Developing Large Pre-trained Model for Breast Tumor Segmentation from Ultrasound Images

Meiyu Li[1], Kaicong Sun[1], Yuning Gu[1], Kai Zhang[1], Yiqun Sun[1], Zhenhui Li[2], and Dinggang Shen[1,3(✉)]

[1] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
dgshen@shanghaitech.edu.cn
[2] Department of Radiology, Yunnan Cancer Hospital, Kunming, China
[3] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. Early detection and diagnosis of breast cancer using ultrasound images are crucial for timely diagnostic decision and treatment in clinical application. However, the similarity between tumors and background and also severe shadow noises in ultrasound images make accurate segmentation of breast tumor challenging. In this paper, we propose a large pre-trained model for breast tumor segmentation, with robust performance when applied to new datasets. Specifically, our model is built upon UNet backbone with deep supervision for each stage of the decoder. Besides using Dice score, we also design discriminator-based loss on each stage of the decoder to penalize the distribution dissimilarity from multi-scales. Our proposed model is validated on a large clinical dataset with more than 10000 cases, and shows significant improvement than other representative models. Besides, we apply our large pretrained model to two public datasets without fine tuning, and obtain extremely good results. This indicates great generalizability of our large pre-trained model, as well as robustness to multi-site data. The code is publicly available at https://github.com/limy-ulab/US-SEG.

Keywords: Large-scale clinical dataset · Deep-supervision · Multi-scale segmentation · Breast ultrasound images

Registration Number: 4319

1 Introduction

Breast cancer is a serious health problem with high incidence and wide prevalence for women throughout the world [1, 2]. Regular screening and early detection are crucial for effective diagnosis and treatment, and hence for improved prognosis and survival rate [3, 4]. Clinical researches have shown that ultrasound imaging is an effective tool for screening breast cancer, due to its critical characteristics of non-invasiveness, non-radiation and inexpensiveness [5–7]. In clinical diagnosis, delineating tumor regions from background is a crucial step for quantitative analysis [8]. Manual delineation always depends on the experience of radiologists, which tends to be subjective and time-consuming [9, 10].

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14226, pp. 89–96, 2023.
https://doi.org/10.1007/978-3-031-43990-2_9

Therefore, there is a high demand for automatic and robust methods to achieve accurate breast tumor segmentation. However, due to speckle noise and shadows in ultrasound images, breast tumor boundaries tend to be blurry and are difficult to be distinguished from background. Furthermore, the boundary and size of breast tumors are always variable and irregular [11, 12]. These issues pose challenges and difficulties for accurate breast tumor segmentation in ultrasound images.

Various approaches based on deep learning have been developed for tumor segmentation with promising results [13–19]. Su et al. [13] designed a multi-scale U-Net to extract more semantic and diverse features for medical image segmentation, using multiple convolution sequences and convolution kernels with different receptive fields. Zhou et al. [14] raised a deeply-supervised encoder-decoder network, which is connected through a series of nested and dense skip pathways to reduce semantic gap between feature maps. In [15], a multi-scale selection and multi-channel fusion segmentation model was built, which gathers global information from multiple receptive fields and integrates multi-level features from different network positions for accurate pancreas segmentation. Oktay et al. [16] proposed an attention gate model, which is capable of suppressing irrelevant regions while highlighting useful features for a specific task. Huang et al. [17] introduced a UNet 3+ for medical image segmentation, which incorporates low-level and high-level feature maps in different scales and learns full-scale aggregated feature representations. Liu et al. [18] established a convolution neural network optimized by super-pixel and support vector machine, segmenting multiple organs from CT scans to assist physicians diagnosis. Pei et al. [19] introduced channel and position attention module into deep learning neural network to obtain contextual information for colorectal tumors segmentation in CT scans. However, although these proposed models have achieved satisfactory results in different medical segmentation tasks, their performances are limited for breast tumor segmentation in ultrasound images due to the low image contrast and blurry tissue boundary.

To address these challenges, we present, to the best of our knowledge, the first work to adopt multi-scale features collected from large set of clinical ultrasound images for breast tumor segmentation. The main contributions of our work are as follows: (1) we propose a well-pruned simple but effective network for breast tumor segmentation, which shows remarkable and solid performance on large clinical dataset; (2) our large pretrained model is evaluated on two additional public datasets without fine-tuning and shows extremely stabilized improvement, indicating that our model has outstanding generalizability and good robustness against multi-site data data.

2 Method

We demonstrate the architecture of our proposed network in Fig. 1. It is based on the UNet [20] backbone. In order to collect multi-scale rich information for tumor tissues, we propose to use GAN [21]-based multi-scale deep supervision. In particular, we apply similarity constraint for each stage of the UNet decoder to obtain consistent and stable segmentation maps. Instead of using Dice score in the final layer of UNet, we also use Dice loss on each of the decoder stages. Besides, we integrate an adversarial loss as additional constraint to penalize the distribution dissimilarity between the predicted

segmentation map and the ground truth. In the framework of GAN, we take our segmentation network as the generator and a convolutional neural network as the discriminator. The discriminator consists of five convolution layers with the kernel sizes of 7×7, 5×5, 4×4, 4×4 and 4×4. Therefore, we formulate the overall loss for the generator, namely the segmentation network, as

$$L_{overall} = \sum_{n=1}^{4} \alpha_n \cdot (1 - 2\left|P^{(n)} \cap G\right| \cdot \left(\left|P^{(n)}\right| + |G|\right)^{-1}) + \beta_n \cdot E_{P^{(n)} \sim P_{SN}(proba)} \log(1 - CN_{\theta_{CN}^n}(SN_{\theta_{SN}}(P^{(n)}))) \quad (1)$$

where *SN* represents the segmentation network and *CN* represents the involved convolutional network. θ_{SN} and θ_{CN} refer to the parameters in the segmentation and convolutional network, respectively. $P^{(n)}$ represents the segmentation maps obtained from the n-th stage in the segmentation network, and *G* refers to the corresponding ground truth. $p_{SN}(proba)$ denotes the distribution of probability maps. $CN_{\theta CN}(SN_{\theta SN}(P^{(2)}))$ represents the probability for the input of CN coming from the predicted maps rather than the real ones. The parameters α_1 is empirically set as 1, α_2, α_3, α_4 are set as 0.1, and β_1, β_2, β_3 and β_4 are set as 0.05. It should be noted that, in UNet, there are 4 stages and hence we employ 4 CNNs for each of them without sharing their weights.

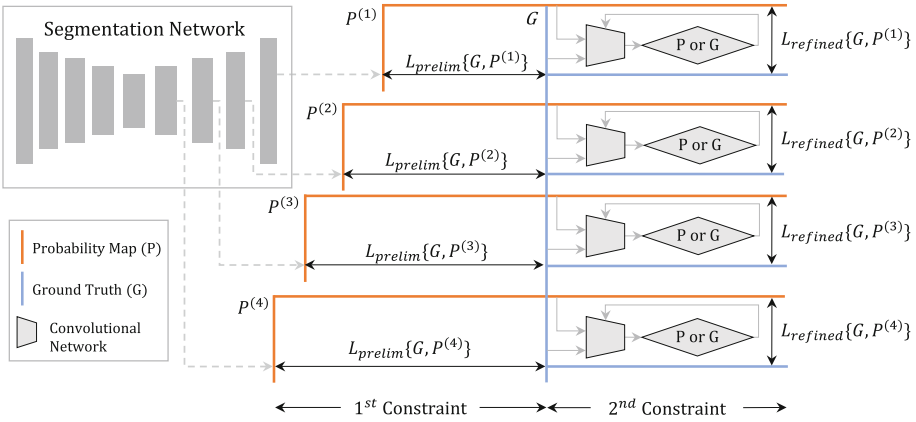

Fig. 1. Overview of the proposed architecture. The first constraint is used to enhance prediction similarity to the standard ones, for preliminary segmentation. The second constraint is used to capture data distribution to maintain consistency in high-dimensional space for map refinement.

Meanwhile, the adversarial loss for each of the CNN is defined as:

$$L_{CN}\left\{G, P^{(2)}\right\} = -\mathrm{E}_{G \sim p_{CN}(truth)} \log(CN_{\theta_{CN}}(G)) - \mathrm{E}_{P^{(2)} \sim p_{SN}(proba)} \log(1 - CN_{\theta_{CN}}(SN_{\theta_{SN}}(P^{(2)}))) \quad (2)$$

where $p_{CN}(truth)$ denotes the distribution of original samples. $CN_{\theta CN}(G)$ represents the probability for the input of CN coming from the original dataset.

In the implementation, we update the segmentation network and all the discriminators alternatingly in each iteration until both the generator and discriminators are converged.

3 Experiments

3.1 Dataset and Implementation Details

We collected 10927 cases for this research from Yunnan Cancer Hospital. Each scan is with resolution of 1×1 mm^2 and size of 512×480. The breast tumors of each case are delineated by three experienced experts. Five-fold cross validation is performed on the dataset in all experiments to verify our proposed network. For external validation, we further test our model on two independent publicly-available datasets collected by STU-Hospital (Dataset 1) [22] and SYU-University (Dataset 2) [23]. In order to comprehensively evaluate segmentation efficiency of our model, Dice Similarity Coefficient (DSC), Precision, Recall, Jaccard, and Root Mean Squared Error (RMSE) are used as evaluation metrics in this work.

Table 1. Quantitative comparison with state-of-the-art segmentation methods on large-scale clinical breast ultrasound dataset.

Method	DSC (%)↑	Precision (%)↑	Recall (%)↑	Jaccard (%)↑	RMSE (mm)↓
DeepRes [24]	75.16 ± 1.13	80.04 ± 1.33	73.52 ± 2.18	61.24 ± 1.53	0.81 ± 0.08
MSUNet [13]	71.84 ± 1.19	67.32 ± 2.04	81.67 ± 1.85	57.10 ± 1.54	0.82 ± 0.10
UNet++ [14]	73.20 ± 0.96	75.98 ± 1.42	73.58 ± 2.00	58.59 ± 1.27	0.84 ± 0.09
SegNet [25]	76.84 ± 0.87	79.08 ± 1.33	77.29 ± 1.73	63.23 ± 1.27	0.77 ± 0.08
AttU-Net [16]	73.46 ± 0.98	80.81 ± 1.33	69.93 ± 1.95	58.94 ± 1.31	0.86 ± 0.09
U^2-Net [26]	78.78 ± 0.91	75.90 ± 1.52	**84.58 ± 1.43**	65.91 ± 1.39	0.72 ± 0.08
UNet 3+ [17]	77.14 ± 0.86	79.45 ± 1.20	77.32 ± 1.72	63.61 ± 1.25	0.77 ± 0.08
Ours	**80.97 ± 0.88**	**81.64 ± 1.30**	82.19 ± 1.61	**68.83 ± 1.38**	**0.68 ± 0.06**

Our proposed algorithm is conducted on PyTorch, and all experiments are performed on NVIDIA Tesla A100 GPU. We use Adam optimizer to train the framework with an initial learning rate of 10^{-4}. The epochs to train all models are 100 and the batch size in training process is set as 4.

3.2 Comparison with State-of-the-Art Methods

To verify the advantages of our proposed model for breast tumor segmentation in ultrasound images, we compare our deep-supervised convolutional network with the state-of-the-art tumor segmentation methods, including DeepRes [24], MSUNet [13], UNet++

[14], SegNet [25], AttU-Net [16], U^2-Net [26] and UNet 3+ [17]. The comparison experiments are carried on a large-scale clinical breast ultrasound dataset, and the quantitative results are reported in Table 1. It is obvious that our proposed model achieves the optimal performance compared with other segmentation models. For example, our method obtains a DSC score of 80.97%, which is 5.81%, 9.13%, 7.77%, 4.13%, 7.51%, 2.19%, and 3.83% higher than other seven models. These results indicate the effectiveness of the proposed model in delineating breast tumors in ultrasound images.

Representative segmentation results using different methods are provided in Fig. 2. The probability maps predicted by our model are more consistent with the ground truth, especially in the tiny structures which are difficult to capture. This verifies the superior ability of the proposed model in maintaining detailed edge information compared with state-of-the-art methods.

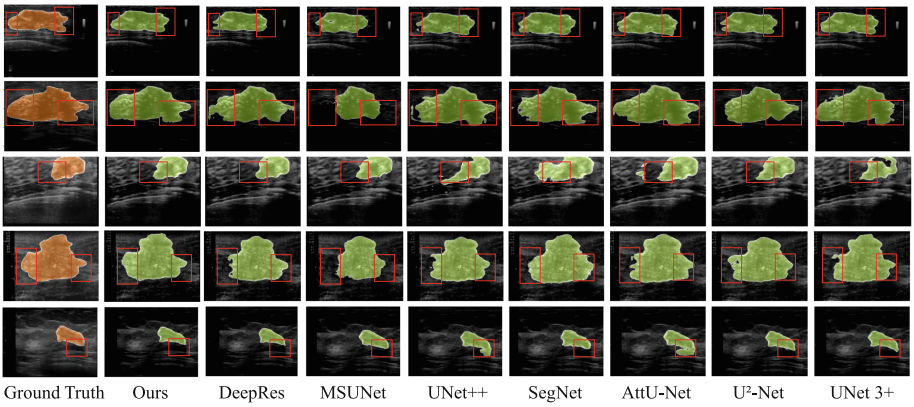

Fig. 2. Segmentation results of five subjects obtained from different models. Red boxes are used to highlight the boundaries which are difficult to segment.

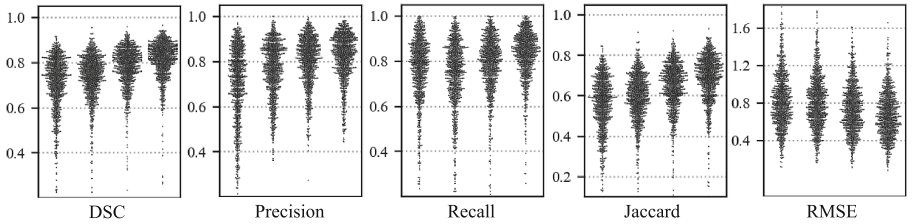

Fig. 3. Five evaluation criteria for total cases from different frameworks: Stage I, Stage II, Stage III, and Stage IV (From left to right in each index).

3.3 Ablation Studies

We demonstrate the efficacy of the deep supervision strategy using ablation studies. Four groups of frameworks (Stage I, Stage II, Stage III and Stage IV) are designed, with the

numerals denoting the level of deep supervision counting from the last deconvolutional layer.

We test these four frameworks on the in-house breast ultrasound dataset, and verify their segmentation performance using the same five evaluation criteria. The evaluation metrics from all cases are presented by the scatter plots in Fig. 3. The obtained quantitative results are shown in Table 2, where Stage IV model achieves the optimal DSC, Precision, Recall, and Jaccard. All these results draw a unanimous conclusion on the relationship between these four frameworks. That is, the segmentation ability of the proposed Stage IV is ameliorated from every possible perspective. Moreover, Stage IV obtains minimal RMSE compared with other three models (0.68 mm vs 0.84 mm, 0.82 mm, 0.75 mm), which means better matching of the predicted maps from Stage IV with the corresponding ground truth. All these comparison results verify the superiority of deep supervision for breast tumor segmentation in ultrasound images.

Table 2. Quantitative results among four groups of segmentation frameworks Stage I, Stage II, Stage III, and Stage IV.

Method	DSC (%)↑	Precision (%)↑	Recall (%)↑	Jaccard (%)↑	RMSE (mm)↓
Stage I	70.52 ± 1.25	70.01 ± 1.77	76.16 ± 2.04	55.72 ± 1.53	0.84 ± 0.09
Stage II	73.43 ± 1.07	76.10 ± 1.30	74.19 ± 2.30	58.99 ± 1.37	0.82 ± 0.08
Stage III	77.67 ± 1.06	79.63 ± 1.43	78.17 ± 1.98	64.42 ± 1.47	0.75 ± 0.07
Stage IV	**80.97 ± 0.88**	**81.64 ± 1.30**	**82.19 ± 1.61**	**68.83 ± 1.38**	**0.68 ± 0.06**

3.4 Performance on Two External Public Datasets

In order to evaluate the generalizability of the proposed model, we introduce external Dataset 1 and Dataset 2 for external validation experiments. Specifically, Dataset 1 and Dataset 2 are used as testing data to evaluate the generalization performance of the models trained on our own dataset without fine tuning, and the corresponding results are shown in Table 3. Promising performance demonstrates outstanding generalization ability of our large pre-trained model, with a DSC score of 81.35% and a Recall of 80.96% on Dataset 1, and a DSC score of 77.16% and a Recall of 93.22% on Dataset 2.

Table 3. External validation performance on two independent publicly-available Dataset 1 and Dataset 2.

Dataset	DSC (%)	Precision (%)	Recall (%)	Jaccard (%)	RMSE (mm)
Dataset 1	81.35 ± 2.04	85.75 ± 2.97	80.96 ± 2.09	70.65 ± 3.07	0.19 ± 0.01
Dataset 2	77.16 ± 2.51	68.34 ± 4.03	93.22 ± 0.65	65.26 ± 3.67	0.42 ± 0.01

4 Conclusion

In this paper, we have developed a large pre-trained model for breast tumor segmentation from ultrasound images. In particular, two constraints are proposed to exploit both image similarity and space correlation information for refining the prediction maps. Moreover, our proposed deep supervision strategy is used for quality control at each decoding stage, optimizing prediction maps layer-by-layer for overall performance improvement. Using a large clinical dataset, our proposed model demonstrates *not only* state-of-the-art segmentation performance, *but also* the outstanding generalizability to new ultrasound data from different sites. Besides, our large pre-trained model is general and robust in handling various tumor types and shadow noises in our acquired clinical ultrasound images. This also shows the potential of directly applying our model in real clinical applications.

Acknowledgment. This work was supported in part by The Key R&D Program of Guangdong Province, China (grant number 2021B0101420006), National Natural Science Foundation of China (grant number 62131015), and Science and Technology Commission of Shanghai Municipality (STCSM) (grant number 21010502600).

References

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics. CA: Cancer J. Clin. **72**(1), 7–33 (2022)
2. Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., et al.: Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast **66**, 15–23 (2022)
3. Berg, W.A., Zhang, Z., Lehrer, D., Jong, R.A., Pisano, E.D., Barr, R.G., et al.: Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. J. Am. Med. Assoc. (JAMA) **307**(13), 1394–1404 (2012)
4. Kalager, M., Haldorsen, T., Bretthauer, M., Hoff, G., Thoresen, S.O., Adami, H.O.: Improved breast cancer survival following introduction of an organized mammography screening program among both screened and unscreened women: a population-based cohort study. Breast Cancer Res. BCR **11**(4), 1–9 (2009)
5. Cheng, H.D., Shan, J., Ju, W., Guo, Y., Zhang, L.: Automated breast cancer detection and classification using ultrasound images: a survey. Pattern Recogn. **43**(1), 299–317 (2010)
6. Kratkiewicz, K., Pattyn, A., Alijabbari, N., Mehrmohammadi, M.: Ultrasound and photoacoustic imaging of breast cancer: clinical systems, challenges, and future outlook. J. Clin. Med. **11**(5), 1165 (2022)
7. Ragab, M., Albukhari, A., Alyami, J., Mansour, R.F.: Ensemble deep-learning-enabled clinical decision support system for breast cancer diagnosis and classification on ultrasound images. Biology (Basel) **11**(3), 439 (2022)
8. Xian, M., Zhang, Y., Cheng, H.-D., Xu, F., Zhang, B., Ding, J.: Automatic breast ultrasound image segmentation: a survey. Pattern Recogn. **79**, 340–355 (2018)
9. Calas, M.J.G., Almeida, R.M.V.R., Gutfilen, B., Pereira, W.C.A.: Intraobserver interpretation of breast ultrasonography following the BI-RADS classification. Eur. J. Radiol. **74**(3), 525–528 (2010)

10. Yap, M.H., Edirisinghe, E.A., Bez, H.E.: Processed images in human perception: a case study in ultrasound breast imaging. Eur. J. Radiol. **73**(3), 682–687 (2010)
11. Jalalian, A., Mashohor, S.B., Mahmud, H.R., Saripan, M.I.B., Ramli, A.R.B., Karasfi, B.: Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clin. Imaging **37**(3), 420–426 (2013)
12. Zhou, Y., Chen, H., Li, Y., Cao, X., Wang, S., Shen, D.: Cross-model attention-guided tumor segmentation for 3D automated breast ultrasound (ABUS) images. IEEE J. Biomed. Health Inform. **26**(1), 301–311 (2021)
13. Su, R., Zhang, D., Liu, J., Cheng, C.: Msu-net: Multi-scale u-net for 2D medical image segmentation. Front. Genet. **12**, 639930 (2021)
14. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
15. Li, M., Lian, F., Guo, S.: Multi-scale selection and multi-channel fusion model for pancreas segmentation using adversarial deep convolutional nets. J. Digit. Imaging **35**, 47–55 (2022)
16. Oktay, O., et al.: Attention U-net: learning where to look for the pancreas. arXiv preprint arXiv:180403999 (2018)
17. Huang, H., et al. (eds.) Unet 3+: a full-scale connected unet for medical image segmentation. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE (2020)
18. Liu, X., Guo, S., Yang, B., Ma, S., Zhang, H., Li, J., et al.: Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks. J. Digit. Imaging **31**, 748–760 (2018)
19. Pei, Y., Mu, L., Fu, Y., He, K., Li, H., Guo, S., et al.: Colorectal tumor segmentation of CT scans based on a convolutional neural network with an attention mechanism. IEEE Access. **8**, 64131–64138 (2020)
20. 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
21. Goodfellow, I.J., et al.: Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, vol. 2, pp. 2672–2680. MIT Press, Cambridge (2014)
22. Negi, A., Raj, A.N.J., Nersisson, R., Zhuang, Z., Murugappan, M.: RDA-UNET-WGAN: an accurate breast ultrasound lesion segmentation using Wasserstein generative adversarial networks. Arab. J. Sci. Eng. **45**(8), 6399–6410 (2020)
23. Huang, Q., Huang, Y., Luo, Y., Yuan, F., Li, X.: Segmentation of breast ultrasound image with semantic classification of superpixels. Med. Image Anal. **61**, 101657 (2020)
24. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. **15**(5), 749–753 (2018)
25. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
26. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M.: U2-Net: going deeper with nested U-structure for salient object detection. Pattern Recogn. **106**, 107404 (2020)

Clinical Applications – Cardiac