Embedding Weighted Feature Aggregation Network with Domain Knowledge Integration for Breast Ultrasound Image Segmentation

Yuxi Liu, Xing An, Longfei Cong, Guohao Dong, and Lei Zhu[✉]

Shenzhen Mindray BioMedical Electronics, Co., Ltd., Shenzhen, China
zhulei@mindray.com

Abstract. Breast cancer is the most common cancer in women, and ultrasound imaging is one of the most widely used approaches for diagnosis due to its non-radioactive process, ease of operation and low cost. Moreover, image segmentation plays a great role in medical image analysis, since it affects the accuracy of computer aided diagnosis (CAD) results. However, the malignant mass of breast in ultrasound images often appears irregular boundary and indistinct margin which is difficult to distinguish from other surrounding tissues. Therefore, breast ultrasound images segmentation is significant for diagnosis, and it has attracted the attention of researchers for many years. In this paper, we propose a weighted feature aggregation network with fusing domain knowledge for two-dimensional breast ultrasound images segmentation. (I) We modify the U-Net by adding a classification branch, in which BI-RADS category information is applied as the classification label. (II) In order to deal with the artifacts in ultrasound, such as posterior shadowing, we conduct Squeeze-and-Excitation (SE) block and aggregation mechanism to compose the up-sampling part in U-Net. (III) We employ the conditional random field (CRF) to optimize segmentation to make the boundaries more continuous and integral after getting the output of U-Net. The experiment conducted on a challenging and representative dataset includes more than three thousand two-dimensional breast ultrasound images. Our method achieves Jaccard Index of 84.9%, Matthew correlation coefficient of 90.9%, and Dice Coefficient of 90.8% in testing which demonstrates the potential clinical value of our work.

Keywords: Segmentation · Breast ultrasound images · Convolutional Neural Network

1 Introduction

Breast cancer is the most common cancer in women, and has a very high mortality rate around the world. Early detection and diagnosis are of great significance for breast cancer treatment, which is beneficial to raise the survival rate. For breast cancer, ultrasound imaging is one of the most efficient and widely used diagnostic methods due to its non-radioactive process, ease of operation and low cost. Additionally, the Breast Imaging Reporting and Data System (BI-RADS) [1] provides standard terminology to describe

breast mass as well as classification system for ultrasound. Moreover, it has been proved that BI-RADS is helpful to doctor's diagnosis and the following therapeutic plan.

Image segmentation is an important procedure in research and clinical practice. Accurate segmentation result benefits a lot to the tasks such as diagnosis and treatment planning. Recently, much attention has been paid to the field of deep learning based segmentation methods. For example, the aggregation methods are employed to boost information flow in proposal-based instance segmentation framework [2]. A fast scanning deep convolutional neural network is proposed to segment the breast tumor region in histopathological images [3]. Moreover, a number of research efforts have been devoted to breast ultrasound images segmentation so as to further improve the performance of computer aided diagnosis (CAD) system for breast cancer. We can broadly divide these works into three major types. Firstly, using convolutional neural network such as FCN and U-Net to segment the mass in breast ultrasound images directly [4]. Secondly, using traditional image processing techniques, i.e. thresholding, clustering and active contour model, to segment breast mass in ultrasound images [5, 6]. Thirdly, this type of method emphasizes on integrating domain knowledge to further improve the accuracy of results with Convolutional Neural Networks (CNN) or traditional methods [7].

It should be noticed that the diagnosis results of CAD are highly correlated with the accuracy of breast mass segmentation performance. However, segmentation for breast ultrasound images is still a challenging problem due to the following three reasons. (I) Recent works on CNN based segmentation methods usually consider breast cancer as benign and malignant, whereas it should be classified into four major types (BI-RADS 2, 3, 4 and 5). In other words, domain knowledge of clinical diagnosis in BI-RADS is not fully used and integrated. Another reason for using BI-RADS category instead of benign and malignant is that the lesions' grade lower than 4a are not recommended for biopsy so that there have no pathological diagnosis results for some cases. (II) Artifacts in breast ultrasound images misleads the algorithm in finding a real mass, especially for malignant mass with a shadow behind the posterior border. (III) The obtained mask of CNN always get a rough border, which is not exact enough to characterize the specific local representation of malignancy, such as microlobulated and spiculated.

In view of these issues, we proposed a U-Net [8] based network to segment breast mass accurately and effectively in ultrasound images. The main contributions of our work are three-fold. (I) A classification branch is added in U-Net with integrated domain knowledge to supervise the detection and segmentation of mass. (II) An embedding weighted aggregation module is introduced to fuse the multi-scale attention information in decoding layers, in order to improve the segmentation performance of malignant mass. (III) A fully connected conditional random field (CRF) module is appended at the end of network, which will further increase the segmentation accuracy of mass with indistinct boundaries.

2   Method

The architecture of the proposed framework is illustrated in Fig. 1. The proposed network is primarily based on U-Net, incorporated with a classification branch to integrate the clinical diagnosis knowledge. In the last fewer layers of the network, weighted feature aggregation module and CRF module are embedded. With the above methods, the

presented work can not only increase the accuracy of mass segmentation with regular boundary and distinct margin, but also improve the performance to detect malignant mass with artifacts in ultrasound images.
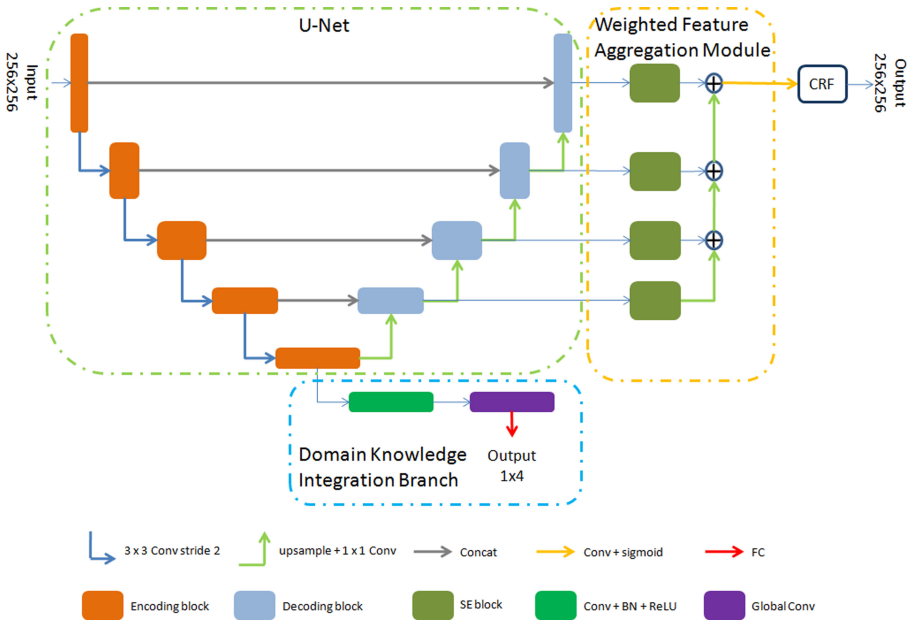

Fig. 1. Illustration of the proposed network, which includes a U-net based encoder-decoder part, a domain knowledge integration branch, a weighted feature aggregation module and a CRF module. 'Conv', 'BN', 'FC', and '+' denotes the convolutional, batch normalization, fully connected layers and addition operation, respectively.

2.1 Domain Knowledge Integration Branch

Inspired by multi-task leaning strategies [9], classification branch is introduced to the U-Net. The generalization ability of breast mass segmentation could be improved by adding a joint learning branch. It is worth noting that BI-RADS divide the mass in breast ultrasound images into several grades according to its possibility of malignancy. Different levels of mass have divergence on the symptom of boundary. Generally, the lower possibility of malignancy the mass is, the more regular and smoother the boundary will be. Therefore, to integrate this domain knowledge, we add a classification branch after the final convolution layer of the top-down path way to predict the BI-RADS category of the mass, instead of using the label of benign and malignant in [7]. As illustrated in Fig. 1, for this classification branch, the inputs go through a stack of 3×3 Conv + BatchNorm + ReLU layers, and then the global convolutional layer generates $1 \times 1 \times C$ feature vector. Finally, a fully connected layer with softmax activation is applied to get the probability for classification, and the softmax cross entropy is used as

the loss function. With the BI-RADS information as the supervised label, breast mass detection and segmentation can achieve a better performance.

2.2 Weighted Feature Aggregation Module

Recently, Feature Pyramid Network (FPN) [10] is one of the most used approaches to solve the multi-scale problem in object detection and segmentation. Usually, feature maps of the same scale are summed up along the channel dimension in FPN module. However, not all features in high-level layers are effective for locating the objects. Moreover, artifacts in ultrasound images cause the boundary of mass unclear or invisible such as posterior shadowing where the area posterior to the mass appears darker. As a result, it is difficult for the CNN network even the clinical experts to find the correct edge of the malignant mass. A novel architectural unit called as Squeeze-and-Excitation (SE) block, which can distinguish the importance of different channels of the neural network is introduced in [11]. Inspired by this view, we propose a weighted feature aggregation module to extract and aggregate features and information from multi-scale layers to optimize the performance of mass segmentation. As illustrated in Fig. 1, output of the last four convolutional layers in the decoding path are fed into the SE block to extract the important information of each layer. The details of SE block can be seen in Fig. 2 and the reduction ratio r is set to be 16. The outputs of SE block in each layer are then passed to a 2×2 upsampling layer and a 3×3 convolutional layer with ReLU activation to keep the same dimension of output image as the feature maps in next stage. Finally, all feature maps are summed up along the channel dimension to form the final inputs of the last step. In short, our aggregation module can extracts useful information and combines it with features in each layer through multi-scale information fusion. Thus, region of the mass and its edge in the final feature maps are emphasized in the network.
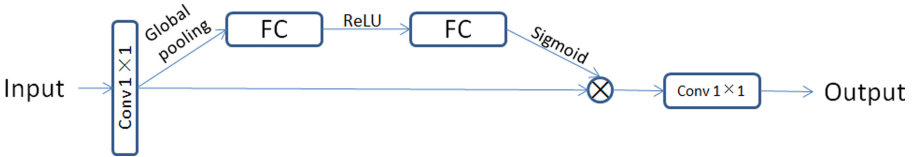

Fig. 2. Illustration of the SE block. '$\times$' denotes multiplication operation.

Finally, a 1×1 convolutional layer followed by sigmoid activation is applied to generate the output. Instead of only using cross-entropy loss for segmentation, we jointly optimize the Dice loss and the cross entropy loss for segmentation task during the training stage, which is defined as:

$$\mathcal{L}_{\text{seg}} = \lambda_1 \mathcal{L}_{\text{Dice}} + \lambda_2 \mathcal{L}_{\text{CE}} \tag{1}$$

where λ_1 and λ_2 are the weights of $\mathcal{L}_{\text{Dice}}$ and $\mathcal{L}_{\text{CE}}$, satisfied with $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$ in this paper. The cross-entropy loss function penalizes pixel classification errors while the dice loss function measures the overlap between the predicted areas and the ground

truth. Finally, the training loss of the whole network is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \alpha \mathcal{L}_{\text{seg}} \tag{2}$$

where α is the hyper-parameter balancing $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{seg}}$. In our experiment, weight α is set to 1.

2.3 CRF Refine Module

In fact, it is common that the malignant mass and tissues around it have similar appearance in breast ultrasound images. That is to say, the margin of the malignant mass often appears indistinct, hence it leads to a decrease in accuracy of the output mask of the mass generated by the network directly. To address this problem, we add a fully connected CRF module at the end of our proposed network. This can improve the continuity and integrity of the contour of the mass by allowing spatial constrains between different objects. Given the probability map from the U-Net and the same size of input ultrasound images, we formulate the final results as the inference from CRF model. The energy function of our CRF model is defined as:

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j) \tag{3}$$

where $\psi_u(x_i)$ is the unary potential term which computed independently for each pixel by a classifier that produces a distribution over the label assignment x_i. And $\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} w^{(m)} k^{(m)}(f_i, f_j)$ is the pairwise energy term measuring likelihood of the neighboring pixel pair where $k^{(m)}(f_i, f_j)$ is a Gaussian kernel, f_i, f_j is the feature vectors for pixel i and j, $w^{(m)}$ are linear combination weights, and $\mu(x_i, x_j)$ is the label compatibility function. We choose the label x_i as our final label and minimize the energy function with 5 iterations based on the mean field approximation algorithm [12].

3 Experiments

Datasets. We conducted experiments on 3341 two-dimensional breast ultrasound images which collected from different hospitals using Mindray Resona 7 Ultrasound Imaging System (Mindray, Shenzhen, China). All the data is reviewed by several experienced ultrasonic physicians and the final diagnosis is obtained by majority voting. There is at least one mass in each image in the dataset. The dataset is divided into six categories, i.e. category 2, category 3, category 4A, category 4B, category 4C, and category5, according to the BI-RADS guideline. There are 702, 883, 358, 356, 291, 753 data for each category, respectively. The data among grades 4A, 4B, and 4C are similar to each other in terms of texture and shape. In most cases, it is not easy to be distinguished even for clinical experts. Moreover, there has a critical imbalance problem of the data, and will cause the divergence in the training stage. Concerning the above problems, we separate the dataset into 4 categories, i.e. category 2, category 3, category 4, and category 5, and randomly split it as training and testing sets in the proportion of 80% and 20%, respectively.

Implementation Details. We automatically cropped all the images with Otsu's thresholding method and only remained the image content, in order to remove the useless regions such as background, probe information, and imaging parameters. Then data augmentations including rotation, shifting, cropping, zooming, and flipping were employed. And the input images were resized to 256×256. The initial weight of the backbone network in our proposed method is Resnet-50 [13], which was pre-trained on ImageNet [14], and the parameters of other layers were randomly initialized. The whole frame work was trained on a NVIDIA Titan Xp GPU with batch size of 16. Adam optimizer with a momentum of 0.9 and a weight decay of 0.001 were used to optimize our models. We trained the network for 100 epochs and stop when the validation loss does not decrease significantly and it took approximate 10 h to train the network on our breast dataset.

Evaluation Metrics. We choose U-Net as the baseline network and adopt Jaccard Index, Matthew correlation coefficient (Mcc), and Dice coefficient for quantitative evaluation. These three metrics are defined as

$$\text{Jaccard Index} = \frac{\text{TP}}{\text{FP} + \text{FN} + \text{TP}} \tag{4}$$

$$\text{Mcc} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt[2]{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \tag{5}$$

$$\text{Dice coefficient} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{6}$$

where TP refers to true positives, FP refers to false positives, TN refers to true negatives, and FN refers to false negatives.

Quantitative Analysis. We report the segmentation results using the evaluation metrics in Table 1. Our model outperforms Mask R-CNN [15] and U-Net in all three evaluation metrics for breast ultrasound images segmentation task. We have trained and tested the model for three times with the same hyper-parameters to eliminate the influence of random factors. Moreover, Fig. 3 presents qualitative results of different methods on five ultrasound images. As shown in the figure, all methods have a great performance on the mass with a smooth and regular boundary in the first row. As for the mass has an irregular and indistinct border, from the second to fourth rows in Fig. 3, segmentation results of Mask R-CNN and U-Net have some parts of missing or over-segmentation in the indistinct areas. The first two methods cannot detect the small mass in the last row while our method detects and segments the mass accurately.

Ablation Study. We conducted a set of ablation experiments to evaluate the contributions and effectiveness of each component of the proposed methods: (i) U-Net (Baseline), (ii) U-Net equipped with domain knowledge branch, (iii) U-Net with domain knowledge branch and aggregation module, (iv) U-Net with aggregation and CRF modules, (v) our proposed method. The results are shown in Table 2. We choose U-Net as the baseline network, which achieves 80.29%, 86.89%, and 86.71% in terms of Jaccard Index, Mcc, and Dice Index, respectively. When we append the proposed domain knowledge

Table 1. Segmentation performance comparison on breast dataset

Method	Jaccard (%)	Mcc (%)	Dice (%)
Mask R-CNN [15]	79.6	86.2	86
U-Net [8]	80.3	86.9	86.7
Ours	**84.9**	**90.9**	**90.8**
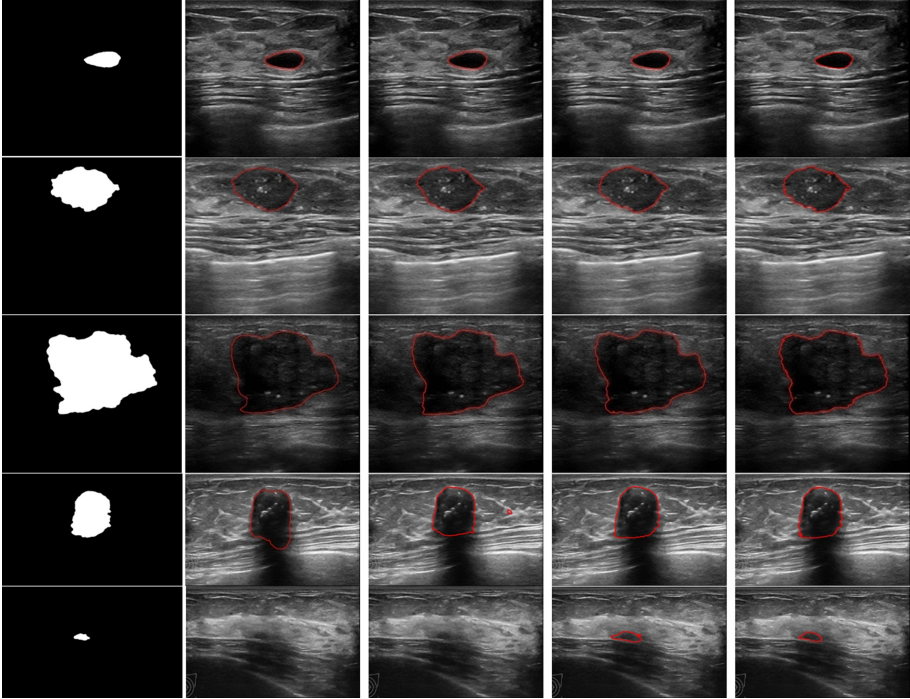
Fig. 3. Qualitative results of segmentation results. From left to right: Ground Truth, Mask R-CNN, U-Net, U-Net + Domain Knowledge and Ours.

integration, it yields results of 81.64%, 88.50%, and 88.59% (Jaccard, Mcc, and Dice). With the weighted aggregation module appended, it brings 0.86%, 0.74%, and 0.49% improvement for Jaccard, Mcc and Dice index. Furthermore, when we adopt all three methods, it significantly improves the performance by 4.56%, 4.04%, and 4.15% for the evaluation metrics compared with the baseline network, which shows the effectiveness of the proposed method.

Table 2. Ablation studies on our network measured by Jaccard Index, Matthew Correlation Coefficient (Mcc) and Dice coefficient.

	Method			Evaluation Metrics		
	DK	Aggregate	CRF	Jaccard (%)	Mcc (%)	Dice (%)
U-Net				80.3	86.9	86.7
	✓			81.6	88.5	88.6
	✓	✓		82.5	89.2	89.1
	✓		✓	83.7	90.0	89.9
		✓	✓	83.4	89.8	89.6
	✓	✓	✓	**84.9**	**90.9**	**90.8**

4 Conclusion

In this paper, we proposed a U-Net based approach for the challenging task of breast mass segmentation on breast ultrasound images. The proposed method takes advantage of both domain knowledge integration and weighted feature aggregation, and obtains an improved performance on malignant mass segmentation. The experiment results demonstrate that our techniques can tackle the segmentation problems of irregular boundary, indistinct margin, and posterior shadowing in breast ultrasound images. Our method provides a fast and accurate ultrasound images processing tool and can be applied to other instance segmentation task in medical field. In the future, the specific BI-RADS information and biopsy results will be investigated and utilized to fine-tune the network when more data is collected.

References

1. D'Orsi, C., et al.: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System (2013)
2. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: IEEE CVPR (2018)
3. Su, H., et al.: Region segmentation in histopathological breast cancer images using deep convolutional neural network. In: IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 55–58 (2015)
4. Yap, M.H., Goyal, M., Osman, F., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inform., 1 (2017). https://doi.org/10.1109/jbhi.2017.2731873
5. Moon, W.K., et al.: Tumor detection in automated breast ultrasound images using quantitative tissue clustering. Med Phys. **41**(4), 042901 (2014)
6. Moraru, L., Moldovanu, S., Biswas, A.: Optimization of breast lesion segmentation in texture feature space approach. Med. Eng. Phys. **36**(1), 129–135 (2014)
7. Cao, Z., et al.: Breast tumor detection in ultrasound images using deep learning. In: Wu, G., Munsell, B.C., Zhan, Y., Bai, W., Sanroma, G., Coupé, P. (eds.) Patch-MI 2017. LNCS, vol. 10530, pp. 121–128. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67434-6_14

8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

9. Caruana, R.: Multitask learning. Mach. Learn. **28**(1), 41–75 (1997)

10. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: IEEE CVPR, pp. 2117–2125 (2017)

11. Hu, J., Shen, L., Albanie, S.: Squeeze-and-excitation networks, pp. 7132–7141 (2018). https://doi.org/10.1109/cvpr.2018.00745

12. Zheng, S., Jayasumana, S., Romera-Paredes, B., et al.: Conditional random fields as recurrent neural networks. In: ICCV, pp. 1529–1537 (2015)

13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)

14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

15. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: 2017 ICCV, pp. 2980–2988 (2017)