8-2021

Breast Ultrasound Image Segmentation Based on Uncertainty Reduction and Context Information

Kuan Huang
Utah State University

BREAST ULTRASOUND IMAGE SEGMENTATION BASED ON UNCERTAINTY REDUCTION AND CONTEXT INFORMATION

by

Kuan Huang

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

of

DOCTOR OF PHILOSOPHY

in

Computer Science

Approved:

Heng-Da Cheng, Ph.D. Major Professor	Lie Zhu, Ph.D. Committee Member
Haitao Wang, Ph.D. Committee Member	Curtis Dyreson, Ph.D. Committee Member
Vicki Allan, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Interim Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2021

ABSTRACT

Breast Ultrasound Image Segmentation Based on Uncertainty Reduction and Context Information

by

Kuan Huang, Doctor of Philosophy

Utah State University, 2021

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

Breast cancer frequently occurs in women over the world. It was one of the most severe diseases and the second common cancer after skin cancer among women in the United States until 2019. Based on the statistical data provided by the American Cancer Society in 2019, the breast cancer incidence rate increased by 0.3% per year from 2012-2016. In contrast, the death rate of breast cancer dropped 40% from 1989 to 2017 because of the more attention to breast cancer. In 2019, the United States was expected to have 268,600 new cases of invasive breast cancer and 48,100 cases of ductal carcinoma in situ (DCIS). Investigated by some organizations, the survival rate of stages 0 and 1 of breast cancer during 2007 and 2013 was close to 100%; however, it lacks apparent symptoms in the early stage of breast cancer. Many patients miss the best chance to cure it because they do not diagnose breast cancer in the early stages.

Breast ultrasound (BUS) imaging is harmless, low cost, portable and effective; therefore, it becomes the most critical approach for breast cancer early detection. However, breast ultrasound (BUS) images are usually of poor quality and low contrast because they contain inherent and speckle noise. These characteristics of BUS images affect the accuracy of diagnosis. Therefore, developing the computer-aided diagnosis (CAD) system for BUS imaging is essential. The CAD system is to help doctors diagnose breast cancer accurately. Breast ultrasound image segmentation is the crucial step in the CAD system. Traditional breast ultrasound image segmentation approaches only focus on tumor segmentation. The reasons are: 1) the final target is to classify tumor into benign or malignant for the CAD systems; therefore, tumor area is important; and 2) detecting other breast tissues is more challenging than detecting tumor area. However, detecting other tissues in the breast such as the skin layer, mammary layer, muscle layer is also important for breast cancer diagnosis. The semantic segmentation of BUS images which can detect tissues and tumors, is important for CAD systems of BUS images.

In this research, there are two major research fields in breast ultrasound image semantic segmentation: 1) reducing uncertainty in semantic segmentation using fuzzy logic; 2) involving context information to optimize segmentation accuracy. The experimental results demonstrate the new approaches obtain the best performance compared with the previous BUS segmentation methods on four datasets.

(130 pages)

PUBLIC ABSTRACT

Breast Ultrasound Image Segmentation Based on Uncertainty Reduction and Context Information

Kuan Huang

Breast cancer frequently occurs in women over the world. It was one of the most serious diseases and the second common cancer among women in 2019. The survival rate of stages 0 and 1 of breast cancer is closed to 100%. It is urgent to develop an approach that can detect breast cancer in the early stages. Breast ultrasound (BUS) imaging is low-cost, portable, and effective; therefore, it becomes the most crucial approach for breast cancer diagnosis. However, BUS images are of poor quality, low contrast, and uncertain. The computer-aided diagnosis (CAD) system is developed for breast cancer to prevent misdiagnosis.

There have been many types of research for BUS image segmentation based on classic machine learning and computer vision methods, *e.g.*, clustering methods, thresholding methods, level set, active contour, and graph cut. Since deep neural networks have been widely utilized in nature image semantic segmentation and achieved good results, deep learning approaches are also applied to BUS image segmentation. However, the previous methods still suffer some shortcomings. Firstly, the previous non-deep learning approaches highly depend on the manually selected features, such as texture, frequency, and intensity. Secondly, the previous deep learning approaches do not solve the uncertainty and noise in BUS images and deep learning architectures. Meanwhile, the previous methods also do not involve context information such as medical knowledge about breast cancer. In this work, three approaches are proposed to measure and reduce uncertainty and noise in deep neural networks. Also, three approaches are designed to involve medical knowledge and long-range distance context information in machine learning algorithms. The proposed methods are applied to breast ultrasound image segmentation.

In the first part, three fuzzy uncertainty reduction architectures are designed to measure the uncertainty degree for pixels and channels in the convolutional feature maps. Then, medical knowledge constrained conditional random fields are proposed to reflect the breast layer structure and refine the segmentation results. A novel shape-adaptive convolutional operator is proposed to provide long-distance context information in the convolutional layer. Finally, a fuzzy generative adversarial network is proposed to reduce uncertainty. The new approaches are applied to 4 breast ultrasound image datasets: one multi-category dataset and three public datasets with pixel-wise groundtruths for tumor and background. The proposed methods achieve the best performance among 15 BUS image segmentation methods on the four datasets.

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to many people who help me and give me support during my five-year Ph.D. program at Utah State University.

Firstly, I would like to thank my advisor: Dr. Heng-Da Cheng, for his patient guidance and advice in my research, studying, and life. He helps me overcome the cross-cultural and new environment at the beginning of my Ph.D. program. He teaches me how to be a researcher by meetings, revising my articles, and giving me suggestions in experiments. Without his persistent help, I will never be able to finish my dissertation.

Secondly, I would also like to thank other members of my Ph.D. committees: Dr. Lie Zhu, Dr. Curtis Dyreson, Dr. Haitao Wang, and Dr. Vicki Allan for their encouragement, comments, and advice for my doctoral studies and research.

Mostly, I wish to express my deepest gratitude to my parents, who bring me into this beautiful world and always teach me to be a strong, honest, brave person. They give me the chance to study abroad, and I will never obtain current achievements without their raising, supporting, and teaching. I want to thank my wife, Meng, who always stands behind me to give me suggestions and encouragement and take care of my life. Her constant love makes this journey brilliant and memorable.

Above all, I must thank many people, such as my previous and current colleagues at Utah State University: Min Xian, Boyu Zhang, Kaige Zhang, Fei Xu, Xiankun Yan, for their help. I also really appreciate my friends: Qun Wang, Cheng Chen, Wanlong Zhao, Yiming Zhao, who have made my life exciting at Utah State University. Thanks to Dr. Ping Xing from the First Affiliated Hospital of Harbin Medical University for providing breast ultrasound images in this research.

Kuan Huang

CONTENTS

Page

ABSTRACT . . . iii

PUBLIC ABSTRACT . . . v

ACKNOWLEDGMENTS . . . vii

LIST OF TABLES . . . xi

LIST OF FIGURES . . . xii

1 INTRODUCTION . . . 1
- 1.1 Background . . . 1
- 1.2 Semantic Segmentation Methods and Uncertainty Reduction Methods . . . 2
- 1.3 Context Information Guided Methods . . . 5
- 1.4 Breast Ultrasound Image Segmentation . . . 7
- 1.5 Outline . . . 10

2 FUZZY SEMANTIC SEGMENTATION FOR BREAST ULTRASOUND IMAGE 11
- 2.1 Overview of the Proposed Architecture . . . 11
- 2.2 Preprocessing . . . 12
- 2.3 The Proposed Fuzzy Block . . . 14
- 2.4 Training Strategy for U-Net with Fuzzy Block . . . 19
- 2.5 Experiment Results . . . 20
  - 2.5.1 Dataset . . . 20
  - 2.5.2 Evaluation Metrics . . . 21
  - 2.5.3 Experiment Details . . . 22
  - 2.5.4 Segmentation Result of U-Net with Fuzzy Block . . . 23
- 2.6 Conclusion . . . 25

3 SPATIAL AND CHANNEL-WISE FUZZY UNCERTAINTY REDUCTION BLOCKS IN DEEP NEURAL NETWORKS . . . 26
- 3.1 Overview of the Proposed Network Structure . . . 27
- 3.2 Spatial-wise Fuzzy Uncertainty Representation and Reduction . . . 27
- 3.3 Channel-wise Fuzzy Uncertainty Representation and Reduction . . . 30
- 3.4 Experiment Results . . . 32
  - 3.4.1 Datasets . . . 32
  - 3.4.2 Experiment Details . . . 33
  - 3.4.3 Metrics . . . 35
  - 3.4.4 Multi-object Semantic Segmentation of BUS Images . . . 36
  - 3.4.5 Semantic Segmentation on Three Public Two-category Datasets . . . 43
- 3.5 Conclusion . . . 46

4 PYRAMID FUZZY UNCERTAINTY REDUCTION NETWORK AND DIRECTION-CONNECTEDNESS FEATURE . . . 48
4.1 Introduction . . . 48
4.2 Pyramid Fuzzy Block . . . 48
4.2.1 Down-sampling Feature Map . . . 49
4.2.2 Fuzzification . . . 49
4.2.3 Uncertainty Representation . . . 50
4.2.4 Uncertainty Reduction . . . 50
4.3 Direction Connectedness Feature . . . 51
4.3.1 Connectedness between Two Adjacent Pixels . . . 52
4.3.2 Connectedness in Four Principal Directions . . . 53
4.3.3 Direction-connectedness Feature Extraction . . . 54
4.3.4 Loss Function . . . 55
4.4 Experimental Results . . . 56
4.4.1 Datasets . . . 57
4.4.2 Training Strategy and Setup . . . 57
4.4.3 Metrics . . . 57
4.4.4 Tumor Segmentation Results on Benchmark Dataset . . . 58
4.4.5 Multi-object Segmentation for BUS Image . . . 59
4.5 Conclusion . . . 62

5 MEDICAL KNOWLEDGE CONSTRAINED CONDITIONAL RANDOM FIELDS 63
5.1 Introduction . . . 63
5.2 Breast-anatomy Constrained Fully Connected CRFs . . . 65
5.3 Label Vector Setting . . . 66
5.4 Experiment Results . . . 68
5.4.1 Datasets, Metrics, and Setup . . . 69
5.4.2 Comparison with Original CRFs . . . 69
5.4.3 Tumor Segmentation Results and Comparison with Previous Non-deep Learning Segmentation Methods . . . 70
5.4.4 Comparison with DC Feature and Medical Knowledge Constrained CRFs . . . 73
5.5 Conclusion . . . 74

6 SHAPE-ADAPTIVE CONVOLUTIONAL OPERATOR AND ITS APPLICATION IN BUS IMAGE SEGMENTATION . . . 76
6.1 Introduction . . . 76
6.2 The Proposed Method . . . 77
6.2.1 Shape-adaptive Convolutional Operator . . . 77
6.2.2 Selecting Pixels for SAC Operator . . . 79
6.2.3 Entire Network Structure and Training Details . . . 81
6.3 Experiment Results . . . 82
6.3.1 Datasets . . . 82
6.3.2 Metrics . . . 83
6.3.3 Parameter k in SAC Operator . . . 83
6.3.4 Positions of k Selected Pixels in SAC Operator . . . 84
6.3.5 Multi-object Segmentation for BUS Images . . . 85

6.3.6 Tumor Segmentation Results Using Public Datasets 86
6.4 Conclusion 87

7 BREAST ULTRASOUND IMAGE SEGMENTATION USING A MULTI-SCALE FUZZY GENERATIVE ADVERSARIAL NETWORK 88
7.1 Introduction 88
7.2 The Proposed Method 89
7.2.1 Overview 90
7.2.2 Fuzzy Attentive Feature Generator 91
7.2.3 Multi-scale Fuzzy Entropy Module 92
7.3 Experiment Results 94
7.3.1 Datasets and Metrics 95
7.3.2 Segmentation Results on Multi-layer Dataset 96
7.3.3 Segmentation Results on Two Public Datasets 96
7.3.4 Experiment Setup 97
7.4 Conclusion 98

8 CONCLUSION AND FUTURE WORK 99

REFERENCES 104

CURRICULUM VITAE 111

LIST OF TABLES

Table Page

2.1 Evaluation results on 325 cases dataset. Evaluation metric is IoU (%). . . . 22

3.1 Dataset properties. 33

3.2 Ablation study on multi-object dataset. SFB: Spatial-wise fuzzy block, CFB: Channel-wise fuzzy block. 38

3.3 Results of multi-object semantic segmentation. Evaluation metric is IoU (%). 42

3.4 Results of two-class semantic segmentation on dataset [1]. 44

3.5 Results of two-class semantic segmentation on dataset [2] and dataset [3]. . 45

4.1 Results of tumor segmentation using dataset [1]. 60

4.2 Evaluation results of multi-object segmentation on BUS images. Evaluation metric using IoU (%). 61

5.1 Evaluation results of breast anatomy constrained CRFs. Evaluation metrics using IoU. 70

5.2 Evaluation results on tumor segmentation. 73

5.3 Comparison with DC feature and medical knowledge constrained CRFs. . . 74

6.1 Results of multi-object segmentation (IoU (%)). 84

6.2 Results on public datasets (%). 86

7.1 Results of multi-layer segmentation (IoU (%)). 95

7.2 Results on public datasets (%). 97

LIST OF FIGURES

Figure Page

1.1 Weird shape tumor and patches similar as tumor areas in BUS images. . . . 2

1.2 Breast anatomy: (a) BUS image; (b) breast anatomy obtained by deep neural networks. . . . 5

1.3 Organization structure of this dissertation. . . . 9

2.1 Flowchart of the two strategies of the fuzzy U-Net: (a) using wavelet; (b) without using wavelet. . . . 12

2.2 Histogram equalization: (a) original images; (b) images after histogram equalization. . . . 13

2.3 Wavelet transform: (a) original images; (b) augmented 3-channel images. . 14

2.4 The membership maps: (a)-(e) the memberships of tumor, fat layer, mammary layer, muscle layer, and background computed by the trainable Gaussian function; (f)-(j) the corresponding memberships computed by the trainable Sigmoid function. . . . 16

2.5 Uncertainty maps: (a)-(e) are the uncertainty maps of tumor, fat layer, mammary layer, muscle layer, and background, which are generated by the trainable Sigmoid membership function; (f)-(j) are generated by the trainable Gaussian membership function. . . . 17

2.6 Overall uncertainty maps: (a) original images; (b) overall categories uncertainty maps generated by using the trainable Sigmoid membership function; (c) overall categories uncertainty maps generated by using the trainable Gaussian membership function. . . . 18

2.7 The fusion of uncertainty maps and input images: (a) original images; (b) 3-channel images with gray-level intensity and wavelet information; (c) resulted images by the Sigmoid function and Eq. (2.9); (d) resulted images by the Gaussian function and Eq. (2.9). . . . 19

2.8 Structure of the proposed U-Net with fuzzy block. . . . 20

2.9 The semantic segmentation results: (a) original images; (b) groundtruths; (c) results of U-Net with gray-level images as the inputs; (d) results of U-Net with 3-channel images as inputs; (e) results of PSPNet; (f) results of Deeplab; (g) results of ResNet-101; (h) results of the proposed fuzzy U-Net with trainable Gaussian membership function and 3-channel images; (i) results of the proposed fuzzy U-Net with trainable Sigmoid membership function and 3-channel images; (j) results of the proposed fuzzy U-Net with trainable Sigmoid membership function and gray-level images. 23

3.1 The proposed SCFURNet: (a) the entire structure; (b) the spatial-wise fuzzy block; (c) the channel-wise fuzzy block. 27

3.2 Channel-wise fuzzy block. 31

3.3 The relation between the number of fuzzy blocks and the segmentation performance. Block number = 1: the fuzzy blocks are applied to the first convolutional block; block number = 2: the fuzzy blocks are applied to the first and second convolutional blocks together; block number = 3: the fuzzy blocks are applied to the convolutional blocks 1, 2, and 3; block number = 4: the fuzzy blocks are applied to the convolutional blocks 1, 2, 3, and 4; block number = 5: the fuzzy blocks are applied to the convolutional blocks 1, 2, 3, 4, and 5. 37

3.4 The relation between the number of fuzzy blocks and the segmentation performance. The fuzzy blocks are applied to the convolutional blocks individually. 38

3.5 Segmentation results of U-Net with ResNet-101 and channel-wise fuzzy block on multi-object dataset. 39

3.6 Segmentation results of U-Net with ResNet-101 and spatial-wise fuzzy block on multi-object dataset. 39

3.7 Visualization results of fuzzy blocks on the multi-object dataset. For each row, we show an input image, an uncertainty map from the spatial-wise fuzzy block; red represents a high value and blue represents a low value in the heatmap. We also provide two channel maps from the outputs of the channel-wise fuzzy block, the results of the original U-Net and the proposed method, and the groundtruths. 41

3.8 Multi-object semantic segmentation of BUS images: (a) original images; (b) groundtruths; (c) results of ResNet-101 + self-attention mechanism; (d) results of attention U-Net; (e) results of ResNet-50; (f) results of ResNet-101; (g) results of Deeplab; (h) results of PSPNet; (i) results of U-Net with wavelet transform; and (j) results of FCN-8s; (k) results of SE-Net (ResNet-101); (l) results of the proposed SCFURNet. 42

3.9 Segmentation results using public dataset: (a) original images; (b) groundtruths; (c) results of ResNet-101 with self-attention mechanism; (d) results of a SE-Net (ResNet-101); (e) results of attention U-Net; (f) results of ResNet-50; (g) results of ResNet-101; (h) results of Deeplab; (i) results of PSPNet; (j) results of U-Net with wavelet transform; and (k) results of FCN-8s; (l) results of proposed SCFURNet. 43

3.10 Small tumor segmentation: (a) original images; (b) groundtruths; (c) results of ResNet-101 with self-attention mechanism; (d) results of a SE-Net (ResNet-101); (e) results of attention U-Net; (f) results of ResNet-50; (g) results of ResNet-101; (h) results of Deeplab; (i) results of PSPNet; (j) results of U-Net with wavelet transform; and (k) results of FCN-8s; (l) results of proposed SCFURNet. 46

4.1 Architectures of the proposed approach: (a) the original U-shape network with VGG-16; (b) the proposed method with pyramid fuzzy block and direction connectedness feature; and (c) the structure of the proposed pyramid fuzzy block. 49

4.2 Uncertainty maps: (a) original BUS images; (b) groundtruths; (c) uncertainty maps calculated by non-pyramid fuzzy block; (d) uncertainty maps calculated by pyramid fuzzy block. Blue represents low uncertainty and red represents high uncertainty. 52

4.3 BUS image layer structure: (a) original BUS image; (b) groundtruths; (c) direction connectedness feature. 53

4.4 The connectedness in the four principal directions: (a) right; (b) up; (c) left; and (d) down. 54

4.5 Direction connectedness feature. Only one round spatial RNN is utilized because it only focuses on the connectedness strength between pixels and the pixels on the boundaries along up, down, left, and right directions. 55

4.6 Segmentation results using benchmark in [1]: (a) original images; (b) groundtruths; (c) results of U-Net; (d) results of ResNet-50; (e) results of ResNet-101; (f) results of Deeplab; (g) results of PSPNet; (h) results of U-Net with wavelet transform; (i) results of FCN-8s; (j) results of U-Net with DSC feature; and (k) results of proposed method. 59

4.7 Semantic segmentation: (a) original images; (b) groundtruths; (c) results of U-Net; (d) results of ResNet-50; (e) results of ResNet-101; (f) results of Deeplab; (g) results of PSPNet; (h) results of U-Net with wavelet transform; (i) results of FCN-8s; (j) results of U-Net with DSC feature; and (k) result of proposed method. 61

5.1 Simulated image to show the context distance among categories. 67

5.2 The coordinates of the labels. 69

5.3 Segmentation results of breast-anatomy constrained fully connected CRFs: (a) original images; (b) groundtruths; (c) results of fuzzy U-Net without CRFs; (d) fine-tuning results using fully connected CRFs; (e) fine-tuning results using the proposed method. 71

5.4 Tumor segmentation results of the proposed method and existing methods: (a) original images; (b) groundtruths; (c) results using [4]; (d) results using [5]; (e) results using [6]; (f) results using [7]; (g) results using [8]; (h) results using the proposed method. 72

5.5 The segmentation results for BUS images without tumors: (a) original images; (b) groundtruths; (c) results using [6]; (d) results using [7]; (e) results using [8]; (f) results of the proposed method. 72

5.6 The segmentation results of the BUS images containing two tumors: (a) original images; (b) groundtruths; (c) results using [6]; (d) results using [7]; (e) results using [8]; (f) results using the proposed method. 74

6.1 Convolutional operators: (a) 3×3 convolutional operator; and (b) SAC operator with 9 pixels. 77

6.2 Selecting pixels for the SAC operator using self-attention coefficient. 79

6.3 (a) The entire network; and (b) the proposed SAC operator. 80

6.4 The relation between segmentation performance and parameter k. 82

6.5 The positions of k pixels involved in the SAC operators: the green pixels are target pixels; red pixels are selected to calculate the convolutions of the target pixels; a1-a4 are based on the k-NN selection approach; a5-a8 are based on the self-attention coefficient selection approach. 83

6.6 Semantic segmentation: (a) original images; (b) groundtruths; (c) results of U-Net with VGG-16; (d) results of U-Net with ResNet-101; (e) results of PSPNet; (f) results of Deeplab; (g) results of FCN-8s; (h) results of ResNet-101 + self-attention mechanism; (i) results of ResNet-101 + deformable convolution; (j) results of ResNet-101 + SAC with k-NN; and (k) results of ResNet-101 + SAC with self-attention coefficient. 85

7.1 An overview of the proposed MSF-GAN. 89

7.2 Illustration of the proposed fuzzy attentive feature generator. 91

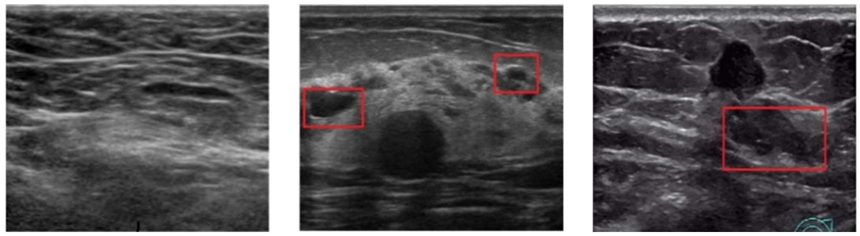
7.3 Illustration of segmentation results on multi-layer dataset. (a) Original BUS image; (b) Groundtruth; (c) U-VGG; (d) U-ResNet; (e) PSPNet; (f) Deeplabv3+; (g) FCN-8s; (h) SegAN; (i) Proposed MSF-GAN. 94

CHAPTER 1

INTRODUCTION

1.1 Background

Breast cancer frequently occurs in women over the world. It was one of the most serious diseases and the second common cancer among women in 2019 [9]. The survival rate of stages 0 and 1 of breast cancer is closed to 100% [10]. It is urgent to develop an approach that can detect breast cancer in the early stages. Breast ultrasound (BUS) imaging is low-cost, portable, harmless, and effective; therefore, it becomes the most important approach for breast cancer diagnosis. However, BUS images are of poor quality, low contrast, and uncertain. The computer-aided diagnosis (CAD) system for BUS image is developed to assist the doctor in diagnosing breast cancer, especially BUS image segmentation in the CAD system. There have been many types of research for BUS image segmentation based on classic machine learning and computer vision methods [11] and deep learning [2]. They have achieved good results in BUS image segmentation, especially for deep learning approaches. However, the previous deep learning approaches do not solve the uncertainty and noise in BUS images and deep learning architectures. Researches [12, 13] show that there exists epistemic and aleatory uncertainty in deep learning architectures and medical images. The main causes of uncertainty are: 1) BUS images are of low quality and contrast; 2) BUS images contain inherent speckle noise and shadows [11]; 3) the BUS images from different machines during different periods have different contrasts, and image intensity changes variously; 4) the characteristics of breasts in different people might be various; 5) there are image patches in BUS images similar to the tumor areas (shown in patches marked by red rectangles in Figs. 1.1 (a), (b)). Besides the uncertainty and noise, the previous deep learning-based BUS image segmentation methods do not involve context information. Here the context information refers to the correlation between pixels such as the breast



(a) (b) (c)

Fig. 1.1: Weird shape tumor and patches similar as tumor areas in BUS images.

anatomy and the characteristics of the BUS image. As shown in Fig. 1.2, BUS images have some characteristics, such as the layer structure in Fig. 1.2 (b), that can improve the segmentation performance.

To handle the problems appearing in the BUS images and increase the performance of BUS image CAD system, segmentation methods for BUS image based on uncertainty reduction and context information are proposed. This research has two major portions: 1) reducing uncertainty in deep neural networks using three fuzzy logic approaches and 2) designing three novel approaches that can reflect context information of BUS image. In the following part, we first review research in semantic segmentation and deep learning; then some research to extract context information in deep convolutional neural networks is listed; finally, existing BUS image segmentation methods are shown.

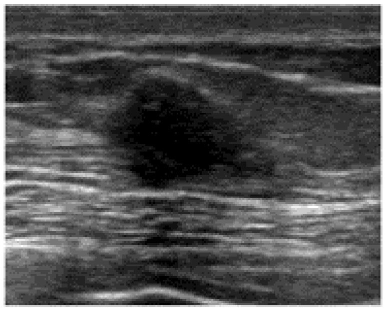
1.2 Semantic Segmentation Methods and Uncertainty Reduction Methods

Deep learning is widely used in nature image semantic segmentation since Long *et al.* adopt fully convolutional networks (FCN) [14]. The reason is that deep learning obtains better results than traditional methods based on the automatically encoded convolutional features. FCN is the first end-to-end deep learning architecture for semantic segmentation; however, it remains several shortcomings, such as blurry boundaries for segmentation results and mis-segmentation. There are three main research directions for deep learning: 1) increasing the receptive field of convolutional kernels, 2) adding more convolutional layers, 3) applying attention mechanisms or recurrent neural networks to convolutional neural networks to extract short-range or long-range context information and reduce uncertainty in the feature map. In [15], He *et al.* develop a residual neural network (ResNet), making the convolutional network much deeper than previous ones. A deeper convolutional layer can extract more useful features. In [16,17], dilated convolution/atrous convolution is applied to avoid the loss of information in the pooling layer. Atrous Spatial Pyramid Pooling (ASPP) [17] and Pyramid Pooling Module (PPM) [18] use spatial multi-scale pooling operation to obtain multi-scale information. In [19], different scales of convolutional filters are utilized in the same convolutional block to obtain multi-scale information and enrich the information in each convolutional block. In [20], a spatial-wise attention block is applied to U-Net. In [21], a channel-wise attention mechanism: Squeeze-and-Excitation Networks (SE-Nets), is proposed and achieved good performance with many network architectures such as ResNet and VGGNet [22]. The previous deep learning-based semantic segmentation methods focus on the network structure, pooling layer, and loss function and achieve good results on many benchmarks. However, few of them discuss reducing uncertainty in feature maps and deep neural networks.

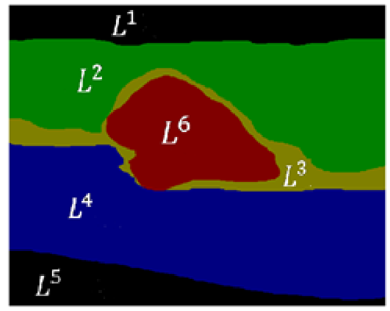
Attention mechanism in convolutional neural networks is popularly used [23] to reduce noise and uncertainty. It assigns the weights to pixels or channels of feature maps to express the importance. In [20], a spatial-wise attention gate is proposed in the decoder of U-Net. Before concatenating the encoder-feature map and the decoder information, the encoder and decoder information are combined to calculate a weight tensor. The weight tensor multiplies with the encoder-feature map. Attention coefficients are bigger in the target areas than those in the background, and the results are better than that of the original U-Net. In [21], the SE-Nets propose a channel-wise attention mechanism. A convolutional operator transforms the feature map in each convolutional block. Then, in each channel, a global average pooling is performed to calculate the mean value of each channel. The results are used as the weight values for the channels in the original feature map. The SE block in SE-Nets is applied to network architectures such as VGG-16, ResNet-101, *etc.*, and achieves good improvement. In [24], both spatial-wise and channel-wise attention mechanisms are applied to the image caption. The network structure follows VGG-19 [22] and ResNet-

152 [15]. In each convolutional block, the weights of spatial-wise attention are based on the original feature map and last sentence context information. The mean value for each channel of the original feature map and last sentence context information is used to calculate the channel-wise attention weights. Another spatial-wise and channel-wise attention FCN [25] is proposed for crowd counting. The network structure follows VGG-16 [22] architecture. The spatial-wise and channel-wise attention weights are computed by the original feature map in the same convolutional block. In spatial-wise attention, the original feature map is input to three 1×1 convolutional kernels. Then, reshaping and transposing operators are applied to the outputs of the 1×1 convolutional kernels to obtain three new features. For channel-wise attention weights, only one 1×1 convolutional kernel is utilized. Then, the output of the convolutional kernel is reshaped and transposed to three different sizes. The attention weights are computed by multiplying and adding three different size features.

The attention mechanism can reduce uncertainty and noise in convolutional feature maps; however, uncertainties are not caused by randomness only and cannot be handled by statistics, probabilities, and attention mechanisms well. Fuzzy logic has been applied to handle the uncertainty successfully. A fuzzy contrast enhancement method is developed [26]. The maximum entropy principle is utilized to map images from spatial domain to fuzzy domain. Then, a fuzzy contrast enhancement algorithm is applied. A fuzzy clustering method is utilized for image segmentation [27]. The fuzzy membership is initialized by k-means clustering. The segmentation cost function is based on the membership of each pixel and the Euclidean distances from the pixels to the cluster center. The fuzzy clustering method achieves better performance than the non-fuzzy version. An edge-detection method based on generalized type-2 fuzzy logic is designed [28]. The membership function is defined using Gaussian generalized type-2 membership functions. In [29], Deng *et al.* firstly propose an adaptive membership function fuzzy neural network for data classification. The memberships are multiplied with the original information. Fuzzy image processing methods can obtain robust results and handle uncertainty and noise well. However, the existing fuzzy neural networks contain shortcomings in fuzzification methods and the combination



(a)



(b)

Fig. 1.2: Breast anatomy: (a) BUS image; (b) breast anatomy obtained by deep neural networks.

of fuzzy information and non-fuzzy information. The physical meanings of fuzzy logic in neural networks should be discussed deeper. Meanwhile, the fuzzy neural networks do not discuss the different sizes of the object.

1.3 Context Information Guided Methods

The deep convolutional neural networks can perform BUS image segmentation. However, the segmentation results are not good because the dataset size is too small, and the network structure is quite deep. Meanwhile, context information such as the correlation between pixels can provide vital information to increase the performance of deep neural networks. Many research discusses the correlation between pixels and channels of feature maps in deep learning architectures in Euclidean dimension and non-Euclidean dimension using matrix multiplication or recurrent neural networks. Also, fully connected conditional random fields (CRFs)/ Markov random fields (MRFs) are often utilized to refine the segmentation results, especially on boundaries, because fully connected CRFs and MRFs discuss the correlation between pixels and provide higher-order information (information from pixels with large Euclidean distance). The correlation between pixels in fully connected CRFs is calculated based on feature values and physical positions.

In [30], Krahenbuhl *et al.* provide an approximation algorithm to fully connected CRFs for multi-object segmentation. The approximation algorithm increases the efficiency of fully connected CRFs and makes it possible for semantic segmentation. In [16, 17], Chen *et al.*

propose the Deeplab structure for nature image semantic segmentation by applying the atrous convolutional operation and atrous spatial pyramid pooling (ASPP). In addition, Deeplab also applies fully connected CRFs to the end of the architecture to refine the segmentation results from deep neural networks. In [31], Zheng *et al.* use a recurrent neural network (RNN) to realize fully connected CRFs proposed in [30]. It makes the Deep learning + CRFs structure become a deep end-to-end architecture, and this structure is typically utilized in semantic segmentation tasks. In [32], Liu *et al.* provide a Markov random field (MRF) method with the mixture of label context to involve context information in deep learning. The MRF is realized by Deep Parsing Network (DPN). Besides the feature values and physical positions, the location relation between objects is utilized to calculate correlations between pixels. For example, people usually sit on a chair, not under a chair. If the deep neural networks classify some pixels under chairs to people, the MRF can avoid this situation.

Besides using CRFs/MRF, many deep learning models can directly provide context information between pixels. In [33], Zhuang *et al.* propose a novel dense related module in deep convolutional neural networks which uses RNN with different skip lengths in spatial directions to aggregate global and local contextual information. In [34], Zhang *et al.* transform the correlation computation method from natural language processing to image processing, called the self-attention mechanism. The self-attention mechanism can provide long-range relations between pixels. In [35], Huang *et al.* continue the research of long-range correlations between pixels based on self-attention. A novel network structure based on RNN is proposed to calculate self-attention coefficients between pixels and reduce complexity compared with the method [34]. In [36], novel context information between pixels rather than self-attention coefficients is proposed, called the direction-aware spatial context (DSC) feature. The DSC feature can measure context information better in the shadow detection task. In [37], the deformable convolution is proposed to obtain non-local information using a weird shape convolutional kernel. Pixels in feature maps are shifted to new locations, and then the standard convolutional operator is applied to the new feature maps.

The pixels used in convolution are not the original neighbor pixels. The deformable convolutional operator can extract context information between pixels not in the Euclidean domain in convolution. These architectures are general for nature images or specific tasks (such as shadow detection tasks in [36]). They do not provide the breast anatomy, which means they cannot perform well for BUS image segmentation. Meanwhile, these architectures contain shortcomings. For example, there are shortcomings for deformable convolution: 1) the shift of each pixel is controlled in a small distance, and the deformability is weak, *i.e.*, the distance of the non-local information is quite small; 2) if the distance of shifting is not controlled, two pixels might overlap or are moved outside of the feature map which causes missing information; 3) it cannot select the number of pixels effectively, and the pixels in convolution are still based on the standard convolutional kernel.

1.4 Breast Ultrasound Image Segmentation

Many approaches have been proposed for BUS image segmentation in the past two decades. Early BUS image segmentation methods utilize classic machine learning and computer vision methods, such as thresholding algorithms, region-growing algorithms, watershed algorithms, graph-based algorithms. There are more and more BUS image segmentation researches based on deep neural network-based algorithms [11, 38, 39].

Classic methods: Xian *et al.* [7] propose a fully automatic breast segmentation method based on graph cut and using the frequency and spatial domain as constraints. In [40], a seeded region growing (SRG) algorithm is proposed which uses an iterative quadtree decomposition and a gradual equipartition algorithm to automatically segment tumor regions. In [41], Bafna *et al.* develop a watershed algorithm including noise removal, binarization, extraction of the region of interests (ROI), and boundary detection using a watershed algorithm. Ilesanmi *et al.* [42] propose a multi-scale superpixel method including a boundary efficient superpixel decomposition and a boundary graph cut segmentation algorithm. Gray level thresholding method and area growing lesion contour detecting method are studied in [43, 44]. The region of interest (ROI) is determined by thresholding, and then a maximization utility function is applied to ROI for obtaining the lesion contour. Moon *et al.* propose a clustering-based breast cancer segmentation method [45]. The method consists of three parts. The first part is quantitative tissue clustering. The tissue within the tumor is different from other tissues. A 3-D mean-shift clustering is used for selecting tumor tissues according to the echogenicities. The fuzzy c-means clustering method divides the segmented regions into four clusters. The morphology and echogenicity features are extracted, and logistic regression is used to classify the benign and malignant tumors. Shan *et al.* propose a fully automatic breast cancer segmentation method based on neutrosophic-l-means clustering [46]. It uses an automatic seed point selection algorithm to generate the ROI and then proposes a novel contrast enhancement method based on the frequency and spatial domain. A clustering method combined with neutrosophic logic, the neutrosophic-l-means (NLM) clustering, is utilized to segment BUS images.

Deep learning-based methods: Deep neural network-based methods have received increasing attention in recent years. LeNet [47] is initially designed for handwritten character recognition and later proved suitable for BUS image segmentation because BUS images are in gray-scale and tumor sizes are relatively small. In [2], patch-based LeNet, U-Net [48], and FCN with AlexNet [49] perform well for BUS image segmentation on two BUS datasets. Shareef *et al.* [50] propose a small tumor-aware network to better segment breast tumors with different sizes by using kernels with three different sizes at each convolutional layer. Lei *et al.* [51] propose a boundary regularized convolutional encoder-decoder network to segmentation anatomical breast layers that are robust to speckle noise and posterior acoustic shadows. They further design a self-co-attention neural network that employs both spatial and channel attention modules to explore contextual relationships in BUS images and achieves better segmentation results [52]. In [53], a medical knowledge constrained deep learn + conditional random fields method is proposed for three-layer BUS image semantic segmentation. In [54], Lee *et al.* propose a deep learning-based BUS image segmentation method using a novel channel attention module with multi-scale grid average pooling. The novel module can extract better global contextual information.

Although most deep learning methods for BUS image segmentation transform existing approaches in nature image segmentation to BUS image segmentation and achieve good results, they do not discuss the uncertainty and the context information in BUS images. Inspired by the success of previous approaches in nature image processing that can use context information to improve image segmentation performance, medical knowledge is involved in BUS image segmentation to improve BUS image segmentation results. As shown in Figs. 1.2 (a) and (b), the BUS images have the following layer structure: 1) on the top is the skin layer; 2) the subcutaneous fat layer is beneath the skin layer; 3) the mammary layer is below the fat layer and followed by the muscle layer. Meanwhile, breast cancer is usually ellipse-shaped and begins from the cells in the mammary layer. In most cases, breast cancer stays inside the mammary layer [55]. In this research, several fuzzy logic-based methods are proposed to reduce uncertainty in convolutional neural networks and BUS images. The breast anatomy is utilized to refine BUS image segmentation; meanwhile, a novel convolutional operator is proposed to solve shortcomings in deformable convolution operator and provide non-local context information in image segmentation.

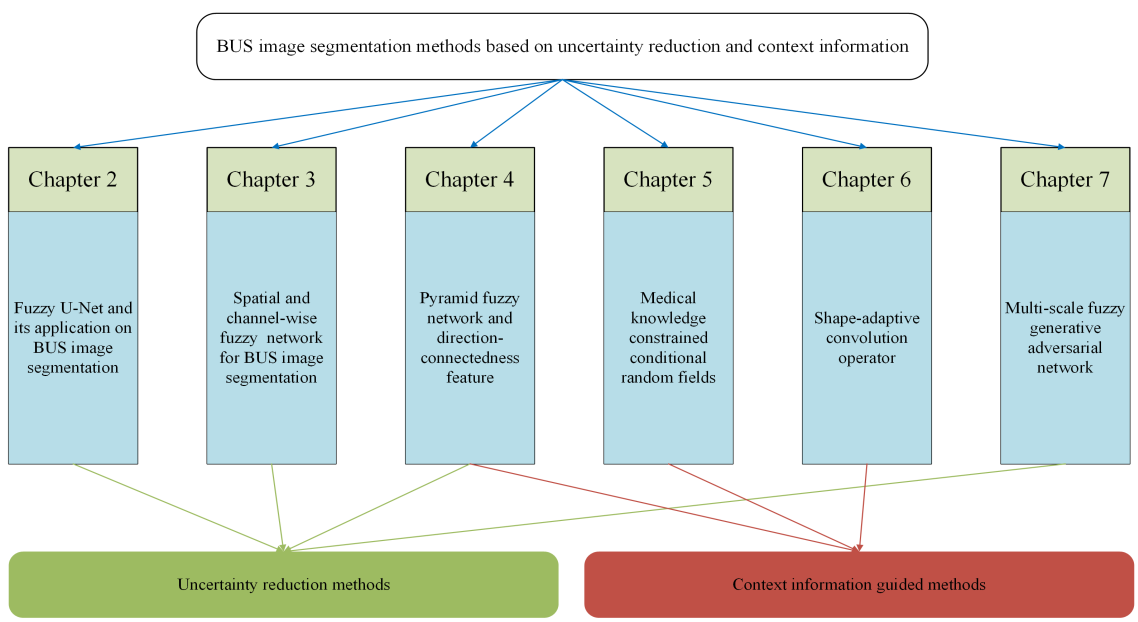


Fig. 1.3: Organization structure of this dissertation.

1.5 Outline

The rest of the dissertation is organized as follows: in Chapter 2, we firstly propose a fuzzy fully convolutional network for BUS image segmentation; in Chapter 3, we improve the fuzzy fully convolutional network by using a novel membership function and an uncertainty representation method; meanwhile, we apply the fuzzy operators to pixels and channels of the feature maps; in Chapter 4, we extend the fuzzy convolutional network in Chapter 3 to a pyramid fuzzy uncertainty reduction network with direction-connectedness feature which can provide breast horizontal layer structure; in Chapter 5, we propose the medical knowledge constrained conditional random fields which can reflect breast anatomy; in Chapters 6, a novel convolutional operator, shape-adaptive convolutional (SAC) operator is proposed, which can provide non-Euclidean domain context information; in Chapter 7, a fuzzy generative adversarial network (GAN) is designed to reduce the uncertainty in the output of the segmentation network and using the adversarial network to help to generate better segmentation result. The future work and conclusion are discussed in Chapter 8. The relation between chapters is shown in Fig. 1.3.

CHAPTER 2

FUZZY SEMANTIC SEGMENTATION FOR BREAST ULTRASOUND IMAGE

In this chapter, we combine fuzzy logic and deep neural network. Inspired by the success of fuzzy logic in image processing, we try to use fuzzy logic to detect and measure uncertainty in feature maps. A trainable membership function is designed to transform BUS images into the fuzzy domain. The uncertainty in the BUS image can be handled well by a novel uncertainty mapping function, and a better semantic segmentation result can be obtained.

2.1 Overview of the Proposed Architecture

The proposed architecture is based on a well-known U-Net [48] with VGG-16 [22]. The flowchart of the proposed approach is shown in Fig. 2.1. We propose a novel fuzzy block (fuzzy block in Fig. 2.1) to refine the input image and the first convolutional feature map. There are two strategies: (1) In Fig. 2.1 (a), the input image is preprocessed by contrast enhancement. Then, wavelet transform is applied. The original image and wavelet information are transformed to the fuzzy domain by membership functions to deal with the uncertainty. Results after reducing uncertainty are input into the first convolutional layer. The obtained feature maps are transformed into the fuzzy domain as well, and the uncertainty is reduced by multiplying uncertainty maps and the corresponding feature maps. (2) In Fig. 2.1 (b), wavelet transform is not utilized. After reducing uncertainty in gray-level intensity and the first convolutional layer, the network can achieve similar performance to that of Fig. 2.1 (a). Two approaches are evaluated by segmentation accuracies and compared with the original fully convolutional network. The details of the proposed architecture are introduced in the following subsections.

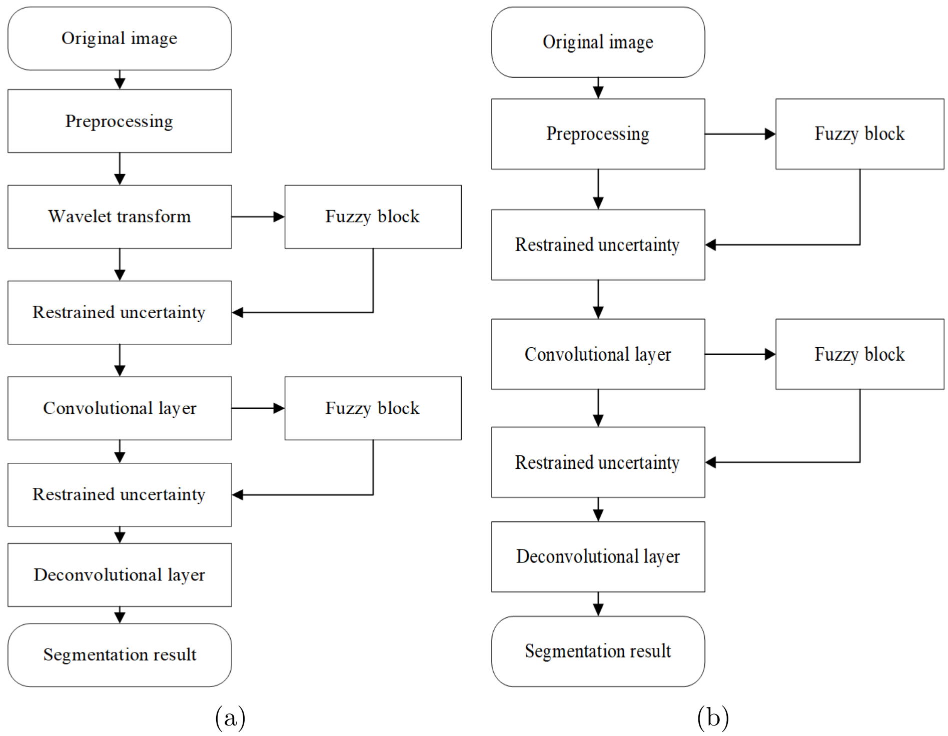


Fig. 2.1: Flowchart of the two strategies of the fuzzy U-Net: (a) using wavelet; (b) without using wavelet.

2.2 Preprocessing

Histogram equalization: The original images are captured in different periods which have different ranges of intensities. It will affect the segmentation results. The histogram equalization is modified to make the input image have the intensity range from 0 to 255 and to conduct contrast enhancement. Histogram equalization is performed on both the training set and testing set. In histogram equalization, the probability of a pixel with intensity θ, $p_z(\theta)$ is computed by Eq. (2.1) [56]:

$$p_z(\theta) = p(z = \theta) = \frac{n_\theta}{n},\ 0 \leq \theta \leq L_z - 1 \tag{2.1}$$

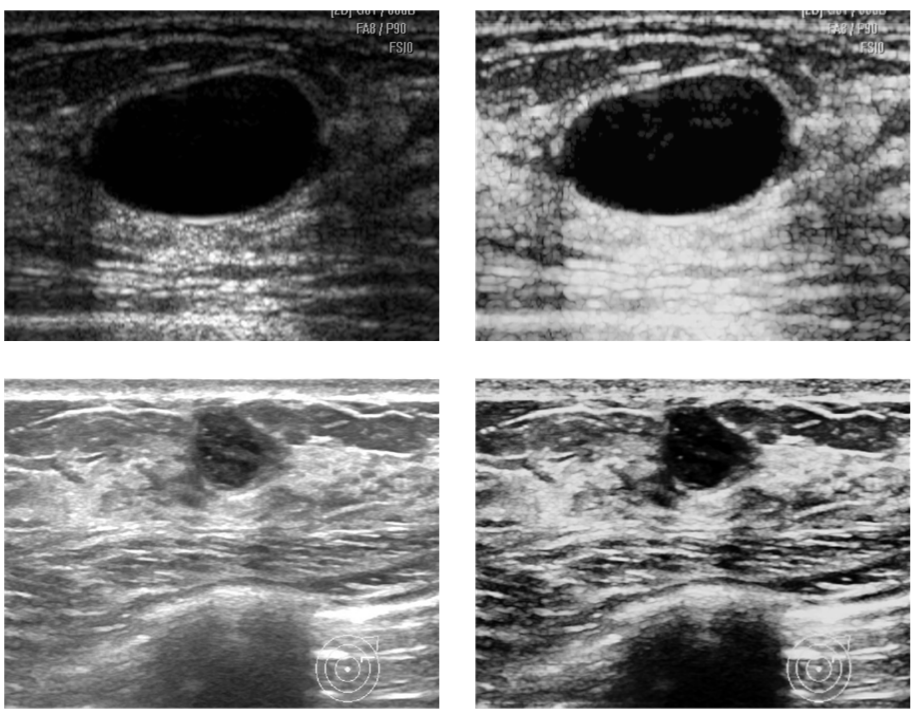
where n_θ represents the number of pixels with intensity θ; L_z is the upper bound of the intensity levels of the image; n represents the total number of pixels. The cumulative distribution function of $p_z(\theta)$ is defined as:

$$cdf_z(\theta) = \sum_{u=0}^{\theta} p_z(\theta) \tag{2.2}$$

The new intensity $h(\theta)$ is computed by:

$$h(\theta) = \left\lfloor \frac{cdf_z(\theta) - cdf_{zmin}}{1 - cdf_{zmin}} \times 255 \right\rfloor \tag{2.3}$$

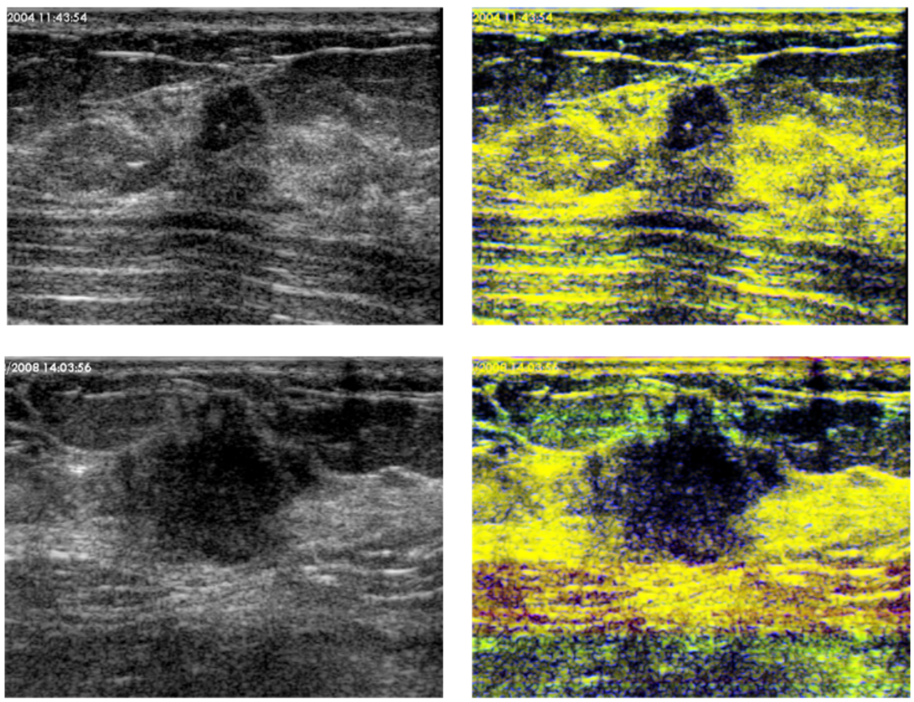
where θ represents the original intensity, and cdf_{zmin} is the minimum non-zero value in the cumulative distribution function. The original images and processed images are shown in Fig. 2.2. After histogram equalization, the images contain higher contrast.



(a) (b)

Fig. 2.2: Histogram equalization: (a) original images; (b) images after histogram equalization.

Wavelet transform: To overcome the small dataset size problem, high-pass filter H and low-pass filter G of wavelet transform are used to obtain the high frequency and low-frequency information. In this research, one level Haar wavelet transformation is applied, and the input image becomes a 3-channel image. The first channel is the original image, the second channel contains the low-frequency coefficients, and the third channel contains the high-frequency coefficients. Fig. 2.3 shows the original images and the augmented images, respectively.



(a) (b)

Fig. 2.3: Wavelet transform: (a) original images; (b) augmented 3-channel images.

2.3 The Proposed Fuzzy Block

The proposed fuzzy block is introduced in this subsection. Similar to the previous fuzzy logic-based image processing methods [26–29], the proposed fuzzy logic contains a fuzzification layer to transform the feature map into the fuzzy domain. Besides the fuzzification layer, the proposed fuzzy block contains an uncertainty representation layer and an uncertainty reduction layer.

Fuzzification layer: The input feature map of the fuzzy block is transformed into the fuzzy domain. Two membership functions are employed: trainable Sigmoid and trainable Gaussian membership functions. Each input node (a pixel in the input feature map) is transformed by the membership function. Let $x_i \in \mathbb{R}^D$ be the input node; i represents the ith pixel; D represents the dimension of the feature. Here the gray-level channel is used as an example to show the membership and uncertainty intuitively. o_i^r represents the output node and r represents the category index. The trainable Sigmoid membership function for the fuzzification layer is computed by Eq. (2.4):

$$o_i^r = \frac{1}{1 + \exp(a_i^r(x_i - b_i^r))}, i = 1, 2, 3, ..., n; r = 0, 1, 2, 3, 4 \tag{2.4}$$

where n represents the number of pixels in the image; r has 5 values: 0 represents the background; 1 represents the tumor; 2 represents the fat layer; 3 represents the mammary layer; 4 represents the muscle layer. $a_i^r \in \mathbb{R}^D$ and $b_i^r \in \mathbb{R}^D$ represent the parameters of the membership function for pixel i. For every category, different pairs of parameters are obtained during training, and the membership of the category is calculated using these parameters. In BUS images, tumor areas have low intensities in the spatial domain, but other layers, such as the mammary layer, have higher intensities. By changing the parameters a_i^r and b_i^r, trainable Sigmoid function can represent the membership of each category. The trainable Gaussian membership function is also used to compare with the trainable Sigmoid membership function to demonstrate the usefulness of fuzzy logic in handling uncertainty. The trainable Gaussian membership function is computed by Eq. (2.5):

$$o_i^r = \exp(-\frac{1}{2}(x_i - \mu_i^r)^{\mathrm{T}} \sigma_i^{r-1}(x_i - \mu_i^r)),\ i = 1, 2, 3, \ldots, n; r = 0, 1, 2, 3, 4 \tag{2.5}$$

where $\mu_i^r \in \mathbb{R}^D$ and $\sigma_i^r \in \mathbb{R}^{D\times D}$ represent the mean and covariance matrix of category r, which are utilized to obtain the memberships of different categories.

The fuzzy memberships are normalized by Eq. (2.6). It makes the summation of memberships in different categories of a pixel become one:

$$\pi_i^r = \frac{o_i^r}{\sum_{r=0}^{4} o_i^r} \tag{2.6}$$

where π_i^r represents the membership for pixel i of category r after normalization.

Heatmaps in Fig. 2.4 are utilized to represent the membership values on gray-level intensity; blue represents low membership value, and red represents high membership value. In Figs. 2.4 (a)-(e), the memberships are computed by the trainable Gaussian membership function on gray-level intensity; and in (f)-(j), the memberships are computed by the trainable Sigmoid membership function on gray-level intensity.

The parameter b_i^r in the trainable Sigmoid membership function is initialized by the mean of the intensities of all training samples in category r. The parameter a_i^r is initialized by 0. The parameter μ_i^r is initialized by the mean of the intensities of all training samples in category r, and σ_i^r is initialized by the covariance of the samples in the same category. All the channels of the input feature map are transformed to the fuzzy domain.

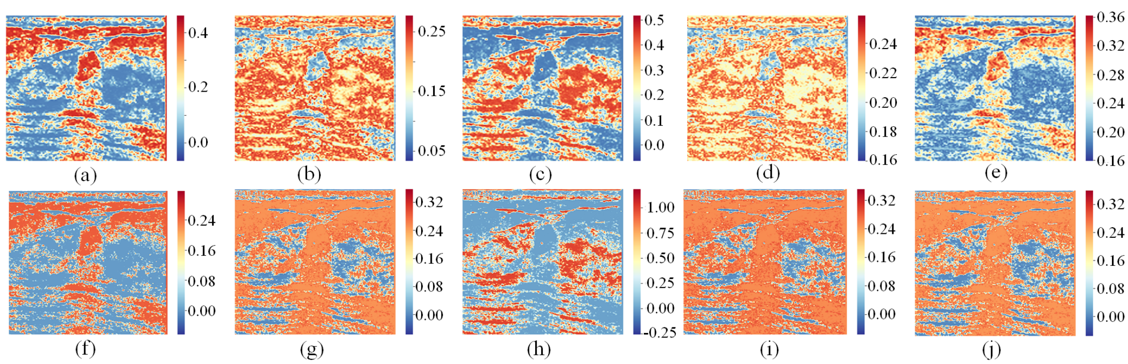


Fig. 2.4: The membership maps: (a)-(e) the memberships of tumor, fat layer, mammary layer, muscle layer, and background computed by the trainable Gaussian function; (f)-(j) the corresponding memberships computed by the trainable Sigmoid function.

Uncertainty representation layer: If the membership of a pixel is close to 1 or close to 0, the uncertainty of the pixel is low. If the membership is around 0.5, the uncertainty is high. It is hard to determine whether the pixel belongs to which category. The inputs of this layer are the fuzzy memberships, and the uncertainties in corresponding categories are computed using Eq. (2.7):

$$u_i^r = \begin{cases} 2 \times \pi_i^r & if\, \pi_i^r < 0.5,\, r = 0, 1, 2, 3, 4 \\ 2 \times (1 - \pi_i^r) & if\, \pi_i^r > 0.5,\, r = 0, 1, 2, 3, 4 \end{cases} \tag{2.7}$$

where π_i^r is the membership of pixel i in the rth category, which is the output of Eq. (2.6). u_i^r represents the uncertainty degree of pixel i in the rth category. The heatmaps of the uncertainty maps on gray-level intensities are generated as shown in Fig. 2.5.

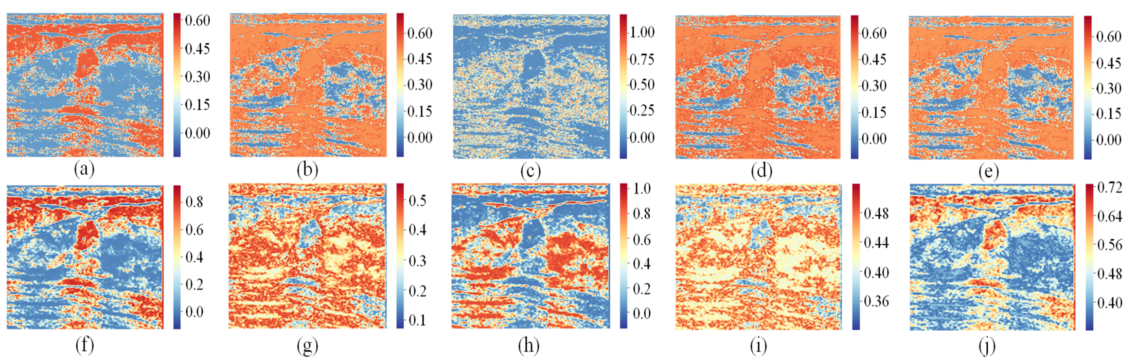


Fig. 2.5: Uncertainty maps: (a)-(e) are the uncertainty maps of tumor, fat layer, mammary layer, muscle layer, and background, which are generated by the trainable Sigmoid membership function; (f)-(j) are generated by the trainable Gaussian membership function.

Heatmaps in Fig. 2.5 show the uncertainty on gray-level intensity in different categories. The red areas have high uncertainties, and blue areas have low uncertainties in corresponding categories. To compute the overall uncertainty, a minimum operation is applied to u_i^r as shown in Eq. (2.8):

$$u_i = \min_r u_i^r,\, r = 0, 1, 2, 3, 4 \tag{2.8}$$

where u_i represents the overall uncertainty degree for pixel i. u_i^r is calculated by Eq. (2.7). The overall categories uncertainty maps on gray-level intensities are shown as the heatmaps in Fig. 2.6.

From Fig. 2.6, it can be observed that the pixels on the boundaries between categories have high uncertainties. After applying the Gaussian membership function, uncertainty in the mammary layer, muscle layer, and fat area can appear clearly. The boundary, mammary layer, and muscle layer have high uncertainties. After applying the Sigmoid membership function, similar results can be obtained.

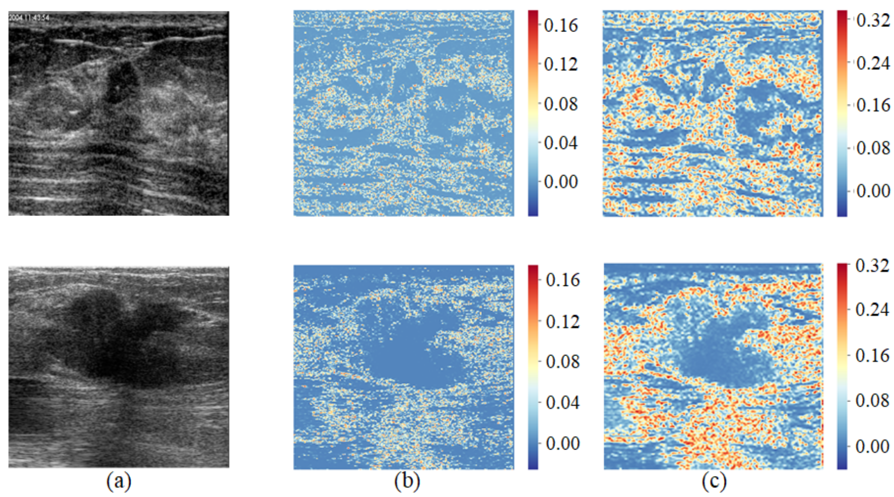


Fig. 2.6: Overall uncertainty maps: (a) original images; (b) overall categories uncertainty maps generated by using the trainable Sigmoid membership function; (c) overall categories uncertainty maps generated by using the trainable Gaussian membership function.

Uncertainty reduction layer: To reduce the uncertainty on the original channel, the overall categories uncertainty maps are fused with the corresponding original channels as shown in Eq. (2.9):

$$x'_i = (1 - u_i) \cdot x_i \tag{2.9}$$

where u_i is the overall categories uncertainty maps obtained by Eq. (2.8), and x_i is the original channels of the input. x'_i the feature after reducing uncertainty. This equation demonstrates if a pixel has high uncertainty, its weight should be reduced.

The results after reducing uncertainty for the input images are shown in Fig. 2.7. The boundary areas in Figs. 2.7 (c) and (d) are more distinct than that in Fig. 2.7 (a).

The resulted image is input to the convolutional layer for obtaining the convolutional feature maps. The base network structure is U-Net with VGG-16. The proposed fuzzy block is applied to the first convolutional feature map as well.

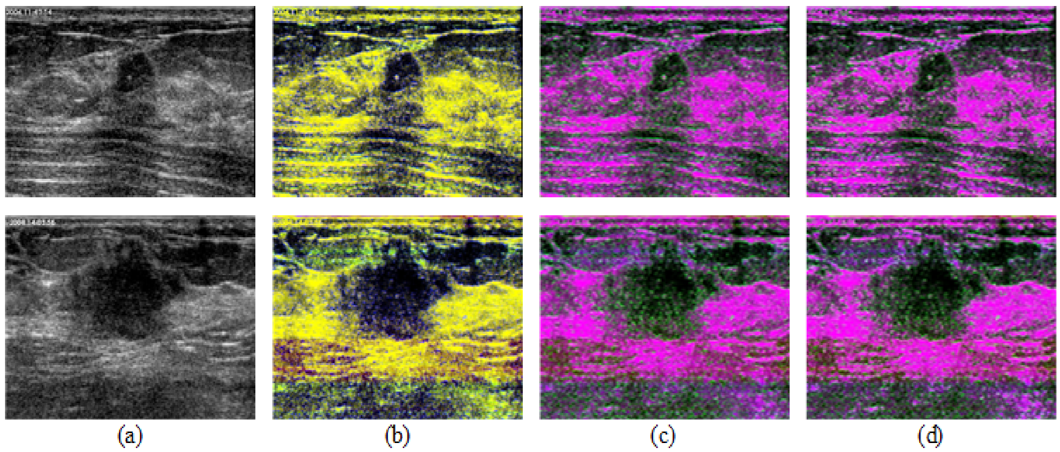


Fig. 2.7: The fusion of uncertainty maps and input images: (a) original images; (b) 3-channel images with gray-level intensity and wavelet information; (c) resulted images by the Sigmoid function and Eq. (2.9); (d) resulted images by the Gaussian function and Eq. (2.9).

2.4 Training Strategy for U-Net with Fuzzy Block

The uncertainty maps are multiplied with the input images, and the results are input to the first convolutional layer. The entire network structure is shown in Fig. 2.8. The output is processed by pixel-wised soft-max, which is defined as [48]:

$$p(x) = \exp(a(x)) / \sum_{r=0}^{4} \exp(a_r(x)) \tag{2.10}$$

where $a(x)$ is the output of the neural network, r represents the class index, and x represents the input pixel. The cross-entropy loss function is computed by the output probability and the label of each pixel:

$$C = -\sum q(x) \log\left(p(x)\right) \tag{2.11}$$

where $q\left(x\right)$ is the pixel label which is background, or tumor area, or fat layer, mammary layer or muscle layer with one-hot encoding. The original parameters in U-Net are initialized randomly. If using trainable Sigmoid membership function, parameter b_i^r in Eq. (2.4) is initialized by the mean of all training samples in category r. Parameter a_i^r in Eq. (2.4) is initialized by 0. The parameters μ_i^r and σ_i^r in Eq. (2.5) are initialized by the mean and covariance of the intensities of all training samples in category r. The training strategy is based on the back-propagation algorithm. All of the functions should be differentiable. Functions in the fuzzy layer using either fuzzy membership functions (trainable Sigmoid function and trainable Gaussian function) are all differentiable. The training strategy is shown in **Algorithm 1**.

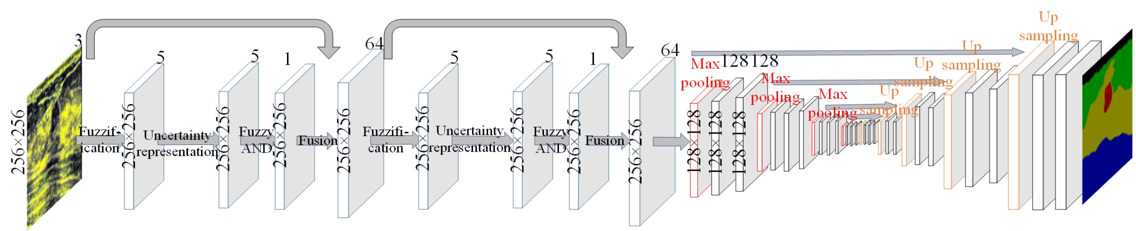


Fig. 2.8: Structure of the proposed U-Net with fuzzy block.

The U-Net with fuzzy block deals with the following issues: 1) it can reduce the uncertainty, and 2) it can solve small sample size problem, and it can even replace the information extension process in [53] (experimental details will be discussed in next section).

2.5 Experiment Results

2.5.1 Dataset

The performance of the proposed U-Net with fuzzy is evaluated by a dataset of 325 cases. Case 1 to 141 are collected over 10 years by the Second Affiliated Hospital of Harbin Medical University using VIVID 7 (GE) and EUB-6500 (Hitachi) imaging systems. Case 142 to Case 325 are collected in recent 3 years by the First Affiliated Hospital of Harbin Medical University using Aixplorer Ultrasound system (SuperSonic Imagine). The resolution of the first 141 cases is 550×450, and the rest 184 images have the resolution of 787×526. Informed

Algorithm 1 Training Strategy for U-Net with Fuzzy Block

Input: M training images: each is resized to 256×256; pixel-wise labels of the M samples; category number r; input channel number D; learning rate η, training epoch number S; batch size P; learning decay rate ϵ; and the parameters β_1, β_2 for Adam method.

Initialization: Parameters in fuzzy layer use the mean and variance of the training samples in each category to initialize. Other parameters are initialized randomly.

for $t = 1, 2, \ldots, S$ **do**
  for $m = 1, 2, \ldots, \frac{M}{P}$ **do**
    Input a batch of images to the network and obtain the error of loss function in Eq. (2.11).
    Compute the weight changing rate $\nabla\omega$ using the back-propagation algorithm and the error of loss function for all the parameters in fuzzy layer and original U-Net. Then, compute the new parameters using Adam method and η, β_1, and β_2.
  end for
  Update the learning rate by the learning decay rate ϵ.
end for

Output: Weight vector of the neural network

consent to the protocol from all patients are acquired. The privacy of the patients is well protected.

An experienced radiologist from the First Affiliated Hospital of Harbin Medical University delineated the boundaries of the layers and tumors. The pixel-wise groundtruths are generated according to the manually delineated boundaries. The proposed U-Net with the fuzzy block is applied and compared with U-Net [48], FCN-8s [14], and [53]. Three state-of-the-art deep learing semantic segmentation methods [15,17,18] are also involved in the comparison.

2.5.2 Evaluation Metrics

Three area metrics are used to evaluate the performance: True Positive Rate (TPR), False Positive Rate (FPR), and Intersection over Union (IoU) [38,57]. The IoU for every category is computed, and the mean over 5 categories IoU is used as the overall performance. The TPR and FPR for the tumor are used to compare the previous tumor segmentation method. Due to the limitation of the number of samples, 10-fold validation is used. The samples are randomly divided into 10 subsets. Each time, 9 of them are used for training, and 1 subset is used for testing. The metrics are computed by Eq. (2.12):

$$\begin{aligned} \text{TPR} =& |A_r \cap A_m|/|A_m| \\ \text{FPR} =& |A_r \cup A_m - A_m|/|A_m| \\ \text{IoU} =& |A_r \cap A_m|/|A_r \cup A_m| \end{aligned} \tag{2.12}$$

where A_r is the region generated by the proposed method or existing methods, and A_m is the region of the groundtruth.

2.5.3 Experiment Details

The proposed fuzzy U-Net is not pre-trained using other datasets. All other compared networks except FCN-8s are not pre-trained on other datasets as well. FCN-8s uses the pre-trained weight parameters on ImageNet [58]. All networks are trained on a computer with Ubuntu 18.04 system, Intel(R) Xeon(R) CPU E5-2620 2.10GHz and 2 NVIDIA GeForce 1080 graphics cards. The batch size is 11. The optimizing method is Adam [59], with an initial learning rate 10^{-4}. The learning decay rate is 5×10^{-4}. The parameter β_1 for Adam method is 0.9, and the parameter β_2 for Adam's method is 0.999. The network weights are initialized randomly. The initialization for parameters in the trainable Sigmoid and Gaussian membership function is introduced in Section 2.4. The implementation is based on the Keras platform with the TensorFlow backend.

Table 2.1: Evaluation results on 325 cases dataset. Evaluation metric is IoU (%).

	Fat	Mammary	Muscle	Background	Tumor	Mean
U-Net [48] with original images	70.34	66.72	66.17	65.91	74.66	68.76
U-Net with 3-channel images [53]	84.05	75.92	74.89	78.35	74.88	77.62
FCN-8s [14] with original images using pretrained model	82.57	75.47	75.53	**78.59**	74.42	77.32
ResNet-101 [15] with original images	81.50	73.41	72.07	74.47	75.29	75.35
PSPNet [18] with original images	82.07	74.40	74.49	77.36	74.75	76.61
Deeplab [17] with original images	78.91	68.71	67.33	73.94	69.04	71.58
Fuzzy U-Net with 3-channel images and Sigmoid membership function	**84.07**	76.01	74.62	78.39	**78.53**	**78.32**
Fuzzy U-Net with 3-channel images and Gaussian membership function	83.47	74.73	73.95	77.51	75.32	70.00
Fuzzy U-Net with original images and Sigmoid membership function	82.56	**76.14**	**74.64**	75.98	77.56	77.38

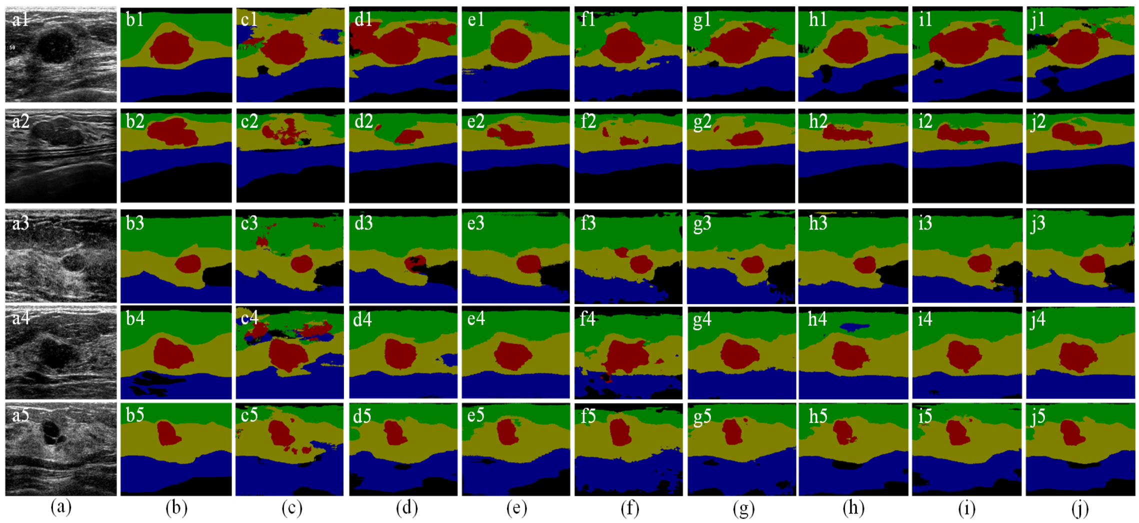


Fig. 2.9: The semantic segmentation results: (a) original images; (b) groundtruths; (c) results of U-Net with gray-level images as the inputs; (d) results of U-Net with 3-channel images as inputs; (e) results of PSPNet; (f) results of Deeplab; (g) results of ResNet-101; (h) results of the proposed fuzzy U-Net with trainable Gaussian membership function and 3-channel images; (i) results of the proposed fuzzy U-Net with trainable Sigmoid membership function and 3-channel images; (j) results of the proposed fuzzy U-Net with trainable Sigmoid membership function and gray-level images.

2.5.4 Segmentation Result of U-Net with Fuzzy Block

To show the effectiveness of the fuzzy logic, the proposed fuzzy operations are applied to U-Net. 3-channel images are defined as images that are augmented by our proposed method (histogram equalization and wavelet transform) in this research. Nine networks are trained: 1) U-Net with gray-level images as inputs; 2) U-Net with 3-channel images as inputs; 3) the proposed fuzzy U-Net with 3-channel images as inputs; trainable Sigmoid membership function is used; 4) the proposed fuzzy U-Net with 3-channel images as inputs; trainable Gaussian membership function is used; 5) to demonstrate the existence of the uncertainty in BUS images and the effectiveness of the fuzzy layer; augmentation on the input images is removed, and the proposed fuzzy U-Net is trained using a gray-level image; only trainable Sigmoid membership function is used; 6) FCN-8s using the original gray-level image as input and pre-trained weight parameters by nature images; 7) PSPNet with gray-level images as inputs; 8) U-shape network with ResNet-101 structure; the input image is gray-level images; 9) Deeplab with gray-level images as inputs.

Fig. 2.9 (b) shows the pixel-wise groundtruths. The black area is the background; the green area is the fat layer; the yellow area is the mammary layer; the blue area is the muscle layer; the red area is the tumor. Fig. 2.9 (c) shows the results of the U-Net with original gray-level images as the inputs, which are the worst. Adding wavelet information can make the segmentation results better in some cases. For example, some misclassified image patches in Fig. 2.9 c3 fat layer (red patches in the green area) are corrected in Fig. 2.9 d3. Comparing Fig. 2.9 c4 and Fig. 2.9 d4, the misclassified patches in the fat layer are corrected as well. The same situation happens in Figs. 2.9 c5 and d5. These experimental results demonstrate that the original gray-level intensity does not work well in the segmentation of BUS images. Adding new features such as the wavelet feature can increase the dimension of the feature, and the segmentation results become better. However, sometimes adding wavelet information can make the results worse. For example, if using gray-level images as inputs, the original U-Net can segment the tumor well in Figs. 2.9 c1 and c3; however, using wavelet transform on input images, the segmentation results become worse in Figs. 2.9 d1 and d3. These results also prove that there exists uncertainty in features, and adding new features can avoid uncertainty in some cases. Meanwhile, our dataset is small. The information is not enough for classification if just using gray-level intensity. Adding wavelet information increases information used in classification because the dimension of the feature increases, and there might be less noise than the gray-level feature. However, new features might cause new uncertainty as well. If adding fuzzy processing and reducing uncertainty in the 3-channel input images, the uncertainty in both gray-level feature and wavelet feature is reduced. The results are better (Figs. 2.9 (g) and (h)). Even if not applying wavelet transform and pre-processing, the fuzzy U-Net can still achieve good results because we reduce the uncertainty in gray-level intensities. In Table 2.1, the evaluation results of all networks are listed. Bold numbers are the corresponding best results. The proposed methods can improve BUS image semantic segmentation. The IoU on the tumor is 78.53% by using fuzzy U-Net with 3-channel images and trainable Sigmoid membership function. It achieves a 4% improvement than that of non-fuzzy U-

Net. The overall IoU over the 5 categories is 78.32% using fuzzy U-Net with 3-channel images and trainable Sigmoid membership function and has a 0.7% improvement than that of the non-fuzzy U-Net. The state-of-the-art methods: Deeplab, PSPNet, and U-Net with ResNet-101, do not achieve good results. The possible reason is lacking training images.

2.6 Conclusion

In this chapter, a novel BUS image semantic segmentation method is proposed. It can achieve good semantic segmentation results. A novel fuzzy block is proposed and applied to U-Net with VGG-16. The fuzzy block can detect and measure the uncertainty of pixels in the input image and the first convolutional feature map. The experimental results demonstrate that the proposed fuzzy U-Net can handle the uncertainty well. The robustness and accuracy of the fuzzy U-Net are better than that of the non-fuzzy U-Net. The proposed method solves the following issues to achieve much better results: 1) it uses fuzzy logic to handle the uncertainty in the original image and feature maps of the convolutional layers; 2) fuzzy approach can provide more information; 3) a novel membership function, trainable Sigmoid function is utilized and achieve better results; 4) uncertainty mapping function is designed and make the combination of fuzzy information and non-fuzzy information more reasonable. There are still three potential improvements: 1) designing better uncertainty representation methods, and 2) applying the fuzzy operators to more convolutional blocks, and 3) reducing uncertainty in channels of the feature map.

CHAPTER 3

SPATIAL AND CHANNEL-WISE FUZZY UNCERTAINTY REDUCTION BLOCKS IN DEEP NEURAL NETWORKS

In the previous chapter, we firstly define the fuzzy block and apply the fuzzy block to the input image and the first convolutional feature map of U-Net with VGG-16. We also conduct experiments on a dataset of 325 BUS images. Here in this chapter, the performance of the fuzzy block is improved. Two novel fuzzy attention mechanisms: the spatial-wise and channel-wise fuzzy blocks, are added to the classic U-shape network with a ResNet-101 network structure. The Spatial and Channel-wise Fuzzy Uncertainty Reduction Network (SCFURNet) is proposed to reduce uncertainty and noise in BUS images and conduct semantic segmentation. The major contributions of this research are: (1) Spatial-wise fuzzy blocks are applied to measure and reduce the spatial uncertainties (spatial dimension), and channel-wise fuzzy blocks are proposed to handle the channel uncertainty (channel dimension); (2) Membership functions in fuzzy blocks are defined by two layers of 1×1 convolutional operator with Sigmoid activation function to increase the non-linearity; the trainable Sigmoid and Gaussian membership functions in Chapter 2 are not used; (3) Fuzzy entropy [60–63] calculated by the memberships of different categories is utilized to measure the uncertainties for pixels and channels instead of the uncertainty map function in Chapter 2, which are defined as the uncertainty degrees. Uncertain pixel and channel are the pixel and channel with higher fuzzy entropies. We conduct more experiments on four datasets: 1) a five-category BUS image dataset with 325 images, and 2) three BUS image datasets with only tumor category (1830 images in total). The proposed approach compares state-of-the-art methods such as U-Net with VGG-16, ResNet-50/ResNet-101, Deeplab, FCN-8s, PSPNet, U-Net with information extension, attention U-Net, and U-Net with the self-attention mechanism.

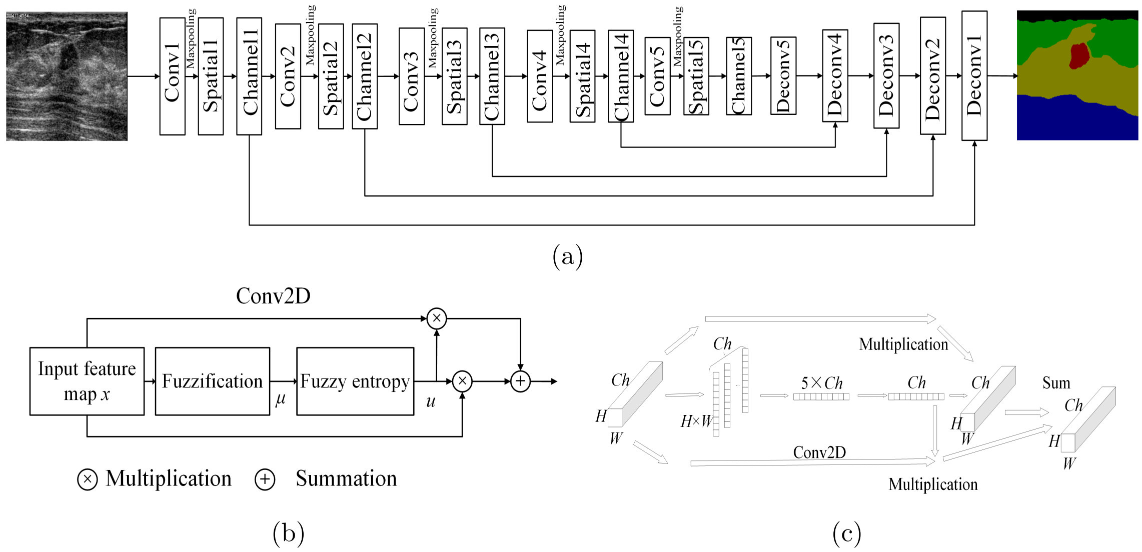


Fig. 3.1: The proposed SCFURNet: (a) the entire structure; (b) the spatial-wise fuzzy block; (c) the channel-wise fuzzy block.

3.1 Overview of the Proposed Network Structure

The proposed SCFURNet contains two major components: 1) the spatial-wise fuzzy uncertainty reduction block and 2) the channel-wise fuzzy uncertainty reduction block. Two fuzzy blocks are applied to the convolutional blocks. The entire network structure is shown in Fig. 3.1 (a). The network is based on U-Net. VGG-16 [22] and ResNet-101 [15] network structures in the convolutional blocks are utilized for comparison. The spatial-wise fuzzy block consists of fuzzification, uncertainty representation, pixel-wise multiplication, and summation. The channel-wise fuzzy block contains reshaping, fuzzification, uncertainty representation, and assigning a weight of each channel.

3.2 Spatial-wise Fuzzy Uncertainty Representation and Reduction

A spatial-wise fuzzy block is utilized to calculate the uncertainty of each pixel and reduce the uncertainty in each convolutional feature map. There are three major components in the spatial-wise fuzzy block: fuzzification, uncertainty representation, and uncertainty reduction. The flowchart of the spatial-wise fuzzy block is shown in Fig. 3.1 (b).

Fuzzification: Each input node from the original feature map is mapped to the fuzzy domain by membership function $f(\cdot)$:

$$\mu_i = f(x_i) \tag{3.1}$$

where $f(\cdot)$ represents the membership function; x_i represents the input node i (here it is a pixel in the input feature map $X \in \mathbb{R}^{H \times W \times Ch}$. H, W, and Ch represent the height, width, and the number of channels of the feature map, respectively); and μ_i represents the memberships of the input node. In some researches [26,29], $f(\cdot)$ is an S-shape function, Sigmoid function, or Gaussian function. In this research, the original features are transformed into fuzzy domain by the trainable Sigmoid membership function:

$$\mu_{ir} = \frac{1}{1 + \exp(\alpha_{ir} x_i + \beta_{ir})} \tag{3.2}$$

where $x_i \in \mathbb{R}^{Ch}$ is the ith pixel in the input feature map. $\alpha_{ir} \in \mathbb{R}^{Ch}$ and $\beta_{ir} \in \mathbb{R}$ are two trainable parameters for the trainable Sigmoid function, and $\mu_{ir} \in \mathbb{R}$ represents the membership in the rth category.

A 1×1 convolutional operation can perform the Sigmoid membership function. In this research, two layers of 1×1 convolutional operator are used as the membership function:

$$\mu_i = Conv1 \times 1(Conv1 \times 1(x_i)) \tag{3.3}$$

where μ_i is the membership vector and $\mu_i = [\mu_{i1}, \mu_{i2}, ..., \mu_{iC}]$; $Conv1\times$ 1 represents the 1×1 convolutional operator; both convolutional operators contain C kernels. Here, two-layer 1×1 convolution is utilized, and it can enable the membership to fit different categories. C is the number of categories. The outputs are normalized by the Soft-max function.

Uncertainty representation: Fuzzy logic is used to handle uncertainty. The memberships express the degrees that the pixel belongs to the categories and can be used to measure the uncertainty. There is an observation for uncertain pixels: if a pixel contains similar memberships of different categories, it is hard to assign to a category. Fuzzy entropy is utilized to reflect such observation, *i.e.*, an uncertain pixel is defined as a pixel with high fuzzy entropy (close to 1); and a certain pixel is defined as a pixel with low fuzzy entropy (close to 0).

For membership vector μ_i, the fuzzy entropy is defined as below [60]:

$$H(\mu_i) = -\frac{1}{\log C} \times \sum_{r=1}^{C} \mu_{ir} \log \mu_{ir} \tag{3.4}$$

where C represents category number; and μ_{ir} represents the membership of category r. If the memberships for all categories are the same, *i.e.*, $\mu_{ir} = 1/C$, the entropy is the highest, *i.e.*, $H(\mu_i) = 1$. In the spatial-wise fuzzy block, the memberships computed in Eq. (3.3) are utilized to calculate the fuzzy entropy as Eq. (3.5):

$$u_i = H(\mu_i) \tag{3.5}$$

where μ_i represents the memberships from Eq. (3.3); and u_i is the uncertainty degree of pixel i, which is in [0, 1], where 0 represents low uncertainty and 1 represents high uncertainty. Every pixel in the input feature map contains the corresponding uncertainty degree. The uncertainty degrees for all pixels consist of the uncertainty map. The uncertainty map has the same size as the input feature map.

Uncertainty reduction: If the uncertainty degree u_i is close to 1, the feature for pixel i generated in the convolutional block is uncertain. If the uncertain degree u_i is close to 0, the feature for pixel i obtained in the convolutional block is useful for the final decision. The features of uncertain pixels should reduce weight in the novel feature map. The features will replace the uncertain pixels to reduce the uncertainty.

As shown in Fig. 3.1 (b), the uncertainty map (u) which consists of uncertainty degrees (u_i) in Eq. (3.5) is utilized as the weight in the combination of the input feature map and a novel feature map:

$$X' = (Conv2D(X) \otimes u) \oplus X \otimes (1 - u) \tag{3.6}$$

where X' represents the novel feature map after reducing uncertainty; $Conv2D$ represents a 2-dimensional 3×3 convolutional operator; $\otimes$ represents the pixel-wise multiplication of matrices, and $\oplus$ represents the pixel-wise summation of matrices. This uncertainty reduction operator indicates that if u is close to 0, *i.e.*, X has low uncertainty, the weights of original features remain high. If u is close to 1, *i.e.*, X has high uncertainty, the weights of original features are reduced and should be replaced. Therefore, a novel feature is extracted by a 3×3 convolutional operator. The new feature map X' is passed to the next operator.

In this section, a novel fuzzification method is utilized to transform the original convolutional feature map into the fuzzy domain. Then, uncertainty is computed using fuzzy entropy. New convolutional features and original features are combined to reduce uncertainties

3.3 Channel-wise Fuzzy Uncertainty Representation and Reduction

The proposed channel-wise fuzzy blocks process the uncertainty in channels after reducing the uncertainty in pixels. Motivated by the channel-wise attention mechanisms [20, 21] and fuzzy logic, the channel-wise fuzzy block utilizes fuzzy entropy to measure the uncertainty degree of the channels of feature maps. An uncertain channel is a channel with higher fuzzy entropy (close to 1). There are also three major components in the channel-wise fuzzy block: fuzzification, uncertainty representation, and uncertainty reduction (Fig. 3.1 (c) and Fig. 3.2).

Fuzzification: Let $X \in \mathbb{R}^{H \times W \times Ch}$ be the input feature map. H and W represent the height and width of the feature map, respectively, and Ch is the number of channels. To calculate the uncertainty degree of each channel, it firstly transforms the input feature map into the fuzzy domain in the channel dimension. It reshapes X to $V \in \mathbb{R}^{HW \times Ch} = [v_1, v_2, ..., v_{Ch}]$, where $v_j \in \mathbb{R}^{HW}$ is the feature vector of channel j. For each v_j, a trainable Sigmoid membership function is utilized to transfer each feature vector v_j to the fuzzy domain:

$$\mu_{jr} = \frac{1}{1 + \exp(\alpha_{jr} v_j + \beta_{jr})} \tag{3.7}$$

where $\mu_{jr} \in \mathbb{R}$ represents the membership of category r for channel j; $\alpha_{jr} \in \mathbb{R}^{HW}$ and

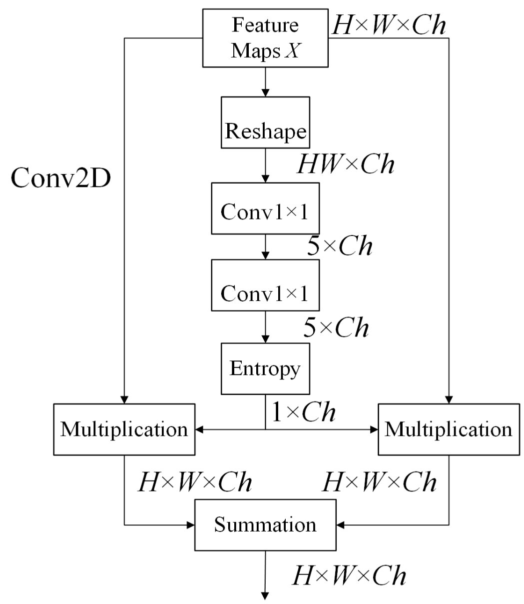


Fig. 3.2: Channel-wise fuzzy block.

$\beta_j \in \mathbb{R}$ are parameters of channel j. The membership is also performed by using two 1×1 convolutional operators with C kernels:

$$\mu_j = Conv1 \times 1(Conv1 \times 1(v_j)) \tag{3.8}$$

where $\mu_j \in \mathbb{R}^C$ represents the membership vector and $\mu_j = [\mu_{j1}, \mu_{j2}, ..., \mu_{jC}]$, and C represents the number of categories. For each channel, there is a membership vector.

Uncertainty representation: After obtaining the memberships, the fuzzy entropy is computed:

$$h_j = -\frac{1}{\log C} \times \sum_{r=1}^{C} \mu_{jr} \log \mu_{jr} \tag{3.9}$$

where $h_j \in \mathbb{R}$ represents the fuzzy entropy of channel j, which measures the uncertainty degree of channel j. Finally, the uncertainty degrees h_j of all channels in the feature map consist of the uncertainty vector $h \in \mathbb{R}^{Ch} = [h_1, h_2, h_3, ..., h_j, ..., h_{Ch}]$.

Uncertainty reduction: Like the spatial-wise fuzzy block, the uncertainty vector h is utilized as the weight vector for the combination of the input feature map and a novel feature map. A 3×3 convolutional operator generates the novel feature map. Each element in h is the weight value of the corresponding channel:

$$X_{Ch} = Conv2D(X) \odot h \oplus X \odot (1 - h) \tag{3.10}$$

where X_{Ch} is the feature map after applying the channel-wise fuzzy block; and $\odot$ represents the channel-wise multiplication. The channel-wise uncertainty reduction operator indicates if h is close to 0, the corresponding channels in the input feature map have low uncertainties, and these channels should contain high weights. If h is close to 1, *i.e.*, the corresponding channels have high uncertainties. The weights of these channels are reduced, and a new feature should replace the input feature.

3.4 Experiment Results

3.4.1 Datasets

To show the effectiveness of the proposed SCFURNet in BUS image semantic segmentation, two kinds of experiments are designed: 1) multi-object (multi-layer) semantic segmentation and 2) binary semantic segmentation (tumor and background). The multi-object semantic segmentation is performed on a dataset having 325 BUS images. The dataset is collected by the Second Affiliated Hospital of Harbin Medical University and the First Affiliated Hospital of Harbin Medical University. An experienced radiologist from the First Affiliated Hospital of Harbin Medical University delineates the boundaries of the four breast layers and tumors. This dataset is the same dataset mentioned in Chapter 2. The pixel-wise groundtruths for five categories: fat layer, mammary layer, muscle layer, tumor, and background are generated according to the manually delineated boundaries. In multi-object semantic segmentation task, the proposed method is compared with state-of-the-art deep learning segmentation methods such as U-Net with VGG-16 [22], U-Net with ResNet-50/ResNet-101 [15], Deeplab V3+ [64], FCN-8s [14], PSPNet [18], and U-Net with information extension [53]. We also compared the proposed methods with some spatial and channel-wise attention modules such as attention U-Net [20], SE-Net [21], and self-attention mechanism [34].

The binary semantic segmentation is performed on three public BUS image datasets [1–3]. Dataset [2] contains 163 BUS images, including 109 benign samples and 54 malignant samples. Dataset [3] contains 780 BUS images, including 437 benign, 210 malignant, and 133 no tumor images. Reference [1] is a BUS image benchmark with 562 images and lists five non-deep learning methods [4–8] for BUS image segmentation. In this task, state-of-the-art semantic segmentation network structures are also applied for comparison. Also, five traditional tumor segmentation methods [4–8] are utilized for comparison. The summary of the four datasets used in experiments is listed in Table 3.1.

Table 3.1: Dataset properties.

	Image Number	Ground Truths
Dataset 1 [2]	163	Tumor/Background
Dataset 2 [3]	780	Tumor/Background
Dataset 3 [1]	562	Tumor/Background
Multi-layer Dataset	325	Fat/Mammary/Muscle/Tumor/Background

3.4.2 Experiment Details

Preprocessing and augmentation: Because of the number limitation of samples, the training samples are augmented by horizontal flip, horizontal shift, vertical shift, rotation, zooming, and shear mapping. The input images are all gray-level images, and intensities are mapped into [-1, 1] by $(x/127.5 - 1)$ [65], where x represents the original intensity. No other augmentation methods are used except U-Net with information extension [53]. The input images are firstly preprocessed by histogram equalization. Then, images are transformed into the wavelet domain. New three-channel images with grey-level intensity in the first channel, wavelet approximation coefficients in the second channel, and wavelet detail coefficients in the third channel are utilized for training the original U-Net.

Experiment environment: All the networks in this chapter are not pre-trained using other datasets. The network weights are initialized randomly. The input image is resized to 128 × 128. The batch size is 12. The optimizing method is the SGD method [66], with a learning rate of 0.001 and momentum of 0.99. All the comparison networks and the proposed method are trained using a computer with Ubuntu 18.04 system, Intel(R) Xeon(R) CPU E5-2620 2.10GHz and 2 NVIDIA GeForce 1080 graphics cards, and each one has 8 Gigabyte memory. The implementation uses PyTorch 1.6.0.

Loss function: The entire network structure is shown in Fig. 3.1 (a). The proposed spatial-wise fuzzy block and channel-wise fuzzy block are applied to five encoders of a U-shape network since the original network has five encoders. For comparing purposes, this paper uses the same number of encoders. The convolutional feature maps from five convolutional blocks are processed by the proposed spatial-wise fuzzy block and channel-wise fuzzy block, sequentially. The encoder of the U-shape network uses VGG-16 and ResNet-101 for comparison. The final layer is the pixel-wised Soft-max:

$$p_r(x) = \frac{\exp(a_r(x))}{\sum_{k=1}^{C} \exp(a_k(x))} \tag{3.11}$$

where x is the input of the network; $a_r(x)$ represents the output of the network; r represents the class index, and C represents the number of categories. The loss function is defined as the summation of category cross entropy loss, and fuzzy entropies from spatial and channel fuzzy blocks:

$$L = L_c + L_s + L_{Ch} \tag{3.12}$$

where the L_c is the classic cross entropy loss function:

$$L_c = -\sum_r l_r(x)\log(p_r(x)) \tag{3.13}$$

where $l(x) \in \mathbb{R}^C$ is the label of x in one-hot encoding. If x is in the rth category, the corresponding rth element in $l(x)$ is 1, and other elements are 0. L_s is computed by the fuzzy entropy (u_i) in the spatial-wise fuzzy blocks in Eq. (3.5). Because the spatial-wise fuzzy block is applied to five convolutional blocks, there are five fuzzy entropy maps from the five convolutional blocks and L_s is defined by the summation of fuzzy entropy maps:

$$L_s = \sum_l \sum_i u_i^l \tag{3.14}$$

where i represents the pixel index, and l represents the index of convolutional blocks. L_{Ch} is computed by the fuzzy entropy (h_j in Eq. (3.9)) in channel-wise fuzzy blocks:

$$L_{Ch} = \sum_l \sum_j h_j^l \tag{3.15}$$

where j represents the channel index.

3.4.3 Metrics

In binary semantic segmentation task, it utilizes metrics in [1] to evaluate the performance. There are five area metrics: True Positive Ratio (TPR), False Positive Ratio (FPR), Jaccard Index (JI), Dice's Coefficient (DS), and Area Error Ratio (AER). The area metrics are defined in the following equation:

$$\begin{aligned}
\mathrm{TPR} =& |A_r \cap A_m| / |A_m| \\
\mathrm{FPR} =& |A_r \cup A_m - A_m| / |A_m| \\
\mathrm{JI} =& |A_r \cap A_m| / |A_r \cup A_m| \\
\mathrm{DS} =& 2|A_r \cap A_m| / |A_r| + |A_m| \\
\mathrm{AER} =& (|A_r \cup A_m| - |A_r \cap A_m|) / |A_m|
\end{aligned} \tag{3.16}$$

where A_r is the set of pixels generated by the proposed method or existing methods, and A_m is the set of pixels in the groundtruths.

In the multi-object semantic segmentation task, Intersection over Union (IoU, also known as the Jaccard Index in the binary task) is a typical metric in semantic segmentation and chosen as the metric. It is computed by:

$$\mathrm{IoU} = |A_r \cap A_m| / |A_r \cup A_m| \tag{3.17}$$

where A_r and A_m are the sets of pixels generated by the algorithms and groundtruths, respectively. Mean IoU (mIoU $= \sum \text{IoU}/C$, and C represents the number of categories) over five categories to evaluate the overall performance.

3.4.4 Multi-object Semantic Segmentation of BUS Images

A dataset with 325 BUS images is utilized, and each of them contains pixel-wise groundtruths of five categories. 10-fold validation is also utilized. The proposed spatial-wise fuzzy blocks and channel-wise fuzzy blocks are applied to U-Net with VGG-16/ResNet-101 as the encoder.

Segmentation Performance and The Number of Fuzzy Blocks: In this subsection, we discuss the relation between the number of fuzzy blocks used in the network and the performance of the segmentation. The U-Net with ResNet-101 is utilized in this research. The proposed spatial-wise fuzzy block and the channel-wise fuzzy block are applied to the encoder of the U-Net with ResNet-101. The ResNet-101 contains 5 convolutional blocks; therefore, we use 5 fuzzy blocks as the maximum number to conduct experiments for comparison. In the first experiment, there is no fuzzy block applied to the U-Net with ResNet-101. In the second experiment, the proposed spatial and channel-wise fuzzy blocks are applied to the first convolutional block. We continue adding the spatial and channel-wise fuzzy blocks to the second, third, fourth, and fifth convolutional blocks and keeping the fuzzy blocks in the previous convolutional blocks.

Fig. 3.3 shows IoU results *vs.* the number of convolutional blocks. When we apply the spatial and channel-wise fuzzy blocks to all five convolutional blocks, the proposed network achieves the best performance on both tumor category and the overall performance. Since the existing structures only have 5 blocks, for comparison purposes, it has a maximum of five fuzzy blocks as well.

To show the increasing performance in Fig. 3.3 is caused by the former fuzzy blocks in deeper convolutional blocks or the combination of the fuzzy blocks and the newly added fuzzy blocks, another experiment is conducted. In this experiment, the fuzzy blocks are added to the five convolutional blocks of ResNet-101 individually. For example, the fuzzy

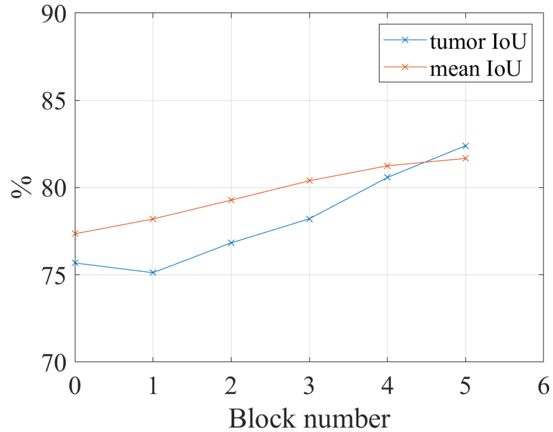


Fig. 3.3: The relation between the number of fuzzy blocks and the segmentation performance. Block number = 1: the fuzzy blocks are applied to the first convolutional block; block number = 2: the fuzzy blocks are applied to the first and second convolutional blocks together; block number = 3: the fuzzy blocks are applied to the convolutional blocks 1, 2, and 3; block number = 4: the fuzzy blocks are applied to the convolutional blocks 1, 2, 3, and 4; block number = 5: the fuzzy blocks are applied to the convolutional blocks 1, 2, 3, 4, and 5.

blocks are added to the second convolutional block of ResNet-101. There is no fuzzy block in convolutional blocks 1, 3, 4, and 5. The experiment results in Fig. 3.4 show a slight increase in performance when applying fuzzy blocks to convolutional blocks 1 to 5. However, the performance cannot outperform the performance of using fuzzy blocks in five convolutional blocks together. When we only add a fuzzy block to the fourth convolutional block, the IoU for tumor is the highest, which is 77.56%. However, when we add fuzzy blocks to all five convolutional blocks, the IoU for tumor is 82.40%. Therefore, the spatial and channel-wise fuzzy blocks are applied to five convolutional blocks in the following experiments.

Ablation Study for Fuzzy Blocks: We employed the spatial-wise fuzzy block (SFB) and the channel-wise fuzzy block (CFB) in five convolutional blocks to reduce the uncertainty in the feature maps. To verify the performance of each fuzzy block, we conduct experiments with different settings in Table 3.2.

As shown in Table 3.2, it compared two convolutional structures: VGG-16 and ResNet-101. Meanwhile, it adopts the spatial-wise fuzzy block and the channel-wise fuzzy block individually in each network. Compared with the U-Net with VGG-16, employing the

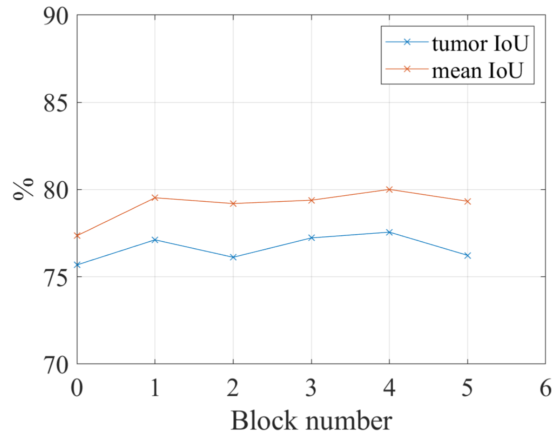


Fig. 3.4: The relation between the number of fuzzy blocks and the segmentation performance. The fuzzy blocks are applied to the convolutional blocks individually.

spatial-wise fuzzy block brings a 1.94% increase in tumor IoU and 2.41% in mean IoU. Meanwhile, employing the channel-wise fuzzy block in U-Net with VGG-16 outperforms the baseline by 0.97% in tumor IoU and 3.68% in mean IoU. When the two fuzzy blocks are used together to the U-Net with VGG-16, the performance further improved to 78.34% in tumor IoU and 79.36% in mean IoU. When changing the convolutional structure to ResNet-101, the performance of using two fuzzy blocks together becomes 82.40% in tumor IoU and 81.67% in mean IoU. The experiment results show that each fuzzy block can reduce uncertainty in the feature maps and increase the tumor segmentation results.

Table 3.2: Ablation study on multi-object dataset. SFB: Spatial-wise fuzzy block, CFB: Channel-wise fuzzy block.

Method	Encoder	SFB	CFB	Tumor IoU	Mean IoU
U-Net	VGG-16			74.66%	75.13%
SCFRNet	VGG-16	✓		76.60%	77.54%
SCFRNet	VGG-16		✓	75.63%	78.81%
SCFRNet	VGG-16	✓	✓	78.34%	79.36%
U-Net	ResNet-101			75.68%	77.35%
SCFRNet	ResNet-101	✓		79.12%	78.67%
SCFRNet	ResNet-101		✓	80.43%	80.12%
SCFRNet	ResNet-101	✓	✓	82.40%	81.67%

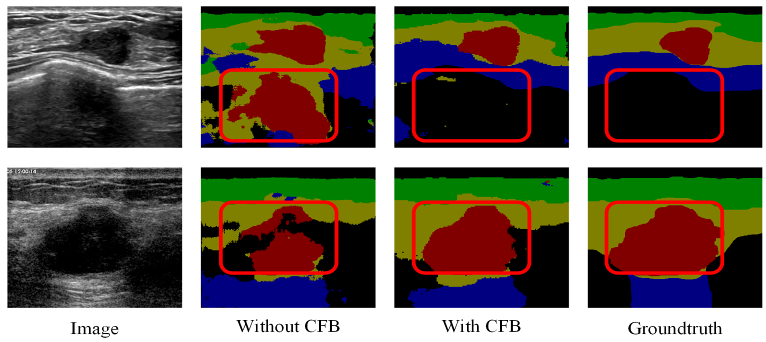


Fig. 3.5: Segmentation results of U-Net with ResNet-101 and channel-wise fuzzy block on multi-object dataset.

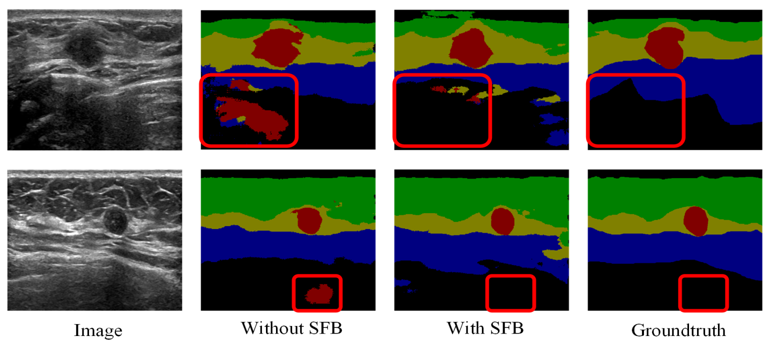


Fig. 3.6: Segmentation results of U-Net with ResNet-101 and spatial-wise fuzzy block on multi-object dataset.

The effectiveness of the proposed spatial and channel-wise fuzzy blocks can be shown in Fig. 3.5 and Fig. 3.6, respectively. The most common misclassification is the tumor area and the background area because both areas contain low intensities. Red rectangles mark the misclassification patches in Fig. 3.5 and Fig. 3.6. They are correctly classified when applied the spatial-wise fuzzy block or channel-wise fuzzy block individually.

Visualization of Fuzzy Blocks: In this part, the uncertainty maps obtained by the spatial-wise fuzzy block and selected channels in the processed feature maps are visualized to help to understand the spatial-wise fuzzy block and the channel-wise fuzzy block.

The spatial-wise fuzzy block is utilized to measure the uncertainty degree of pixels in the input feature map and reduce the effect of the uncertain pixels. Therefore, the uncertainty map generated in the spatial-wise fuzzy block can show the uncertain pixels and corresponding uncertainty degrees (refer to Fig. 3.7). For example, in the first row, the areas marked by red rectangles are background and tumor areas. They have similar intensities. In the uncertainty map, these areas are high uncertainty areas. The original U-Net misclassifies the background area; however, the proposed method can correct it (shown in columns 5 and 6). In the second row and third row, the tumor areas are also marked as the uncertain areas, *i.e.*, the original U-Net cannot handle these areas. The heatmaps indicate that the proposed spatial-wise fuzzy block can find uncertain areas of the input feature maps, and it can also measure the uncertainty degree of the pixels.

For channel-wise fuzzy block, it is hard to give an understandable visualization about the uncertainty map directly because each channel of the input feature map only contains an uncertainty value. Instead, we show some processed channels to see whether they highlight clear semantic areas. In Fig. 3.7, we display the 39th and 21st channels of each feature map after employing a channel-wise fuzzy block. We can see that in the 21st channel of the feature map, the highlighted areas are in the mammary layers. The 39th channel of the feature map highlights the area of the tumor. However, some areas in other categories contain high response in the 39th channel of the feature maps (such as the muscle layer in the first and third rows and the fat layer in the second row). These results indicate that the proposed fuzzy blocks can help to generate feature maps with clear semantic information; however, there still exists uncertain areas.

Semantic Segmentation Results on Dataset with 325 Images: The segmentation results on the multi-object dataset are in Fig. 3.8. Fig. 3.8 (b) shows the pixel-wise groundtruths: the green areas are fat layers; the yellow areas are mammary layers; the blue areas are muscle layers; the red areas are tumors, and the black areas are background areas. The results in Fig. 3.8 (i) are obtained when the input images are the three-channel images. The three-channel images are with gray-level intensity in the first channel, wavelet approximation coefficients in the second channel, and wavelet detail coefficients in the third channel. The network structure is the U-shape network with ResNet-101. The results in Fig. 3.8 (f) are obtained when the images are the original gray-level images and the network structure is the same as the network used in Fig. 3.8 (i). Comparing Figs. 3.8 (i) and (f), the tumor

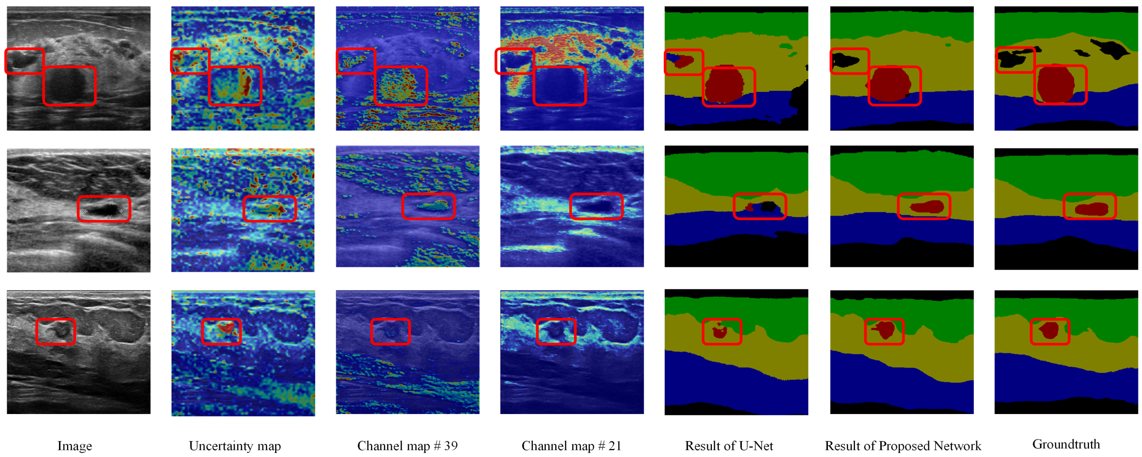


Fig. 3.7: Visualization results of fuzzy blocks on the multi-object dataset. For each row, we show an input image, an uncertainty map from the spatial-wise fuzzy block; red represents a high value and blue represents a low value in the heatmap. We also provide two channel maps from the outputs of the channel-wise fuzzy block, the results of the original U-Net and the proposed method, and the groundtruths.

segmentation results in Figs. 3.8 i2 and i4 are better than that in Figs. 3.8 f2 and f4. However, the results in Figs. 3.8 i1 and i3 are not improved. The experiment results of using wavelet feature in the input layer prove that involving wavelet feature cannot handle some misclassification such as the background area and tumor area because they contain similar feature value in both wavelet domain and space domain.

SCFURNet generated new convolutional features. New convolutional features and original convolutional features are combined according to the uncertainty degrees in pixels and channels. It reduces the effect of uncertain pixels and uncertain channels. This mechanism overcomes the drawback in Fig. 3.8 (i). For example, in Fig. 3.8 f3, the original U-Net with ResNet-101 can segment the tumor. In Fig. 3.8 i3, when adding wavelet features, the segmentation results of tumors and the mammary layer become worse. Other network structures also do not handle these images well. The performances are shown in Table 3.3. Bold numbers represent the corresponding best results. The IoU increases 6.72% in tumor segmentation compared with that of the original U-Net with ResNet-101. It achieves a 7.52% improvement in IoU in tumor segmentation compared with that of the U-Net with ResNet-101 and wavelet transform. The proposed method achieves 4.27% and 4.05%

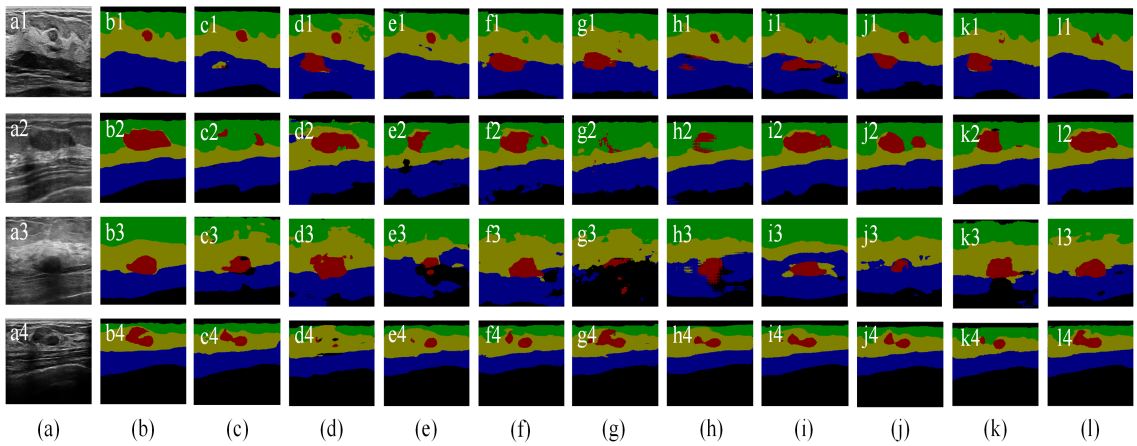


Fig. 3.8: Multi-object semantic segmentation of BUS images: (a) original images; (b) groundtruths; (c) results of ResNet-101 + self-attention mechanism; (d) results of attention U-Net; (e) results of ResNet-50; (f) results of ResNet-101; (g) results of Deeplab; (h) results of PSPNet; (i) results of U-Net with wavelet transform; and (j) results of FCN-8s; (k) results of SE-Net (ResNet-101); (l) results of the proposed SCFURNet.

improvements in overall mIoU compared with that of the U-Net with gray-level intensity and wavelet transform, respectively. The proposed method achieves the best performance in tumor segmentation and the best overall performance among all methods. The overall performance indicates that the proposed SCFURNet can handle misclassification caused by similar feature values of different layers because the proposed method can reduce the weights of the similar features of different layers and add novel features.

Table 3.3: Results of multi-object semantic segmentation. Evaluation metric is IoU (%).

	Fat	Mammary	Muscle	Background	Tumor	Mean
ResNet-50	82.58	73.98	73.08	77.23	76.34	76.64
ResNet-101	82.50	74.41	75.69	77.47	75.68	77.35
FCN-8s	82.57	75.47	75.53	78.59	74.42	77.32
PSPNet	82.07	74.40	74.49	77.36	74.75	76.61
Deeplab	78.91	68.71	67.33	73.94	69.04	71.58
Attention U-Net	83.99	77.61	75.69	77.99	76.26	78.31
SE-Net	80.91	75.21	71.23	76.57	75.90	75.96
Self-attention	82.53	76.23	75.91	80.29	78.81	78.75
[53]	84.05	75.92	74.89	78.35	74.88	77.62
SCFURNet	**84.72**	**79.84**	**77.39**	**83.98**	**82.40**	**81.67**

* Bold numbers are the best results.

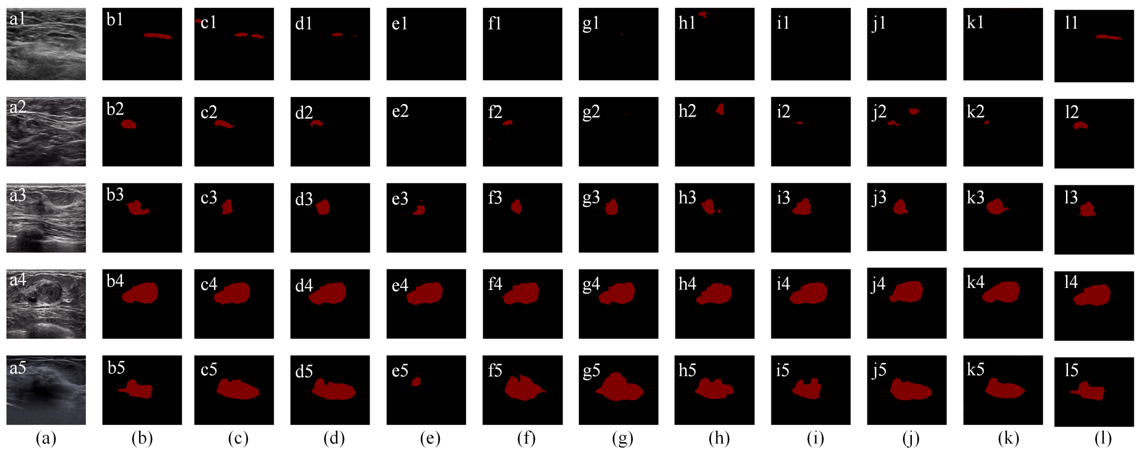


Fig. 3.9: Segmentation results using public dataset: (a) original images; (b) groundtruths; (c) results of ResNet-101 with self-attention mechanism; (d) results of a SE-Net (ResNet-101); (e) results of attention U-Net; (f) results of ResNet-50; (g) results of ResNet-101; (h) results of Deeplab; (i) results of PSPNet; (j) results of U-Net with wavelet transform; and (k) results of FCN-8s; (l) results of proposed SCFURNet.

3.4.5 Semantic Segmentation on Three Public Two-category Datasets

Overall Performance on Three Public Datasets: The proposed spatial-wise fuzzy block and channel-wise fuzzy block are applied to a U-Net with ResNet-101 network because it achieves better results compared with U-Net with VGG-16 in Subsection 3.4.4. All other compared deep networks such as ResNet-50, ResNet-101, and FCN-8s are trained to segment tumors in these three datasets. Because of the limited number of samples (the total number of samples for 3 datasets is only 1505), 10-fold validation is utilized: (1) each of the three datasets is divided into 10 groups randomly; (2) pick 9 groups of each dataset as the training set and the rest 1 group as the testing set; and (3) the final evaluation metrics are calculated by the average of 10 experiments.

Fig. 3.9 shows the segmentation results using the three two-category datasets [1–3]. Fig. 3.9 (a) shows the original images and (b) shows the groundtruths. Figs. 3.9 a1-a3 contain small tumors hard to discriminate from the shadow and background areas. There are some tumors with blurry boundaries in Figs. 3.9 a4 and a5. The results of the proposed SCFURNet are shown in Fig. 3.9 (l). It achieves impressive improvements, especially for small tumors (l1-l3) compared with that of the compared deep learning methods. Some

Table 3.4: Results of two-class semantic segmentation on dataset [1].

	TPR	FPR	JI	DS	AER
	Semi-Automatic Methods				
[4]	0.82	0.13	0.73	0.84	0.31
[5]	0.84	0.07	0.79	0.88	0.23
	Fully-Automatic Methods				
[7]	0.81	0.16	0.72	0.83	0.36
[6]	0.81	1.06	0.60	0.70	1.25
[8]	0.67	0.18	0.61	0.71	0.51
Deeplab	0.89	0.11	0.82	0.89	0.22
ResNet50	0.92	0.08	0.86	0.92	0.16
ResNet101	0.92	0.10	0.85	0.91	0.18
FCN8s	**0.94**	0.10	0.86	0.92	0.16
PSPNet	0.93	0.09	0.86	0.92	0.16
Attention U-Net	0.92	0.09	0.85	0.91	0.17
SE-Net	0.92	0.10	0.85	0.91	0.18
Self-attention	0.91	0.07	0.86	0.92	0.15
[53]	0.92	0.09	0.86	0.92	0.16
SCFURNet	**0.94**	**0.06**	**0.88**	**0.93**	**0.14**

* Bold numbers are the best results.

tumors such as a4 and a5 in Fig. 3.9 contain high uncertainties in the boundary areas. It is even hard for human to detect these tumors. In Fig. 3.9 a4, all the methods achieve good results; however, the proposed SCFURNet achieves the best result. In Fig. 3.9 l5, the proposed SCFURNet achieves the best results than the compared deep learning methods (Figs. 3.9 c5-l5). The evaluation metrics for different methods on the dataset [1] are listed in Table 3.4. Five non-deep learning methods [4–8] are also involved in the comparison using this dataset. Results in Table 3.4 show: (1) deep learning methods obtain improvements compared with traditional BUS image segmentation methods listed in [1]; (2) some famous deep learning architectures such as Deeplab, PSPNet, do not obtain improvements for dataset [1] and the possible reason is the limited number of the samples; and (3) the proposed method achieves the best results since it can solve the small target problems and uncertainties in the boundary areas.

The evaluation metrics for datasets [2] and [3] are shown in Table 3.5. The proposed method achieves the best results among all evaluation metrics compared with state-of-the-art deep learning methods on three public datasets except the FPR and AER on the

Table 3.5: Results of two-class semantic segmentation on dataset [2] and dataset [3].

		TPR	FPR	JI	DS	AER
Dataset [2]	Deeplab	63.68%	36.06%	52.93%	61.91%	72.38%
	ResNet50	81.29%	36.58%	68.70%	76.94%	55.29%
	ResNet101	83.58%	34.40%	71.43%	79.45%	50.82%
	FCN8s	82.72%	41.14%	67.50%	76.87%	58.42%
	PSPNet	81.08%	40.42%	69.77%	78.24%	59.34%
	Attention-UNet	83.58%	34.40%	71.43%	79.45%	50.82%
	Self-attention	82.58%	**26.39%**	73.83%	81.37%	**33.81%**
	SE-Net	79.23%	36.75%	70.90%	79.10%	35.12%
	[53]	81.19%	31.63%	71.48%	80.21%	48.44%
	SCFURNet	**84.70%**	44.69%	**73.27%**	**81.08%**	59.99%
Dataset [3]	Deeplab	59.88%	39.39%	49.65%	59.39%	79.52%
	ResNet50	78.45%	49.39%	67.09%	76.36%	68.94%
	ResNet101	79.40%	46.02%	69.26%	77.90%	66.62%
	FCN8s	74.23%	46.69%	63.16%	73.03%	72.63%
	PSPNet	77.11%	46.65%	65.21%	74.75%	69.54%
	Attention-UNet	77.52%	38.67%	67.81%	76.77%	60.92%
	Self-attention	79.02%	29.30%	71.49%	78.46%	55.50%
	SE-Net	78.40%	38.95%	68.30%	77.24%	60.55%
	[53]	78.07%	42.37%	68.43%	76.96%	64.30%
	SCFURNet	**79.86%**	**22.01%**	**72.14%**	**80.51%**	**42.15%**

* Bold numbers are the best results.

dataset [2]. The self-attention mechanism in ResNet-101 obtains lower FPR and AER on the dataset [2]. Lower FPR and AER indicate that non-local context information provided by the self-attention mechanism can help to reduce errors in segmentation. However, the proposed method achieves the best overall performance by reducing uncertainty in pixels and channels. The proposed network also achieves the best results among all evaluation metrics on the dataset [3]. Finally, the proposed method achieves 2.03%, 1.84%, and 2.85% in Jaccard Index on three public BUS datasets compared with that of the original U-shape network with ResNet-101, respectively.

Small Tumor Segmentation: In this part, the effectiveness of the proposed method on small tumor segmentation is shown. Some patches in the BUS images contain similar feature values to that of the tumor areas. For example, the patch in Fig. 3.10 a1 marked by the red rectangle is the background area and has low intensity. The tumor area in Fig. 3.10 a1 is small and close to the red rectangle area. Some methods misclassify the red

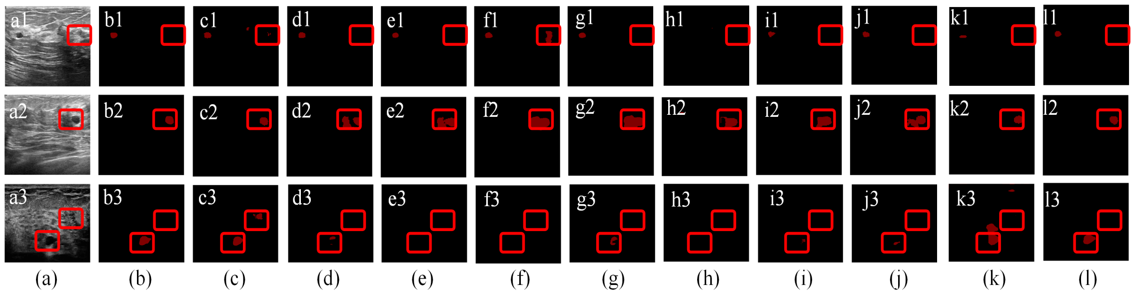


Fig. 3.10: Small tumor segmentation: (a) original images; (b) groundtruths; (c) results of ResNet-101 with self-attention mechanism; (d) results of a SE-Net (ResNet-101); (e) results of attention U-Net; (f) results of ResNet-50; (g) results of ResNet-101; (h) results of Deeplab; (i) results of PSPNet; (j) results of U-Net with wavelet transform; and (k) results of FCN-8s; (l) results of proposed SCFURNet.

rectangle area to the tumor, such as the U-Net with ResNet-101 in Fig. 3.10 f1. When applying the proposed spatial and channel-wise fuzzy block to U-Net with ResNet-101, the misclassification is solved by reducing the original feature values and involving new features. In Fig. 3.10 a2, the tumor is located near a patch with low intensity. Most of the previous methods misclassify the image to the tumor. This leads to a bigger segmented tumor area than the groundtruth (Figs. 3.10 d2-j2). It is because the small tumor contains similar feature values with noise patches or background patches. However, the proposed method achieves the best results in small tumor images; therefore, it can achieve the best overall performance on all datasets.

3.5 Conclusion

In this Chapter, the proposed fuzzy block is extended to two kinds of fuzzy blocks, and they are applied to the U-shape network with ResNet-101, and the proposed network is applied to BUS image semantic segmentation. The novel network achieves 2.03%, 1.84%, and 2.85% improvements in the Jaccard Index using three public BUS datasets than the original U-shape network with ResNet-101. The proposed method obtained IoU increases 6.72% in tumor segmentation and 4.27% in the overall performance in the five-category BUS dataset compared with that of the original U-shape network with ResNet-101. The proposed method achieves the best results due to the following reasons: (1) The proposed spatial and channel-wise fuzzy blocks can locate uncertain pixels and uncertain channels in feature maps and can reduce the influence of uncertain pixels and channels; (2) By reducing the uncertainty in feature maps, some patches having similar features with that of tumor areas can be classified correctly, especially for small tumors; (3) The fuzzy entropy of memberships can measure the uncertainty degree of pixels and channels accurately.

CHAPTER 4
PYRAMID FUZZY UNCERTAINTY REDUCTION NETWORK AND DIRECTION-CONNECTEDNESS FEATURE

4.1 Introduction

In this chapter, we continue to improve the performance of the fuzzy block in deep neural networks. To reflect the uncertainty of feature maps in different resolutions and provide BUS anatomy information, a novel deep learning architecture that can reduce uncertainty in feature maps and provide breast anatomy information is proposed. The entire network structure is shown in Fig. 4.1 (b). The U-shape network with VGG-16 [22] (Fig. 4.1 (a)) is chosen as the base network because it has achieved good performance in BUS image segmentation [2, 53]. Two novel structures are proposed and added to the original U-shape network: (1) a pyramid fuzzy uncertainty block which can resize input feature map to two different resolutions and reduce uncertainty; objects in different resolutions have different uncertain areas; (2) a new feature extraction block which can involve the context information, *i.e.*, BUS image layer structure information, using the connectedness between pixels and the boundary pixels in the up, down, right, left directions is calculated. The details of these two parts are introduced in Section 4.2 and Section 4.3.

4.2 Pyramid Fuzzy Block

The pyramid fuzzy block contains three portions: down-sampling part, fuzzification part, and uncertainty representation part (Fig. 4.1 (c)). In the down-sampling section, the input feature maps are down-sampled into two different resolutions. Then, the feature maps in various resolutions are fuzzified by a trainable Sigmoid function. The uncertainty degree of each pixel is calculated by fuzzy entropy for memberships of different categories. In this research, a pixel with a high fuzzy entropy value is treated as an uncertain pixel.

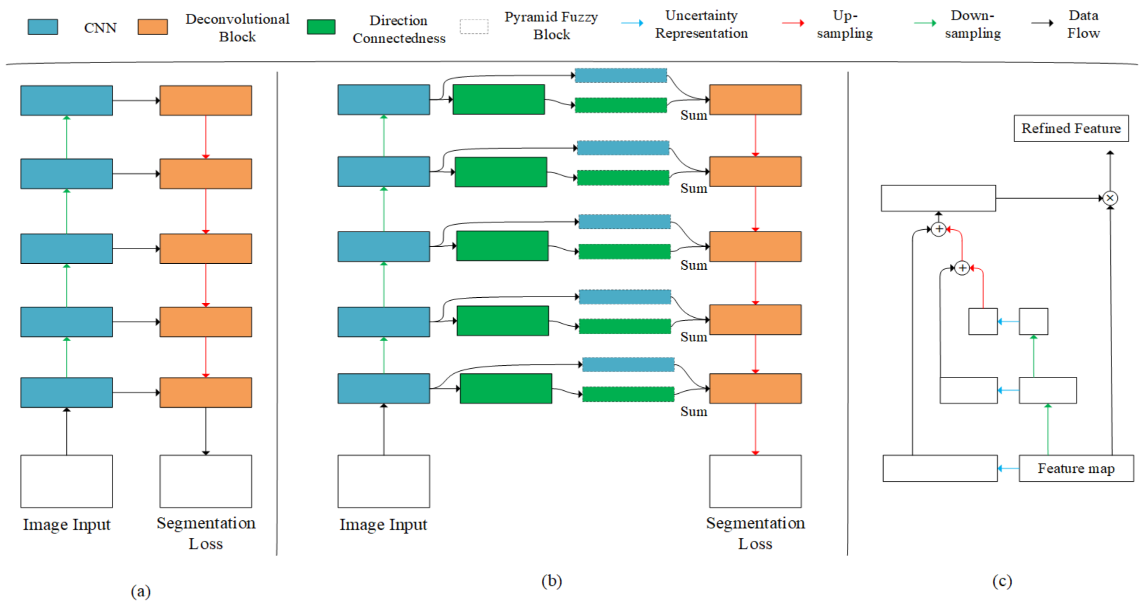


Fig. 4.1: Architectures of the proposed approach: (a) the original U-shape network with VGG-16; (b) the proposed method with pyramid fuzzy block and direction connectedness feature; and (c) the structure of the proposed pyramid fuzzy block.

4.2.1 Down-sampling Feature Map

X^0 represents the input feature map where $\mathrm{X}^0 \in \mathbb{R}^{M \times N \times D}$; M and N represent the width and length of input feature maps; D represents the channel number. X^0 is down-sampled twice to $\mathrm{X}^1 \in \mathbb{R}^{M/2 \times N/2 \times D}$ and $\mathrm{X}^2 \in \mathbb{R}^{M/4 \times N/4 \times D}$ (green arrows in Fig. 4.1 (c)). An object in X^1 and X^2 has 1/2 and 1/4 of the original size in X^0, respectively. Therefore, the network can learn uncertainty for different sizes of one object, which a non-pyramid version cannot handle.

4.2.2 Fuzzification

After down-sampling, the feature maps in three resolutions $\mathrm{X}^l = [\mathrm{X}^0, \mathrm{X}^1, \mathrm{X}^2]$ are transformed into fuzzy domain by the trainable Sigmoid membership function in Eq. (4.1):

$$\mu_{ik}^l = \frac{1}{1 + \exp(\alpha_{ik}^l x_i^l + \beta_{ik}^l)} \tag{4.1}$$

where $x_i^l \in \mathbb{R}^D$ is the feature vector of the ith pixel in X^l. $\alpha_{ik}^l \in \mathbb{R}^D$ and $\beta_{ik}^l \in \mathbb{R}$ are two trainable parameter vectors for the trainable Sigmoid function in the kth category. μ_{ik}^{l} represents the memberships in the kth category. Two 1×1 convolutional operators with Sigmoid activation function perform the trainable Sigmoid membership function:

$$\mu_i^l = Conv1 \times 1(Conv1 \times 1(x_i^l)) \tag{4.2}$$

where $\mu_i^l \in \mathbb{R}^C$ represents the membership vector and $\mu_i^l = [\mu_{i1}^l, \mu_{i2}^l, ..., \mu_{iC}^l]$; C represents the number of categories; the inner 1×1 convolutional operator contains 64 kernels; the outer 1×1 convolutional operator has C kernels. In this chapter, the multi-layer 1×1 convolution can increase the nonlinearity of the membership function. Meanwhile, it is easier to realize in convolutional networks, and the time complexity can be reduced.

4.2.3 Uncertainty Representation

After getting the memberships of different categories, the uncertainty degree of each pixel should be calculated by the memberships. There is an observation: if a pixel contains the same memberships of different categories (take binary segmentation as an example: the membership of background is 0.5 and the membership of foreground is 0.5), this pixel is hard to assign a category. This observation can be represented by fuzzy entropy. In this paper, the uncertainty degree is measured by fuzzy entropy [60] of memberships. The uncertainty degree in three resolutions (X^0, X^1, and X^2) is calculated by the following equation [60]:

$$u_i^l = -\frac{1}{\log C} \times \sum_{k=1}^{C} \mu_{ik}^l \log \mu_{ik}^l \tag{4.3}$$

where μ_{ik}^l represents the membership of category k for pixel i; and u_i^l represents the uncertainty degree for pixel i.

4.2.4 Uncertainty Reduction

In the previous subsection, the uncertainty degrees of each pixel in three resolutions are calculated by the fuzzy entropies of the memberships, and three uncertainty maps

$\mathrm{u}^l = (\mathrm{u}^0 \in \mathbb{R}^{M \times N}, \mathrm{u}^1 \in \mathbb{R}^{M/2 \times N/2}, \mathrm{u}^2 \in \mathbb{R}^{M/4 \times N/4})$ are composed of uncertainty degrees (u_i^l) of pixels. u^l has the same size as the feature map X^l. u^l is in range [0, 1]. 0 means low uncertainty degree, and 1 means high uncertainty degree. The uncertainty maps of three resolutions are combined by pixel-wise summation and up-sampling (Fig. 4.1 (c)). The final uncertainty map is computed by Eq. (4.4):

$$\mathrm{u} = Up\_sampling((Up\_sampling(\mathrm{u}^2) \oplus \mathrm{u}^1) \oplus \mathrm{u}^0) \tag{4.4}$$

where u represents the final uncertainty map for the input feature map X^0 and it is normalized to [0, 1] (0 means low uncertainty degree and 1 means high uncertainty degree). $\oplus$ represents the pixel-wise summation of matrices. The final uncertainty map is utilized as the weight of the original feature map and pixel-wise multiplication is utilized to combine the uncertainty map u and the input feature map X^0:

$$\mathrm{X}' = \mathrm{X}^0 \otimes (1 - \mathrm{u}) \tag{4.5}$$

where X' represents the feature map after uncertainty reduction; and $\otimes$ represents the pixel-wise multiplication of matrices. This operation expresses that if the pixels have high uncertainty degrees (u is close to 1), they have low weights; and if the pixels have low uncertainty degrees (u is close to 1), they have high weighs.

The uncertainty maps calculated by the pyramid fuzzy block for input images can be found in Fig. 4.2 (d). The pyramid fuzzy block can find the high uncertain areas in background and tumor in (Figs. 4.2 d1 and d2), which indicates that it is hard to classify background area and tumor. However, the non-pyramid version detects low uncertainty in those areas (Figs. 4.2 c1 and c2). The background area and tumor contain similar low intensity (shown in Fig. 4.2 (a)); therefore, the pyramid fuzzy block obtains better uncertainty maps.

4.3 Direction Connectedness Feature

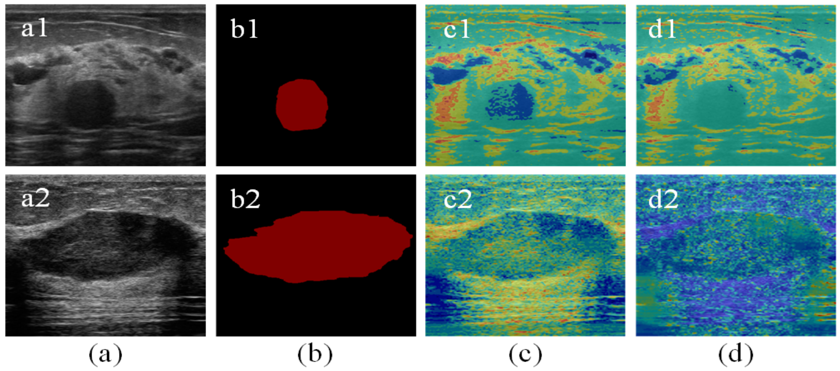


Fig. 4.2: Uncertainty maps: (a) original BUS images; (b) groundtruths; (c) uncertainty maps calculated by non-pyramid fuzzy block; (d) uncertainty maps calculated by pyramid fuzzy block. Blue represents low uncertainty and red represents high uncertainty.

As introduced in Section 1.3, the BUS images contain a layer structure, shown in Fig. 1.2 There are skin layer, fat layer, mammary layer, muscle layer, and background from the top to bottom of the BUS image. Connectedness is one of the most important global topological properties and has been applied to many image segmentation approaches [67–69]. The connectedness between pixels not on the boundary and pixels on the boundary is important. For example, if two pixels are in one horizontal line (such as the red line in Figs. 4.3 (a) and (b)), they contain a similar high connectedness strength to pixels on the left and right boundary in the same horizontal line. They also contain similar connectedness strength to the pixels on the up and bottom boundary in the vertical direction because the layer structure is fixed in the vertical direction of the BUS image. Therefore, in this research, connectedness between pixels and boundary pixels in the four principal directions (up, down, right, and left) is utilized as the context feature to describe breast anatomy. This subsection is organized as follows: (1) the connectedness between two adjacent pixels is defined firstly, and (2) the connectedness between pixels and pixels on the boundary along four principal directions is defined based on the connectedness between two adjacent pixels.

4.3.1 Connectedness between Two Adjacent Pixels

To express the connectedness strength between two adjacent pixels, it is computed by

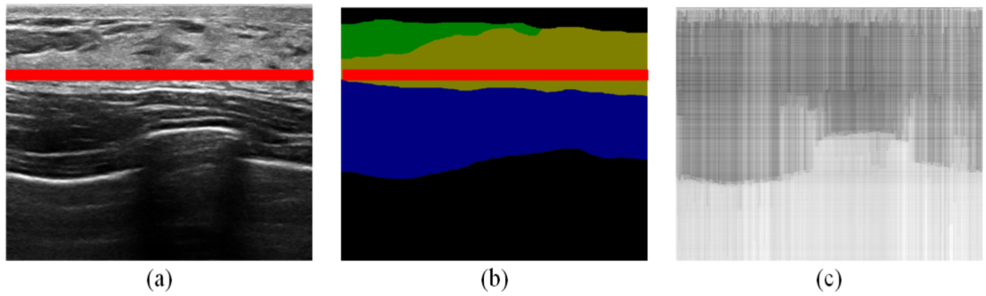


Fig. 4.3: BUS image layer structure: (a) original BUS image; (b) groundtruths; (c) direction connectedness feature.

the feature values of these two neighboring pixels. Let f_i and f_j represent the feature values of two adjacent pixels i and j, and the connectedness strength γ is defined as [67]:

$$\gamma = \exp(-\|f_i - f_j\|^2/2\delta^2) \tag{4.6}$$

where δ is the parameter to control the strength of the connectedness; $\|f_i - f_j\|^2$ represents the Euclidean distance between the feature values.

4.3.2 Connectedness in Four Principal Directions

As shown in Fig. 4.4, the boundary pixels are the pixels in the brown color. The pixels on the left boundary of the BUS image (Fig. 4.4 (a)) only calculate the connectedness strength to the pixels along the horizontal right direction. The pixels on the bottom boundary (Fig. 4.4 (b)) only calculate the connectedness strength along the vertically up direction. It is similar for the left and down directions (Figs. 4.4 (c) and (d)). There is only one path to get the pixels inside the image from pixels on the boundary. Take the left direction as an example: Let $x_{i,j}$ represent a pixel in the image, and $x_{i,1}$ represent the pixel on the left boundary in the ith row and let $\Gamma_{i,j}^{left}$ represent the connectedness strength between $x_{i,j}$ and $x_{i,1}$ [67]:

$$\Gamma_{i,j}^{left} = \min(\Gamma_{i,j-1}^{left}, \exp(-\|f_{i,j-1} - f_{i,j}\|^2/2\delta_{left}^2)) \tag{4.7}$$

where $f_{i,j-1}$ and $f_{i,j}$ are feature values for pixels $x_{i,j-1}$ and $x_{i,j}$; $\Gamma_{i,1}^{left}$ is initialized as 1. The connectedness strength in the other three principal directions can be expressed similarly [67]:

$$\begin{aligned}\Gamma_{i,j}^{right} &= \min(\Gamma_{i,j+1}^{right}, \exp(-\|f_{i,j+1} - f_{i,j}\|^2/2\delta_{right}^2)) \\ \Gamma_{i,j}^{up} &= \min(\Gamma_{i+1,j}^{up}, \exp(-\|f_{i+1,j} - f_{i,j}\|^2/2\delta_{up}^2)) \\ \Gamma_{i,j}^{down} &= \min(\Gamma_{i-1,j}^{down}, \exp(-\|f_{i-1,j} - f_{i,j}\|^2/2\delta_{down}^2))\end{aligned} \tag{4.8}$$

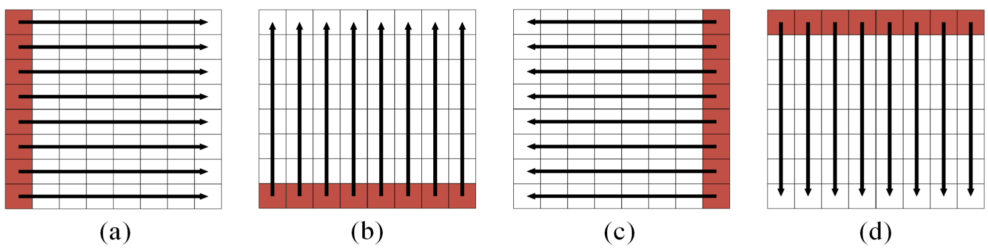


Fig. 4.4: The connectedness in the four principal directions: (a) right; (b) up; (c) left; and (d) down.

After calculating the connectedness strengths along four directions, the connectedness strengths are added together to obtain the breast structure context information:

$$\Gamma_{i,j} = \Gamma_{i,j}^{left} + \Gamma_{i,j}^{right} + \Gamma_{i,j}^{up} + \Gamma_{i,j}^{down} \tag{4.9}$$

If the gray-level intensity is the feature, the breast structure context information $\Gamma_{i,j}$ is shown in Fig. 4.3 (c). This information can display the layer structure, therefore it is applied to U-shape network to represent breast structure directly.

4.3.3 Direction-connectedness Feature Extraction

To use the breast layer structure in a deep neural network, the direction-connectedness (DC) feature is defined, and a DC feature extraction block is proposed (Fig. 4.5). In the DC feature extraction block, the input feature map is input to two paths. In the first path, the input feature map is processed by a 1×1 convolutional operator. Then, a spatial RNN is utilized to extract the connectedness between boundary pixels and each pixel in the four principal directions (shown in Eqs. (4.7) and (4.8)). The parameters δ_{left}, δ_{right}, δ_{up} and δ_{down} are recurrent translation parameters in the spatial RNN. The connectedness strengths in four directions are added together by pixel-wise summation (Eq. (4.9)). The second path for the input feature map is three convolutional operators. The output feature map combines with the connectedness strength in the first path by a pixel-wise multiplication. The combination is the DC feature.

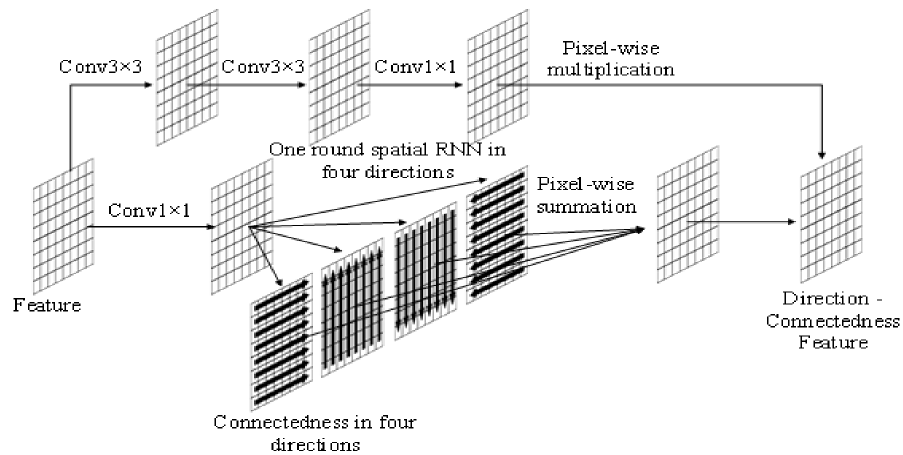


Fig. 4.5: Direction connectedness feature. Only one round spatial RNN is utilized because it only focuses on the connectedness strength between pixels and the pixels on the boundaries along up, down, left, and right directions.

4.3.4 Loss Function

The entire network is shown in Fig. 4.1 (b). It shows that the feature map in each convolutional block has three paths: the first one is the next convolutional block to obtain new convolutional features; the second one is a direction-connectedness module to get direction-connectedness feature; the third one is a pyramid fuzzy block to reduce uncertainty. Both convolutional features and DC features are reduced uncertainty by the proposed pyramid fuzzy blocks. The convolutional features and DC features are combined by pixel-wise summation. At the end of the network, there is a segmentation task. The loss function for the segmentation task is defined on the summation of category cross-entropy loss for image segmentation and fuzzy entropies from pyramid fuzzy blocks:

$$L_{overall} = L + L_c + L_d \tag{4.10}$$

where L represents the classic cross entropy loss:

$$L = -\sum l(x)\log(p(x)) \tag{4.11}$$

where x represents the input pixel; $l(x)$ is the label of x in one-hot encoding. $l(x)$ is a vector with C elements, where C represents the number of categories. If x is in the kth category and the corresponding kth element in $l\left(x\right)$ is 1 and other elements are 0; $p\left(x\right)$ is the predicted vector and each element represents the probability for the corresponding category.

There are five convolutional blocks and five direction connectedness feature extraction blocks. All the ten features are input to the pyramid fuzzy block to compute the uncertainty degree by Eq. (4.4). L_c and L_d in Eq. (4.10) are the summation of fuzzy entropy of convolutional features and direction connectedness features:

$$\begin{aligned} L_c &= \sum_q \sum_i \mathrm{u}_i^q \\ L_d &= \sum_r \sum_i \mathrm{u}_i^r \end{aligned} \tag{4.12}$$

where u_i^q and u_i^r are the ith pixel for uncertainty degree map u computed by Eq. (4.4) using the convolutional features and direction connectedness features; q represents the block index of convolutional block and r represents the block index of direction connectedness block.

4.4 Experimental Results

4.4.1 Datasets

Two datasets are utilized to evaluate the proposed method: (1) a BUS image benchmark [1] which contains pixel-wise groundtruths only for tumors; (2) a multi-object BUS dataset image which contains five categories: fat layer, mammary layer, muscle layer, tumor, and background. The First dataset contains 562 images. The multi-object BUS dataset contains 325 images that are the one utilized in Chapter 2 and Chapter 3. To show the effectiveness of the proposed network, eight state-of-the-art semantic segmentation network structures, such as U-Net [48] with VGG-16, U-Net with ResNet-50/ResNet-101 [15], Deeplabv3+ [64], FCN-8s [14], PSPNet [18], FCN with information extension [53], and U-Net with the direction-aware spatial context (DSC) features [36] are compared with the proposed method.

4.4.2 Training Strategy and Setup

The training images are augmented by horizontal flip, horizontal shift, vertical shift, rotation, zooming, and shear mapping. They are all gray-level images and mapped to the intensity range [-1, 1] by (x/127.5-1) [65]. No other pre-processing method is used.

In this research, a computer with Ubuntu 18.04 system, Intel(R) Xeon(R) CPU E5-2620 2.10GHz, and 2 NVIDIA GeForce 1080 graphics cards is used. The network weights are initialized randomly. The batch size is 8. The optimizing method is the Adam method [59] with an initial learning rate 10^{-4}, and learning decay rate is 5×10^{-4}. The parameter β_1 for Adam method is 0.9, and the parameter β_2 for Adam's method is 0.99. All the networks (compared and proposed) are not pretrained on other datasets. The implementation is based on the Keras platform with the TensorFlow backend.

4.4.3 Metrics

In the benchmark dataset, five area metrics introduced in Subsection 3.4.3: True Positive Ratio (TPR), False Positive Ratio (FPR), Jaccard Index (JI), Dice's Coefficient (DS), and Area Error Ratio (AER) are utilized to evaluate the performance of all methods. For the multi-object dataset, the Intersection over Union (IoU) and mean IoU (mIoU) introduced in Subsection 3.4.3 are utilized.

4.4.4 Tumor Segmentation Results on Benchmark Dataset

All other compared state-of-the-art deep learning methods such as ResNet-50, ResNet-101, FCN-8s, and U-Net with the direction-aware spatial context (DSC) features are trained in the same experiment environment (such as initial methods, batch size, training epoch, *etc.*). 10-fold validation is utilized: (1) all images are divided into 10 groups randomly; (2) pick 9 groups as the training set and the rest 1 group as the testing set; and (3) the final results are calculated by the average of 10 experiments.

Fig. 4.6 shows the segmentation results of the benchmark [1]. Fig. 4.6 (a) is the original image; Fig. 4.6 (b) is the groundtruths and the red areas in Figs. 4.6 (b)-(k) are the tumors. Fig. 4.6 a1 contains a small tumor. This tumor is like black strip structures in the muscle layer and fat layer. The previous network structures cannot handle this tumor (shown in Figs. 4.6 c1-i1). In Fig. 4.6 a2, the back area in the muscle layer contains a similar feature to that of the tumor. The original U-Net (Fig. 4.6 c2), U-Net with ResNet-50 (Fig. 4.6 d2), Deeplab (Fig. 4.6 f2), U-Net with wavelet transform (Fig. 4.6 h2), FCN-8s (Fig. 4.6 i2) misclassify the black area in the muscle into the tumor. The proposed method can solve this by involving breast anatomy and fuzzy uncertainty reduction. The U-Net with DSC feature can also segment this tumor well because it also involves spatial context features. However, involving the DSC feature in U-Net cannot handle some cases such as Figs. 4.6 j1 and j5. The proposed method can handle this tumor better than DSC because DSC is designed for shadow detection, and the proposed method can reflect breast anatomy and reduce uncertainty. Adding wavelet information in the input image can also solve the mis-segmentation in Fig. 4.6 a2 and Fig. 4.6 a5. However, it still cannot handle Fig. 4.6 a2 and Fig. 4.6 a3 well. It proves that adding other features can solve some mis-segmentation; however, there still exists uncertainty in the new feature. The proposed method can overcome this by using a pyramid fuzzy block to reduce the uncertainty in new features. The proposed method obtains the best results in all samples in Figs. 4.6 a1-a5.

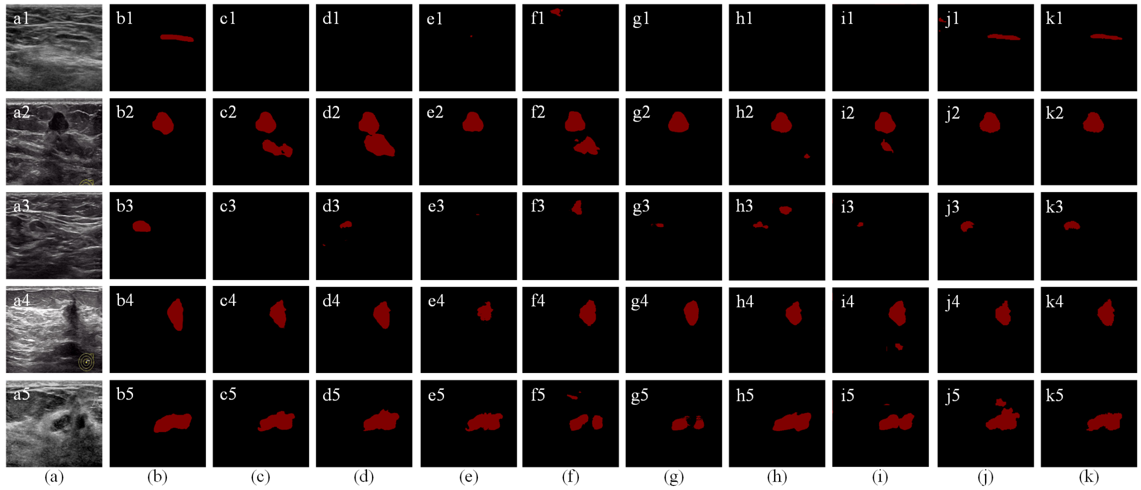


Fig. 4.6: Segmentation results using benchmark in [1]: (a) original images; (b) groundtruths; (c) results of U-Net; (d) results of ResNet-50; (e) results of ResNet-101; (f) results of Deeplab; (g) results of PSPNet; (h) results of U-Net with wavelet transform; (i) results of FCN-8s; (j) results of U-Net with DSC feature; and (k) results of proposed method.

The evaluation metrics are listed in Table 4.1. Five non-deep learning methods [4–8] mentioned in the [1] are also adopted for comparison. Bold numbers are the corresponding best results. Table 4.1 shows that the proposed method achieves the best evolution metrics. Although FCN-8s achieves slightly higher TPR, it obtains much worse FPR and other metrics. Deep learning approaches achieve much better results than non-deep learning methods.

4.4.5 Multi-object Segmentation for BUS Image

The proposed method is also applied to multi-object BUS image segmentation in a dataset with 325 images. 10-fold validation is still utilized in the experiment because of the limited number of training samples. The state-of-the-art deep neural network structures mentioned in the previous chapter are utilized to compared with the proposed method as well. In this research, five breast layers are segmented by different network structures. Fig. 4.7 (a) shows four BUS images, and the corresponding groundtruths for these BUS images are in Fig. 4.7 (b). The green areas are fat layers; the yellow areas are mammary layers; the blue regions are muscle layers; the red areas are tumors, and the background areas are

Table 4.1: Results of tumor segmentation using dataset [1].

	TPR	FPR	JI	DS	AER
	Semi-Automatic Methods				
[4]	0.82	0.13	0.73	0.84	0.31
[5]	0.84	**0.07**	0.79	0.88	0.23
	Fully-Automatic Methods				
[7]	0.81	0.16	0.72	0.83	0.36
[6]	0.81	1.06	0.60	0.70	1.25
[8]	0.67	0.18	0.61	0.71	0.51
U-Net	0.92	0.09	0.86	0.92	0.17
Deeplab	0.89	0.11	0.82	0.89	0.22
ResNet50	0.92	0.08	0.86	0.92	0.16
ResNet101	0.92	0.10	0.85	0.91	0.18
FCN-8s	**0.94**	0.10	0.86	0.92	0.16
PSPNet	0.93	0.09	0.86	0.92	0.16
[53]	0.92	0.09	0.86	0.92	0.16
DSC	0.91	0.10	0.84	0.91	0.18
Proposed	0.93	**0.07**	**0.87**	**0.93**	**0.15**

black.

In Fig. 4.7, it can be seen that the results of the original U-Net (Figs. 4.7 c1-c4), ResNet-50/RestNet-101 (Figs.4.7 d1-d4 and e1-e4), Deeplab (Figs. 4.7 f1-f4), PSPNet (Figs. 4.7 g1-g4), and FCN-8s (Figs. 4.7 i1-i4) are not very good. When adding wavelet information in the input layer, the U-Net can obtain better results in Figs. 4.7 h1 and h4. However, adding wavelet information in the input layer can make the segmentation result worse. For example, in Fig. 4.7 h2, the mammary layer is segmented wrongly into the muscle layer. Also, in Fig. 4.7 h3, part of the tumor is segmented into the muscle layer and background wrongly. Such problems do not exist in the original U-Net (Figs. 4.7 c2 and c3). When adding the DSC feature in U-Net, the segmentation results become better (Figs. 4.7 j1-j4). However, it does not solve the problem in Fig. 4.7 j1 (the green patch under tumor). It involves more mistakes in Fig. 4.7 j3 than the original U-Net with the gray-level image as input (Fig. 4.7 c3). These experiment results prove that new features can increase the feature dimension and solve some mis-segmentation. However, new features contain new uncertainty and noise and might make it hard to classify those pixels. The proposed method adds a new direction connectedness feature to the original U-Net. This feature obtains breast anatomy information. Moreover, both newly added features and original convolutional feature in U-Net are processed by fuzzy operations in pyramid fuzzy blocks. Uncertainty of pixels in both features is reduced weights in the combination of two features. The feature used to make the final decision is reduced. Hence, the segmentation results for the proposed method are the best (Figs. 4.7 k1-k4).

Table 4.2: Evaluation results of multi-object segmentation on BUS images. Evaluation metric using IoU (%).

	Fat	Mammary	Muscle	Background	Tumor	Mean
U-Net	70.34	66.72	66.17	65.91	74.66	68.76
ResNet-50	82.58	73.98	73.08	77.23	76.34	76.64
ResNet-101	81.50	73.41	72.07	74.47	75.29	75.35
FCN-8s	82.57	75.47	**75.53**	78.59	74.42	77.32
PSPNet	82.07	74.40	74.49	77.36	74.75	76.61
Deeplab	78.91	68.71	67.33	73.94	69.04	71.58
[53]	84.05	75.92	74.89	78.35	74.88	77.62
DSC	83.86	76.38	74.95	77.25	78.07	78.10
Proposed	**84.45**	**76.90**	75.48	**79.35**	**79.63**	**79.16**

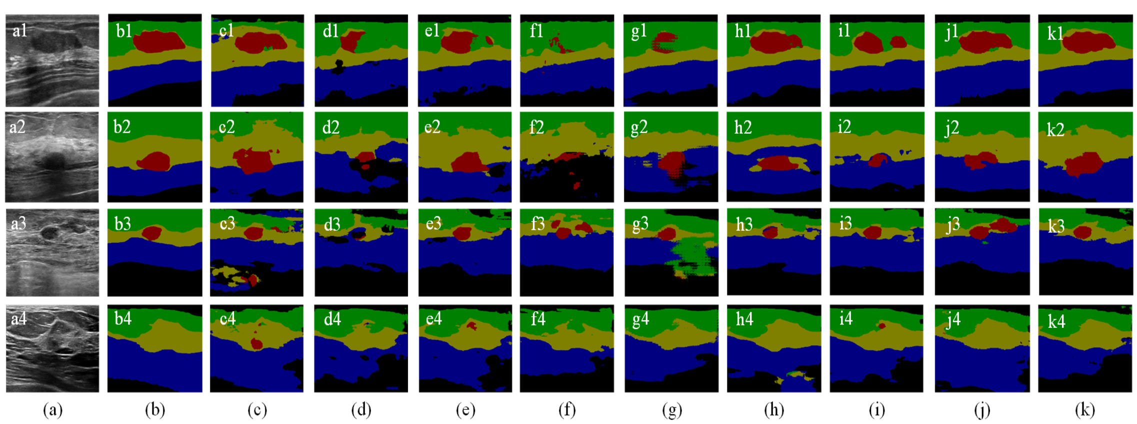


Fig. 4.7: Semantic segmentation: (a) original images; (b) groundtruths; (c) results of U-Net; (d) results of ResNet-50; (e) results of ResNet-101; (f) results of Deeplab; (g) results of PSPNet; (h) results of U-Net with wavelet transform; (i) results of FCN-8s; (j) results of U-Net with DSC feature; and (k) result of proposed method.

The evaluation metrics for multi-object BUS image segmentation are shown in Table 4.2. Bold numbers are the best results. The proposed method obtains 4.97% IoU improvement in the tumor category and 10.4% overall mean IoU improvement compared with the original U-Net with the gray-level image as input. The state-of-the-art deep learning architectures such as ResNet, PSPNet, and Deeplab do not obtain good results using this dataset because there are not enough training samples to train the complex network structures.

4.5 Conclusion

This chapter presents a novel network structure using fuzzy logic and spatial context information applied to BUS image semantic segmentation. The proposed method achieves the best overall performance in binary semantic segmentation and multi-object semantic segmentation compared with eight state-of-the-art deep learning architectures. It achieves improvement because of the following reasons: (1) The proposed pyramid fuzzy block can find the uncertain pixels and reduce their weights in different resolutions; therefore, the proposed fuzzy block can provide different-scale uncertainty information. (2) The connectedness strength between inside pixels and boundary pixels along the left, right, up, and down directions can represent breast anatomy better than previous context features. Finally, the proposed direction connectedness feature is combined with the original convolutional feature, and the novel feature can obtain better segmentation results.

CHAPTER 5

MEDICAL KNOWLEDGE CONSTRAINED CONDITIONAL RANDOM FIELDS

5.1 Introduction

The proposed DC feature in Chapter 4 can provide breast anatomy. Besides the DC feature, a novel context information term in conditional random fields is proposed in this study. Conditional random fields (CRFs) are widely used in nature image semantic segmentation combined with deep learning and non-deep learning approaches. Segmentation is modeled as a minimization energy function of CRFs (Eq. (5.1)) [30]. In the energy function of CRFs, there are two terms: unary term (θ_i in Eq. (43)) and pairwise term (θ_{ij} in Eq. (5.1)). The unary term is the segmentation probability map obtained from a unary classifier such as a deep neural network, support vector machine.

$$E(\mathrm{X}) = \sum_i \theta_i(x_i) + \sum_i \sum_j \theta_{ij}(x_i, x_j) \tag{5.1}$$

where y_i represents the label for pixel i. The first term in Eq. (5.1), $\theta_i\left(y_i\right) = -\log P\left(y_i\right)$ is a unary potential function. In this research, the unary potential function is provided by the U-shape network with fuzzy blocks introduced in Chapter 2. The segmentation probability map $P\left(y_i\right)$ contains the probability belong to each category for each pixel. The segmentation result of the unary classifier is not reflected by other pixels, which means there is no context information. The pairwise term in the energy function of CRFs discusses the relation between different pixels. The pairwise term is defined as:

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \sum_m \omega_m k_m(f_i, f_j) \tag{5.2}$$

where $\mu\left(y_i, y_j\right) = 1$, if $y_i \neq y_j$, and $\mu\left(y_i, y_j\right) = 0$ if $y_i = y_j$, which is known as the Potts model. This coefficient shows if two pixels are in the same category, the energy is minimum. k_m is a Gaussian kernel, where f_i and f_j are the features of pixels i and j. ω_m is the combined weight of the mth Gaussian kernel. There are two Gaussian kernels in [30]. In the first Gaussian kernel, the feature is defined on the physical position and the color feature of the pixels. In this research, the color feature (RGB) represents the gray-level information in the R channel, the approximation coefficient in the G channel, and the high-frequency information of wavelet transform in the B channel. If the input image is not preprocessed, only intensity combined with the position is used. The second Gaussian kernel is only defined on the positions of pixels. The detail of the pairwise potential function is shown in Eq. (5.3) [30]:

$$\sum_m \omega_m k_m(f_i, f_j) = \omega_1 \mathrm{exp}(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}) + \omega_2 \mathrm{exp}(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}), m = 1, 2 \quad (5.3)$$

where p_i represents the position of the ith pixel, and I_i represents the color feature of the ith pixel. Eq. (5.3) indicates that the pairwise term in the energy function of CRFs can compute the correlation between pixels. The correlation is dependent on position and feature.

In this work, breast anatomy is represented by the pairwise term of CRFs. In Fig. 1.2 (b), the BUS image contains 6 different areas: skin, fat, mammary, muscle, background, and tumor. The skin layer is treated as background because the number of samples containing the skin layer is small. However, due to the position, the skin layer is different from the retro-muscle background area. To make the context of different layers more reasonable, the skin layer is treated as a pre-fat background area. Finally, the contexts of the pre-fat background area, fat layer, mammary layer, muscle layer, retro-muscle layer, and breast tumor are used. V_i is defined to represent the category of pixel i assigned by a deep neural network, $V_i \in \{L^1, L^2, L^3, L^4, L^5, L^6\}$. $L^1, L^2, L^3, L^4, L^5, L^6$ represent the pre-fat background area, fat layer, mammary layer, muscle layer, retro-muscle layer, and breast tumor, respectively (Fig. 1.2 (b)). The label vector $V_i \in \{L^1, L^2, L^3, L^4, L^5, L^6\}$ is applied to the pairwise term of CRFs to reflect breast anatomy.

5.2 Breast-anatomy Constrained Fully Connected CRFs

As discussed in Section 1.3, breast cancer usually begins in the mammary layer. However, some pixels in the fat layer and muscle layer might be classified into wrong categories; besides, pixels in the muscle layer have similar intensity levels to that of the pixels in the mammary layer, which may also cause misclassification. Medical knowledge can overcome misclassification. After locating the positions of the fat layer, mammary layer, and muscle layer, the context information can be used to prevent the wrongly classified patches in each layer. The original fully connected CRFs contain the energy function in Eqs. (5.2) and (5.3). The Gaussian kernel in Eq. (5.3) consists of pixel positions and color features. To involve breast anatomy, the category of pixel i assigned by the neural network, which is defined as V_i, is treated as another feature and a new Gaussian kernel based on V_i, and position of the pixel, p_i is utilized. The new energy function contains three terms:

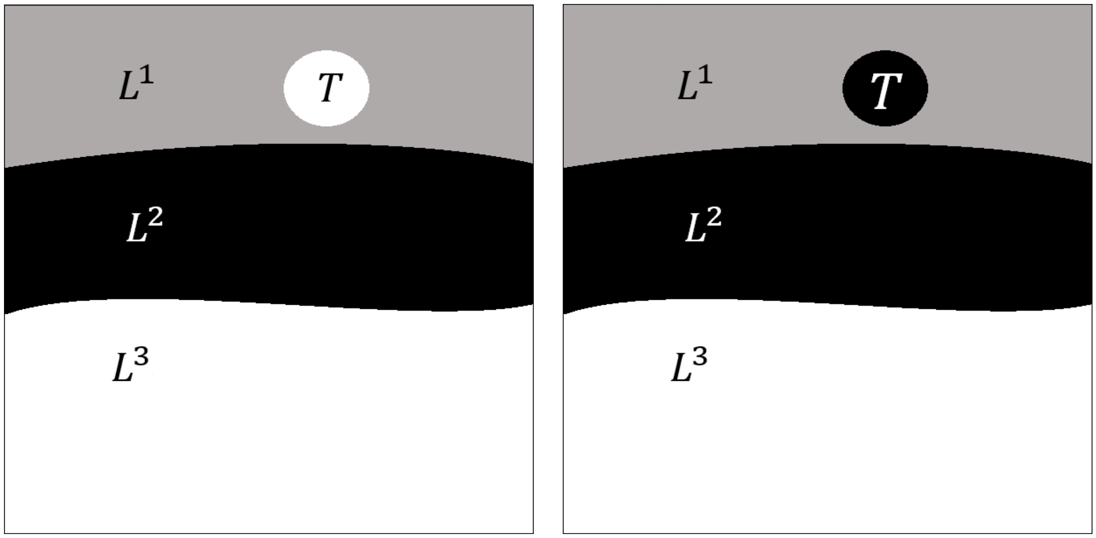
$$\begin{aligned}\sum_m \omega_m k_m(f_i, f_j) = \omega_1 \text{exp}(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}) + \omega_2 \text{exp}(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}) \\ + \omega_3 \text{exp}(-\frac{\|p_i - p_j\|^2}{2\sigma_\tau^2} - \frac{\|V_i - V_j\|^2}{2\sigma_\lambda^2}), m = 1, 2, 3\end{aligned} \tag{5.4}$$

where $\text{exp}(-\frac{\|p_i - p_j\|^2}{2\sigma_\tau^2} - \frac{\|V_i - V_j\|^2}{2\sigma_\lambda^2})$ is the Gaussian kernel of the layer context information, and V_i, and V_i, represent categories of pixel i and j assigned by the deep neural network. σ_τ and σ_λ are the parameters of CRFs.

Here, two distances on context are defined: 1) context distance between two pixels, $\|V_i - V_j\|^2$ where i and j represent pixel i and j in the image, and 2) context distance between two categories, $\|L^s - L^t\|^2$ where s and t represent the category index. In this research, $1 \leq s,\ t \leq 6,\ i \in \mathbb{Z}$. $\|L^s - L^t\|^2$ is the Euclidean distance of the two category vectors. $\|V_i - V_j\|^2$ is the context distance between the category of pixel i and category of pixel j. For example, if the pixel i is in category L^1, and pixel j is in the category L^2, $\|V_i - V_j\|^2$ equals to $\|L^1 - L^2\|^2$. The value of the label vector $V_i \in \{L^1, L^2, L^3, L^4, L^5, L^6\}$ is defined in the following chapter.

5.3 Label Vector Setting

To demonstrate how to utilize the context distance between pixels and context distance between and categories and their effectiveness on BUS image segmentation, a simulated image is shown in Fig. 5.1. In Fig. 5.1 (a), L^1, L^2, and L^3 represent three categories. T represents a wrongly classified patch, which should be in L^1, but assigned to L^3 by the unary classifier. If the context distances among three categories are set as the context distance between L^1 and L^2 equals to the context distance between L^2 and L^3; the context distance between L^1 and L^2 is greater than the context distance between L^1 and L^3; then pixels in T has the chance to be corrected into L^1. Here, four pixels are chosen to demonstrate how it works: 1) pixel i in area T; 2) pixel h in L^1 area; 3) pixel v in L^2 area; 4) pixel j in L^3 area. Pixel i is in area T and area T is now in category L^3. Pixel v is in area L^2, so the context distance between pixel i and pixel v equals the context distance between categories L^3 and L^2 as introduced in the previous paragraph, *i.e.*, $\|V_i - V_v\|^2 = \|L^3 - L^2\|^2$. For other pixels, the situations are the same, *i.e.*, $\|V_i - V_h\|^2 = \|L^3 - L^1\|^2$; $\|V_i - V_j\|^2 = \|L^3 - L^3\|^2 = 0$. Therefore, $\|V_i - V_v\|^2 = \|L^3 - L^2\|^2 > \|V_i - V_h\|^2 = \|L^3 - L^1\|^2 > \|V_i - V_j\|^2 = \|L^3 - L^3\|^2$ because of the assumption made before. Meanwhile, $\|p_i - p_j\|^2 > \|p_i - p_v\|^2 > \|p_i - p_h\|^2$, where p_i, p_h, p_v and p_j are the position of these pixels in Eq. (5.4). Hence, the pixels in area T have smaller context distances with pixels in L^1 than that in L^2, and the pixels in area T have smaller space distances with the pixels in L^1 than that in L^2. Even if the pixels in area T have zero context distances with the pixels in L^3 (*i.e.*, they are in the same category), they have smaller space distances with the pixels in L^1 than that in L^3. Therefore, the pixels in area T still can be classified into category L^1. In Fig. 5.1 (b), the pixels in area T are wrongly classified into category L^2 and the pixels in area T have the same context distances as the pixels in L^1 and L^3, but they have smaller space distances with the pixels in L^1 than that in L^3. If the pixels in area T have zero context distances with the pixels in L^2, their space distances with the pixels in L^1 are smaller than that with the pixels in L^2. Therefore, the pixels in area T still have the chance to be classified into L^1 by properly setting weight ω_3 and parameters σ_τ and σ_λ in Eq. (5.4).



(a) (b)

Fig. 5.1: Simulated image to show the context distance among categories.

The BUS images (Figs. 1.2 (a) and (b)) are like the simulated examples. The context distances between the categories can be classified into three classes (Fig. 1.2 (b)) in the BUS images: 1) two layers are neighbors to each other (D_1), *e.g.*, fat layer (L^2) and mammary layer (L^3); 2) two layers are separated by another layer (D_2), such as the fat layer (L^2) and muscle layer (L^4); 3) two layers are separated by two layers (D_3), such as the fat layer (L^2) and retro-muscle background area (L^5):

$$
\begin{aligned}
D_1 &= \|L^i - L^{i+1}\|^2, 1 \leq i \leq 4, i \in \mathbb{Z} \\
D_2 &= \|L^i - L^{i+2}\|^2, 1 \leq i \leq 3, i \in \mathbb{Z} \\
D_3 &= \|L^i - L^{i+3}\|^2, 1 \leq i \leq 2, i \in \mathbb{Z}
\end{aligned}
\tag{5.5}
$$

The relations among them are:

$$D_1 > D_2 > D_3 \tag{5.6}$$

The reason for setting such relations among them (Eq. (5.6)) is to follow the situation in the simulated example. The relations encourage a clear boundary and void wrongly classified patches like T in Fig. 5.1. L^1 and L^5 have high space distance while their context distance is not considered because the high space distance plays a more important role in the Gaussian kernel. After defining the context distances among the five layers, the context distances between the tumor and five layers could be defined. The tumor (L^6) usually locates in the mammary layer (L^3). Sometimes, the mammary layer above the tumor or bellow tumor is very thin; and the tumor seems to be in the fat layer (L^2) or muscle layer (L^4). The context distance between tumor (L^6) and mammary layer (L^3) should be the largest because it encourages a clear boundary between tumor and mammary layer. The context distances between the tumor (L^6) and fat layer (L^2) or muscle layer (L^4) should be the second largest, which gives the chance to correct some wrongly classified patches in these layers. The context distance between tumor (L^6) and the background (L^1 and L^5) should be the smallest. Because some background areas are likely classified as the tumor, and such a situation should be voided (refer to Fig. 5.1). The relationships are shown in Eq. (5.7):

$$\|L^6 - L^3\|^2 > \|L^6 - L^2\|^2 \approx \|L^6 - L^4\|^2 > \|L^6 - L^1\|^2 \approx \|L^6 - L^5\|^2 \tag{5.7}$$

The category vectors L^1, L^2, L^3, L^4, L^5, and L^6 should satisfy the constraints in Eqs. (5.5)-(5.7) to realize the medical anatomy constraints. By solving Eqs. (5.5)-(5.7), $L^1 = \{61.2, 20, 15\}$, $L^2 = \{25, 37.1, 0\}$, $L^3 = \{40, 0, 0\}$, $L^4 = \{55, 37.1, 0\}$, $L^5 = \{18.8, 20.7, 15\}$, and $L^6 = \{40, 30, 26.5\}$. For $D_1 \approx 40$, $D_2 \approx 30$, $D_3 \approx 23$, $\|L^6 - L^3\|^2 = 40$, $\|L^6 - L^2\|^2 \approx \|L^6 - L^4\|^2 \approx 30$, $\|L^6 - L^1\|^2 \approx \|L^6 - L^5\|^2 \approx 26$. The relations among context labels are shown in Fig. 5.2. If a pixel is classified into category $L^s, s = 1, 2, 3, 4, 5, 6$, a category map will be created and the corresponding pixel in the category map will be assigned by the value of L^s. The category map is used as another feature in Eq. (5.4)

By setting the label vectors with these values, the proposed CRFs energy function encourages two pixels whose space distance and context distance are both small to be in the same category. It will remove some wrongly classified patches. The mean-field approximate algorithm [30] is used to solve the fully connected CRFs energy minimization problem.

5.4 Experiment Results

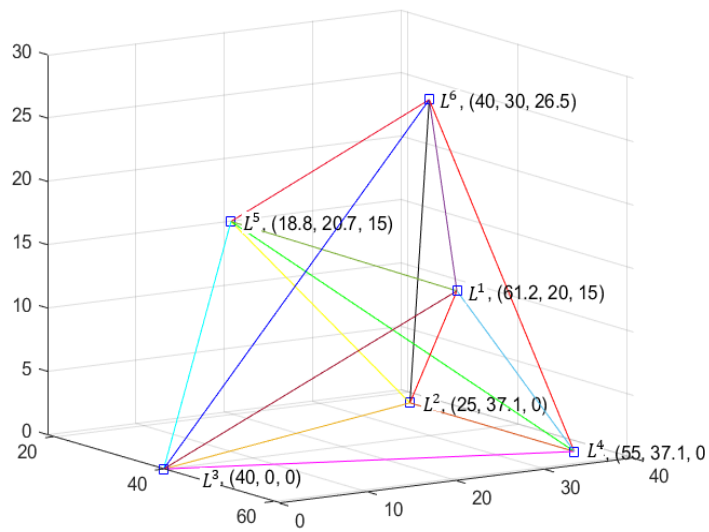


Fig. 5.2: The coordinates of the labels.

5.4.1 Datasets, Metrics, and Setup

The same computer, datasets, and metrics with the Chapter 2 are utilized. The U-Net with VGG16 and the fuzzy block in the Chapter 2 is utilized to provide an initial segmentation map for CRFs. The CRFs parameters $\omega_1 = 1$, $\omega_2 = 2$, $\omega_3 = 1$, $\sigma_\alpha^2 = 60$, $\sigma_\beta^2 = 10$, $\sigma_\gamma^2 = 3$, $\sigma_\tau^2 = 3$, and $\sigma_\lambda^2 = 2$. They are determined by experiments, and the medical context label and the context distance relation are shown in Fig. 5.2. The implementation is based on MATLAB (R2018b, MathWorks Inc., MA).

5.4.2 Comparison with Original CRFs

The fuzzy U-Net in Chapter 2 is utilized to provide the unary term of CRFs. Breast anatomy constrained fully connected CRFs uses the medical context information. The original fully connected CRFs and the approximation algorithm in [30] are employed to optimize the energy function of CRFs. It has three effects: 1) correct the wrongly classified pixels; 2) make the boundaries between layers more accurate; 3) increase the overall segmentation performance. The segmentation results are shown in Fig. 5.3. In Table 5.1, the output of fuzzy U-Net, the refined results of original CRFs, and proposed CRFs are utilized for comparison.

Table 5.1: Evaluation results of breast anatomy constrained CRFs. Evaluation metrics using IoU.

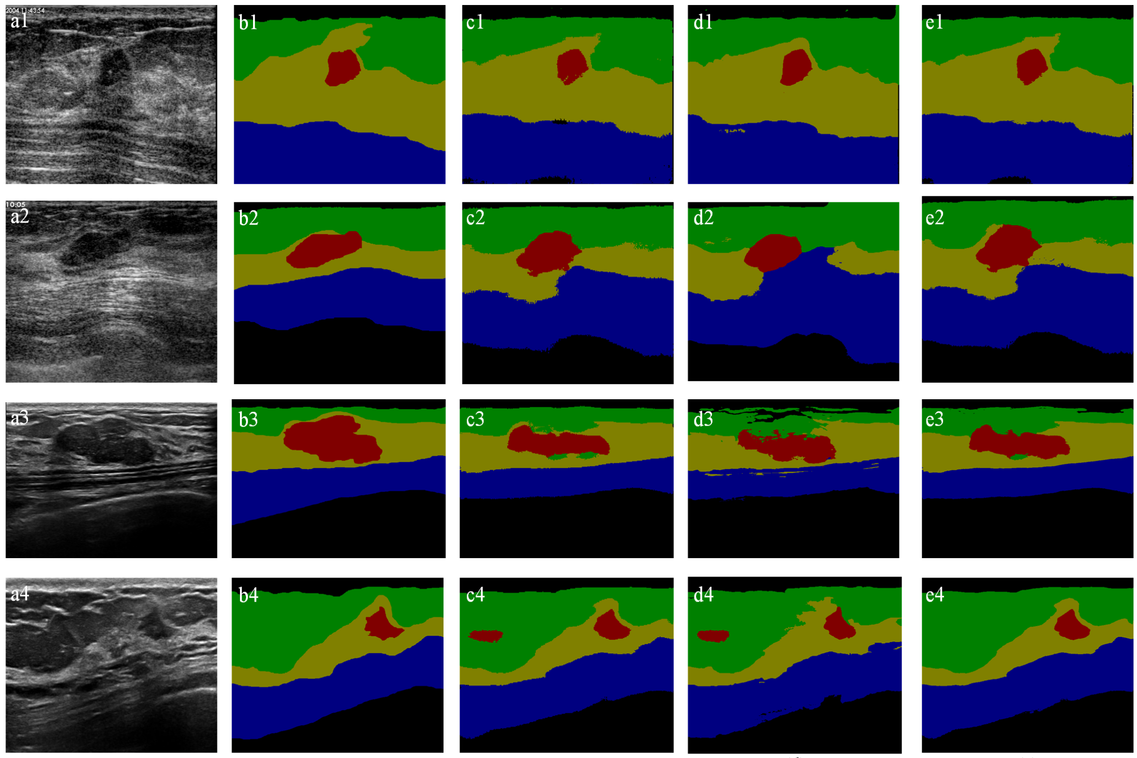
	Fat	Mammary	Muscle	Background	Tumor	Mean
Fuzzy U-Net + CRFs	81.52	**78.63**	75.24	76.48	79.32	78.24
Fuzzy U-Net	84.07	76.01	74.62	78.39	78.53	78.32
Fuzzy U-Net + Proposed CRFs	**85.06**	77.24	**78.66**	**80.09**	**81.29**	**80.47**

Figs. 5.3 c1-c4 are the segmentation results for the proposed fuzzy U-Net. Figs. 5.3 d1-d4 are the fine-tuning results of the original fully connected CRFs and Figs. 5.3 e1-e4 are the fine-tuning results of the proposed CRFs. Comparing with Figs. 5.3 c1, d1, and e1, the original CRFs fine-tune the tumor boundary and make it close to the groundtruth (Fig. 5.3 b1). However, the original CRFs make the boundary between the mammary layer and the fat layer worse. The same situation happens in Figs. 5.3 d2-d4. In Fig. 5.3 d2, the boundary of the tumor is smoother than the result of fuzzy U-Net; however, the muscle layer grows into the mammary layer using the original CRFs, and in Fig. 5.3 d3, the background area and fat layer interlace each other. In Figs. 5.3 c4 and d4, there are pixels in the fat layer classified into the tumor. The original CRFs fail to correct the mis-classification patch. The proposed CRFs utilize the medical context constraints to overcome such a problem (Figs. 5.3 e1-e4).

Table 5.1 shows the IoU of each category and the overall mean IoU. Bold numbers are the corresponding best results. The proposed method achieves 81.29% of IoU for tumors, and 80.47% of overall IoU. In the results of both tumor and overall IoU, the proposed method achieves about 2% improvements than that of the original CRFs.

5.4.3 Tumor Segmentation Results and Comparison with Previous Non-deep Learning Segmentation Methods

This subsection compares the proposed fuzzy U-Net + breast-anatomy constrained CRFs architecture with five non-deep learning methods. The existing non-deep learning



(a) (b) (c) (d) (e)

Fig. 5.3: Segmentation results of breast-anatomy constrained fully connected CRFs: (a) original images; (b) groundtruths; (c) results of fuzzy U-Net without CRFs; (d) fine-tuning results using fully connected CRFs; (e) fine-tuning results using the proposed method.

methods only focus on breast cancer segmentation, while semantic segmentation methods work on multi-object segmentation. In this subsection, the proposed method and the methods in [4–8] are compared. The semi-automatic BUS image segmentation methods [4,5] are studied, in which the regions of interest (ROIs) are given, and the methods could segment the tumor areas automatically. The fully automatic BUS image segmentation methods are studied [6–8]. The tumor segmentation results are shown in Fig. 5.4.

In Fig. 5.4, the semi-automatic segmentation methods (Figs. 5.4 (c) and (d)) obtain good results. Semi-automatic segmentation methods are helpful when doctors focus on specific areas and operate with the CAD systems interactively. Existing fully automatic segmentation methods get worse results since the performance of these methods relied on the individual dataset. They can obtain good performance only using their own datasets and need a massive number of training samples. The proposed method can achieve the

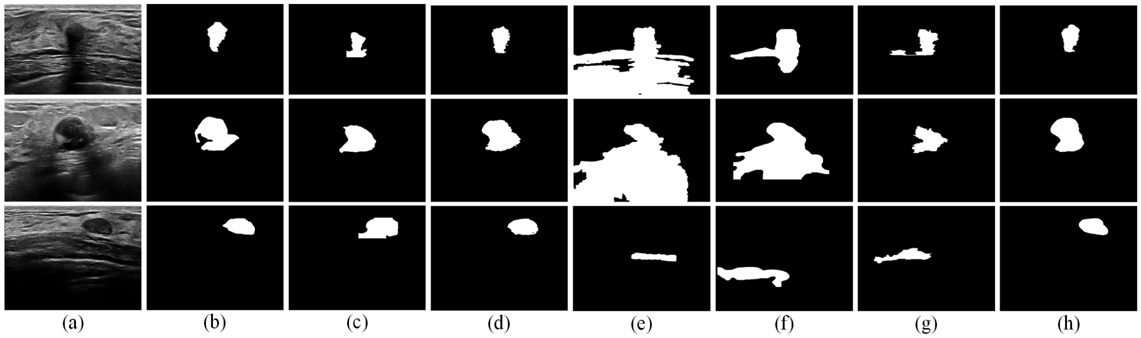


Fig. 5.4: Tumor segmentation results of the proposed method and existing methods: (a) original images; (b) groundtruths; (c) results using [4]; (d) results using [5]; (e) results using [6]; (f) results using [7]; (g) results using [8]; (h) results using the proposed method.

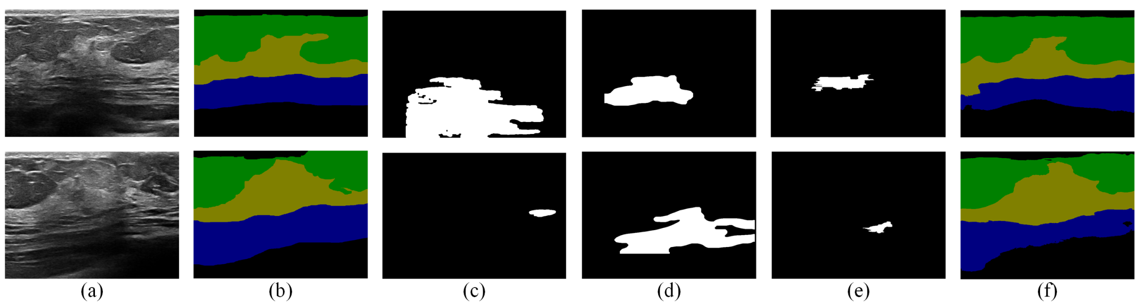


Fig. 5.5: The segmentation results for BUS images without tumors: (a) original images; (b) groundtruths; (c) results using [6]; (d) results using [7]; (e) results using [8]; (f) results of the proposed method.

best result even on a small dataset, and its robustness is much higher than that of other comparison methods.

Table 5.2 shows that the proposed method achieves the best results among all methods in comparison (Bold numbers are the corresponding best results). Furthermore, the proposed method can process the BUS images without tumors. The previous fully automatic methods could not solve such a problem; since all of the compared previous methods are based on the prerequisite, there is only one tumor in the image. As shown in Fig. 5.5, the two samples do not contain tumors. Figs. 5.5 (c)-(e) are the results of the previous fully automatic methods [6–8]. The white areas in the results are the tumors by the three methods, *i.e.*, they do not work well. In Fig. 5.5 (f), the proposed method can classify the layers in the BUS images well.

Table 5.2: Evaluation results on tumor segmentation.

	TPR	FPR	IoU
	Semi-Automatic Method		
Method [5]	83.01%	9.65%	79.74%
Method [4]	84.37%	17.51%	72.65%
	Fully-Automatic Method		
Method [6]	75.94%	43.84%	63.94%
Method [7]	83.55%	83.28%	65.22%
Method [8]	78.05%	15.43%	73.45%
Proposed	**90.33%**	**9.00%**	**81.29%**

Existing fully automatic segmentation methods [6–8] cannot solve multi-tumor cases as well. In Fig. 5.6, three BUS images are not in our dataset, and each image contains 2 tumors. The first image is collected by a doctor of the First Affiliated Hospital of Harbin Medical University; the second one is found in a public dataset [57]; the third one is from in [70]. In Fig. 5.6, the existing methods (Figs. 5.6 (c)-(e)) can only detect one tumor for each image, *i.e.*, they cannot obtain good results for containing more than one tumor; however, the proposed method can (Fig. 5.6 (f)).

Comparing the results of classic BUS image segmentation methods with the proposed method, here are some conclusions: 1) Classic methods depend on manually selected features; however, deep learning methods can automatically encode the convolutional features. Convolutional features perform better than manually selected features. 2) Compared classic methods depend on some assumptions. For example, the BUS images must contain a tumor, and the tumor must locate at the center of the whole image. That is why these classic methods cannot handle BUS images containing more than one tumor or no tumor. The proposed method reduces the uncertainty of features in the U-Net; therefore, the proposed method obtains the best result.

5.4.4 Comparison with DC Feature and Medical Knowledge Constrained CRFs

To compare the effectiveness of reflecting breast anatomy using the DC feature in Chapter 4 and medical knowledge constrained CRFs in this chapter, the comparison of the

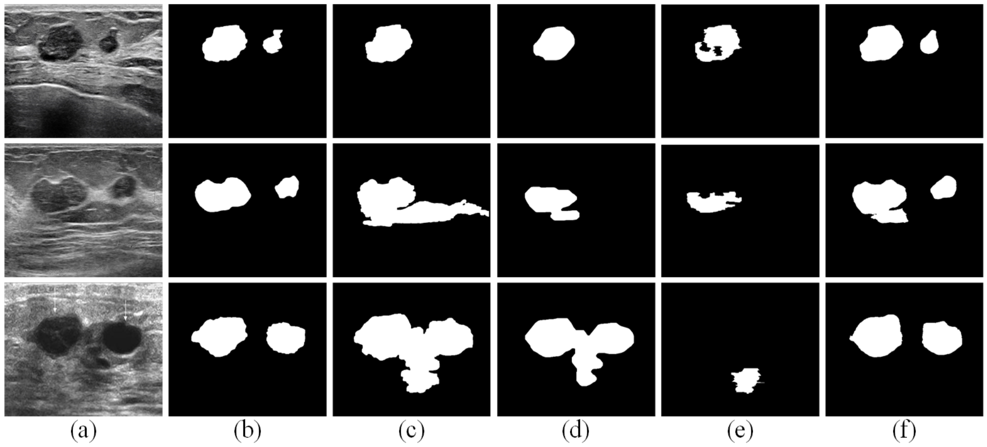


Fig. 5.6: The segmentation results of the BUS images containing two tumors: (a) original images; (b) groundtruths; (c) results using [6]; (d) results using [7]; (e) results using [8]; (f) results using the proposed method.

DC feature and medical knowledge constrained CRFs is made. In this experiment, the unary classifier for medical knowledge constrained CRFs is still the fuzzy U-Net proposed in Chapter 2. The experiment results are shown in Table 5.3. Bold numbers represent better results. The fuzzy U-Net + medical knowledge constrained CRFs architecture achieves better results in all categories and the overall performance. The experiment results demonstrate that 1) medical knowledge constrained CRFs can provide the order of layers in BUS image besides layer structure; however, DC feature only provides layer structure; 2) the medical knowledge constrained CRFs can refine the segmentation results of deep learning methods.

Table 5.3: Comparison with DC feature and medical knowledge constrained CRFs.

	Fat	Mammary	Muscle	Background	Tumor	Mean
Proposed CRFs	**85.06**	**77.24**	**78.66**	**80.09**	**81.29**	**80.47**
DC feature	84.45	76.90	75.48	79.35	79.63	79.16

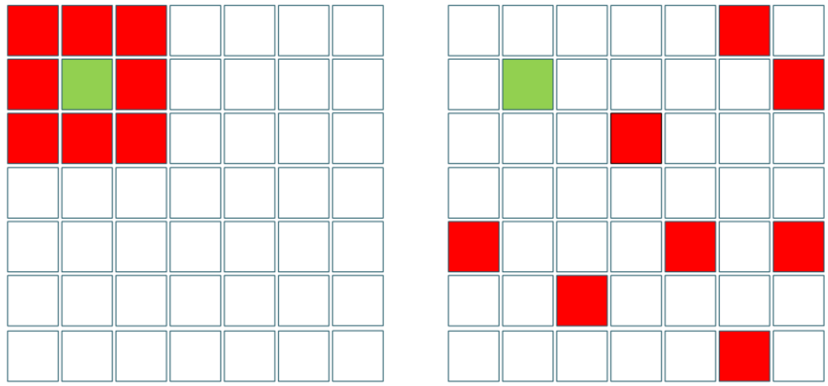
5.5 Conclusion

In this chapter, breast anatomy constrained conditional random fields are proposed to fine-tune the segmentation result from a deep convolutional neural network. In the breast anatomy constrained conditional random fields, the order of the breast tissue layers of the human breast is modeled in the pairwise-term of the energy function of the conditional random fields. The order of the breast tissue layers can be reflected during the minimization of the energy function. Experiment results on the 325-image BUS image dataset show that the proposed breast anatomy constrained conditional random fields can refine the segmentation results from deep convolutional networks. The proposed conditional random fields also obtain better results than the original conditional random fields. The deep learning + CRFs method outperforms non-deep learning BUS image segmentation methods. Therefore, the proposed conditional random fields can successfully increase the performance of the conditional random fields and provide more accurate segmentation results based on the correlation of pixels and breast anatomy.

CHAPTER 6
SHAPE-ADAPTIVE CONVOLUTIONAL OPERATOR AND ITS APPLICATION IN BUS IMAGE SEGMENTATION

6.1 Introduction

In Chapters 4 and 5, the breast anatomy is applied to machine learning algorithms to provide more information. The information is based on medical knowledge rather than image features and automatically encoding deep features. Automatically encoding deep features is one of the most important reasons leading to the success of deep convolutional neural networks. There are a lot of studies on obtaining better convolutional features and providing context information in deep learning; however, they do not discuss the higher-order information in the features. Higher-order information is the information from pixels with high Euclidean distances to the target pixel (also called non-local information). The non-local information can provide important information in pixel-wise classification. There is research for obtaining non-local information such as the non-local network [71], self-attention mechanism [34], criss-cross attention [35], *etc.* These non-local operators are investigated using the self-attention mechanism and calculate the correlation between one pixel with all other pixels in the feature map through matrix multiplication. The correlation is utilized as the attention coefficient of the feature map, which means they do not merge the non-local information. Deformable convolution [37] is the first research that tries to merge non-local information in the convolutional operator; however, the convolution pixels are still based on small Euclidean distances. In [72], a dynamic graph convolutional network is used for citation network and social media classifications. The graph structure is updated during the period that the network is trained by the k nearest neighbor (k-NN) algorithm and k-means cluster in the feature domain, and it is only specific for graph data. To extract more efficient convolutional features for BUS images, we propose a novel convolutional operator,



(a) (b)

Fig. 6.1: Convolutional operators: (a) 3×3 convolutional operator; and (b) SAC operator with 9 pixels.

a shape-adaptive convolutional (SAC) operator, to extract the features rather than in the Euclidean space. The positions of the pixels are selected by algorithms which means they can be from non-local positions.

6.2 The Proposed Method

The original 2-dimensional (2D) convolutional operator is reviewed. Then, the SAC operator is designed and compared with the original convolutional operator. Two approaches to select pixels for the SAC operator are discussed. Finally, the entire network structures and training strategy are presented.

6.2.1 Shape-adaptive Convolutional Operator

In Fig. 6.1 (a), the standard 3×3 convolutional operator is shown. The standard 3×3 convolutional operator only merges information from pixels with the shortest Euclidean distance (eight neighbors). However, pixels with larger Euclidean distances (Fig. 6.1 (b)) might also contain important information. In order to define the proposed SAC operator clearly, the standard convolutional operator is reviewed. For an input feature map $M \in \mathbb{R}^{H \times W \times C}$ and a convolutional operator $w \in \mathbb{R}^{S \times S \times T}$, H and W represent the height and width of the input feature map; C donates the number of channels; S represents the kernel size, and T represents the number of kernels in the operator. The original convolutional operation is defined as:

$$Out\left(i,j,t\right)=\sum_{C_h=0}^{C-1}\sum_{p=0}^{S-1}\sum_{q=0}^{S-1}M\left(i+p-\lfloor S/2\rfloor,j+q-\lfloor S/2\rfloor,C_h\right)w\left(p,q,t\right)+bias\left(t\right)\quad(6.1)$$

where $i=0,1,2,...,H-1$; $j=0,1,2,...,W-1$; $t=0,1,...,T-1$; $Out\left(i,j,t\right)$ represents the convolutional result for the pixel in the ith row and the jth column using the tth kernel. The original convolutional operator is a 2D cross-correlation operator of the input feature map and convolutional kernel. The convolutional kernel will slide through the entire feature map. In this work, we design a novel convolution operator that can select convolutional pixels effectively and can extract the higher-order information well.

Suppose for a pixel in the feature map, there are k selected pixels (the approaches for selecting the k pixels will be discussed in Subsection 6.2.2), and an adjacent matrix $Adj\in\mathbb{R}^{C\times HW\times k}$ is defined to store the feature values of the selected k pixels; HW represents the number of pixels in the input feature map. A 1×1 convolutional operator $w\in\mathbb{R}^{T}$ is performed on the adjacent matrix:

$$Q\left(C_h,u,t\right)=\sum_{k_h=0}^{k-1}Adj\left(C_h,u,k_h\right)w\left(t\right)\quad(6.2)$$

where $C_h=0,1,2,...,C-1$; $u=0,1,2,...,HW-1$; and $t=0,1,...,T-1$; $Q\left(C_h,u,t\right)$ represents the convolutional result using the tth kernel. The 1×1 convolutional operator does not contain the bias. The final convolutional result is computed by:

$$Out\left(u,t\right)=\sum_{C_h=0}^{C-1}Q\left(C_h,u,t\right)+bias\left(t\right)\quad(6.3)$$

where $u=0,1,2,...,HW-1$; and $t=0,1,...,T-1$. $Out\left(u,t\right)$ represents the final convolutional result of the uth pixel using the tth kernel.

$Out\left(u,t\right)\in\mathbb{R}^{HW\times T}$ is reshaped to the size of $Out\left(i,j,t\right)\in\mathbb{R}^{H\times W\times T}$ and passed to the next operation. The combination (Eq. (6.4)) of Eq. (6.2) and Eq. (6.3) is equivalent to the original 2D convolutional operator (Eq. (6.1)) if k pixels are selected by the closest

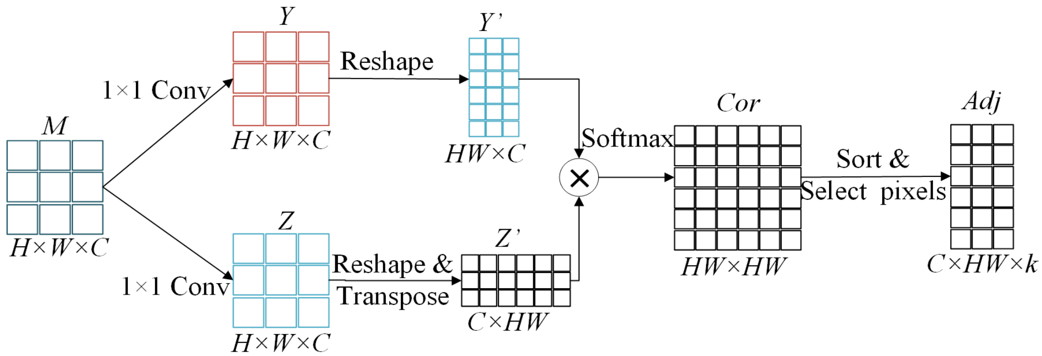


Fig. 6.2: Selecting pixels for the SAC operator using self-attention coefficient.

Euclidean distance. Here we choose 9 pixels as an example. The closest 9 pixels to the target pixel are the target pixel and its neighbors (marked by green and red in Fig. 6.1 (a), respectively).

$$Out\,(u,t) = \sum_{C_h=0}^{C-1} \sum_{k_h=0}^{k-1} Adj\,(C_h,u,k_h)\,w\,(t) + bias\,(t) \tag{6.4}$$

where $u = 0, 1, 2, ..., HW - 1$; and $t = 0, 1, ..., T - 1$.

6.2.2 Selecting Pixels for SAC Operator

Two approaches are utilized for selecting k pixels in the SAC operator: the k-NN algorithm and the self-attention coefficient calculation [34]. In the first approach, we choose the k pixels according to feature values; in the second approach, each pixel calculates the self-attention coefficients between other pixels in the feature map; and the k pixels which have the highest coefficients are selected.

***K* nearest neighbor:** Input feature map $M \in \mathbb{R}^{H\times W\times C}$ is reshaped to $M' \in \mathbb{R}^{HW\times C}$. $M'\,(u,:)$ and $M'\,(v,:)$ represent the uth and vth rows in M'; the L_2 norm between the feature vector $M'(u,:)$ and $M'(v,:)$ are calculated:

$$D(u,\ v) = ||M'(u,:) - M'(v,:)||^2 \tag{6.5}$$

where $u = 0, 1, \ldots, HW - 1$; $v = 0, 1, \ldots, HW - 1$; and $D\,(u, v)$ represents the L_2 norm between the uth pixel and the vth pixel in the input feature map. The indexes of the smallest

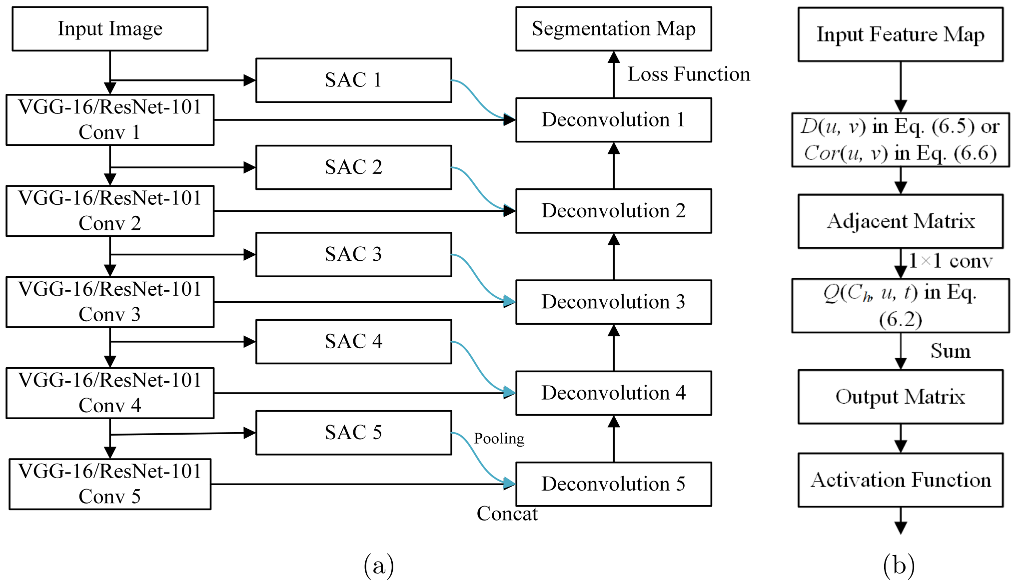


Fig. 6.3: (a) The entire network; and (b) the proposed SAC operator.

k values are found, and their feature values are stored in adjacent matrix $Adj \in \mathbb{R}^{C\times HW\times k}$.

Self-attention coefficient: The k pixels are selected based on the self-attention coefficient calculated by the self-attention mechanism [34] without attention multiplication. In Fig. 6.2, the input feature map M is input into two 1×1 convolutional operators to generate two new maps $Y, Z \in \mathbb{R}^{H\times W\times C}$. Then, Y and Z are reshaped to $Y' \in \mathbb{R}^{HW\times C}$ and $Z' \in \mathbb{R}^{C\times HW}$. We perform a matrix multiplication of matrices Y' and Z' and apply a Soft-max to calculate the spatial self-attention coefficient matrix Cor.

$$Cor(u,v) = \frac{\exp\left(Y'(u,:)\cdot Z'(:,v)\right)}{\sum_{v=0}^{HW-1}\exp(Y'(u,:)\cdot Z'(:,v))} \tag{6.6}$$

where $u = 0, 1, ..., HW - 1$; $v = 0, 1, ..., HW - 1$; $Y'(u,:)$ represents the uth row of Y'; $Z'(:,v)$ represents the vth column of Z'. $Cor(u,v)$ measures the correlation between the uth pixel and the vth pixel in the input feature map. If two pixels contain similar feature values, the self-attention coefficient between them is high. Then, the adjacent matrix $Adj \in \mathbb{R}^{C\times HW\times k}$ is created based on the feature values of the k pixels with the highest self-attention coefficients for each pixel.

6.2.3 Entire Network Structure and Training Details

The proposed SAC operator (Fig. 6.3 (b)) is applied to the U-Net with VGG-16 [22] and the U-Net with ResNet-101 [15] (Fig. 6.3 (a)), respectively, for comparison. The input image and feature maps from the convolutional blocks 1-4 are processed by the proposed SAC operators. The output feature maps of SAC operators are concatenated with the feature maps of the convolutional blocks. The procedure of the proposed SAC operator is shown in **Algorithm 2**.

Experiment environment: The loss function is defined as a category cross-entropy loss. The weights in the proposed network are initialized randomly. The batch size is 8. The input images are resized to 128×128. The optimizing method is the Stochastic Gradient Descent (SGD), with a learning rate of 0.001 and momentum of 0.99. The details for selecting k in the SAC operator will be discussed in the experiment section. The training epoch number is set to 80. The experiments are conducted using a computer with Ubuntu 18.04 system, Intel(R) Xeon(R) CPU E5-2620 2.10 GHz, and 8 NVIDIA GeForce 1080 graphics cards. The implementation is based on PyTorch 1.6.0. and 10-fold validation is utilized in the experiments.

Algorithm 2 SAC Algorithm

Input: Input feature map $M \in \mathbb{R}^{H\times W\times C}$; the number of pixels used to calculate convolution (k); the number of output channels (the number of filters, T).

Initialization: The weights and bias are initialized by the uniform distribution.

1: Compute the relation matrix ($D(u,v)$ or $Cor(u,v)$ in Subsection 6.2.2) by Eq. (6.5) or Eq. (6.6).

2: Compute $Adj \in \mathbb{R}^{C\times HW\times k}$ by sorting $D(u,v)$ or $Cor(u,v)$, and selecting feature values of the top k pixels.

3: Perform 1×1 convolution on Adj; the kernel size is $1\times 1\times T$; and the intermediate result is $Q\,(C_h,u,t)$.

4: Perform summation in the t dimension of $Q\,(C_h,u,t)$; the result is $Out\,(u,t)$.

5: Add bias to $Out\,(u,t)$.

6: Reshape the result to $Out\,(i,j,t) \in \mathbb{R}^{H\times W\times T}$.

Output: Output feature map $Out\,(i,j,t) \in \mathbb{R}^{H\times W\times T}$.

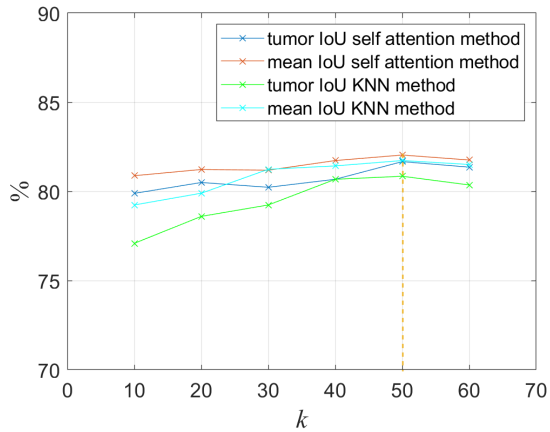


Fig. 6.4: The relation between segmentation performance and parameter k.

6.3 Experiment Results

6.3.1 Datasets

The proposed network is applied to three datasets. The private multi-object BUS image dataset collected by a cooperative hospital contains 325 BUS images mentioned and utilized in the previous chapters. This dataset contains pixel-wise groundtruths for 5 categories: fat layer, mammary layer, muscle layer, background area, and tumor. We also utilize two public binary BUS image datasets for further evaluating the proposed method. The binary BUS image datasets only contain pixel-wise groundtruths for the tumors and background areas. These two datasets are also mentioned in Table 3.1. The first binary dataset [2] contains 163 BUS images, including 109 benign and 54 malignant images; and the second dataset [3] contains 780 BUS images, including 437 benign, 210 malignant, and 133 no tumor images.

The non-local operator self-attention mechanism [34] and the deformable convolution [37] are also applied to the input feature map and 4 convolutional feature maps of U-Net with ResNet-101, respectively, for comparison. Moreover, five state-of-the-art semantic segmentation methods: U-Net with VGG-16, U-Net with ResNet-101, Deeplab V3+ [64], FCN-8s [14], and PSPNet [18] with ResNet-101, are compared with the proposed approach as well. All methods are not pre-trained using other datasets.

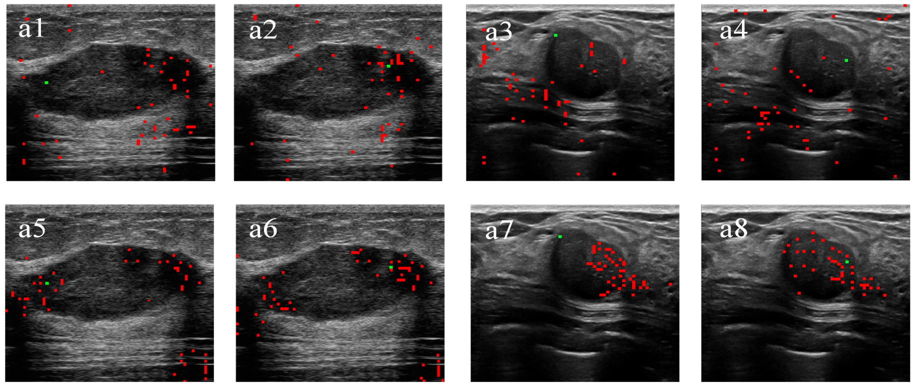


Fig. 6.5: The positions of k pixels involved in the SAC operators: the green pixels are target pixels; red pixels are selected to calculate the convolutions of the target pixels; a1-a4 are based on the k-NN selection approach; a5-a8 are based on the self-attention coefficient selection approach.

6.3.2 Metrics

A metric popularly used in semantic segmentation, Intersection over Union (IoU, also known as the Jaccard Index in the binary task), is utilized to evaluate the performance of multi-object segmentation tasks. Mean IoU (mIoU $= \sum \text{IoU}/C_n$, and C_n represents the number of categories) over five categories is utilized to evaluate the overall performance.

Five area metrics: True Positive Ratio (TPR), False Positive Ratio (FPR), Jaccard Index (JI), Dice's Coefficient (DS), and Area Error Ratio (AER) [1] are utilized to evaluate the performance on two public binary datasets.

6.3.3 Parameter k in SAC Operator

This part discusses the relationship between segmentation results and parameter k (the number of selected pixels). We choose the multi-object dataset to train the proposed network. The parameter k for five SAC operators is the same. The proposed SAC + ResNet-101 is utilized. Parameter k is changed from 10 to 60, and the step size is 10. The segmentation performances are shown in Fig. 6.4. When $k = 50$, the mean IoU and the IoU for tumors are the highest for both the k-NN selection approach and self-attention approach. Therefore, the parameter k is set to 50 for both self-attention coefficient and k-NN selection approach in Subsection 6.3.5 and 6.3.6.

Table 6.1: Results of multi-object segmentation (IoU (%)).

	Fat	Mammary	Muscle	Background	Tumor	Mean
VGG-16	70.34	66.72	66.17	65.91	74.66	68.76
ResNet-101	81.50	73.41	72.07	74.47	75.29	75.35
FCN-8s	82.57	75.47	75.53	78.59	74.42	77.32
PSPNet	82.07	74.40	74.49	77.36	74.75	76.61
Deeplab	78.91	68.71	67.33	73.94	69.04	71.58
Self-attention	82.53	76.23	75.91	80.29	78.81	78.75
Deformable Convolution	81.49	75.87	74.12	79.34	76.77	77.52
Method1	82.64	75.67	75.81	80.23	78.34	78.54
Method2	84.10	78.58	76.18	83.61	80.03	80.50
Method3	85.52	80.06	77.34	84.18	79.74	81.37
Method4	**86.18**	**80.65**	**78.17**	**84.69**	**81.07**	**82.15**

* Method1: VGG-16 + SAC with k-NN, Method2: VGG-16 + SAC with self-attention coefficient, Method 3: ResNet-101 + SAC with k-NN, Method4: ResNet-101 + SAC with self-attention coefficient.

6.3.4 Positions of k Selected Pixels in SAC Operator

In this part, the positions of the k selected pixels to calculate the SAC operators are displayed. The input of the first SAC operator is the original image; all the results in Fig. 6.5 are from the first SAC operator. In Fig. 6.5, two images are chosen to exhibit the pixels used in the SAC operator. The self-attention coefficient approach and k-NN selection approach are compared, and $k = 50$ for both approaches.

Fig. 6.5 shows that the proposed SAC operator can select non-local pixels for convolution; however, the ordinary convolutional operator can only select local pixels. In Figs. 6.5 a1-a4, the selected pixels are based on k-NN and feature values. Two different pixels (green pixels in Figs. 6.5 a1-a4) in the same category in an image have similar correlated pixels. However, the selected correlated pixels are distributed in tumor areas, mammary layer, and background, *i.e.*, they cause segmentation errors. The pixels selected by the self-attention coefficient (Figs.6.5 a5-a8) are mostly distributed inside the tumor areas, which means the self-attention coefficient can select the correlated pixels in the correct category. Even for two different pixels (green pixels in Figs. 6.5 a5-a8) in the same category in one image, they have similar distributions of the correlated pixels.

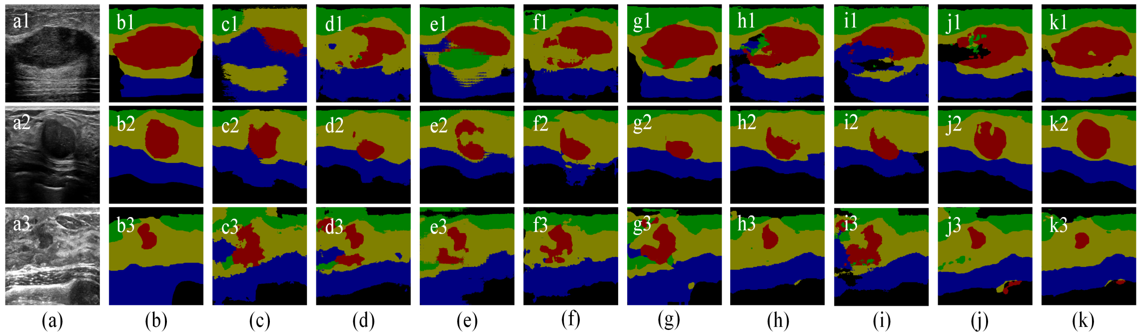


Fig. 6.6: Semantic segmentation: (a) original images; (b) groundtruths; (c) results of U-Net with VGG-16; (d) results of U-Net with ResNet-101; (e) results of PSPNet; (f) results of Deeplab; (g) results of FCN-8s; (h) results of ResNet-101 + self-attention mechanism; (i) results of ResNet-101 + deformable convolution; (j) results of ResNet-101 + SAC with k-NN; and (k) results of ResNet-101 + SAC with self-attention coefficient.

6.3.5 Multi-object Segmentation for BUS Images

Fig. 6.6 shows the segmentation results using a multi-object dataset. The proposed approach can segment the BUS images better because the SAC operator can utilize the correlated pixels determined correctly in convolution for each pixel in BUS images. These correlated pixels can provide non-local context information. The performance of using the self-attention coefficient is better than that of using k-NN in the SAC operator (refer Fig. 6.6 j1 and k1) because the correlated pixels provided by the self-attention coefficient are in the correct categories; meanwhile, the correlated pixels for different pixels in the same category are similar. Applying the self-attention mechanism to ResNet-101 can reduce misclassification because the self-attention coefficient provides better non-local context information. However, the deformable convolution does not perform well because the pixels used in convolution are still according to the short Euclidean distances with the target pixel.

The evaluation results can be found in Table 6.1. Bold numbers are the corresponding best results. The SAC with self-attention coefficient in the ResNet-101 achieves the best IoUs in all categories and the overall performance. Applying the self-attention mechanism can increase performance. However, the deformable convolution does not increase the performance as much as the proposed method or self-attention mechanism. The result indicates that the deformable convolution only contains limited ability in providing non-local

Table 6.2: Results on public datasets (%).

Datasets	Methods	TPR	FPR	JI	DS	AER
[2]	VGG-16	79.30	45.84	68.16	76.40	66.54
	Deeplab	63.68	36.06	52.93	61.91	72.38
	ResNet101	83.58	34.40	71.43	79.45	50.82
	FCN8s	82.72	41.14	67.50	76.87	58.42
	PSPNet	81.08	40.42	69.77	78.24	59.34
	Self-attention	82.58	**16.39**	73.83	81.37	33.81
	Deformable Convolution	84.11	37.15	71.86	79.92	53.04
	Proposed1	87.51	27.16	76.40	82.26	38.67
	Proposed2	**88.21**	21.23	**77.90**	**83.21**	**32.12**
[3]	VGG-16	78.66	41.98	68.77	77.37	63.32
	Deeplab	59.88	39.39	49.65	59.39	79.52
	ResNet101	79.40	46.02	69.26	77.90	66.62
	FCN8s	74.23	46.69	63.16	73.03	72.63
	PSPNet	77.11	46.65	65.21	74.75	69.54
	Self-attention	79.76	**35.26**	69.74	78.46	55.50
	Deformable Convolution	78.11	40.20	68.88	77.35	62.09
	Proposed1	81.05	39.11	71.63	78.27	51.07
	Proposed2	**82.56**	37.12	**72.12**	**79.12**	**45.01**

* Proposed1: ResNet-101 + SAC with k-NN, Proposed2: and ResNet-101 + SAC with self-attention coefficient. Bolds are the corresponding best results.

information.

6.3.6 Tumor Segmentation Results Using Public Datasets

Two public BUS image datasets are utilized to evaluate the proposed method as well. The proposed SAC + ResNet-101 is utilized because it achieves better results than the proposed SAC + VGG-16 using the multi-object dataset. The overall segmentation performance can be found in Table 6.2. Bold numbers are the best results. The proposed SAC with self-attention coefficient achieves the best metrics except for FPR compared with other methods; however, using the self-attention mechanism in ResNet-101 obtains lower FPR. Lower FPR indicates that using the self-attention mechanism in the ResNet-101 can reduce errors in segmentation results by using non-local context information; however, the proposed method can achieve better overall performance by combining non-local information in convolution. The proposed SAC + self-attention coefficient obtains 6.47% and 2.86% increases of JI scores compared with that of the U-Net with ResNet-101 on datasets [2] and [3], respectively. The evaluation results indicate that the self-attention coefficient can provide effective context information; however, using the self-attention coefficient in selecting convolutional pixels and merging the long-distance information is better than just using it as an attention weight.

6.4 Conclusion

In this chapter, we introduce a shape-adaptive convolutional operator to BUS image segmentation. Compared with the DC feature in Chapter 4 and medical knowledge constrained CRFs in Chapter 5, they provide breast medical knowledge to BUS image segmentation. The SAC operator can extract non-local information better than that in the Euclidean space. Non-local information is extracted by two methods: 1) k nearest neighbor, and 2) the self-attention coefficient calculation. Experimental results demonstrate that the two methods can find better-correlated pixels in feature space for the target pixel, and they can provide more useful information. Moreover, the SAC operator can select pixels effectively and avoid losing pixels during the deformation of the convolutional kernel. The proposed method achieves significant improvement on three BUS image datasets. The proposed SAC + self-attention coefficient obtains a 5.78% increase of IoU in tumor category on a multi-object dataset and 6.47% and 2.86% increases of JI scores on datasets [2] and [3] compared with that of the U-Net with ResNet-101.

CHAPTER 7
BREAST ULTRASOUND IMAGE SEGMENTATION USING A MULTI-SCALE FUZZY GENERATIVE ADVERSARIAL NETWORK

7.1 Introduction

In the previous chapters, many BUS image segmentation approaches have been reviewed. These approaches can be divided into five categories: thresholding algorithms, region-growing algorithms, watershed algorithms, graph-based algorithms, and deep neural network-based algorithms [11], [38], [39]. We have been proposed five methods for BUS segmentation in Chapter 2 to Chapter 6. These methods can solve uncertainty in the channels and pixels in the feature maps and involve breast context information in machine learning algorithms. To further improve the performance of classic segmentation networks, Generative Adversarial Network (GAN) is involved in BUS image segmentation. In the previous research, researchers propose a Generative Adversarial Network (GAN) [73] which employs an adversarial network to guide the segmentation network to generate more accurate segmentation results. In [74], a semantic segmentation method including a convolutional semantic segmentation network along with an adversarial network is proposed to eliminate inconsistencies between groundtruth maps and predicted segmentation results. Xue *et al.* [75] further propose an adversarial network with multi-scale L_1 loss for image segmentation that can learn features in different scales and capture contextual relationships to boost the segmentation accuracy. Han *et al.* [76] propose a semi-supervised generative adversarial network with a dual-attentive-fusion block to enhance discrimination for BUS image segmentation.

Despite the good performance of the above methods, they do not consider the uncertainty in BUS images. In this study, we proposes a novel multi-scale fuzzy generative adversarial network (MSF-GAN) for BUS image segmentation that uses uncertainty maps

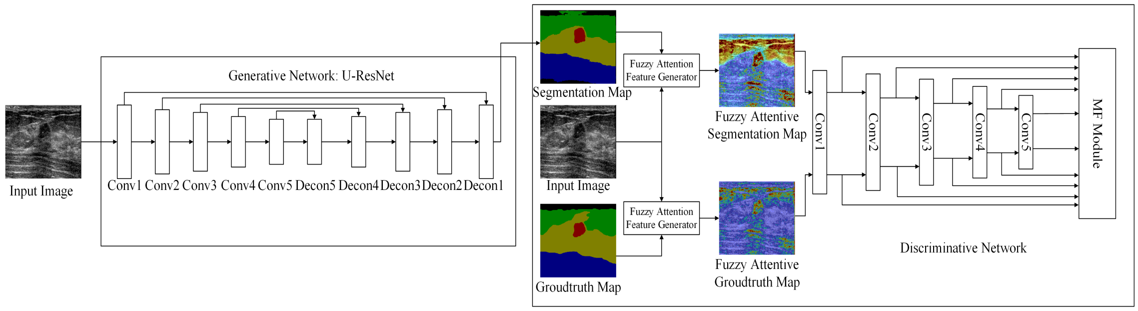


Fig. 7.1: An overview of the proposed MSF-GAN.

to train the discriminative network. Inspired by reference [75], the proposed MSF-GAN consists of a generative network (G-net) and a discriminative network (D-net) which respectively minimize and maximize the loss functions. The output of G-net is a segmentation map. The proposed MSF-GAN employs a fuzzy attentive feature generator and a multi-scale fuzzy entropy (MSF) module, which can transform the segmentation maps and groundtruth maps into the fuzzy domain to measure uncertainty. The multi-scale fuzzy entropy (MSF) module can distinguish the difference in uncertainty maps from two inputs and help to train a better segmentation network. The major contributions of the proposed approach are: (1) Design a novel MSF-GAN for BUS image segmentation that outperforms six state-of-the-art deep neural network-based methods on three BUS datasets in terms of five metrics. (2) Design a fuzzy attentive feature generator to generate fuzzy attentive feature maps for the segmentation maps generated by G-net and ground-truth maps. (3) Design an MSF module to measure the uncertainty in segmentation maps and groundtruth maps and calculate a multi-scale L_1 loss on uncertainty maps to help to train the segmentation network.

7.2 The Proposed Method

The proposed MSF-GAN consists of a G-net for the generation of pixel-wise segmentation maps and a D-net for guiding G-net to generate more accurate segmentation maps. Between G-net and D-net, a fuzzy attentive feature generator is employed to transform the segmentation maps and groundtruth maps to the fuzzy domain and then generate fuzzy attentive feature maps for them. D-net incorporates an MSF module to calculate multi-scale L_1 loss on uncertainty maps extracted from fuzzy attentive maps. In this subsection, we first present the architecture of MSF-GAN, then present the fuzzy attentive feature generator, and finally present the MSF module.

7.2.1 Overview

The architecture of the proposed MSF-GAN is illustrated in Fig. 7.1. MSF-GAN employs a U-ResNet (a U-shape network with ResNet-101 [15] as its backbone) as its G-net to generate pixel-wise segmentation results, denoted as segmentation maps. All input BUS images are first resized to 128×128 and then fed into G-net. A segmentation map of size $128 \times 128 \times C$ is generated for an input BUS image, where C represents the total number of categories. Each pixel contains C values in the segmentation map, and each element represents the probability to the corresponding category. Then, we use a fuzzy attentive feature generator that takes an original BUS image and its groundtruth map as inputs to compute a fuzzy attentive groundtruth map. Similarly, we compute a fuzzy attentive segmentation map by an original BUS image and its segmentation map. The fuzzy attentive feature generator will be introduced in Subsection 7.2.2 in detail. The D-net is composed of five convolutional layers with kernels of size 4×4, stride 2, padding 1, and ReLU activation function. It takes a fuzzy attentive groundtruth map and a fuzzy attentive segmentation map as two inputs and calculates a multi-scale L_1 loss on their uncertainty maps, which will be introduced in Subsection 7.2.3. The objective of G-net is to generate accurate segmentation maps, and the objective of D-net is to distinguish the uncertainty of the segmentation maps and groundtruth maps. For an input BUS image, if the uncertainty map of the segmentation map is very close to the uncertainty map of the groundtruth map, then it is hard for D-net to discriminate them. In contrast, if the uncertain map of the segmentation map is not close to the uncertain map of the groundtruth map, it means there still exists uncertainty in the segmentation map. The goal is to make G-net generate very accurate segmentation maps which contain similar uncertainty maps to the groundtruth maps. In this study, we enhance the discriminating ability of the D-net by using a fuzzy

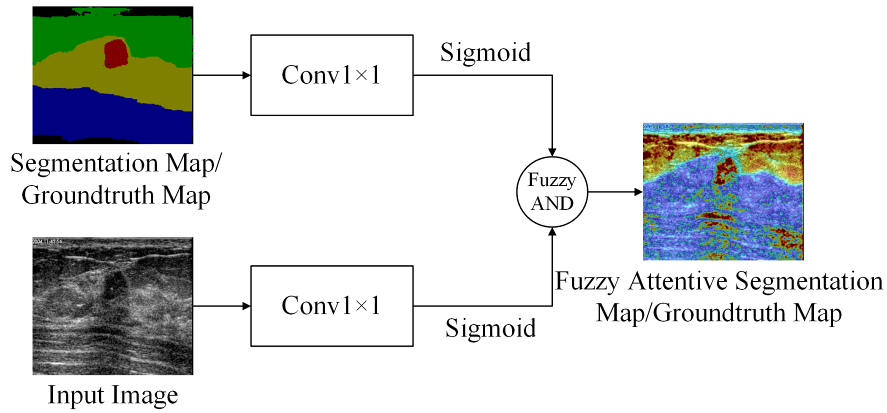


Fig. 7.2: Illustration of the proposed fuzzy attentive feature generator.

attentive feature generator and a multi-scale L_1 loss calculated on uncertainty maps and therefore force G-net to generate more accurate segmentation maps that are very close to the groundtruth maps.

7.2.2 Fuzzy Attentive Feature Generator

The target for fuzzy attentive feature generator is to transform the input of the discriminative network to the fuzzy domain. Fig. 7.2 illustrates the proposed fuzzy attentive feature generator. It takes a pair of an original BUS image and its segmentation map generated by the G-net , or a pair of an original BUS image and its groundtruth map as inputs. Specifically, for an original image, its segmentation map and its groundtruth map are individually transformed into the fuzzy domain by a convolutional operator with a kernel size of 1×1 and sigmoid function as activation function. The operation of fuzzification can be represented by:

$$F_x \; = Conv \; 1 \times 1 \; (x) \tag{7.1}$$

where x can be an original BUS image of size 128×128, a segmentation map generated by G-net of size $128 \times 128 \times C$, or a groundtruth map of size $128 \times 128 \times C$, where C is the total number of categories. After fuzzification, x is transformed into F_x of size $128 \times 128 \times C$. Then, we respectively perform a fuzzy AND operator on a pair of the fuzzified original image (denoted as F_o) and fuzzified segmentation map (denoted as F_{pre}), and on a pair of F_o and the fuzzified groundtruth map (denoted as F_{gt}) to generate a fuzzy attentive groundtruth map FA_{gt} and a fuzzy attentive segmentation map FA_{pre}. This operation can be represented by:

$$FA_{pre} = \min(F_o, F_{pre}) \tag{7.2}$$

$$FA_{gt} = \min(F_o, F_{gt}) \tag{7.3}$$

where min is the AND operator in fuzzy logic that performs a pixel-wise minimization operation on its two inputs. FA_{pre} and FA_{gt} are of size $128 \times 128 \times C$. Different from reference [75] that directly uses groundtruth map masked images and segmentation map masked images as the inputs of D-net, we first generate three types of fuzzified maps and then compute two fuzzy attentive maps by using them as the inputs of D-net. We can train D-net better by using these fuzzy attentive maps to extract multi-scale features and calculate a multi-scale L_1 loss on uncertainty maps extracted from these fuzzy attentive maps because through a non-linear transformation of the fuzzification and fuzzy AND operator in fuzzy feature generator, the fuzzy features are more discriminable than the non-fuzzy features and we can also measure uncertainty on fuzzy features.

7.2.3 Multi-scale Fuzzy Entropy Module

In D-net, five convolutional layers with kernels of different sizes are used to extract multi-scale features on the input fuzzy attentive groundtruth map FA_{gt} and fuzzy attentive segmentation map FA_{pre}. These features are then fed into the proposed MSF module to calculate a multi-scale L_1 loss on uncertainty maps, which are calculated via FA_{gt} and FA_{pre}. By training a powerful D-net to better discriminate the uncertainty map of FA_{gt} and that of FA_{pre}, G-net is forced to generate more accurate segmentation maps. As shown in Fig. 7.1, D-net takes a fuzzy attentive groundtruth map FA_{gt} and a fuzzy attentive segmentation map FA_{pre} as the inputs, then employs five convolutional layers to extract multi-scale features. Let L denote the total number of convolutional layers in D-net (here $L = 5$). Let $f^l(FA_{pre})$ and $f^l(FA_{gt})$ denote the feature map extracted by the l-th layer of

D-net, respectively. Then it performs a 1×1 convolution with ReLU activation function on $f^l(FA_{pre})$ and $f^l(FA_{gt})$ respectively to transform their channel number to C to calculate fuzzy entropy. The transformed feature maps are denoted as:

$$T^l_{pre} = Conv1 \times 1 \ (f^l(FA_{pre})) \tag{7.4}$$

$$T^l_{gt} = Conv1 \times 1 \ (f^l(FA_{gt})) \tag{7.5}$$

Then, it calculates the fuzzy entropy on T^l_{pre} and T^l_{gt} to represent their uncertainty maps respectively:

$$E^l_{pre}(i,j) = -\frac{1}{\log C}\sum_{c=1}^{C} T^l_{pre}(i,j,c) \cdot \log T^l_{pre}(i,j,c) \tag{7.6}$$

$$E^l_{gt}(i,j) = -\frac{1}{\log C}\sum_{c=1}^{C} T^l_{gt}(i,j,c) \cdot \log T^l_{gt}(i,j,c) \tag{7.7}$$

where $T^l_{pre}(i,j,c)$ and $T^l_{gt}(i,j,c)$ represent the values of the i-th row, j-th column and c-th channel of T^l_{pre} and T^l_{gt}, respectively. It then computes a multi-scale L_1 loss on the uncertainty maps $E^l_{pre}(i,j)$ and $E^l_{gt}(i,j)$ by:

$$\min_{\theta_G}\max_{\theta_D} \mathcal{L}(\theta_G,\theta_D) = \frac{1}{N}\sum_{n=1}^{N} \ell_{mae}(E^{l,n}_{pre}(i,j), E^{l,n}_{gt}(i,j)) \tag{7.8}$$

where θ_G and θ_D denote the parameters of G-net and D-net respectively; N denotes the total number of training images; $E^{l,n}_{pre}$ and $E^{l,n}_{gt}$ denote the uncertainty map extracted by the l-th layer on the n-th training image, respectively. ℓ_{mae} is the Mean Absolute Error (MAE) (L_1 loss), defined as:

$$\ell_{mae}(E^l_{pre}, E^l_{gt}) = \frac{1}{L}\sum_{l=1}^{L} \left\| E^l_{pre} - E^l_{gt} \right\|_1 \tag{7.9}$$

The loss $\mathcal{L}$ in Eq. (7.8) can capture a rich contextual relationship between pixels by using the multi-scale uncertainty maps E^l_{pre} and E^l_{gt} generated by different convolutional layers. During the training of MSF-GAN, we minimize $\mathcal{L}$ with respect to the parameters θ_G

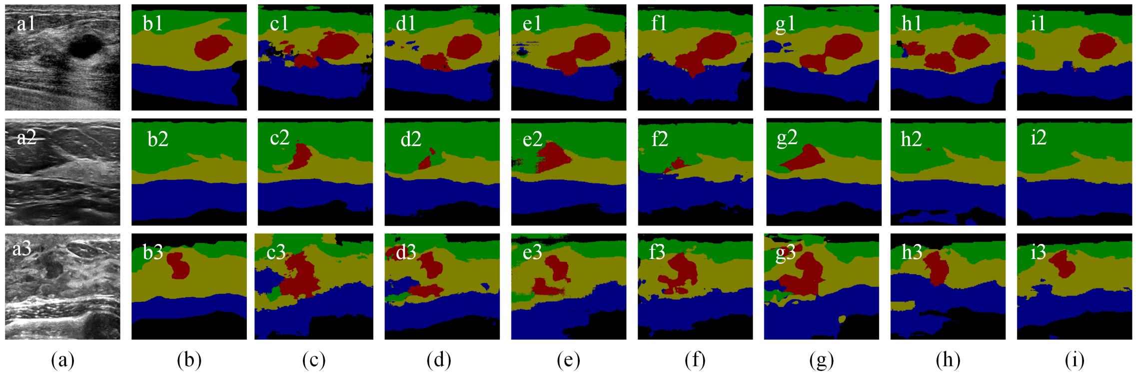


Fig. 7.3: Illustration of segmentation results on multi-layer dataset. (a) Original BUS image; (b) Groundtruth; (c) U-VGG; (d) U-ResNet; (e) PSPNet; (f) Deeplabv3+; (g) FCN-8s; (h) SegAN; (i) Proposed MSF-GAN.

of G-net, while maximizing it with respect to the parameters θ_D of D-net. The objective of G-net is to generate accurate segmentation maps that contain similar uncertainty to groundtruth maps so that $\mathcal{L}$ is minimized. The uncertainty is represented by fuzzy entropy. In contrast, the objective of D-net is to distinguish the uncertainty of segmentation maps from the uncertainty of groundtruth maps and therefore force G-net to generate accurate segmentation maps. When D-net is powerful enough, it can distinguish these two kinds of uncertainty maps very well so that $\mathcal{L}$ is maximized. To implement this strategy, we train G-net and D-net in an alternating scheme: first, fix G-net and train D-net to maximize $\mathcal{L}$, and then fix D-net and train G-net to minimize $\mathcal{L}$. During the training procedure, both G-net and D-net are becoming more and more powerful. By using fuzzy attentive feature maps and the multi-scale L_1 loss computed from these fuzzy attentive feature maps, the discriminating ability of D-net is further enhanced compared with [75]. Therefore, our more powerful D-net can better guide G-net to generate more accurate segmentation maps close to groundtruth maps.

7.3 Experiment Results

7.3.1 Datasets and Metrics

We evaluate the performance of the proposed MSF-GAN on three datasets: a multi-layer dataset (introduced in Subsection 2.5.1 and Subsection 3.4.1), and two public datasets mentioned in Table 3.1: Dataset 1 [2] and Dataset 2 [3]. The multi-layer dataset is a private dataset consisting of 325 images with a mean image size of 500×300 pixels. The groundtruth annotations include four breast anatomical layers (fat layer, mammary layer, muscle layer, background layer) and tumors. Dataset 1 and dataset 2 are two public BUS datasets where groundtruth annotations only separate tumors and background. Dataset 1 has 163 images with a mean image size of 760×570 pixels, where most of the images contain small tumors. Dataset 2 has 780 images with a mean image size of 500×500 pixels where tumors are in different sizes. In total, there are 1268 images used for evaluation.

We further compare the segmentation performance of MSF-GAN and six state-of-the-art deep neural network-based methods on above mentioned three BUS datasets. The six compared methods are: U-Net [48] with ResNet-101 [30] as its backbone (denoted as U-ResNet), U-Net with VGG-16 [22] as its backbone (denoted as U-VGG), FCN-8s [14], SegAN [75], PSPNet [18], and Deeplabv3+ [64]. We use five metrics for the evaluation. They are: True Positive Ratio (TPR), False Positive Ratio (FPR), Intersection over union (IoU), Dice's Coefficient (DS) (also known as F1-score), and Area Error Ratio (AER).

Table 7.1: Results of multi-layer segmentation (IoU (%)).

	Fat	Mammary	Muscle	Background	Tumor	Mean
U-ResNet	81.50	73.41	72.07	74.47	75.29	75.35
U-VGG	70.34	66.72	66.17	65.91	74.66	68.76
FCN-8s	82.57	75.47	75.53	78.59	74.42	77.32
SegAN	81.68	75.89	72.53	81.69	77.23	77.80
PSPNet	82.07	74.40	74.49	77.36	74.75	76.61
Deeplab	78.91	68.71	67.33	73.94	69.04	71.58
MSF-GAN	**83.11**	**77.05**	**73.11**	**81.98**	**78.50**	**78.75**

* Bold values are the best results for the corresponding classes.

7.3.2 Segmentation Results on Multi-layer Dataset

Table 7.1 compares the performance of MSF-GAN and six compared methods on the multi-layer dataset in terms of IoU. Bold values are the best results for the corresponding classes. On this dataset, MSF-GAN achieves the best segmentation result for all classes in terms of IoU. Specifically, it achieves the highest mean IoU value of 78.75% among five classes, including fat layer, mammary layer, muscle layer, background, and tumor. It should be noticed that the proposed MSF-GAN outperforms SegAN, which is also a GAN-based network using a multi-scale L_1 loss, for all classes in terms of IoU.

Fig. 7.3 presents segmentation results of MSF-GAN and six compared methods for three representative BUS images in the multi-layer dataset. For image a1 in the top row containing a tumor and a small tumor-like region, all of six compared methods mistakenly segment the tumor-like region while MSF-GAN correctly segments the tumor region with the highest mean IoU value of 71.33%. For image a2 in the middle row containing no tumor, all six compared methods mistakenly segment a tumor region while MSF-GAN correctly generates a segmentation result without tumor. Among all methods, GAN-based networks SegAN (h2) and MSF-GAN (i2) outperform other non-GAN-based networks. The second-best method SegAN achieves a mean IoU value of 81.73%, while the MSF-GAN achieves the best mean IoU value of 87.81%. For image a3 in the bottom row containing an irregular tumor without a clear contour, for the tumor region, four non-GAN-based methods (c3 to g3) fail to give an accurate segmentation result, and two GAN-based networks (h3 and i3) give an accurate segmentation result close to groundtruth. SegAN fails to produce accurate segmentation results for other layers and has a mean IoU value of 60.40%. MG-GAN achieves the highest mean IoU value of 77.90%. As shown in Table 7.1 and Fig. 7.3, the proposed fuzzy attentive feature generator and multi-scale L_1 loss calculated on multi-scale uncertainty maps are efficient to enhance the discriminating ability of D-net and force G-net to generate more accurate segmentation results.

7.3.3 Segmentation Results on Two Public Datasets

Table 7.2 compares the performance of MSF-GAN and six state-of-the-art methods on

Dataset 1 and Dataset 2 in terms of TPR, FPR, IoU, DS, and AER. MSF-GAN has the highest TPR value of 84.57%, the highest IoU value of 73.30%, and the highest DS value of 81.58% on Dataset 1. MSF-GAN achieves the best performance in terms of IoU, FPR, DS, and AER and a comparable TPR value on Dataset 2. Specifically, it improves the second-best method by 2.69%, 2.68%, 19.15%, and 14.22% for IoU, DS, FPR, and AER, respectively. The experiment results show that the proposed MSF-GAN can perform well on public datasets. It contains high robustness and segmentation ability.

Table 7.2: Results on public datasets (%).

Datasets	Methods	TPR	FPR	IoU	DS	AER
Dataset 1 [2]	U-ResNet	83.58	**34.40**	71.43	79.45	**50.82**
	U-VGG	79.30	45.84	68.16	76.40	66.54
	FCN-8s	82.72	41.14	67.50	76.87	58.42
	SegAN	81.13	49.96	70.11	78.05	68.83
	PSPNet	81.08	40.42	69.77	78.24	59.34
	Deeplab	63.68	36.06	52.93	61.91	72.38
	MSF-GAN	**84.57**	40.31	**73.30**	**81.58**	55.73
Dataset 2 [3]	U-ResNet	**79.40**	46.02	69.26	77.90	66.62
	U-VGG	78.66	41.98	68.77	77.37	63.32
	FCN-8s	74.23	46.69	63.16	73.03	72.63
	SegAN	76.23	25.95	69.21	77.83	49.71
	PSPNet	77.11	46.65	65.21	74.75	69.54
	Deeplab	59.88	39.39	49.65	59.39	79.52
	MSF-GAN	78.34	**20.98**	**71.12**	**79.99**	**42.64**

7.3.4 Experiment Setup

To ensure a fair comparison, we set these parameters to be the same for all compared methods. All experiments are conducted on Ubuntu 18.04 system with Intel(R) Xeon(R) CPU E5-2620 2.00 GHz and two NVIDIA GeForce 1080Ti graphics cards with 11 Gigabyte memory. An Adam optimizer with learning rate = 0.0002, β_1 = 0.9, and β_2 = 0.99 is used for training. The batch size is set as 12, and the number of training epochs is set as 80. The initial weights are initialized randomly. Input images are augmented by horizontal flip, horizontal shift, vertical shift, rotation, zooming, and shear mapping before fed into networks. We employ 10-fold cross-validation to evaluate the performance of MSF-GAN and six compared methods.

7.4 Conclusion

In this chapter, we propose a novel MSF-GAN method for BUS image segmentation consisting of a generative network and a discriminative network. MSF-GAN employs a fuzzy attentive feature generator to extract fuzzy attentive feature maps respectively from segmentation maps generated by the generative network and from groundtruth maps and then uses an MSF module to extract multi-scale uncertainty maps from these fuzzy attentive feature maps to calculate a multi-scale L_1 loss that can capture the rich contextual relationship among pixels. By using the fuzzy attentive feature generator and the multi-scale L_1 loss calculated on uncertainty maps, the discriminating ability of the discriminative network is enhanced and can better guide the generative network to generate more accurate segmentation results. The proposed MSF-GAN outperforms six state-of-the-art deep neural network-based methods in terms of TPR, FPR, IoU, DS, and AER on three BUS datasets.

CHAPTER 8

CONCLUSION AND FUTURE WORK

Breast cancer is one of the most serious diseases affecting women's health all over the world. The incident rate of breast cancer keeps increasing in recent years. Many researchers focus on the early detection of breast cancer. Many medical imaging approaches are applied to clinical diagnosis, such as magnetic resonance imaging (MRI), X-ray, computed tomography (CT) imaging, and ultrasound imaging. Since the development of computer science technology and the Internet, there are many computer-aided diagnosis systems, especially the rapid growth of deep learning in recent years. The performance of computer-aided diagnosis systems increases significantly after applying deep learning.

Deep learning is essential to the field of computer vision and pattern recognition. Automatically encoding deep features is one of the most important reasons for the success of deep convolutional neural networks. There are a lot of studies on obtaining better convolutional features. The first one is increasing the depth of the convolutional neural networks. However, based on the backpropagation [77] training method, it is hard to train a network if the depth increases. A residual neural network (ResNet) [15] is developed, making the deeper network more trainable than before. Atrous Spatial Pyramid Pooling (ASPP) [17] and Pyramid Pooling Module (PPM) [18] used spatial multi-scale pooling operation to obtain multi-scale information. In [19], different scales of convolutional filters were utilized in the same convolutional block to get multi-scale information and enrich the information in each convolutional block. These researches hope to reflect the information of targets in different sizes in the same convolutional block. Attention mechanisms such as [20,21,34] can reduce uncertainty and provide context information in deep convolutional neural networks. However, deep learning approaches still have some shortcomings. If we apply deep learning algorithms to medical image analysis, there should be some prior medical knowledge that can help to increase the performance. Here are some shortcomings for deep learning algorithms: (1) existing attention mechanisms can only solve uncertainties caused by randomness; other uncertainty and noise are not discussed in the previous methods; (2) although there are researches related to providing context information in deep learning or classic machine learning methods, context information about breast cancer and breast ultrasound image is not involved; (3) context information from non-local pixels can be reflected by recurrent neural networks, and non-local operators; however, non-local information is not mentioned in convolutional operator; (4) current deep learning methods are like "black box" methods, which means we are hard to understand the convolutional features and why they are useful for final classification. In this dissertation, some of the shortcomings of deep learning are discussed, and it is applied to breast ultrasound image segmentation.

In Chapter 2, we firstly design a fuzzy block in deep convolutional neural networks, and we try to solve uncertainty and noise in convolutional neural networks by fuzzy logic. This research is inspired by previous studies of fuzzy logic in computer vision. However, previous fuzzy methods do not define uncertainty clearly. In this research, a trainable Sigmoid membership function and a trainable Gaussian membership function are utilized to transform the input feature map of the fuzzy block to the fuzzy domain. We define a pixel whose membership to a category is close to 0.5 as an uncertainty pixel. By this definition, we design an uncertainty mapping function. After measuring the uncertainty degree of each pixel, an attention mechanism is utilized to reduce the weight of uncertain pixels. The proposed fuzzy block contains the fuzzification part, uncertainty mapping part, and reducing uncertainty part. The fuzzy block is applied to the input image and the first convolutional feature map.

In Chapter 3, we continue the research on reducing uncertainty in convolutional feature maps. In this chapter, an improved fuzzy block is proposed. In the fuzzification part of the fuzzy block, a 1×1 convolutional operator with a Sigmoid activation function is utilized to replace the trainable Sigmoid and Gaussian membership functions in Chapter 2. In the uncertainty mapping part, fuzzy entropy is chosen to represent the uncertainty degree because fuzzy entropy has successfully represented uncertainty in previous research.

Meanwhile, not only the uncertainty for each pixel is measured, but also the uncertainty for each channel is measured. Therefore, spatial-wise and channel-wise fuzzy blocks are utilized in all five convolutional blocks of deep neural networks to reduce uncertainty.

In Chapter 4, new research on uncertainty reduction is conducted. Meanwhile, a novel context feature: the direction-connectedness (DC) feature, is designed to reflect breast horizontal layer information in deep neural networks. A pyramid fuzzy block is proposed to measure the uncertainty of objects in different scales by a pyramid structure in one fuzzy block. The input feature map of the fuzzy block is down-sampled to two different resolutions. The fuzzy operators designed in Chapter 3 are applied to the two new resolution feature maps and the original resolution feature map. Since one object has different sizes in different images, the pyramid structure can detect the uncertainty by using various resolutions of feature maps at one time. The proposed DC feature can provide breast horizontal layer structure by calculating the connectedness between one pixel and the boundary pixels in horizontal left, right, and vertical up, down directions. The DC feature can provide more information that can help to segment breast cancer well.

In Chapter 5, a deep learning + medical knowledge constrained CRFs architecture is proposed for BUS image segmentation. Breast anatomy is modeled as a constrained term in the energy function of CRFs. The initial segmentation results from a deep neural network are utilized to provide the breast anatomy information. Each pixel in the segmentation map is given a label vector. Values of the label vectors are specifically defined for different categories. The context information (the order of the breast layers) is provided during the energy function optimization. The medical knowledge constrained CRFs can refine the segmentation results from deep neural networks and involve breast anatomy information.

In Chapter 6, a shape-adaptive convolutional operator is proposed. Besides the breast horizontal layer structure and the order of breast layers provided in Chapter 4 and Chapter 5, novel context information from non-local pixels is involved in convolutional operators. The original 2-dimensional (2D) convolutional operator is a cross-correlation operator of the input feature map and convolutional kernel. The kernel is a square shape, which means it can merge information in the neighbor pixels of the target pixel. However, pixels that are not closed to the target pixel in Euclidean distance might still contain correlation. The novel shape-adaptive convolutional operator extends the original 2D convolutional operator to the general version. The pixels used in convolution are not neighbor pixels, and they are calculated through k nearest neighbor algorithm or self-attention mechanism, which means the shape of the kernel is not square shape. The novel convolutional kernel can provide contextual information based on the feature domain and self-attention coefficient domain. The proposed architecture outperforms six state-of-the-art deep learning methods on a 325-image dataset and 2 public datasets.

In Chapter 7, to further explore a novel segmentation approach for BUS image segmentation and investigate the ability of Generative Adversarial Network (GAN) in BUS image segmentation, a novel multi-scale fuzzy generative adversarial network (MSF-GAN) is proposed for BUS image segmentation. Two novel modules are designed and applied to a classic multi-scale loss GAN. The first module is a fuzzy feature generator, which takes the segmentation maps from the generative network + original input image or the groundtruth map + original input image as inputs. The fuzzy feature generator can transform the input information to the fuzzy domain and generate input for the adversarial network by the fuzzy logic operator. The second module is a multi-scale fuzzy (MSF) entropy module in the adversarial network, which distinguishes the uncertainty of fuzzy feature maps generated from segmentation maps and groundtruth maps from five convolutional layers using L_1 loss. The proposed method can measure and compare the uncertainty in the segmentation maps and groundtruth maps and help the generative network generate more accurate segmentation maps. The proposed network outperforms six state-of-the-art deep learning methods on three datasets.

We develop several approaches for BUS image segmentation, and we also provide new methods in deep learning which can reduce uncertainty and provide context information. However, we acknowledge that there are remaining challenges in BUS image CAD systems and deep learning: (1) only breast layer structure and breast anatomy are utilized as context information in our researches; there is other important context information in BUS image which can also improve the performance of CAD systems; (2) our researches are focus on BUS image segmentation; however, the classification of malignant and benign tumor [78,79] and BI-RADS [80,81] are more important because the final object of CAD systems is breast cancer diagnosis; they are not discussed in this dissertation; (3) the interpretability of deep learning such as convolutional feature visualization [82], and the correlation convolutional feature and classification [83] results are not mentioned. In the future, our research will be on (1) involving more context information using deep learning or classic machine learning methods, which can increase the current BUS image segmentation and classification methods; (2) developing other higher-order information (information from long distance pixels) extraction methods and explore the influence of higher-order information in classification decision; (3) conducting researches in the classification of malignant and benign tumor and BI-RADS classification which is the key to CAD systems; (4) exploring the interpretability of deep learning in BUS image segmentation and classification (such as the correlation between convolutional features and the classification decision) which can help to understand and improve deep learning approach; (5) trying to apply uncertainty and context information researchers to nature image processing and develop some general algorithms for computer vision.

REFERENCES

[1] 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)," *arXiv preprint arXiv:1801.03182*, 2018.

[2] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.

[3] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[4] B. Liu, H.-D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu, "Probability density difference-based active contour for ultrasound image segmentation," *Pattern Recognition*, vol. 43, no. 6, pp. 2028–2042, 2010.

[5] Y. Liu, H.-D. Cheng, J. Huang, Y. Zhang, and X. Tang, "An effective approach of lesion segmentation within the breast ultrasound image based on the cellular automata principle," *Journal of digital imaging*, vol. 25, no. 5, pp. 580–590, 2012.

[6] J. Shan, H. Cheng, and Y. Wang, "Completely automated segmentation approach for breast ultrasound images using multiple-domain features," *Ultrasound in medicine & biology*, vol. 38, no. 2, pp. 262–275, 2012.

[7] 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.

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

[9] American Cancer Society. (2019) How common is breast cancer. [Online]. Available: https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html

[10] C. E. DeSantis, S. A. Fedewa, A. Goding Sauer, J. L. Kramer, R. A. Smith, and A. Jemal, "Breast cancer statistics, 2015: Convergence of incidence rates between black and white women," *CA: a cancer journal for clinicians*, vol. 66, no. 1, pp. 31–42, 2016.

[11] Q. Huang, Y. Luo, and Q. Zhang, "Breast ultrasound image segmentation: a survey," *International journal of computer assisted radiology and surgery*, vol. 12, no. 3, pp. 493–507, 2017.

[12] C. F. Baumgartner, K. C. Tezcan, K. Chaitanya, A. M. Hötker, U. J. Muehlematter, K. Schawkat, A. S. Becker, O. Donati, and E. Konukoglu, "Phiseg: Capturing uncertainty in medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2019, pp. 119–127.

[13] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34–45, 2019.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[20] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[23] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning.* PMLR, 2015, pp. 2048–2057.

[24] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in

Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5659–5667.

[25] J. Gao, Q. Wang, and Y. Yuan, "Scar: Spatial-/channel-wise attention regression networks for crowd counting," *Neurocomputing*, vol. 363, pp. 1–8, 2019.

[26] H.-D. Cheng and H. Xu, "A novel fuzzy logic approach to contrast enhancement," *Pattern recognition*, vol. 33, no. 5, pp. 809–819, 2000.

[27] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *computerized medical imaging and graphics*, vol. 30, no. 1, pp. 9–15, 2006.

[28] P. Melin, C. I. Gonzalez, J. R. Castro, O. Mendoza, and O. Castillo, "Edge-detection method for image processing based on generalized type-2 fuzzy logic," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 6, pp. 1515–1525, 2014.

[29] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 1006–1012, 2016.

[30] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," *Advances in neural information processing systems*, vol. 24, pp. 109–117, 2011.

[31] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1529–1537.

[32] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1377–1385.

[33] Y. Zhuang, F. Yang, L. Tao, C. Ma, Z. Zhang, Y. Li, H. Jia, X. Xie, and W. Gao, "Dense relation network: Learning consistent and context-aware representation for semantic image segmentation," in *2018 25th IEEE international conference on image processing (ICIP)*. IEEE, 2018, pp. 3698–3702.

[34] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.

[35] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 603–612.

[36] X. Hu, L. Zhu, C.-W. Fu, J. Qin, and P.-A. Heng, "Direction-aware spatial context features for shadow detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7454–7462.

[37] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 764–773.

[38] 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.

[39] A. E. Ilesanmi, U. Chaumrattanakul, and S. S. Makhanov, "Methods for the segmentation and classification of breast ultrasound images: a review," *Journal of Ultrasound*, pp. 1–16, 2021.

[40] H. Fan, F. Meng, Y. Liu, F. Kong, J. Ma, and Z. Lv, "A novel breast ultrasound image automated segmentation algorithm based on seeded region growing integrating gradual equipartition threshold," *Multimedia Tools and Applications*, vol. 78, no. 19, pp. 27 915–27 932, 2019.

[41] Y. Bafna, K. Verma, L. Panigrahi, and S. P. Sahu, "Automated boundary detection of breast cancer in ultrasound images using watershed algorithm," in *Ambient communications and computer systems*. Springer, 2018, pp. 729–738.

[42] A. E. Ilesanmi, O. P. Idowu, and S. S. Makhanov, "Multiscale superpixel method for segmentation of breast ultrasound," *Computers in Biology and Medicine*, vol. 125, p. 103879, 2020.

[43] Q. Huang, F. Yang, L. Liu, and X. Li, "Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis," *Information Sciences*, vol. 314, pp. 293–310, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002002551400810X

[44] J. Cheng, D. Ni, Y. Chou, J. Qin, C. Tiu, Y. Chang, C. Huang, D. Shen, and C. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific Reports*, vol. 6, Apr. 2016, copyright: Copyright 2016 Elsevier B.V., All rights reserved.

[45] W. K. Moon, C.-M. Lo, R.-T. Chen, Y.-W. Shen, J. M. Chang, C.-S. Huang, J.-H. Chen, W.-W. Hsu, and R.-F. Chang, "Tumor detection in automated breast ultrasound images using quantitative tissue clustering," *Medical physics*, vol. 41, no. 4, p. 042901, 2014.

[46] 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.

[47] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[50] B. Shareef, M. Xian, and A. Vakanski, “Stan: Small tumor-aware network for breast ultrasound image segmentation,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1–5.

[51] B. Lei, S. Huang, R. Li, C. Bian, H. Li, Y.-H. Chou, and J.-Z. Cheng, “Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network,” *Neurocomputing*, vol. 321, pp. 178–186, 2018.

[52] B. Lei, S. Huang, H. Li, R. Li, C. Bian, Y.-H. Chou, J. Qin, P. Zhou, X. Gong, and J.-Z. Cheng, “Self-co-attention neural network for anatomy segmentation in whole breast ultrasound,” *Medical image analysis*, vol. 64, p. 101753, 2020.

[53] K. Huang, H.-D. Cheng, Y. Zhang, B. Zhang, P. Xing, and C. Ning, “Medical knowledge constrained semantic breast ultrasound image segmentation,” in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 1193–1198.

[54] H. Lee, J. Park, and J. Y. Hwang, “Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 7, pp. 1344–1353, 2020.

[55] D. Ramsay, J. Kent, R. Hartmann, and P. Hartmann, “Anatomy of the lactating human breast redefined with ultrasound imaging,” *Journal of anatomy*, vol. 206, no. 6, pp. 525–534, 2005.

[56] Y.-T. Kim, “Contrast enhancement using brightness preserving bi-histogram equalization,” *IEEE transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, 1997.

[57] A. Rodtook and S. S. Makhanov, “Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer,” *Journal of Visual Communication and Image Representation*, vol. 24, no. 8, pp. 1414–1430, 2013.

[58] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[59] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[60] S. Al-Sharhan, F. Karray, W. Gueaieb, and O. Basir, “Fuzzy entropy: a brief survey,” in *10th IEEE International Conference on Fuzzy Systems.(Cat. No. 01CH37297)*, vol. 3. IEEE, 2001, pp. 1135–1139.

[61] O. A. Kittaneh, M. A. U. Khan, M. Akbar, and H. A. Bayoud, “Average entropy: A new uncertainty measure with application to image segmentation,” *The American Statistician*, vol. 70, no. 1, pp. 18–24, 2016. [Online]. Available: https://doi.org/10.1080/00031305.2015.1089788

[62] H. Maassen and J. B. Uffink, "Generalized entropic uncertainty relations," *Physical Review Letters*, vol. 60, no. 12, p. 1103, 1988.

[63] S. Wehner and A. Winter, "Entropic uncertainty relations—a survey," *New Journal of Physics*, vol. 12, no. 2, p. 025009, 2010.

[64] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[65] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade.* Springer, 2012, pp. 9–48.

[66] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[67] 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.

[68] K. C. Ciesielski, P. A. Miranda, A. X. Falcão, and J. K. Udupa, "Joint graph cut and relative fuzzy connectedness image segmentation algorithm," *Medical image analysis*, vol. 17, no. 8, pp. 1046–1057, 2013.

[69] J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *Graphical models and image processing*, vol. 58, no. 3, pp. 246–261, 1996.

[70] M. Alice. (2018) Breast ultrasound cancer vs benign. [Online]. Available: https://canceroz.blogspot.com/2017/02/breast-ultrasound-cancer-vs-benign.html

[71] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[72] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao, "Dynamic hypergraph neural networks." in *IJCAI*, 2019, pp. 2635–2641.

[73] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[74] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint arXiv:1611.08408*, 2016.

[75] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3, pp. 383–392, 2018.

[76] L. Han, Y. Huang, H. Dou, S. Wang, S. Ahamad, H. Luo, Q. Liu, J. Fan, and J. Zhang, "Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network," *Computer methods and programs in biomedicine*, vol. 189, p. 105275, 2020.

[77] A. Van Ooyen and B. Nienhuis, "Improving the convergence of the back-propagation algorithm," *Neural networks*, vol. 5, no. 3, pp. 465–471, 1992.

[78] W.-C. Shia and D.-R. Chen, "Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine," *Computerized Medical Imaging and Graphics*, vol. 87, p. 101829, 2021.

[79] D.-R. Chen, R.-F. Chang, C.-J. Chen, M.-F. Ho, S.-J. Kuo, S.-T. Chen, S.-J. Hung, and W. K. Moon, "Classification of breast ultrasound images using fractal feature," *Clinical imaging*, vol. 29, no. 4, pp. 235–245, 2005.

[80] C. Balleyguier, S. Ayadi, K. Van Nguyen, D. Vanel, C. Dromain, and R. Sigal, "Birads™ classification in mammography," *European journal of radiology*, vol. 61, no. 2, pp. 192–194, 2007.

[81] S. Obenauer, K. Hermann, and E. Grabbe, "Applications and literature review of the bi-rads classification," *European radiology*, vol. 15, no. 5, pp. 1027–1036, 2005.

[82] Q.-s. Zhang and S.-C. Zhu, "Visual interpretability for deep learning: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 27–39, 2018.

[83] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao *et al.*, "Interpretability of deep learning models: a survey of results," in *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation.* IEEE, 2017, pp. 1–6.

CURRICULUM VITAE

Kuan Huang

435-512-8519 | kuan.huang0624@gmail.com | linkedin.com/in/kuan-huang

Personal Statement

I am a Ph.D. candidate in the Department of Computer Science at Utah State University (USU). I obtained my bachelor's degree from Harbin Institute of Technology (HIT). My major research field is developing novel semantic segmentation methods to segment breast ultrasound (BUS) images.

Research Interests

- Computer Vision
- Deep Learning
- Pattern Recognition
- Medical Image Analysis
- Fuzzy Logic

Education

- Ph.D., Computer Science, Utah State University, Logan, Utah, USA, Adviser: Heng-Da Cheng, Professor, May 2021 (GPA: 3.98/4.0).
- B.Eng., Measurement Control Technique & Instruments, Harbin Institute of Technology, Harbin, China, July 2016 (GPA: 3.78/4.0).

Professional Work Experience

- Internship: Harbin Electric Corporation, Harbin, China, Spring 2015.
- Teaching Assistant: Utah State University, Logan, USA, 2016-2021.

Research Experience

My research interests focus on improving deep learning algorithms and applying deep learning approaches to breast ultrasound (BUS) image segmentation:

- **Fuzzy deep convolutional neural network:** Uncertainty and noise were reduced by fuzzy logic in position and channel dimension. The feature map was resized to different resolutions to reduce uncertainty by a pyramid fuzzy uncertainty reduction block. The proposed methods achieved improvements compared with state-of-the-art methods on a dataset with 325 images and a public dataset.
- **Involving breast anatomy in segmentation:** (1) The theory of connectedness and recurrent neural network were used to compute the connectedness between each pixel and boundary pixels in the horizontal and vertical directions, which was defined as direction-connectedness (DC) feature. The DC feature could reflect the breast anatomy. DC feature was added to our fuzzy deep learning architecture for BUS image and outperformed eight state-of-the-art deep learning methods. (2) Medical knowledge constrained CRFs were proposed, which applied the BUS image layer structure as the new pairwise term in the energy function of CRFs.
- **Shape-adaptive convolutional operator:** A novel convolutional operator whose shape was decided based on selection of convolutional pixels was proposed. The selection was based on feature values or self-attention coefficients. The novel convolutional operator could involve higher order information rather than in Euclidean domain.
- **Future work:** The loss function will be controlled by the fuzzy membership calculated in the convolutional layers. Moreover, novel approaches to select convolutional

pixels in a shape-adaptive convolutional operator will be studied. The experiments will be extended from biomedical images to nature images.

Services

- Served as a reviewer for Journal of Computer Science and Technology, Medical Image Analysis, Neural Computing and Applications, Neurocomputing, IEEE Winter Conference on Applications of Computer Vision, IEEE International Conference on Pattern Recognition, and IEEE International Conference on Image Processing.
- Served as a volunteer for Freescale Cup National Undergraduate Smart Car Contest: 2013 Summer.
- Being an IEEE graduate student member.
- Met with tenure-track candidates during their on-site interviews at Utah State University.

Honors and Award

- Travel grand of the 24th International Conference on Pattern Recognition, Beijing, 2018
- Outstanding Individual Student Scholarship, School of Electrical Engineering, HIT, 2014
- Excellent Students of Harbin Institute of Technology, HIT, 2014
- Excellent Students of Harbin Institute of Technology, HIT, 2013
- Third prize of Chinese Mathematics Competitions, Heilongjiang Province, 2013
- Excellent Volunteer in Freescale Cup National College Students Intelligent, HIT, 2013
- Second prize in Project Design for the freshman, School of Electrical Engineering, HIT, 2013

- Outstanding Individual Student Scholarship, School of Electrical Engineering, HIT, 2013
- Third prize in Adolescents Science and Technology Innovation Contest, China, 2012

Publications

1. **Published Papers:**

   - [*J*] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, B. Zhang, Semantic Segmentation of Breast Ultrasound Image with Fuzzy Deep Learning Network and Breast Anatomy Constraints, accepted by Neurocomputing (Impact Factor = 4.438), 2021.
   - [*C*] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, Shape-Adaptive Convolutional Operator for Breast Ultrasound Image Segmentation, accepted by IEEE International Conference on Multimedia and Expo 2021 (ICME 2021).
   - [*C*] M. Xu, K. Huang, Q. Chen, X. Qi, MSSA-Net: Multi-scale Self-attention Network for Breast Ultrasound Image Segmentation, accepted by IEEE International Symposium on Biomedical Imaging 2021 (ISBI 2021).
   - [*C*] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, B. Zhang, Semantic Segmentation of Breast Ultrasound Image with Pyramid Fuzzy Uncertainty Reduction and Direction Connectedness Feature, in 2020 25th International Conference on Pattern Recognition (ICPR 2020).
   - [*J*] Q. Yu, K. Huang, Y. Zhu, X. Chen, W. Meng, Preliminary results of computer-aided diagnosis for magnetic resonance imaging of solid breast lesions, in Breast Cancer Research and Treatment (Impact Factor = 3.831), 2019.
   - [*C*] K. Huang, H. D. Cheng, Y. Zhang, B. Zhang, P. Xing, C. Ning, Medical Knowledge Constrained Semantic Breast Ultrasound Image Segmentation, in 2018 24th International Conference on Pattern Recognition (ICPR 2018).

- [*C*] 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 International Conference on Pattern Recognition (ICPR 2018).

2. **Papers Under Review:**

- [*C*] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, MSF-GAN: Multi-Scale Fuzzy Generative Adversarial Network for Breast Ultrasound Image Segmentation, submitted to the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2021).
- [*C*] K. Huang, M. Xu, X. Qi, NGMMs: Neutrosophic Gaussian Mixture Models for Breast Ultrasound Image Classification, submitted to the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2021).
- [*J*] K. Huang, Y. Zhang, H. D. Cheng, P. Xing, SCFURNet: Spatial and Channel-wise Fuzzy Uncertainty Reduction Network for Breast Ultrasound Image Semantic Segmentation, submitted to Medical Image Analysis (Impact Factor = 11.148).
- [*J*] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, K. Huang, B. Zhang, J. Ding, C. Ning, Y. Wang, submitted to IEEE Transactions on Medical Imaging.
- [*J*] B. Zhang, Y. Zhang, H. D. Cheng, M. Xian, S. Gai, O. Cheng, K. Huang, Computer-Aided Knee Joint Magnetic Resonance Image Segmentation-A Survey, submitted to IEEE Transactions on Biomedical Engineering (TBME).