the three components was the final saliency value of each pixel. The drawbacks were shared as [2] due to the nature of the mapping which directly transferred image features into saliency values by using a unified mapping and the strategy of "winner-take-all". Xu *et al*. [3] proposed a general bottom-up saliency estimation model that integrated the robust hypotheses: the global contrast, adaptive center-bias, boundary constraint and the smoothness term based on color statistic. The model generated a local contrast map and utilized its weighted center as the adaptive center instead of the fixed image center, which was much more robust when the objects were far

away from the image center. The model was flexible, and the global optimum could be reached by using the primal-dual interior-point method. However, the model could not deal with low contrast or gray-level images; furthermore, it always located a salient region and could not handle images without salient objects. Recently, Xu *et al.* [44] proposed a novel hybrid framework for tumor saliency estimation. In the framework, it integrated the background map, foreground map [26] and adaptive center-bias. However, it shared the same drawback as [3] that the data term in the objective function only penalized pixels with nonzero saliency values; and the equality constraint forced the summation of all saliency values to be 1 that led to at least one relative salient object in every image (Fig. 4.2 (c)).

To overcome the above challenges, we propose a novel optimization-based approach for estimating tumor saliency map of BUS image. First, we construct a novel cost function that penalizes the inconsistency between image features and saliency values for both salient and non-salient pixels. By doing so, the equality constraint [3, 44] can be eliminated, and the new approach does not output salient regions for every BUS image. Second, breast anatomy is modeled by using Neutro-Connectedness theory [29, 30] and
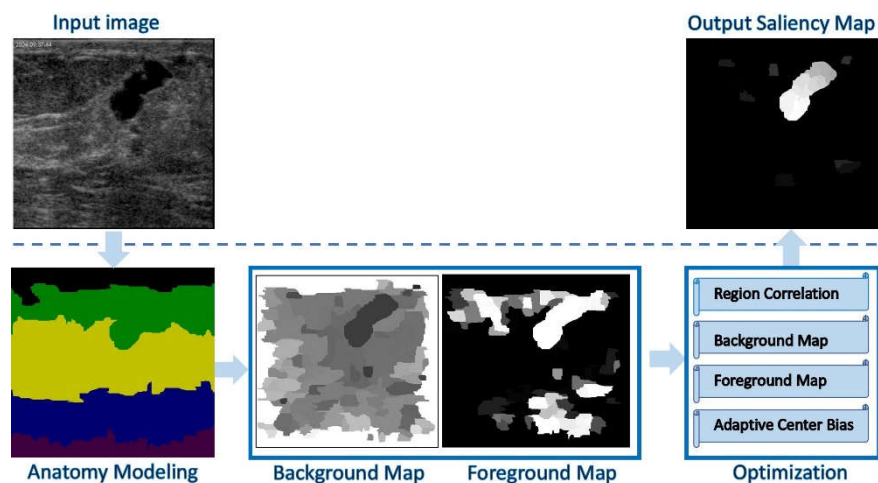
Fig. 4.3. Pipeline of the proposed approach.

applied as non-local context information to solve the problem of outputting wrong salient regions for BUS images with dark shadows (see Figs. 4.6-4.7). The tumor regions will have higher connectedness than that of the background in the low contrast images. The results will be much more reliable by utilizing the breast anatomy knowledge, and it makes the shadow layer with high rate be background; especially, for the images having large tumors. The framework of the proposed method is shown in Fig. 4.3.

4.2 Problem formulation

In the proposed approach, tumor saliency estimation (TSE) is formulated as a Quadratic Programming (QP) problem, and it focuses on solving the problems in existing approaches by building a united optimization-based framework that incorporates robust cognitive hypothesis, e.g., the adaptive center-bias, and region-based correlation hypothesis, and the background and foreground cues.

Let $\{R_i\}_{i=1}^{i=N}$ be a set of image regions generated by a quick shift algorithm [45s], and $S = (s_1, s_2, \cdots, s_N)^T$ be a vector of saliency values, where s_i denotes the saliency value of the ith region and $s_i \in [0, 1]$. The TSE problem is formulated as

$$
\begin{aligned}
\text{minimize} \quad & E(S) = \alpha E^{data}(S) + E^{smooth}(S) \\
\text{subject to} \quad & 0 \leq s_i \leq 1, i = 1,2,\cdots,N; \\
& B^T S = 0, B = (b_1, b_2, \cdots b_N)^T, b_i = \{0,1\}
\end{aligned} \tag{30}
$$

where the data term E^{data} models the background cue, foreground cue and adaptive center-bias cue; and the smoothness term E^{smooth} models the region-based correlation; α balances the influence of the two terms; the equality constraint $B^T S = 0$ is only applied to the mask border regions; and b_i is 1 if the ith region is adjacent to the image border, and 0 otherwise.

$$E^{data}(S) = E^{fg}(S) + E^{bkg}(S) \tag{31}$$

$$E^{fg}(S) = S^T \cdot (-(\ln(D) + \beta \ln(W))) \tag{32}$$

$$E^{bkg}(S) = (1 - S)^T \cdot (-\ln(T)) \tag{33}$$

In Eq. (31), E^{fg} defines the cost of assigning non-zero saliency value to each region, and E^{bkg} defines the cost of assigning zero value to a region. In previous optimization-based approaches [44], only E^{fg} was defined, and no explicit cost was given for outputting zero saliency values. In order to avoid the configuration of all zero saliency values for the entire image, a constraint $\sum s_i = 1$ was defined to force the output to have at least one salient region for every image. This is one of the major drawbacks of previous approaches and makes them unable to deal with BUS images without tumors. In order to overcome the drawback, E^{bkg} is defined in the data term of the cost of assigning zero value to an image region. This strategy can avoid the zero-configuration problem, because all zeros will lead to a high penalty if a salient region (tumor) exists in the image; and it outputs all zeros only when no tumor exists. In Eqs. (32)-(33), $W = (w_1, w_2, \cdots, w_N)^T$ is the foreground map, and defines the probability of each image region to be a tumor region; $D = (d_1, d_2, \cdots, d_N)^T$ is the distance map, and d_i defines the distance between the ith region and the adaptive center; and β balances the contribution of the two terms. T denotes the background map, and defines the probability of an image region to be a background region. The definitions of W, D and T will be given in section 4.2.2

$$E^{smooth}(S) = \sum_{i=1}^{N} \sum_{j=1}^{N} (s_i - s_j)^2 r_{ij} Dist_{ij} \tag{34}$$

E^{smooth} in Eq. (34) defines the penalty on similar regions with different saliency values. The terms r_{ij} and $Dist_{ij}$ define the similarity and spatial distance between the ith and the jth regions, respectively.

The problem defined by Eqs. (30) – (34) is a typical QP problem with linear equality and inequality constraints. The original problem can be rewritten as follows:

$$minimize\ f_0(S) = \alpha \sum_{i=1}^{N} -s_i \ln(d_i) + \gamma \sum_{i=1}^{N} -s_i \ln(w_i) + \\ \alpha \sum_{i=1}^{N} -(1-s_i)\ln(t_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}(s_i - s_j)^2 r_{ij}\ Dist_{ij}$$

$$subject\ to\ \ 0 \le s_i \le 1, i = 1,2,\cdots,N; \\ B^T S = 0, B = (b_1, b_2, \cdots b_N)^T, b_i = \{0,1\}$$

(35)

where $\gamma = \alpha\beta$, refer Eqs. (30) and (32).

4.3 Data term

The data term is composed of three major components: foreground map (W), distance map (D), and background map (T). W models priors of general tumor appearance; D models the adaptive center-bias hypothesis; and T is defined as the weighted connectedness between image and border regions. The definitions of the three parts are guided by breast anatomy.

(a) BUS image   (b) Breast Anatomy
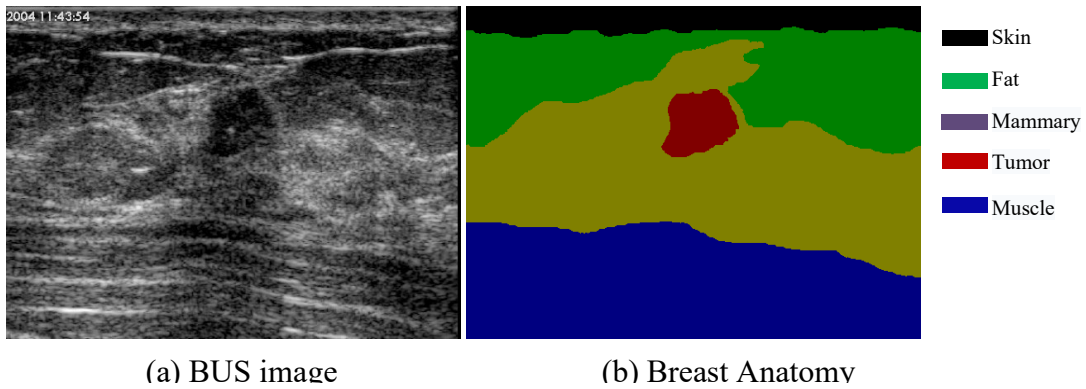
Fig. 4.4. An example of breast anatomy.

4.3.1 Breast anatomy modeling using Neutro-Connectedness.

Breast anatomy represents the structure of the breast and is useful for breast tumor detection and classification in clinical practice. Breast contains four primary layers: skin layer, fat layer, mammary layer, and muscle layer [47]. BUS image regions in different layers have different appearances (e.g, intensities and textures); and breast tumor mainly exists in the mammary layer (see Fig. 4.4). Breast anatomy was utilized in [2,26] for tumor segmentation. In [2], Shao identified two horizontal lines to remove the fat and muscle by applying phase congruency [46] and Otsu's thresholding. However, it was difficult to identify the two horizontal lines accurately; in some cases, part of the tumor could be divided into the fat region. In this work, we propose a new Neutro-Connectedness (NC) [29] based framework that models breast anatomy by incorporating region similarity and image depth. It decomposes BUS images into 3 to 5 layers.

There are two components in NC: the degree of connectedness t and confidence of connectedness c, $NC(i,j) = [t(i,j), c(i,j)]$ where i and j indicate the ith and jth pixel or region, respectively. Image regions from the same layer have strong connectedness (e,g., high t and c values), and from different layers have weak connections. NC builds on the following three fundamental concepts:

(1) NC of two adjacent regions i and j. The degree of connectedness of two adjacent regions is defined as their similarity, noted as μ_t; and the degree of confidence is defined as the homogeneity between them, noted as μ_c.

(2) NC of a path. The degree of connectedness of a path is defined as the minimum value of μ_t along the path, and degree of confidence is the minimum μ_c value along the path.

(3) NC of any two regions. The degree of connectedness is defined by the strongest path between the two regions. It uses the confidence of the corresponding path as the degree of connectedness confidence of the two regions.

NC computation generates NC maps that demonstrate the degrees of connectedness and confidence among image regions, and NC trees that show regions' structure. Each NC tree contains a group of image regions that share common properties. All NC trees form an natural decomposition of an image. In this work, NC is applied to decompose BUS images into different layers; and we redefine the NC of two adjacent regions by utilizing the region similarity and image depth. The depth term penalizes the growth of NC trees along the vertical direction to avoid the cross-layer expanding.

$$u_t(i,j,k) = r_{ij}(i,j) \cdot o_{ik}(i,k) \tag{36}$$

$$\mu_c(i,j) = min\ (h(i), h(j)) \tag{37}$$

$$r_{ij}(i,j) = exp\ (-|I(i) - I(j)|/\sigma_1^2) \tag{38}$$

$$o_{ik}(i,k) = exp\ (-|row(i) - row(k)|/\sigma_2^2) \tag{39}$$

In Eq. (36), r_{ij} denotes the similarity between the ith and jth regions, and o_{ik} is the normalized depth difference between the ith region and the root region (k) of a NC tree. In Eq. (37), $h(\cdot)$ defines the homogeneity of a region [29, 30]. In Eq. (38), $I(i)$ and $I(j)$ are the normalized intensities of the ith and jth regions, respectively; $row(i)$ denotes the row index of the ith region center. σ_1^2 and σ_2^2 control the shapes of the two exponential functions. σ_1^2 is 0.5 the same as [44]. Without the depth term in Eq. (39), the layer of the region i will be determined by the connectedness values between region i and different root regions. For illustration, the ith region is at the bottom of the image, and the root region k is at the top of the image, and the jth layer is between them. If the connectedness between
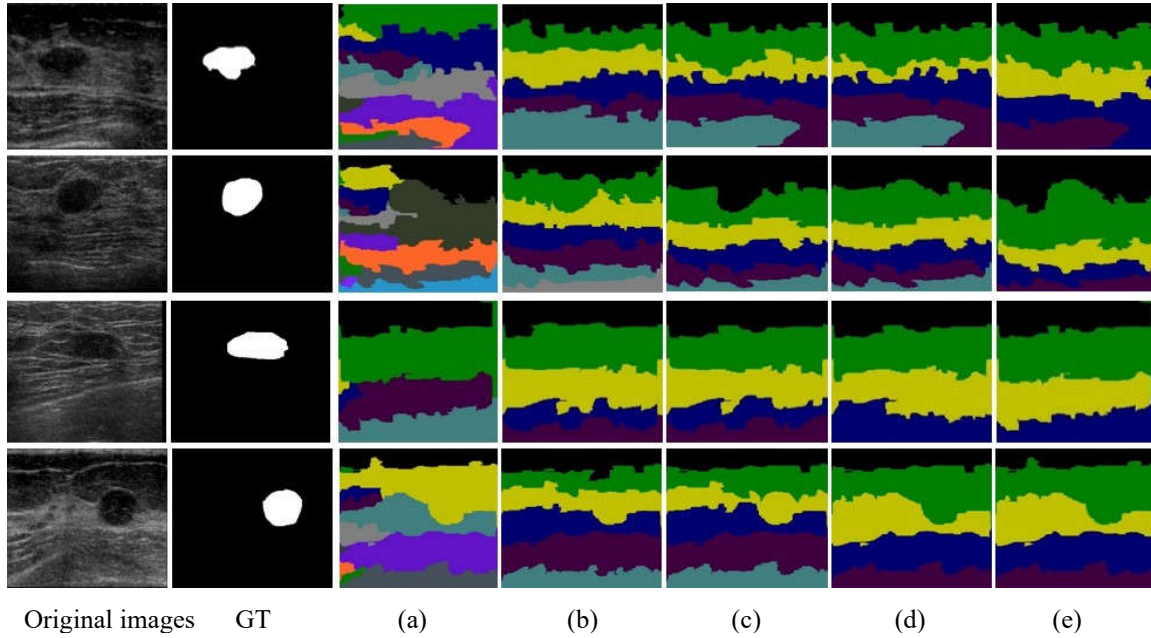
Fig. 4.5. Effectiveness of different σ_2^2. (a) $\sigma_2^2 = 0.2$, before merging;(b)- (e): with $\sigma_2^2 = 0.05, 0.1, 0.15$ and 0.2, after merging, respectively. From top to bottom, the same color indicates the same layer.

the ith region and the root region k is the maximum among the connectednesses of the ith region with the root regions, and the ith region would be expanded into the layer of the root region k. In such a case, the NC tree expanded by the root region k would cross the group j, which does not match the breast anatomy. The expanding of the layer along the vertical direction should be avoided, but only along the horizontal direction using Eq. (39). σ_2^2 controls the span of the layer, and larger σ_2^2 will result in fewer layers. The initial σ_2^2 is chosen by applying values from 0.1 to 0.5 with a step size of 0.1 on the training dataset (refer section **4.3.2**) to generate the layers. Eighty percent of the training dataset images were decomposed into three to five layers when $\sigma_2^2 = 0.2$, and the other images were decomposed into 6 or more layers. However, we can control the layer number to be 3 to 5 by initializing σ_2^2 as 0.2 and updating it adaptively. If the number of layers is greater than five, decrease σ_2^2 by 0.05; and increase σ_2^2 by 0.05, otherwise.

After computing the NC of two adjacent regions, the connectednesses of a path and

between any two regions can be calculated easily. The left boundary regions of an image are set as the roots for generating NC trees. All the regions on a tree are in a group (layer). If a layer cannot cover more than 75% of the image width, it will be merged into its nearest layer. The effectiveness of the merging step with different σ_2^2 is shown in Fig. 4.5 (a) and (e). Note that each generated image layer is composed of a group of image regions that have high connectedness with each other; those regions have high probability from the same biological tissue layer, but the generated image layer is not the biologic tissue layer.

4.3.2 Foreground map (FG) generation.

The foreground map measures regions' possibilities to be tumor regions. We propose a two-stage strategy to generate the foreground map by using both image appearance and breast anatomy. The Z-shaped function is used for each layer to emphasize regions with low intensities. Layer's location generated in the last section Breast anatomy modeling using Neutro-Connectedness is employed to reduce the impact of the dark regions from the fat and shadow regions. The Z-shaped function [26] is utilized; however, the parameters a, b, and c in Eq. (40) are chosen adaptively for different layers of different images, and the intensities are mapped in [0,1].

$$Z(I(i); a, b, c) = \begin{cases} 1, & I(i) \leq a \\ 1 - \frac{(I(i)-a)^2}{(c-a)(b-a)}, & a < I(i) \leq b \\ \frac{(I(i)-c)^2}{(c-a)(c-b)}, & b < I(i) \leq c \\ 0, & I(i) > c \end{cases} \tag{40}$$

Firstly, the global values of a, noted as a_g, and c, noted as c_g, were calculated using all layers of an image. According to the experiments on the training dataset (see section **4.6.2**), we choose a_g and c_g as following:

$$a_g = i, subject\ to\ \frac{\Sigma_{I(j)\le i}(j)}{\Sigma_j(j)} = \varepsilon_1$$

$$c_g = i, subject\ to\ \frac{\Sigma_{I(j)\le i}(j)}{\Sigma_j(j)} = \varepsilon_2 \qquad (41)$$

$$j = 1,2,\cdots N$$

where $I(j)$ is the intensity of the region j, respectively. $\varepsilon_1 = 1/10$ and $\varepsilon_2 = 6/10$ based

on experiments, respectively.

Table 4.1 Algorithm 1: Foreground map (FG) generation

Input: $\{I^i\}_{i=1}^{layerNum}$
Output: FG, $layerW$
1.Calculate global a_g and c_g
2. **for** i from 1 to $layerNum$ **do**
    Generate local parameters a_i, b_i , and c_i;
    Check whether it's a dark layer or a normal layer.
    $a_i = \min(a_g, a_i)$, $c_i = \min(c_g, c_i)$, $b_i = mean(I^i < c_i)$
    Apply a_i, b_i , and c_i to (40) to generate the initial FG_i;
  end for
3.Separate the layers into three groups and update the values of $loopS$ and $loopE$.
4.To reduce the shadow layer's influence, assign the intensities of the shadow
  layer's regions to be the highest as the new I'; update a_g, and c_g.
5. **for** i from $loopS$ to $loopE$ **do**
    if it is a dark layer indicating that a big tumor exists
      $a_i = \max(a_g, a_i), c_i = \max(c_g, c_i)$, $b_i = (a_i + c_i)/2$;
    else
      $a_i = \min(a_g, a_i), c_i = \min(c_g, c_i), b_i = mean(I^i < c_i)$.
  Apply the new local parameters to (40) to generate the FG in the layer and
  assign the layer weight using (43).
    end for
6.Normalize $layerW$ to [0,1], and the final FG is the dot product of FG and (1-
  $layerW$).

Since the fat layer is darker than other regions, if using unified parameters *a, b* and

c in Eq. (40) to generate the foreground map, it will make the fat layer have the highest

value and miss part of the tumor in some cases (refer Fig. 4.6(c)). Therefore, it refined the

FG maps for the layers by introducing local $a_i, b_i,$ and c_i adaptively for each layer. Where

a_i and c_i are generated in the same way as global values based on the ith layer; and b_i is

the mean of the intensities less than c_i. If b_i is small, it indicates that most of the regions in the layer have low intensities due to the following two reasons: (1) if the layer locates at the bottom or top, it has a high probability that the layer is a non-mammary dark layer (e.g., fat layer); (2) otherwise, it is a large tumor in the mammary layer with high probability. If a_i is larger than the global c_g, it indicates that most of the regions in the layer have high intensities; therefore, it has a very low probability to contain a tumor. The condition $b_i - a_g < 0.1(c_g - a_g)$ is used to determine whether there is a dark layer, and the condition $a_i > c_g$ is used to check if the layer is a smooth bright layer in **Algorithm 1**. The step 2 of Algorithm 1 is the first stage to generate the initial FG maps.

The tumor-like regions may also exist in the top and/or bottom layers, and there may be more than one dark layers. Thus, to exclude tumor-like regions, we separate layers into the bottom, top, and middle groups, and assign different weights to the layers in different groups to refine the foreground map. Based on observation, if the distance between the ith layer and the layers in middle group is larger, the ith layer belongs to the top or bottom group, and has no tumor with higher probability. The middle group of different images could contain vary number of layers (1 to 5). If the middle part contains fewer layers, each layer contains a tumor with higher probability. Therefore, in the second stage, it refined the FG maps of the dark layers in the middle group and initialized the weights for the top and bottom groups as (42) and the middle group as (43).

$$layerW_i^{LT} = layerW_i^{LB} = \max\left(\left(i - \left\lceil \frac{layerNum}{2} \right\rceil\right)^2, 1\right); \tag{42}$$

$$layerW_i^{LM} = \exp\left(\frac{-\sqrt{layerNum}}{2(loopE - loopS + 1)}\right); \tag{43}$$

where $layerW_i^L$ indicates the weight for the ith layer locating in L, $L \in \{LT, LB, LM\}$ which indicates the layer in the top (LT), bottom (LB) and middle (LM), respectively. $layerNum$ is the number of total layers in the image.

Separate the layers into three groups and update the values of $loopS$ and $loopE$ by the strategy: $loopE$ is the beginning layer of the bottom group and initialized as $layerNum$. $loopE$ is updated to be $layerNum - 1$ if the last two layers are dark layers and other dark layers between $loopS$ and $layerNum$ -2 exist. $loopS$ is the ending layer of the top group and initialized as 1. $loopS$ is updated to be $loopS$ +1 if the first layer is a dark layer and other dark layers between $loopS$ and $loopE$ -1 exist. Assign the layer weight using (42). If there is only one dark layer or two continuous dark layers exist, keep the layers as mammary-like layers. Assign the layer weight using (43). Update the $loopS$ and $loopE$. The details of the foreground map generation are described in **Algorithm 1**.

Fig. 4.6 shows some comparison examples of the proposed method and the method [26]. As shown in the Fig. 4.6 (c), the boundaries of the tumors in FG maps generated by [26] were missed which were caused by the impact of the shadow layers. The proposed method reduces the impact of the shadow layers by integrating the layers' information, and generates more accurate FG maps (see Fig. 4.6 (d)). In Figs. 4.6 (a) and (c), some parts of the tumors were missed due to the effect of the global parameters. In such a case, some
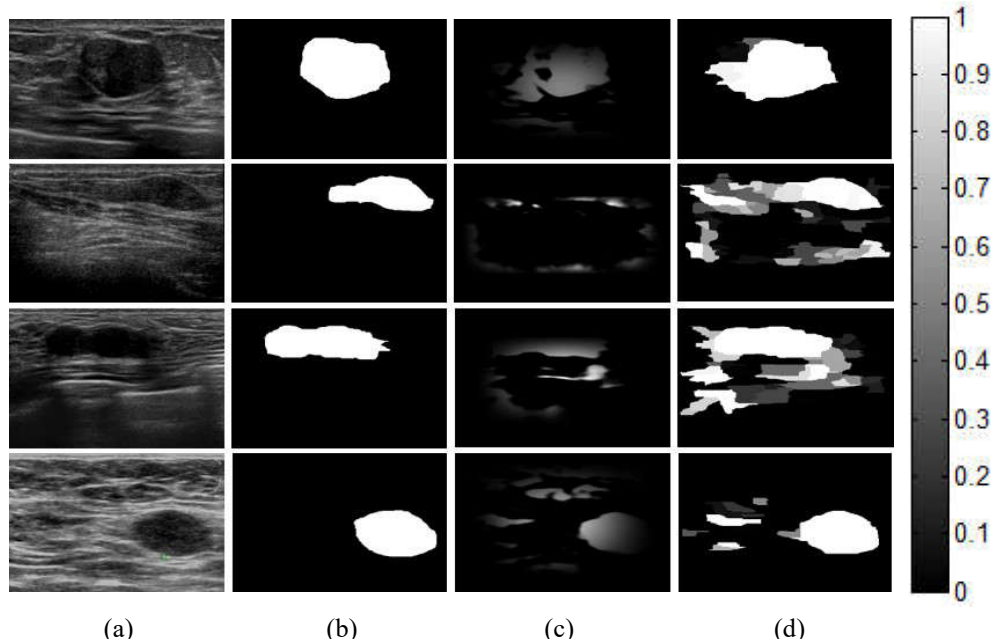
Fig. 4.6. Examples of FG generation. (a) Original BUS images; (b) the ground truths; (c) the foreground maps by [5]; and (d) the final foreground maps of the proposed approach. The region with high intensity belongs to the tumor with the higher probability, vice versa.

regions have much higher intensities than that in the tumor, and applying the unified global parameters in Z function makes these regions get low values in FG maps. As shown in Figs. 4.6 (a) and (d), the proposed method solves this problem by combining the local and global parameters of each layer.

4.3.3 Distance map generation

Traditional saliency estimation models usually use the image center as an important visual cue to estimate the saliency map. However, it will fail when objects are far away from the center. The approach in [3] solved this problem on natural images by estimating the adaptive center (AC) using weighted local contrast map; but the local contrast map was sensitive to noise and could not achieve good performance on BUS images. In this section, we define the AC as the weight center of the foreground map.

$$AC = \frac{\Sigma_{x,y}(x,y)W(x,y)}{\Sigma_{x,y}(x,y)}$$
$$x = 1,2,\cdots M, \text{ and } y = 1,2,\cdots, N$$
(44)

where $W(x,y)$ is the value of pixel (x, y) in the foreground map, and M and N are the number

of image rows and columns, respectively.

The AC distance vector will force the regions far away from the AC to gain small

saliency value and is defined as $D = (d_1, d_2, \cdots, d_N)^T$

$$d_i = exp\left(-\|(x,y)_i - AC\|_2 / \sigma_3^2\right)$$
(45)

where $(x,y)_i$ is the normalized coordinates of the ith region's center. $\|\cdot\|_2$ is the l_2 norm;

and σ_3^2 is set to 0.1 by experiments.

4.3.4 Background map (BG) generation

Boundary connectivity is an effective prior utilized in many visual saliency

estimation models [3, 19, 35-38]. Most models define the boundary connectivity by using

the shortest path between the local regions and the boundary. However, such connectivity

could not handle noisy data well. The degree of confidence domain in NC is very useful

for avoiding the fake connectedness caused by uncertainty, such as noise. As the particular

characteristic that no tumor is touching the border, it sets the border regions as the

background seeds to generate the NC map using the algorithm in [29], noted $\boldsymbol{nc_i}$ as the NC

value of the ith region in the NC map. The higher $\boldsymbol{nc_i}$ indicates the higher probability that

the region belongs to background.

We define the value T in the BG map as follows:

$$t_i = nc_i{}^2 \times layerW_j$$
(46)

where the ith region belongs to the jth layer, and $layerW_j$ is the jth layer's weight.

Fig. 4.7 shows some comparable samples. The connectedness based on graph
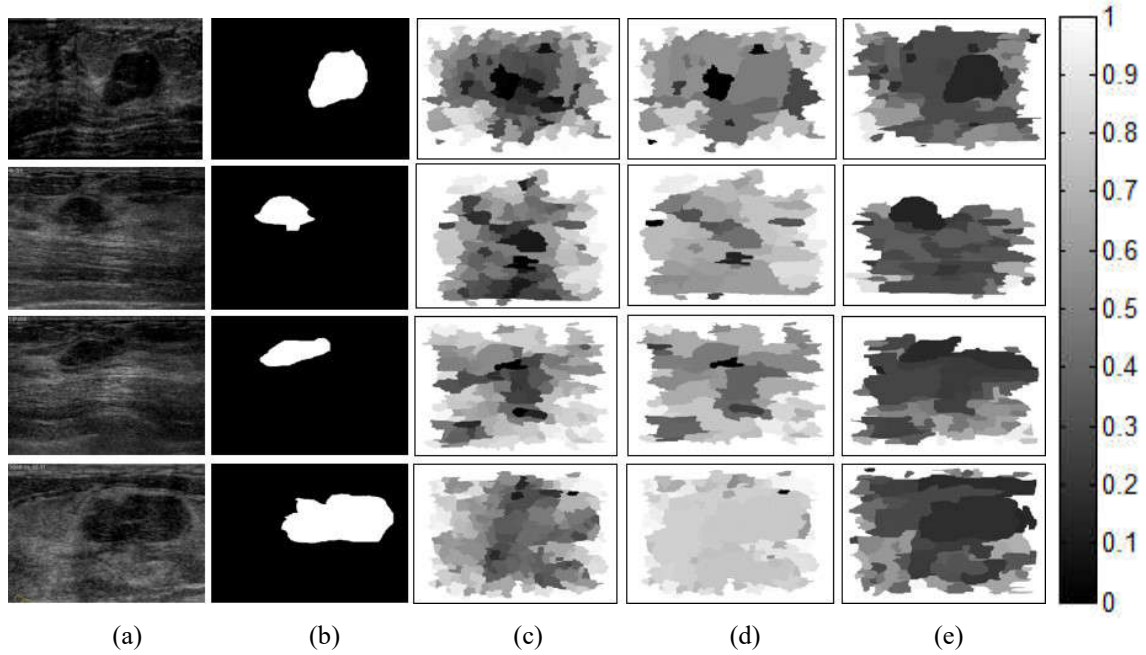
Fig.4.7. Examples of BG generation. (a) Original BUS images; (b) the ground truth; (c) obtained by graph shortest path[48]; (d) obtained by [44] without breast anatomy; and (e) obtained by the proposed method. The region with high intensity belongs to the background with higher probability, vice versa.

shortest path failed to handle the BUS images with too small or too large tumors (see the 2nd-4th rows of Fig. 4.7 (c)), or poor quality with noise (see the 1st row). The results generated by NC without layers' information will make the tumor regions have higher connectedness than the background regions in the low contrast images (see the 2nd and 4th rows of Fig. 4.7 (d)). Moreover, the maps generated by the NC method are much smoother than that of graph shortest path method. The BG result will be much more reliable by utilizing NC with the layers' information, and it makes the shadow layer with a high rate to be background; especially, for the images having large tumors.

4.4 Smoothness term

We utilize regions' feature correlation to force similar regions to have similar saliency values. Specifically,

$$r_{ij} \cdot Dist_{ij} = r_{ij} \cdot exp\left(-\left\|(x,y)_i - (x,y)_j\right\|_2 / \sigma_1^2\right) \tag{47}$$

where r_{ij} measures the similarity of regions i and j; and $Dist_{ij}$ is defined based on the spatial distance between the ith and the jth regions; and $\|\cdot\|_2$ is the l_2 norm.

4.5 Optimization

The primal-dual method is applied to optimize the proposed QP problem, and the global optimal can be achieved [49]. There are three steps to generate the optimization solution: (1) modify the Karush-Kuhn-Tucker (KKT) conditions and obtain the dual, prime and centrality residuals; (2) obtain the primal-dual search direction; and (3) update S and the dual variables. The details of the optimization are described as follows:

The inequality constraints can be rewritten as a set of functions:

$$f_k(S) = -S_k \le 0, k = 1,2,\cdots,N$$
$$f_k(S) = S_{k-N} - 1 \le 0, k = N+1, N+2, \cdots, 2N \tag{48}$$

where N is the number of image regions, and S_k is the saliency value of the kth region. We write all inequality constraints in a matrix:

$$f(S) = \begin{bmatrix} f_1(S) \\ f_2(S) \\ \vdots \\ f_{2N}(S) \end{bmatrix} = \begin{bmatrix} -S \\ S-1 \end{bmatrix}_{2N \times 1} \tag{49}$$

The derivative matrix is

$$Df(S) = \begin{bmatrix} \nabla f_1(S)^T \\ \nabla f_2(S)^T \\ \vdots \\ \nabla f_{2N}(S)^T \end{bmatrix} = \begin{bmatrix} -E \\ E \end{bmatrix}_{2N \times N} \tag{50}$$

where E is the identity matrix.

The dual residual is

$$r_d = \nabla f_0(S)^T + Df(S)^T\lambda + vB$$
$$= -\alpha\ln(C) - \gamma\ln(W) + \alpha\ln(T)$$
$$+ \sum_{i=1}^{N}\sum_{j=1}^{N} 4 \times (s_i - s_j)r_{ij}\,Dist_{ij} + \begin{bmatrix}-E\\E\end{bmatrix}^T \lambda + vB \tag{51}$$

where α and γ balance the three terms defined in Eqs. (1) and (3); and vectors $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_{2N})^T$ and v are the dual feasible parameters.

The primal residual is

$$r_p = \begin{bmatrix} O^T S - 1 \\ B^T S \end{bmatrix} \tag{52}$$

where O is a 2N-by-1 vector, and all the values are 1s.

The centrality residual is

$$r_c = -diag(\lambda)f(S) - (1/g)O \tag{53}$$

where g is the step size, and initialized as 1.

The partial derivatives of r_d, r_p and r_c with respect to variables S, λ and v are as follows:

$$\frac{\partial r_d}{\partial S} = \begin{cases} \sum_{i=1}^{N} 4 \times \left(\sum_{j=1}^{N}(r_{ij} \times Dist_{ij}) - r_{ii} \times Dist_{ii},\, if\ i = j \right. \\ \sum_{i=1}^{N}\sum_{j=1}^{N} 4 \times (s_i - s_j)r_{ij}\,Dist_{ij},\, if\ i \neq j \end{cases} \tag{54}$$

$$\frac{\partial r_c}{\partial S} = -diag(\lambda) \times \begin{bmatrix}-E\\E\end{bmatrix}, \frac{\partial r_p}{\partial S} = B^T \tag{55}$$

$$\frac{\partial r_d}{\partial \lambda} = \begin{bmatrix}-E\\E\end{bmatrix}^T, \frac{\partial r_c}{\partial \lambda} = -diag(f(s)), \frac{\partial r_p}{\partial \lambda} = 0_{1 \times 2N} \tag{56}$$

$$\frac{\partial r_d}{\partial v} = B, \frac{\partial r_c}{\partial v} = 0_{2N \times 1}, \frac{\partial r_p}{\partial v} = 0 \tag{57}$$

In each iteration, the Newton step $(\Delta S, \Delta\lambda, \Delta v)$ is obtained by solving Eq. (58) using the partial derivatives in Eqs. (54) - (57).

$$\begin{bmatrix} \frac{\partial r_d}{\partial S} & \frac{\partial r_d}{\partial \lambda} & \frac{\partial r_d}{\partial v} \\ \frac{\partial r_c}{\partial S} & \frac{\partial r_c}{\partial \lambda} & \frac{\partial r_c}{\partial v} \\ \frac{\partial r_d}{\partial v} & \frac{\partial r_p}{\partial \lambda} & \frac{\partial r_p}{\partial v} \end{bmatrix} \begin{bmatrix} \Delta S \\ \Delta \lambda \\ \Delta v \end{bmatrix} = - \begin{bmatrix} r_d \\ r_c \\ r_p \end{bmatrix} \tag{58}$$

The variables S, λ and v are updated using the following equations.

$$S^{k+1} = S^k + g^k \times \Delta S, \lambda' = \lambda + g^k \times \Delta \lambda,$$
$$v' = v + g^k \times \Delta v \tag{59}$$

In Eq. (59), g^k is the step size and updated by using the line search method in each iteration; g^0 and S^0 are initialized as 1 and $(1/N)(1,1,\cdots,1)^T$, respectively. The dual residual, primal residual, and centrality residual are updated in each iteration, and the optimization stops when the sum of the l_2 norms is less than 10^{-6}.

4.6 Experimental Results

4.6.1 Dataset, Metrics and setting

The newly proposed method was validated on a dataset containing 562 BUS images from a public benchmark [27] and a private dataset of 96 BUS images without tumors [44]. The two datasets are collected from the Second Affiliated Hospital of Harbin Medical University, the Affiliated Hospital of Qingdao University, and the Second Hospital of Hebei Medical University under different types of ultrasound devices, including GE VIVID 7(General Electric Healthcare, Chicago, IL, USA), GE LOGIQ E9(General Electric Healthcare, Chicago, IL, USA), Hitachi EUB-6500(Hitachi Medical Systems, Chiyoda, Japan), Philips iU22(Philips Healthcare, Amsterdam, Netherlands), and Siemens ACUSON S2000(Siemens Healthineers Global, Munich, Germany). The ultrasound images from Harbin Medical University were collected by GE VIVID 7 and Hitachi EUB-6500; the images from Qingdao University were collected by GE LOGIQ E9 and Philips

iU22; and Hebei Medical University used Siemens ACUSON S2000 to collect the ultrasound images. The ground truths of the images were generated by four experienced radiologists [27]. Informed consent to the protocol was obtained from the involved patients and all images in the datasets are de-identified. All experiments are conducted by using Matlab (R2018a, MathWorks Inc., MA) on a Windows-based PC equipped with a dual-core (3.6 GHz) processor and 8 GB memory.

Metrics of saliency estimation are the same as in section 2.4.1.

Parameter setting. All the experiments are based on the parameters: $\alpha = 4, \gamma = 40$.

4.6.2 Parameters tuning

Values of α and γ. As presented in the section **4.2**, the detection framework has 4 major parts. Applying one of the data terms cannot always provide the correct information to generate the saliency map (see Figs. 4.6-4.7.). The tuning parameter $\boldsymbol{\alpha}$ controls the relative impact of the data term and smoothness term on the optimization. And $\boldsymbol{\gamma}$ controls the balance effect of foreground cue and background cue. It evaluates the performance of the proposed method with α ranging from 0 to 100 and γ ranging from 0 to 3000, using randomly selected subset of 60 images from the 562 images (Xian et al. 2018b). There are three stages to choose the parameters. In the first stage, it makes the step size of α and γ be 50 and 500 respectively, and obtains the range of each parameter which can achieve better P-R curve performance and *MAE* value if the P-R curve is similar. In the second stage, the α step size is 20, and the γ step size is 100. And in the third stage, the α step size is 2, and the γ step size is 20.
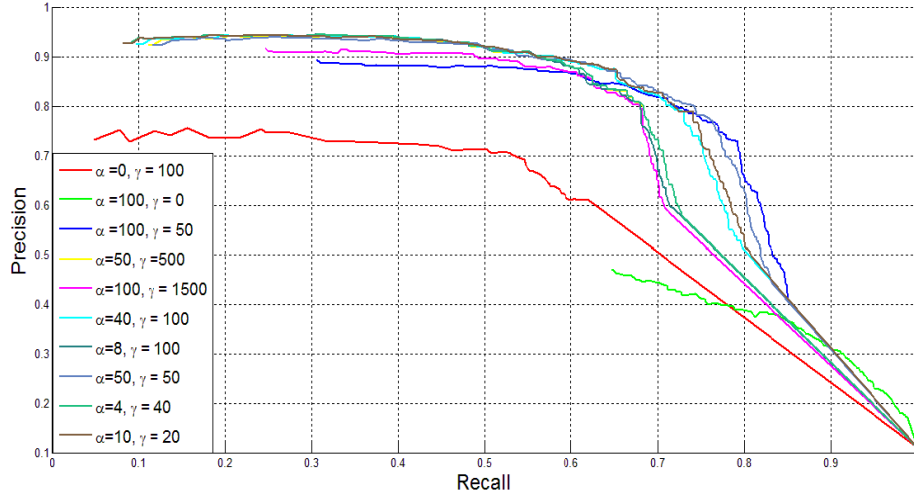
Fig. 4.8 The parameters α and γ tuning.

As shown in Fig. 4.8, the proposed approach achieves much better performance when the value of γ is much bigger than that of α; and when the value of α is less than 10, and γ is less than 50, the performances are similar on the P-R curves; therefore, based on the minimum *MAE,* $\alpha = 4$ and $\gamma = 40$.

Values of a and c. As presented in the section **4.3**, the value a in the Z function will make the regions with lower intensities have higher values because of the low-intensity appearance of the tumor, and value c will make the regions with high intensities have low values. We choose a and c based on tuning ε_1 and ε_2 in (41) using the training dataset with 60 images. We evaluate the P-R curve performances with ε_1 ranging [0.1, 0.4] and ε_2 ranging [0.5, 0.8] with the step size 0.1 and choose the parameters which can achieve the best P-R curve performance. Based on our experiments, the performances are competitive when ε_1 is fixed at less than the 0.2, and ε_2 varies in [0.5, 0.8]; and the best P-R curve was obtained when ε_1 is 0.1 and ε_2 is 0.6.
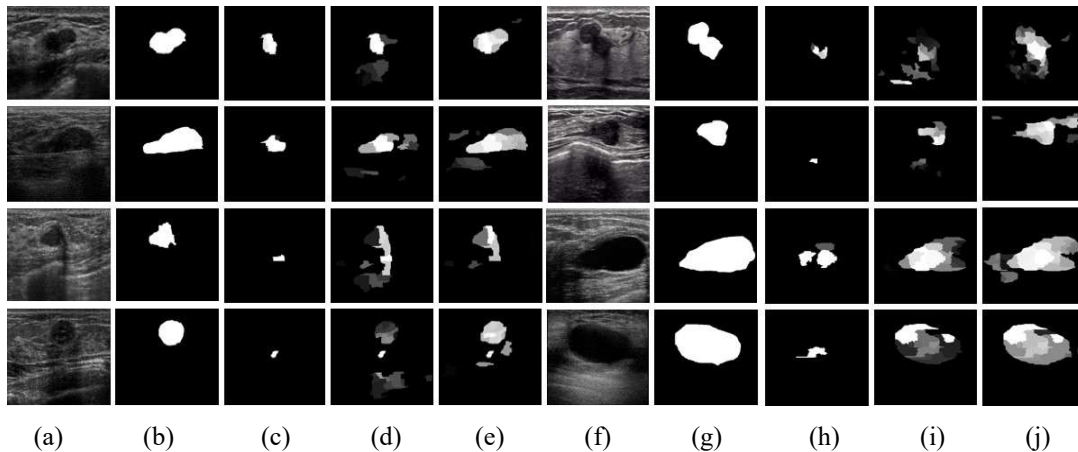
Fig. 4.9. Effectiveness of the breast anatomy. (a) and (f) Original BUS images; (b) and (g) the ground truths; (c) and (h) without layers' information in both terms, OUR_NL; (d) and (i) without layers' information in BG term, OUR_NL_BG ; and (e) and (j) the saliency map with layers' information in both terms.

4.6.3 The effectiveness of the breast anatomy

Here, it compares the methods without the layers' information in the FG and BG generation on the images in the dataset [27] remaining 502 images. In our methods, α and γ are set to 4 and 40 respectively. As shown in Fig. 4.9, the proposed method without layers' information in both terms, abbreviated as OUR_NL, will fail to locate the tumor or most parts of the tumor (see Figs. 4.9 (c) and (h)); and the proposed method without layers' information in the BG term, abbreviated as OUR_NL_BG, will locate the objects much more accurately than OUR_NL. However, it will miss some parts of the tumor (see the 1st and 2nd rows of Figs. 4.9 (d) and (i)) and cannot concentrate the high saliency values on the salient objects (see the 3rd and 4th rows of Figs. 4.9(d) and (i)). The overall performances of OUR_NL, OUR_NL_BG, and OURS in Figs. 4.13-4.14 demonstrate that the proposed method with the layers' information in the two terms is more robust than that without it.
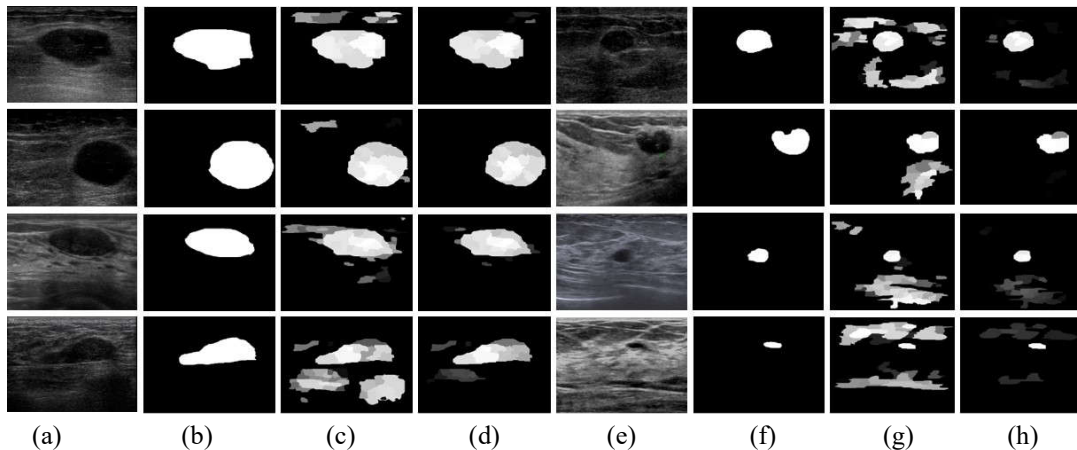
Fig. 4.10. Effectiveness of the new objective function. (a) and (e) original images; (b) and (f) the ground truths; (c) and (g) saliency maps obtained by OUR_OPT; and (d) and (h) saliency maps generated by the proposed method.

4.6.4 The effectiveness of the new objective function

We illustrate the effectiveness of the new objective function by two category samples. 1) on the dataset [27] remaining 502 images, apply FG and BG generated by the proposed method as the weighted map and NC map to the objective function in [44] which is one-way penalty objective function, abbreviated as OUR_OPT. Fig. 4.10 shows some comparable samples results. The OUR_OPT method always locates the tumor position correctly and generates good saliency map on the images with large tumors (see Figs. 4.10 (a) and (c)). However, it will make the other non-tumor regions have high saliency values on the images with small tumors (see Figs. 4.10 (e) and (g)). The method with the proposed objective function can concentrate high saliency values on the tumor regions and low values on the other regions. In addition, we apply the new method to the image without tumor, and compare the result with that of RRWR [17], SMTD [2], HFTSE [44] and OUR_OPT. The sample results are shown in Fig. 4.11. The saliency maps are normalized to [0, 1]. Fig.4.12 shows that the two penalty terms optimization framework can generate much more accurate saliency map than that of others.

<div align="center">(a)   (b)   (c)   (d)   (e)</div>
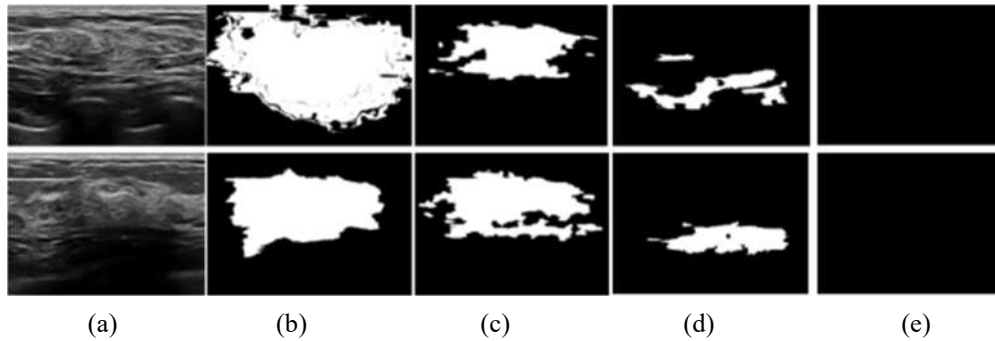
Fig. 4.11. Effectiveness of the new objective function on images without tumors. (a)-(e): original images, the saliency maps generated by [17],[2],[44], the new proposed approach, respectively. The ground truths will be an image with all pixels having intensities 0. The region with higher intensity indicates that the region with higher probability belongs to a tumor.

4.6.5 Overall performance

The proposed method is compared with most recently published methods SMTD [2], OMRC [3], MR [19], RRWR [17], HFTSE [44] and three models generated by the proposed method with different components in the optimization framework on the benchmark [27] remaining 502 images. RRWR, MR and OMRC are the bottom-up models and achieve good performances on the natural images. SMTD is the directly mapping method for tumor saliency estimation, and HFTSE is an optimization model to determine the existence of tumor and estimates tumor saliency for the image having tumors. OUR_NL is the two-penalty objective function with FG and BG maps in HFTSE; and OUR_NL_BG is the two-penalty objective function with the layered FG and the BG maps in HFTSE; and OUR_OPT is a one-penalty objective function with the layered FG and BG in HFTSE.

Fig. 4.12 shows the comparison results of the nine models. The proposed method and other two models OUR_NL_BG, OUR_OPT can locate the tumors accurately; especially, for the image with the big or small tumors. OUR_NL_BG model, without the layers' information in the BG terms, can generate a similar saliency map as the proposed
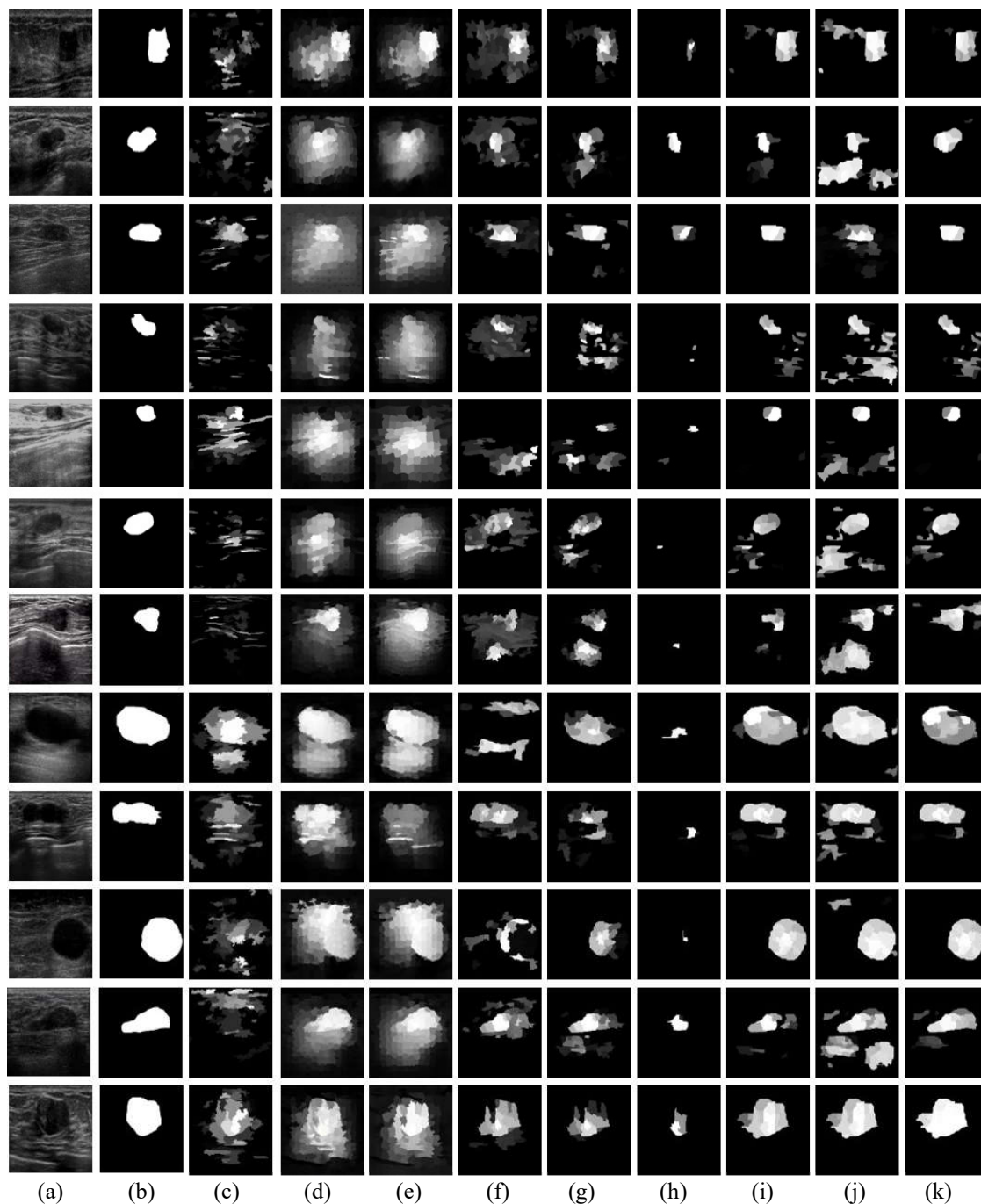
Fig. 4.12 Visual effects of detecting saliency maps by nine methods. (a) original images; (b) the ground truths; (c)-(k): the saliency maps generated by [3], [17], [19], [2],[44], OUR_NL, OUR_NL_BG,

method; however, it would miss some part of the tumor as described in section **4.3.2** *Background map generation*. OUR_OPT model can highlight the non-tumor regions as well as the tumor regions. Therefore, this model can achieve higher recall ratio, as shown

in Fig. 4.14. OUR_NL model, without the layers' information in both terms, will force the regions with very high value in FG and very low value in BG to have high saliency value. Thus, it will make a very small part of the object to be the salient object. It will cause a higher precision ratio but lower recall ratio as shown in Fig. 4.14. HFTSE would miss parts of large tumors and miss the entire object in the image with low contrasts (see the 5th row of Fig. 4.12). SMTD would miss the object in the images with very big or very small tumors (see the 5th, 8th and 10th rows of Fig. 4.12) and make the surround dark regions have high saliency values. OMTC, MR and RRWR worked better on the image with large tumors than that with small ones, even missed small tumors (see the 5th row in Fig. 4.12); moreover, these methods made the background regions around the tumors have higher saliency values. This situation will make these methods have higher recall ratios but lower precision ratios.

The overall performances of the nine models are shown in Figs. 4.13-4.14 using the metrics *MAE* values, $F_{meansure}$ values, and P-R curves. As shown in Fig. 4.13, the proposed method, noted as OURS, achieves a competitive P-R curve and the highest $F_{meansure}$ and the lowest *MAE*. As discussed, SMTD, MR, RRWR and HFTSE can obtain relatively high average recall ratios, but the precision ratios and F-measures are quite low. The reason is that these methods make the tumor and its surrounding background have high saliency values. OUR_NL only highlights a small part of the tumor as a salient object and can achieve the highest precision ratio and the lowest recall ratio. As shown in Figs. 4.13-4.14, OUR_NL_BG generates a much better overall performance than the model OUR_NL. It indicates that adding the layers' information on FG term will have a significant impact on improving the performance. Moreover, the model Ours with layers'

information on both terms generates better overall performance than the model OUR_NL_BG. It suggests that using layers' information on BG has a positive effect, but not as much as on FG.
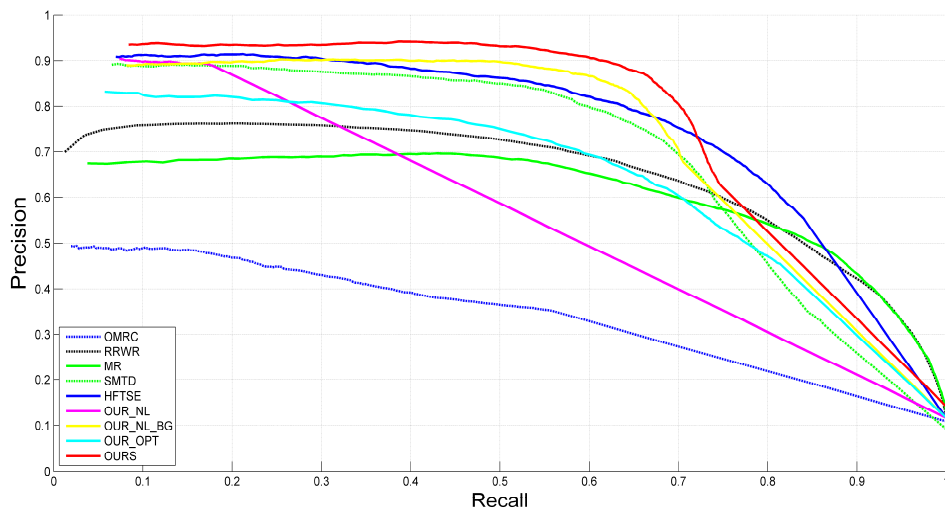
Fig 4.13. Precision-Recall curves of applying the nine models.
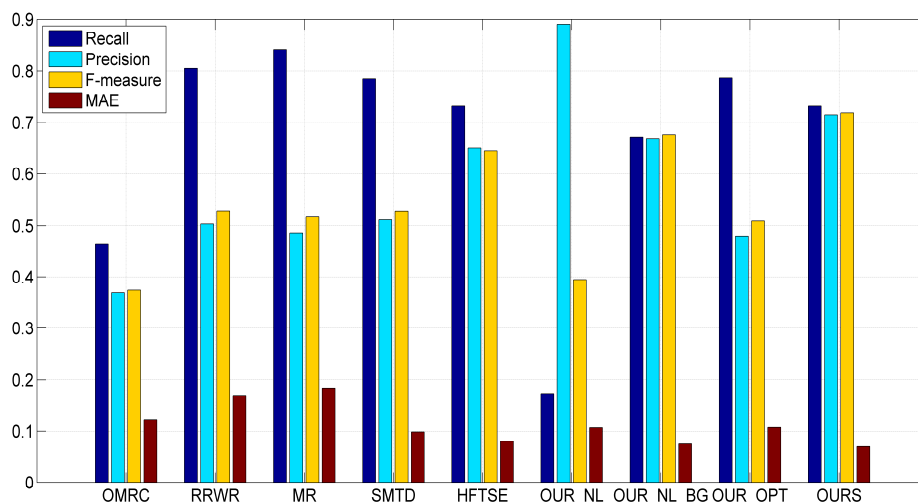
Fig. 4.14. The $F_{meansure}$, mean precision and recall ratios, and *MAE* of applying the nine models.

CHAPTER 5

BREAST ANATOMY ENRICHED TUMOR SALIENCY ESTIMATION

In this chapter, a new VSE approach is studied to detect breast tumors, which utilized a deep neural network to generate semantic breast anatomy. A new background map generation method weighted by the semantic probability and spatial distance was proposed to improve the performance.

5.1 VSE via deep neural network

Many automatic BUS segmentation approaches have been studied [25-34, 50-51]. Domain-related knowledge was utilized to locate tumors automatically in traditional methods. However, the performances of the models were instable due to collected images under various sources and periods using different machines with various qualities of the images, such as low contrast, more artifacts, etc. [50-51] proposed BUS segmentation models based on deep neural networks and [51] demonstrated that the CNN models could achieve much better performance than the traditional models. However, the two challenges existed: 1) no enough BUS image data available for training; 2) segmentation results completely based on the training dataset and the deep network.

Visual saliency estimation (VSE) measures the probabilities of human attention attracted by different image regions, which is essential and accessible for detecting the objects and achieving automatic segmentation [3,9,13,38,51,52]. Recently, CNN-based models are widely used to detect visual saliency with a specific task. Three strategies are
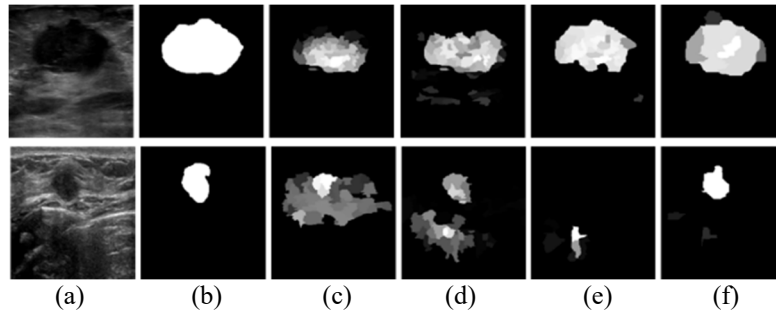
Fig. 5.1. Tumor saliency detection examples. (a) original images; (b)the ground truths; (c)-(f) the saliency maps generated by [2], [44], [55] and the proposed method, respectively.

employed in the most of CNN-based modes: 1) utilize more than one deep neural network to generate the saliency maps [51,52]; 2) integrate the high-level semantic knowledge by the deep neural network and low-level hand-craft features or visual saliency hypothesis [13,52]; 3) refine the object boundary in the final step [13]. The studies showed that CNN-based models generated much better performance than the bottom-up models.

For BUS images, many VSE methods have been investigated [1,2,44,55]. Examples are shown in Fig. 5.1. [55] presented a novel unsupervised framework to estimate the tumor saliency based on integrating breast anatomy modeling. It decomposed the BUS image into several horizontal layers by Neutro-Connectedness (NC) theory, which would make the regions with strong connectedness gather into the same layer. However, the generated layers cannot present semantic anatomy information. In this paper, we propose a novel TSE top-down model. Firstly, we utilize U-Net [50] to generate the initial four semantic breast anatomy layers (skin, fat, mammary, and muscle layers) [51]. Then it refines the wrong breast anatomy layers by combining the non-semantic decomposing layers based on NC theory (refer [29]). The final saliency maps are generated by the
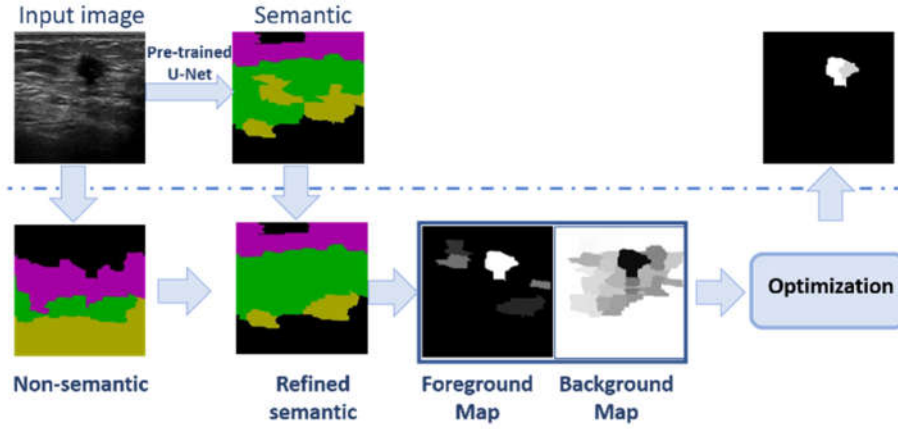
Fig. 5.2. The pipeline of the proposed model.

optimization framework integrating foreground cue, background cue, adaptive-center bias, and region-based correlation.

The pipeline of the proposed approach is shown in Fig. 5.2.

5.2 The proposed method

The proposed approach generates the tumor saliency map by the existing united optimization-based framework [55] integrating robust cognitive hypotheses, e.g., the adaptive center-bias, and region-based correlation, and the background and foreground cues. The saliency map is $S = (s_1, s_2, \cdots, s_N)^T$ which is a vector of saliency values, and s_i denotes the saliency value of the ith region and $s_i \in [0, 1]$. N is the number of superpixels generated by [45]. The optimization formulation is:

$$minimize\ E(S) = S^T\big(-(\alpha \ln(C) + \beta \ln(F))\big) +$$
$$\gamma(1 - S)^T(-\ln(T)) + \sum_{i=1}^{N}\sum_{j=1}^{N}(s_i - s_j)^2 r_{ij}\, D_{ij}$$

$$\text{subject to}\ \ 0 \leq s_i \leq 1, i = 1, 2, \cdots, N;$$
$$B^T S = 0, B = (b_1, b_2, \cdots b_N)^T,\ b_i = \{0,1\}$$
$$(60)$$

$$r_{ij} = exp\left(-\left|I'_i - I'_j\right|/\sigma_1^2\right) \tag{61}$$

$$D_{ij} = \exp\left(-\left\|rc_i - rc_j\right\|_2 / \sigma_2^2\right) \tag{62}$$

In Eq. (60), the term $T = (t_1, t_2, \cdots, t_N)^T$ is the background map, and larger t_i indicates the ith region belonging to the background with higher probability; the term $C = (c_1, c_2, \cdots, c_N)^T$ defines the coordinate distances between the regions' centers and the adaptive-centers, and larger c_i value indicates that the region is closer to the adaptive-center; the term $F = (f_1, f_2, \cdots, f_N)^T$ is the foreground map, and larger f_i indicates the higher probability of the ith region belonging to the foreground, and the terms r_{ij} and D_{ij} define the similarity and the spatial distance between the ith and the jth regions, respectively. The term $(1\text{-}S)^T(-\ln(T))$ defines the cost on the background map and forces the regions with smaller values in the background map to have higher values in the saliency map; specifically, assigns 1 to the regions with 0 values in the background map due to the higher penalty. The term $S^T(-\ln(C))$ defines the cost of the adaptive-center bias and forces the regions with larger distances to have smaller values in the saliency map; specifically, assigns 0 to the regions with the largest distance from the adaptive center. The term $S^T(-\ln(F))$ defines the cost of the foreground map and forces the regions with smaller value in foreground map to have smaller values in the saliency map; specifically, assigns 0 to the regions with 0 values in the foreground map due to the high penalty. The quadratic term models the region-based correlations which force similar regions with similar saliency values. Parameters α, β, and γ are used to balance the impact of each component. the equality constraint $B^T S = 0$ is applied to all the regions; b_i is 1 if the ith region is adjacent

to the image border, and 0 otherwise. In Eqs. (61) and (62), I'_i and rc_i are the intensity and region center of the ith region, respectively. $|\cdot|$ is the l_1 norm, $\|\cdot\|_2$ is the l_2 norm, and $\sigma_1^2 = \sigma_2^2 = 0.5$ by [55].

5.2.1 Semantic breast anatomy (SBA) map generation

Initial SBA map. The breast contains four primary layers: skin layer, fat layer, mammary layer, and muscle layer. Regions in different layers have different appearances, and the tumor always exists in the mammary layer. Due to the limitation of the number of training data, it is a challenge to generate accurate tumor segmentation results based on CNN.

The proposed approach utilizes the well-known U-Net [51], which consists of fully convolutional encoder and decoder sub-networks with skip connections. [52] demonstrated that the U-Net could generate good performance on a limited BUS images dataset for producing the initial SBA map. The number of convolutional filters in the network is (32, 32, 64, 64, 128). The input images have dimensions of 256×256 pixels, noted as *I*. The segmentation result of U-Net, *SA,* has dimension 256×256, and the segmentation probability map, *SP,* has dimension 4×256×256. $SP_{k,i,j}$ denotes the pixel in *SP* and indicates the probability of pixel *I(i,j)* belonging to the *k*th category. The value $SA_{i,j}$ is $k = \max(SP_{k,i,j}|_{k=0}^3)$. The pixel-based maps *I, SP* and *SA* are converted into the region-based maps I', SP' and SA' using the region-based optimization framework. The label with the largest value of the labels of each superpixel will be the region label. SA'_i indicates the

skin, fat, mammary, and muscle layer, and i is 1, 2, 3, 4, respectively. More details will be

discussed in section 5.3.2.

Table 5.1 Algorithm 2: Refine SBA map

Input: *SA′, NCL; SA′* contains 4 layers*, NCL* contains 3-5 layers by [55].
Output: *NSA;*

1. nl is the maximum layer of *NCL*
2. For each layer in *SA '*, check the validation of the layer by whether covering more than 75% columns of the image [55]
3. $NSA_1 = SA'_1 \cap NCL_1 - SA'_2 \cup SA'_3 \cup SA'_4$;
4. $NSA_5 = SA'_1 \cap NCL_{nl} - SA'_2 \cup SA'_3 \cup SA'_4$
5. **if** $nl = 3$
6. $NSA_2 = NCL_1 - SA'_1 \cup SA'_3 \cup SA'_4$;
7. $NSA_4 = NCL_{nl} - NSA_5 \cup SA'_3 \cup SA'_2$
8.   **if** $|SA'_1|/N > 0.5$ and $NSA_4 = \emptyset$ and SA'_4 is valid
9. $NSA_5 = \emptyset$
10. **if** $nl = 4$, Do the lines 12-22;
11. **if** $nl = 5$, change NCL_1 into NCL_2 ; and change NCL_{nl} and SA'_4 into NCL_{nl-1}, Do line 12-22
12. **if** the mammary layer SA'_3 is valid,
13.   **if** SA'_2 is valid and $SA'_3 \cap NCL_2 \neq \emptyset$,
14. $NSA_2 = (NCL_1 - SA'_1) \cup (SA'_2 \cap NCL_2)$
15.   **else**: $NSA_2 = NCL_1 - SA'_1 \cup SA'_3 \cup SA'_4$
16.   **if** SA'_4 is valid and $SA'_3 \cap NCL_{nl-1} \neq \emptyset$,
17. $NSA_4 = (NCL_{nl} \cap SA'_1) \cup SA'_4 - SA'_3$
18.   **else**: $NSA_4 = NCL_{nl} - NSA_5 \cup SA'_3$
19.   **else**
20. $NSA_2 = NCL_1 - SA'_1 \cup SA'_3 \cup SA'_4$;
21. $NSA_4 = NCL_{nl} - NSA_5 \cup SA'_3 \cup SA'_2$
22. The rest of the region will be in NSA_3

Refine SBA map. Based on observation, U-Net generates acceptable SBA map in

most cases, even using small training dataset. However, some anatomical layers missed

most of the parts and cross-layer (one layer is divided into more than one parts with no

connections by other layers) appears in some cases. It refines the original SBA map by the

NC anatomical map, named *NCL,* which decomposes the BUS image into several

horizontal layers by NC and the regions in the same layer with strong connectedness [55].

Algorithm 2 is to refine the initial SBA map by *NCL*. The refined map contains 5 layers;

the 1 and 5 layers are skin layers due to the skin layer existing in the top and bottom in some cases.

After refining, the new semantic anatomy (*NSA*) maps, are kept the same as *SA* in most of the cases (see the 1st row of Fig. 5.3). The refinement is to avoid cross-layer and incomplete layer, and to keep the high recall ratio on the mammary layer (see Fig. 5.3). It will reduce the probability of missing the tumor (see 2nd-5th rows of Fig. 5.3(c)) and recover the incomplete layer generated by the deep learning models (see Fig.5.3(e)).

5.2.2 Foreground map (FG) generation

The foreground map (FG) measures image regions' possibilities to be tumor regions. [55] proposed two algorithms to identify the dark/shadow layers and generate a foreground map for each layer, and it produced good results; especially, on the images with

(a)　　(b)　　(c)　　(d)　　(e)　　(f)　　(g)
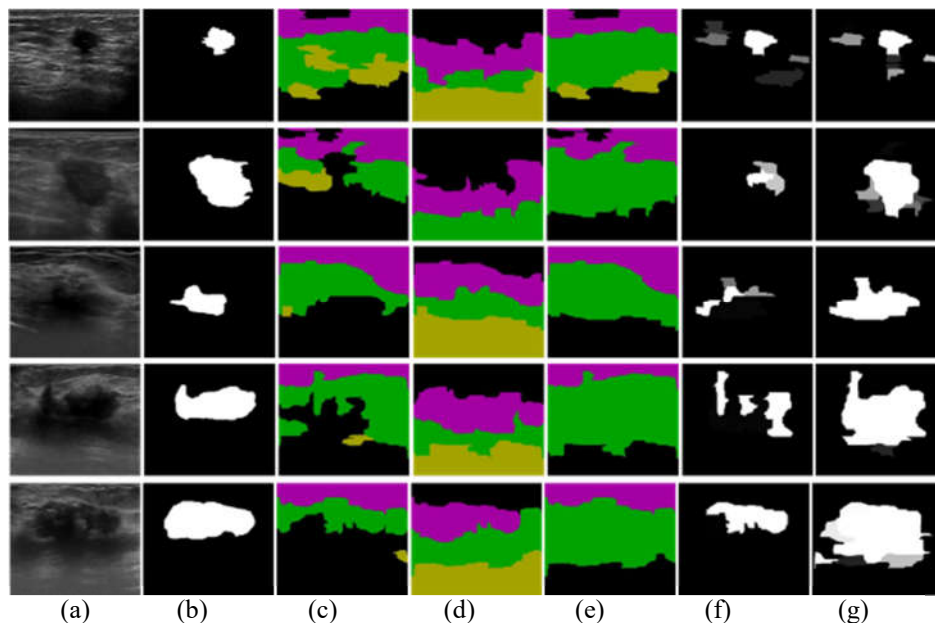
Fig.5.3 The visual effects of refining SBA maps. (a) original images; (b)ground truths; (c) SBA maps generated by U-Net;   (d) the non-semantic layers generated by [55]; (e) the refined SBA maps; (f) the FG based on (c); (g)the FG based on (e).

large and/or small tumors. We applied the algorithm in [55] to output a flag with three values. If $flag$ =-1, it indicates a smooth layer (most of the regions in the layers with high intensities); If $flag$ =1, it indicates a dark/shadow layer; otherwise, it is a normal layer. We adopt the same algorithm to identify the dark layer and employ the Z-function to generate the FG. The effectiveness is shown as Figs. 5.3 (f) and (g).

5.2.3. Adaptive-center distance map generation

[3] proposed the adaptive-center bias instead of the fixed image center bias which estimated the adaptive center (AC) using weighted local contrast map on natural images. [44, 55] demonstrated the effectiveness of generating the AC by weighted foreground map on BUS images. In this approach, we adopt the method to generate the AC and the distance map C. $c_i = \exp\left(-\|rc_i - AC\|_2/\sigma_3^2\right)$ where the rc_i is the center coordinate of the ith region; $\|\cdot\|_2$ is the l_2 norm and $\sigma_3^2 = 0.1$ [55].

5.2.4. Background map (BG)

Boundary connectivity is an effective prior utilized in many visual saliency estimation models [1, 27, 29, 44, 45, 55]. [44, 55] has demonstrated the boundary connectivity based on NC theory which calculates neutron connectedness between the regions and the boundary regions [29] is effective to avoid noisy data and generate much smoother and more accurate background map on BUS images. Therefore, we generate the NC map by the algorithm in [29] and note nc_i as the NC value of the ith region in the NC map. The value of the ith region in the initial BG map is defined as $t_i = nc_i^2$.

Meanwhile, we define the layer weight based on the region-based semantic probability maps SP' (refer the section **5.2.1**). The initial weight for each layer is defined as the mean value of the probabilities that regions in the layer belong to the mammary layer. If the mammary layer is valid, we assign $rW_{k,i} = \max(SP'_{3,i}, LayerW_k)$; otherwise, assign all the regions with weight 1.

$$LayerW_k = \left.\sum_{i \in NSA_k} SP'_{3,i} \middle/ |NSA_k| \right. \tag{63}$$

where $LayerW_k$ is the probability of the kth layer belonging the mammary layer; $SP'_{3,i}$ is the predicted probability of the ith region belonging to the mammary layer; $rW_{k,i}$ is the probability of ith region in the kth layer belonging to the mammary layer.

To avoid the isolate region in the non-mammary layer with very small nc value, we define the final value of the ith region in the initial BG map weighted by the probability of mammary layer and the distance from the AC, noted as $t_i = 1 - (1 - nc_i^2) \times rW_{k,i} \times c'_i$. c'_i is assigned to be 1 if $flag = 1$ in the mammary layer and $c_i > 0.5$ or $flag = 0$ in the non-mammary layers and $c_i \geq 0.75$; otherwise $c'_i = c_i$. The T should be normalized. The new defined background map will avoid the situation that some isolating non-tumor regions obtain the lowest values and decrease the saliency values of the tumor regions a lot even the tumor regions gains the highest values in the foreground map (see Fig. 5.4). The effectiveness of the new background map will be discussed in section 5.3.3.

| (a) | (b) | (c) | (d) | (e) | (f) | (g) |
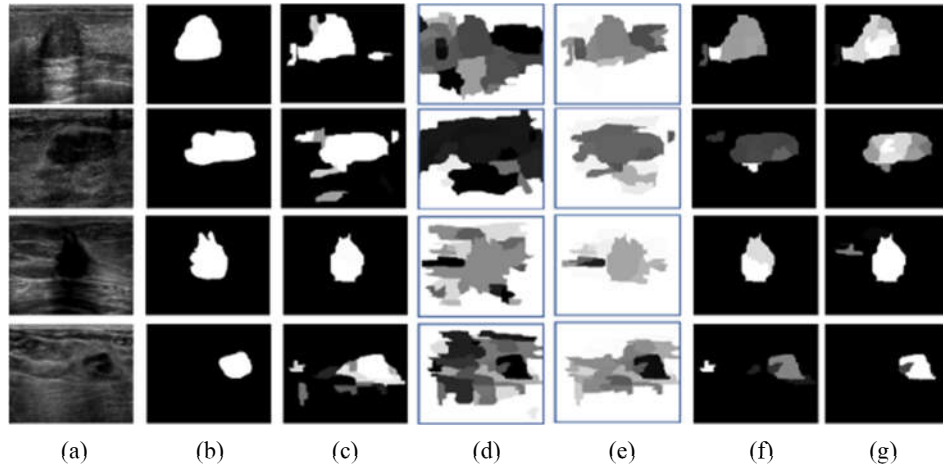
Fig 5.4　The effects of different components in the objective function. (a) original images; (b) ground truths ; (c) the FG; (d) the BG map in [55]; (e) the proposed BG; (f) the saliency map based on (d); (g) the saliency map using new BG.

5.2.5. Optimization

The optimization framework is similar as that in section **4.2**, therefore, we utilize the same optimization method with the same initial and stop conditions as section **4.5**.

The Fig.5.4 shows the final optimal saliency maps generated with different components in the objective function. The model with the BG in [55] will decrease the tumor saliency values when non-tumor regions gain the lowest value in BG (see Fig.5.4(g)). The overall performance will be discussed in section **5.3.3**.

5.3 Experimental Results

5.3.1 Datasets, metrics and setting

We train and test the U-Net using a dataset with 325 images. 229 images in the dataset contain breast tumors, and other 96 images have no tumors [51]. The training and validation dataset are random chosen 90% images from the total dataset (the images with

tumor and without tumor are 90%, respectively), and the ratio of training set and validation

set is 8:2. The rest 10% dataset is used as test dataset. The training set is that the number

of epochs is 100, and the batch size is 5, and learning rate is 0.1.

We validate the performance of the newly proposed TSE method using a dataset

containing 562 BUS images from a public benchmark [27]. Due to tuning the parameters

in Eq. (60), we randomly choose 60 images as a training dataset, and the rest is utilized to

evaluate the overall performance.

Metrics of saliency estimation are the same as section **2.4.1.**

Parameter setting: all the experiments are based on $\alpha = 10, \beta = 51, \gamma = 6$.

5.3.2 Parameters tuning

As presented in section **5.2.1**, there are four components in the objective function.

Therefore, the parameters α, β, and γ are utilized to balance the impact of each components and

generate better performance. The tuning parameter α controls the relative impact of the adaptive-

center term, and β controls the relative impact of the foreground map, and γ controls the relative

impact of the background map. The larger value of α, β and γ indicates the corresponding term

has greater impact on the performance. We evaluate the performance of the proposed method on

randomly selected subset of 60 images from the 562 images [27] under different parameters and

choose the range of each parameter which could obtain the better P-R curve and *MAE* value if the

P-R curve is similar. Since the objective function in the paper is similar with that in [55], we adopt

the initial ranges for the three parameters based on the experiments in [55] that the performances
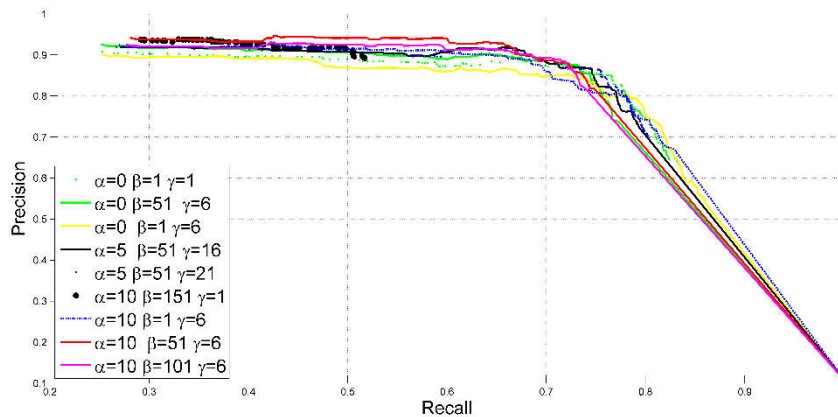
Fig. 5.5. The parameters tuning.

are similar when the parameter on the foreground map is less than 50, and the parameter on the

background map is less than 10. Therefore, we range α from 0 to 10 with step size 5, β from 1 to

151 with step size 50 and γ from 1 to 21 with step size 5. As shown in Fig.5.5, the P-R curves are

competitive under most of the parameter combination, and it achieves a better P-R curve and *MAE*

when α is 10, and β is 51, and, γ is 6.

5.3.3 The overall performance of the proposed method

The proposed model is compared with most recently published TSE methods

SMTD [2], HFTSE [44], TBAM [55], and two models generated by the proposed method

with the background map generated by different strategies on the 502 images. SMTD,

HFTSE, and TBAM are the bottom-up VSE models with the specific breast tumor

appearance knowledge. SMTD defined a unified global contrast mapping to estimation the

tumor saliency. HFTSE proposed an optimization TSE model after determining the

existence of tumor. The proposed method is noted as OURs, with the local contrast strategy

generating the FG map and the new proposed BG map. OUR_ BG1 is the optimization

(a)   (b)   (c)   (d)   (e)   (f)   (g)
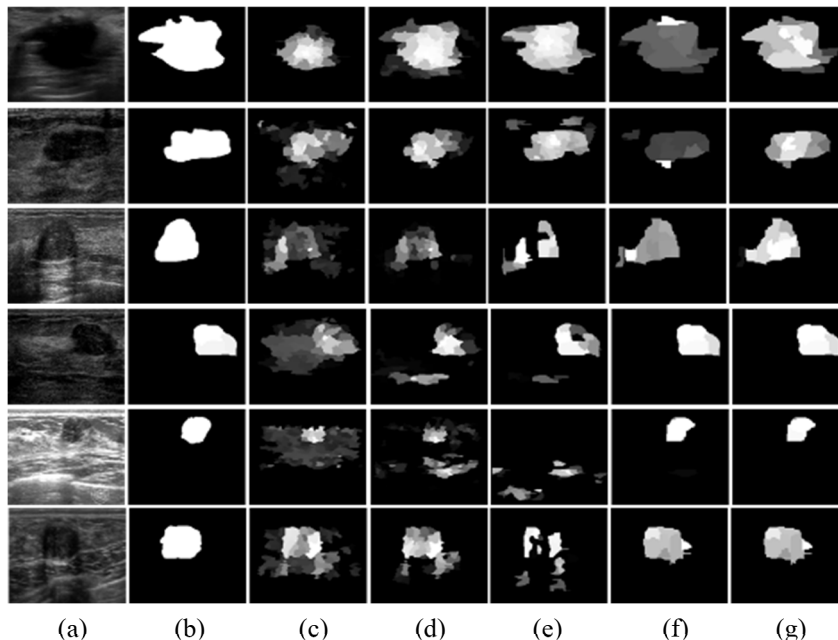
Fig. 5.6 The visual effects of detecting the saliency maps by the five models. (a) original images; (b)ground truths; (c)-(g) the saliency maps generated by [2], [44], [55] OUR_BG1 and OURs, respectively.

model with the BG map generated by HFTSE in Eq. (60) and the FG map with local contrast strategy.

The comparison visual effects of detecting saliency map by the five models are shown in Fig. 5.6. OUR_ BG1 with the background map generated by TBAM obtains a similar saliency map on most of the cases that the tumor regions gain the lowest values in the background map. However, it will only highlight the non-salient regions and decrease the saliency values of the tumor regions when there are some isolating non-tumor regions obtaining the lowest values in the background map (see Fig. 5.4 and the 1st -3rd rows of Fig. 5.6). SMTD would miss the parts of object in the images with very big (see the 1st and 2nd rows of Fig. 5.6) and make the surround dark regions have high saliency values (see the 3rd -6th rows of Fig. 5.6). This situation will make the model SMTD could achieve high

recall ratio but low precision ratio (see Fig. 5.7). HFTSE would miss parts of large tumors and make the surrounding tumor regions have high saliency values (see the 2nd -6th rows of Fig. 5.6). TBAM model generated accurate saliency maps when detecting the correct layer with tumor (see 1st ,2nd and 4th rows of Fig. 5.6), especially for the images with large or small tumors, but it failed in the cases that the tumor in the top or bottom part of the images(see 2nd row of Fig. 5.1 and 5th row of Fig.5.6).
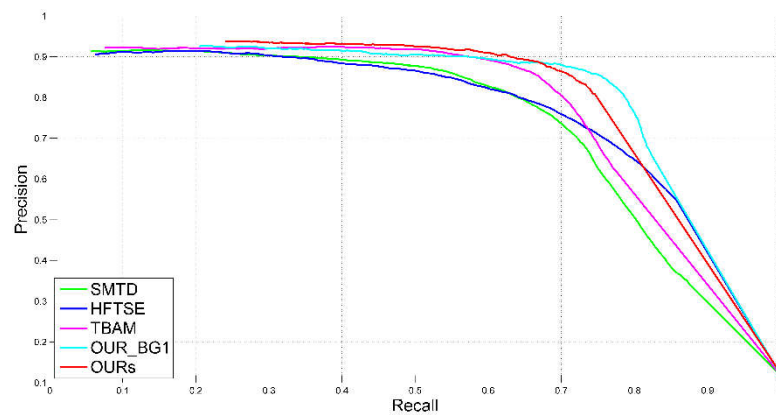
Fig. 5.7 The P-R curves of the five models.
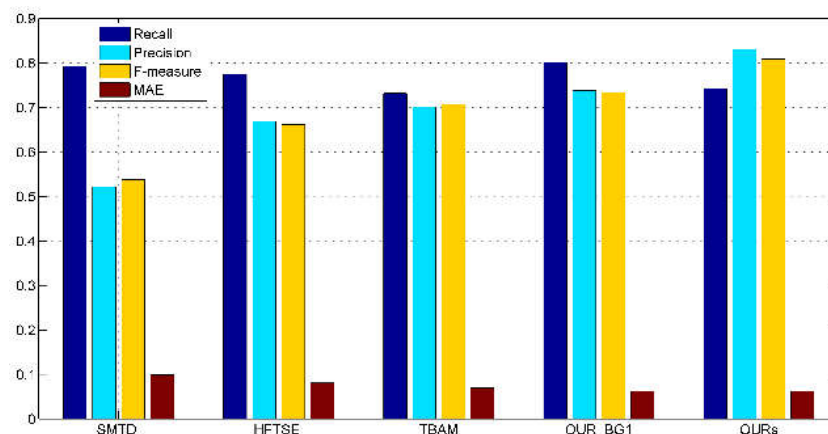
Fig. 5.8 The metrics MAE and $F_{meansure}$ values of the five models.

The overall performances of the seven models are shown in Fig. 5.7-5.8. As shown in Fig.5.7-5.8, the proposed model OURs achieves the best P-R curve and lowest *MAE*,

highest $F_{meansure}$ values. As discussed, SMTD and HFTSE achieved relative high recall

ratio and quite low precision ratio due to highlight the surrounding tumor regions as well

as tumor regions. TBAM achieved a balance metrics MAE and $F_{meansure}$ values, but it

failed in the cases when detecting the tumor layer wrong. This situation will make the

model hardly achieve better overall performance without guided by semantic knowledge.

OUR_BG1 generates the competitive P-R curve as OURs. However, OURs achieves much

better $F_{meansure}$ values, which indicates a better background map generation will improve

the TSE performance a lot.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This dissertation is comprised of my main work in Visual Saliency Estimation

(VSE) and applications. The main contributions are summarized as follows.

(1) The new VSE model in natural images formulates saliency estimation as a

quadratic program (QP) problem based on robust hypotheses. First, we propose an adaptive

center-based bias hypothesis to replace the most common image center-based center-bias.

It calculates the weighted center by utilizing local contrast, which is much more robust

when the objects are far away from the image center. Second, we model smoothness term

on saliency statistics of each color. It forces the pixels with similar colors to have similar

saliency statistics. The proposed smoothness term is more robust than the smoothness term

based on region dissimilarity when the image has a complicated background or low

contrast.

(2) The proposed hybrid framework for tumor saliency estimation (TSE) is to detect

the breast tumor on the ultrasound images by modeling radiologists' attention mechanism.

The TSE framework involves two steps: tumor existences and tumor saliency estimation.

The decision tree is utilized to determine tumor existences based on the foreground map

information. In the second step, the optimization framework integrates both high-level

domain-knowledge (the Neutro-Connectedness (NC) map) and robust low-level saliency

assumptions (the adaptive-center and the correlation) to improve the performance.

(3) We propose a novel optimization model to estimate tumor saliency for BUS images by integrating breast anatomy knowledge. Breast anatomy modeling solves the missing boundary problem caused by the shadows. Meanwhile, applying the combination of global and local parameters in each layer solves the problem of missing parts of large tumors. Extensive experiments demonstrate that more accurate foreground and background maps are generated when the images have large or small tumors. Moreover, we propose a new objective function to handle BUS images without tumors. The new objective function could be applied to other optimization frameworks to perform natural image saliency estimation, object detection, and segmentation as well. The strategy of modeling the anatomy knowledge by Neutro-connectedness theory can be applied to model tissue relationships in medical images of many tasks such as thyroid tumor detections, liver tumor detections, lung tumor detections, and abdominal small organ detections.

(4) We propose a novel TSE model guided by the semantic breast anatomy knowledge for BUS images. In the novel model, the non-semantic breast anatomy modeling is integrated to solve the cross-layer and incomplete mammary layer in the semantic anatomy map generated by U-Net. The strategy is effective when the semantic information could not have been generated accurately due to limited data or unknown data. A new background map generation method is proposed to improve the performance, which is weighted by the semantic probability and spatial distance on the mammary layer.

In the future, I will focus on improving the performance of tumor saliency estimation based on the deep neural network. Moreover, we will also explore the possibility

of applying Visual Saliency Estimation models to other biomedical image data, such as
Computed Tomography Images, X-ray Images, Magnetic Resonance Imaging, and
Biomedical Sequence Images.

REFERENCES

[1] Y. Xie, K. Chen, J. Lin, An Automatic Localization Algorithm for Ultrasound Breast Tumors Based on Human Visual Mechanism, Sensors-Basel, 17 (2017) 1101.

[2] H. Shao, Y. Zhang. A saliency model for automated tumor detection in breast ultrasound images. Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, 2015.

[3] F. Xu, M. Xian. Unsupervised saliency estimation based on robust hypotheses. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016.

[4] L. Itti, C. Koch, E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI 20 (11): 1254-1259, 1998.

[5] E. Niebur and C. Koch, Computational Architectures for Attention. The Attentive Brain, pp. 163–186. Cambridge, Mass.: MIT Press, 1998.

[6] X. Hou. and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, 2007.

[7] C. Guo, Q. Ma, L.-M. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In CVPR 2008.

[8] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In CVPR 2009.

[9] M.-M Cheng, G.-X. Zhang, N.-J. Mitra, X.-L. Huang and S.-M. Hu. Global contrast based salient region detection. In CVPR, 2011.

[10]   K.-Y. Chang, T.-L. Liu, H.-T. Cheng and S.-H. Lai. Fusing generic objectness and visual saliency for salient object detection. In ICCV, 2011.

[11]   X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In CVPR, 2012.

[12]   H.Jiang, J. Wang, Z.-J. Yuan, Y. Wu, N.-N Zheng and S.-P. Li. Salient object detection: A discriminative regional feature integration approach. In CVPR, 2013.

[13]   J. Kim, D. Han, Y.-W. Tai and J.-M. Kim. Salient region detection via high-dimensional color transform. In CVPR, 2014.

[14]   J. Li, Y. Tian, L. Duan, and T. Huang, Estimating visual saliency through single image optimization, IEEE Signal Processing Letters, vol. 20, no. 9, pp. 845–848, 2013.

[15]   B. Jiang, L. Zhang, H.-C. Lu, C. Yang and M.-H. Yang. Saliency detection via absorbing markov chain. In ICCV, 2013.

[16]   W. Zhang, Q. Xiong, W.-R. Shi and S.-H. Chen. Region saliency detection via multi-feature on absorbing Markov chain. The Visual Computer, 2015.

[17]   C. Li, Y. Yuan, W. Cai, Y. Xia. Robust saliency detection via regularized random walks ranking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015.

[18]   P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. International Journal of Computer Vision 59(2): 167-181, 2004.

[19]   W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency Optimization from Robust Background Detection. In CVPR, 2014

[20]   S. Alpert, M. Galun, R. Basri, and A. Brandt, Image segmentation by probabilistic bottom-up aggregation and cue integration, In CVPR, 2007.

[21]   Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, In CVPR, 2013.

[22]   Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In ACM Multimedia, pp. 815–824, 2006.

[23]   S. Goferman, L. Zelnik-Manor and A. Tai. Context-aware saliency detection. PAMI 34(10): 1915-1926, 2012.

[24]   R. Siegel, D. Naishadham, A. Jemal, Cancer Statistics, 2012, CA-Cancer J. Clin., 62 (2012) 10-29.

[25]   H.D. Cheng, J. Shan, W. Ju, Y.H. Guo, L. Zhang, Automated breast cancer detection and classification using ultrasound images: A survey, Pattern Recognt., 43 (2010) 299-317.

[26]   M. Xian, Y. Zhang, and H.D. Cheng. Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains, Pattern Recognition, vol. 48, no. 2, pp. 485– 497, 2015.

[27]   M. Xian, Y. Zhang, H. D. Cheng, F. Xu, K. Huang, B. Zhang, J. Ding, C. Ning, Y. Wang, A Benchmark for Breast Ultrasound Image Segmentation (BUSIS), https://arxiv.org/submit/2126838.

[28]   Z. Hao, Q. Wang, Y. K. Seong et al., "Combining CRF and multi-hypothesis detection for accurate lesion segmentation in breast sonograms," in MICCAI, 2012, pp. 504-511.

[29]   M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, and J. Ding, "Neutro-connectedness cut," IEEE Transactions on Image Processing, vol. 25, no. 10, pp. 4691-4703, 2016.

[30]   M. Xian, "Neutro-Connectedness Theory, Algorithms and Applications," Utah State University, 2017.

[31]   M. Xian, H. Cheng, and Y. Zhang, "A Fully Automatic Breast Ultrasound Image Segmentation Approach Based on Neutro-Connectedness." In ICPR, 2014, pp. 2495-2500.

[32]   J. Shan, H. D. Cheng, and Y. X. Wang, "Completely Automated Segmentation Approach for Breast Ultrasound Images Using Multiple-Domain Features," Ultrasound Med. Biol, vol. 38, no. 2, pp. 262-275, Feb, 2012.

[33]   J. Shan, H. Cheng, and Y. Wang, "A novel segmentation method for breast ultrasound images based on neutrosophic l-means clustering," Medical Physics, vol. 39, no. 9, pp. 5669-5682, 2012.

[34]   Z. Hao, Q. Wang, H. Ren et al., "Multiscale superpixel classification for tumor segmentation in breast ultrasound images," in IEEE ICIP, 2012, pp. 2817-2820.

[35]   J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, F. Wu, Background prior-based salient object detection via deep reconstruction residual, IEEEbTrans. Circuits Syst. Video Technol. 25 (8) (2015) 1309-1321

[36]   X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, in: Proc. 14th IEEE Int. Conf. Comput.Vis., Sydney, NSW, Australia, 2013, pp. 2976-2983.

[37]   Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: Proc. 12th Eur. Conf. Comput. Vis., Firenze, Italy, 2012, pp.29-42.

[38]   J. Wang, H. Lu, X. Li, N. Tong, and W. Liu, Saliency detection via background and foreground seed selection, Neurocomputing, 2015, 152, pp. 359-368

[39]   C. Yang, L. Zhang, H.-C. Lu, R. Xiang and M. -H. Yang. Saliency detection via graph-based manifold ranking. In CVPR, 2013

[40]   T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, Learning to detect a salient object, PAMI, 33(2), pp. 353–367, 2011.

[41]   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 (2018) 394–424. doi:10.3322/caac.21492.

[42]   H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du, Approaches for automated detection and classification of masses in mammograms, Pattern Recognit. 39 (2006) 646–668. doi:10.1016// j.patcog.2005.07.006.

[43]   M. Xian, Y. Zhang, H.D. Cheng, F. Xu, B. Zhang, J. Ding, Automatic breast ultrasound image segmentation: A survey, Pattern Recognit. 79 (2018) 340–355. doi:10.1016/ j.patcog.2018.02.012.

[44]   F. Xu, M. Xian, Y. Zhang, K. Huang, H.D. Cheng, B. Zhang, J. Ding, C. Ning, Y. Wang, A hybrid framework for tumor saliency estimation, in: 2018 24th Int. Conf. Pattern Recognit., IEEE, 2018: pp. 3935–3940. doi:10.1109/ICPR.2018.8545599.

[45]   A. Vedaldi, S. Soatto, Quick shift and kernel methods for mode seeking, in: Comput. Vis. – ECCV 2008, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008: pp. 705–718. doi:10.1007/978-3-540-88693-8_52.

[46]   P. Kovesi, Image features from phase congruency, Videre J. Comput. Vis. Res. 1(3) (1999) 1–26.

[47]   S.H.P. A. Thomas Stavros, Cynthia L. Rapp, Breast anatomy: the basis for understanding sonography, in: Breast Anat. Basis Underst. Sonogr., Lippincott Williams & Wilkins, Philadelphia, 2004: pp. 56–108.

[48]   P.J. Toivanen, New geodosic distance transforms for gray-scale images, Pattern Recognit. Lett. 17 (1996) 437–450. doi:10.1016/0167-8655(96)00010-4.

[49]   S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, Cambridge, 2004, ch. 4, pp. 136-156.

[50]   O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 224–241.

[51]   K. Huang, H. D. Cheng, Y. Zhang, B. Zhang, P. Xing, and C. Ning, "Medical Knowledge Constrained Semantic Breast Ultrasound Image Segmentation," International Conference on Pattern Recognition, 2018, vol. 2018-August, pp. 1193–1198.

[52]   L. Wang, H. Lu, X. Ruan, and M. H. Yang, "Deep networks for saliency detection via local estimation and global search," IEEE Conference on Computer Vision and Pattern Recognition, vol. 07-12-June-2015, pp. 3183–3192.

[53]   G. Lee, Y. W. Tai, and J. Kim, "Deep saliency with encoded low-level distance map and high-level features," IEEE Conference on Computer Vision and Pattern Recognition, vol. 2016-December, pp. 660–668.

[54]   A. Vakanski, M. Xian, P. Freer. "Attention Enriched Deep Learning Model for Breast Tumor Segmentation in Ultrasound Images.," Ultrasound in Medicine & Biology, 2020 (Accepted).

[55]   F. Xu, Y. Zhang, M. Xian, H. D. Cheng, B.Zhang, J. Ding, C. Ning, and Y.Wang, "Tumor Saliency Estimation for Breast Ultrasound Images via Breast Anatomy Modeling," arXiv preprint arXiv:1906.07760, 2019.

CURRICULUM VITAE

Fei Xu
(2020)

EDUCATION

Ph.D., School of Computer Science   **April 2020**
Utah State University, US.
Advisor: Hengda Cheng, professor.
GPA: 4.0

M.S., School of Computer Science and Technology   **July 2011**
Harbin Institute of Technology (HIT), Harbin, China.
GPA ranks in top 10% of 197 students.

B.S., School of Computer Science and Technology   **July 2009**
Northeast Normal University, Changchun, China.
GPA (3.91) ranked in top 2% of 190 students.

RESEARCH INTERESTS

- Artificial Intelligence
- Data analysis
- Computer vision
- Image processing

HONORS, CERTIFICATES, ANDAWARDS

- Doctoral Student Researcher Award, Computer Science, Utah State University, 2019
- Teaching Assistant Certificate, Utah State University, 2014
- Best Employee, China Mobile, China, 2013
- Project Model, China Mobile, China, 2013
- ISTQB Certified Tester, China Mobile, China, 2013
- Project Management Training, China Mobile, China, 2013
- Qualification Certificate of Computer and Software Technology Proficiency, China Mobile, China, 2013
- People Fellowship, Top Grade, Harbin Institute of Technology, China, 2011
- People Fellowship, Second Grade, Harbin Institute of Technology, China, 2010
- National Scholarship, Northeast Normal University, China, 2009
- People Fellowship (three times), Northeast Normal University China, 2005 – 2009

WORK EXPERIENCE

- Tencent (Forbes Global 2000, 2018: 105th), Shenzhen, China   Summer 2010
  R&D engineer in the Data Sorting Analysis Group
  Improved the efficiency of previous data sorting algorithms; built a benchmark to evaluate the performance of clustering methods; and proposed an efficient method for duplicate data removal.

- Harbin Institute of Technology, Harbin, China   September 2009 - June 2011
  Research Assistant
  Data retrieval and sorting using random projection. Proposed a new method to transfer variable-length vectors into vectors with constant number of low dimensions by using random projection. The proposed approach was applied to image retrieval and sorting.

- China Mobile (Forbes Global 2000, 2018: 25th), Shenzhen, China   2011 - 2014
  Software Engineer
  Worked as a project lead to design and evaluate a new automatic test system which improved the test efficiency significantly and generated one patent; worked as a core member of the Business Integration Test Project; and worked as a core member of the functional test of the online Mobile Mall.

RESEARCH EXPERIENCE

Utah State University, Logan, UT   September 2014 – April 2020

Research Assistant

(1) Theory and algorithms for data topology modeling

• Built the generalized Neutro-Connectedness (NC) based on local similarity and inhomogeneity, which enabled the NC to be independent of top-down prior from the task domain and to have wide potential applications like data analysis, image segmentation, saliency estimation, object detection, etc.;

• Implemented NC computation algorithm that takes polynomial time to output NC values and NC forest simultaneously. The NC forest explored the topological structure of datasets.

(2) Optimization frameworks for object detection and segmentation

• Modeled psychological and cognitive hypotheses in a unified optimization framework to simulate the human visual attention mechanism to detect salient objects. Extensive

experiments demonstrate that our method outperformed ten state-of-the-art methods on three public benchmark datasets.

• Two optimization frameworks integrating low-level image features and high-level topological structure of biomedical image were proposed to detect breast tumor. The proposed models achieved the best performance on the public dataset among the state-of-the-art methods.

(3) Multi-instance learning (MIL)with local weighted citation-KNN
MIL was proposed to solve learning problems with incomplete information about labels of training data. We model two categories of image object detection as a MIL problem, in which each image is viewed as a bag and nonoverlap sub-regions on the image are considered as instances. We proposed the local weighted citation-KNN approach that decides the label of each test image by considering the local distribution of samples (voters). In comparison with Citation-KNN and other methods, the proposed algorithm demonstrates competitive classification accuracy and adaptability.

(4) Domain-enriched deep neural networks for data analysis
Integrated both high-level semantic priors and low-level breast ultrasound information in a Quadratic Programming (QP)framework with equality and inequality constraints. The proposed model outperformed 4 state-of-the-art methods on a public dataset with 562 images.

Harbin Institute of Technology, Harbin, Heilongjiang, China

Research Assistant   September 2009 - June 2011
Image retrieval and sorting by using random projection. We proposed a new method to transfer the original high-dimensional feature vector into features with constant number of low dimensions by using random projection. The proposed approach was applied to image retrieval and sorting

TEACHING EXPERIENCE

Teaching Assistant, Utah State University
Collaborated on class development, tutored students during office hour or upon requests and graded all assignments and projects for:
  * Python Introduce(CS1440),   Fall 2018 - Spring 2020
  * Operating Systems and Concurrency(CS3100),   Spring 2018
  * Computer Vision: Foundations and Applications (CS5680),   Fall 2017
  * Computer Systems, Organization and Architecture(CS 3810),   Spring 2015-2017,
                                                                        Fall 2015
  * Foundations of Computer Science Lecture(CS 1030),   Fall 2016
  * C++ Programming (CS1410),   Fall 2014

Instructor, Utah State University   Fall 2014
Worked as lab instructor for C++ Programming s(CS1410) and given 3 lectures about sorting technology.

Lab Instructor, Harbin Institute of Technology   Spring 2011
Lab instructor for undergraduate C++ programming course. Lectured and tutored for the weekly lab session. Developed course projects and graded assignments, exams and projects.

Systematic Teaching Training, Northeast Normal University, China   2005 - 2009
Obtained the Teacher Qualification Certificate after the completion of courses in the following areas:
- Teacher and Teaching Methodology
- The Basic Theory of Curriculum and Pedagogy
- Adolescent Learning and Development
- Mental Health and Counselling
- Philosophy of Education

VOLUNTEER

Lecturer at The Second Middle School of Fusong, Jilin Province, China   2008 - 2009
Prepared lectures and taught basic computer skill (Microsoft Word, PowerPoint, Excel, et al.). 3 classes every week, and 45 students per class.

PROGRAMMING SKILLS

- Proficient: C++, C, Matlab
- Familiar: Python

PROFESSIONAL SERVICE

Reviewed manuscripts for
- Pattern Recognition
- IEEE Transactions on Image Processing
- Neurocomputing

Publications

Under Review

1. [C] F. Xu, Y. Zhang, M. Xian, H. D. Cheng,B. Zhang,J. Ding, C. Ning,and Y.Wang. Breast Anatomy Enriched Tumor Saliency Estimation, submitted to International Conference on Pattern Recognition (ICPR), 2020.

2. [J] F. Xu, M. Xian,Y. Zhang, H. Cheng,B. Zhang,J. Ding, C. Ning,and Y.Wang. Tumor Saliency Estimation for Breast Ultrasound Images via Breast Anatomy Modeling, submitted, 2020.

3. [J] M. Xian, Y. Zhang, H. Cheng, F. Xu, K. Huang, B. Zhang, J. Ding, C. Ning, and Y.Wang, A Benchmark for Breast Ultrasound Image Segmentation (BUSIS), under review of *IEEE Transaction on Medical Imaging*, 2020.

Published Publication

4. [C] F. Wei, J. Ding, C. Ning, F. Xu, M. Xian, and Y. Zhang, Texture analysis and imbalanced data processing for papillary thyroid microcarcinoma detection, in BIBE 2018; International Conference on Biological Information and Biomedical Engineering, 2018, pp. 1-4.

5. [C] F. Xu, M. Xian, Y. Zhang, K. Huang, H. D. Cheng, B. Zhang, A Novel Hybrid Framework for Tumor Saliency Estimation, in: International Conference on Pattern Recognition (ICPR),2018.

6. [J] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, B. Zhang, and J. Ding, Automatic breast ultrasound image segmentation: A survey, *Pattern Recognition,* vol. 79, pp. 340-355, 2018.

7. [C] M. Xian, F. Xu, H. D. Cheng, Y. Zhang, and J. Ding, EISeg: Effective interactive segmentation, in International Conference on Pattern Recognition (ICPR), 2016, pp. 1982-1987.

8. [J] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, J. Ding, Neutro-Connectedness Cut, *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4691-4703, 2016

9. [C] F. Xu, M. Xian, H. D. Cheng, Unsupervised saliency estimation based on robust hypotheses, in IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1-6.

10. [J] J. Ding, H. Cheng, M. Xian, Y. Zhang, and F. Xu, Local-weighted Citation-kNN algorithm for breast ultrasound image classification, *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 24, pp. 5188-5193, 2015.

11. [C] J. Ding, M. Xian, H. D. Cheng, Y. Zhang, and F. Xu, An algorithm based on LBPV and MIL for left atrial thrombi detection using transesophageal echocardiography, in Proceedings International Conference on Image Processing, ICIP, 2015, vol. 2015-December, pp. 42244227.

12. [C] H. Shao, Y. Zhang, M. Xian, H. D. Cheng, F. Xu, J. Ding, A saliency model for automated tumor detection in breast ultrasound images, in IEEE International Conference Image Processing (ICIP), 2015, pp. 1424-1428.

13. [J] F. Xu, J. F. Liu, and B.Y. Zhang and R .Wu and W. Wei, Research on clustering of natural scene images with texts based on random projection, *Application Research of Computers*, Vol.28, No. 12, pp. 4730-4733, 2011.

14. [Thesis] F. Xu, Research on Clustering of Natural Scene Images with Texts Based on Random Projection, Thesis, 2011.