

Consistent Learning-Based Breast Tumor Segmentation and Its Application in Sentinel Lymph Node Metastasis Prediction

Fengjun Zhao¹, Kaiming Huang¹, Zhipeng Sun¹, Xin Chen², Xiaowei He¹, Bin Wang^{1,*} and Cao Xin^{1,*}

Abstract— Accurate staging of lymph nodes provides crucial diagnostic information for breast cancer patients, where segmentation is of great importance by localizing and visualizing the breast tumor of interest. Nevertheless, current segmentation methods perform average when facing large span of tumor sizes, degraded image quality, blurred tumor boundaries, and resulting noise during manual annotation. Therefore, we develop a Multi-scale RepVGG-based Segmentation Network (MPSegNet) to segment breast tumor from MR images. In particular, we construct a consistent learning framework for the MPSegNet to alleviate the impact of noisy labels upon segmentation results. The rationale is that different views covering the same breast tumors are supposed to generate identical segmentation predictions. Then, we predict SLN metastasis given segmented breast tumors, where we evaluate the relationships between the predictive performance and tumor segmentations under different consistencies. The results show the superiority of our method over other state-of-the-art methods. A high consistency among multiple views can boost the segmentation performance during consistent learning. However, the optimal segmentation does not produce the best SLN metastatic prediction results, implying that the dependence of classification upon segmentation needs to be elaborately investigated further.

Clinical Relevance— This study facilitates more accurate segmentation of breast tumors with consistent learning, and provides an initial analysis between tumor segmentation and subsequent prediction of SLN metastasis, which has potential significance for the precise medical care of breast cancer patients.

I. INTRODUCTION

Breast cancer is the most common malignant tumor threatening women all over the world [1]. Accurate staging of lymph nodes provides important diagnostic and prognostic information in the management of patients with breast cancer [2]. Sentinel lymph node (SLN) represents the first affected drainage site in the event of tumor spread, which is clinically determined by biopsy. However, this invasive procedure not only has complications, such as pain, paresthesia and arm swelling, but also suffers from underestimation if not covering the tissue with metastatic tumor cells. Radiomics with noninvasive magnetic resonance imaging (MRI) facilitates the comprehensive analysis of breast tumors. In particular, various radiomic signatures have then been derived to predict breast cancer SLN metastasis in MRI [3-6], including T_1 -weighted imaging, T_2 -weighted imaging, diffusion-weighted imaging (DWI) and dynamic contrast-

enhanced MRI. However, radiomics generally requires a complicated procedure consisting of tumor segmentation, feature extraction, feature selection and prediction model construction. Of note, segmentation can help localize and visualize the breast tumor of interest, which is essential for precise diagnosis and subsequent treatment.

Over the past decade, state-of-the-art deep learning (DL) models such as fully convolutional network (FCN) [7], encoder-decoder structure [8], attention-based network [9] have been widely used in medical image segmentation. Nevertheless, accurate segmentation of breast tumors from MRI remains challenging due to large span of tumor sizes, degraded image quality (e.g. motion artifacts), blurred tumor boundaries, and resulting noise during manual annotation. To cope with these challenges, modified DL models such as cascaded multi-scale encoder-decoder [10] and atrous spatial pyramid pooling (ASPP) [11] were performed to improve the segmentation for different sizes of tumors. Also, active contour models were integrated into DL models to focus on tumor boundary segmentation [12].

Despite showing promising performance, the above DL methods on one hand require complicated network architectures and thus consume many computational resources. Therefore, we propose a consistent learning-based model termed as Multi-scale RepVGG-based Segmentation Network (MPSegNet) to segment breast tumor from MR images. The conventional convolution in our MPSegNet is replaced by a lightweight RepVGG block [13], which prevents overfitting and improves the performance of the DL network. A convolutional block attention module (CBAM) [14] is inserted between two adjacent RepVGG blocks to highlight the salient features. On the other hand, it is difficult for DL models to eliminate the noise included in the human annotated labels. Shi *et al.* [15] have tried to distill supervision information from both pixel and image levels to solve low-quality labeling issue. Inspired by this study, we construct a consistent learning framework based on the developed MPSegNet to alleviate the impact of noisy labels upon segmentation results. The rationale is that different views covering the same breast tumors are supposed to generate identical segmentation predictions. Finally, we perform the prediction of SLN metastasis given the segmented breast tumors, where the relationships between the predictive performance and segmentations under different consistencies were well investigated.

This work was supported in part by Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (22JP087), China Postdoctoral Science Foundation (2019M653717), and Shaanxi International S&T Cooperation Program (2021KW-55).

*Corresponding author: wbin@nwu.edu.cn; caoxin@nwu.edu.cn.

¹ Xi'an Key Lab of Radiomics and Intelligent Perception, School of Information Science and Technology, Northwest University, Xi'an, Shaanxi 710069, China.

² Department of Radiology, Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi 710007, China.

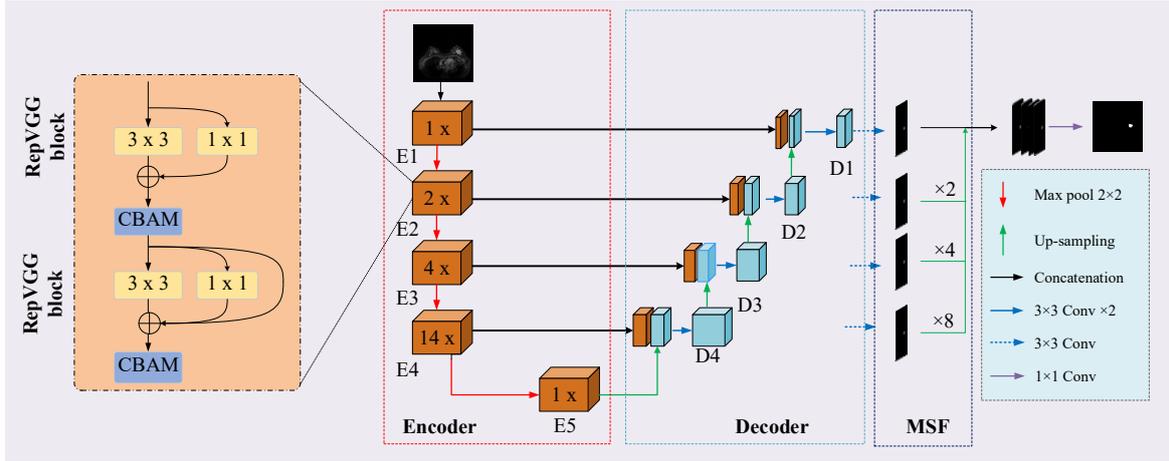


Figure 1. Overview of the proposed MPSEgNet model consisting of an encoder, a decoder and a multi-scale fusion (MSF) module.

II. METHODS

A. Segmentation Model

As shown in Fig. 1, the proposed MPSEgNet consists of an encoder, a decoder and a multi-scale fusion (MSF) module. The encoder has 22 RepVGG blocks that are grouped into 1, 2, 4, 14 and 1 corresponding 5 stages of feature abstraction. Each block encompasses two parallel convolutions with 3×3 and 1×1 kernel sizes, respectively, and one identity mapping. Note that the first block in each stage omits the identity mapping. The CBAM is inserted between two adjacent RepVGG blocks to improve the feature representation. The decoder has four upsampling and fusion steps. Each step upsamples the feature maps and concatenates them with the skip-connected feature maps from the encoder, followed by two 3×3 convolutions and ReLU activation. In this way, four fused feature maps denoted as D1, D2, D3 and D4 are sequentially generated. The MSF module generates four side output probability maps given D1-D4 by applying a 3×3 convolution and ReLU mapping. Then, we upsample these probability maps by different rates (8, 4, 2 and 1), and then concatenate them together. After performing 1×1 convolution and sigmoid activation, we finally obtain the output segmentation probability map with the same size as the input image.

B. Consistent Learning

We also perform consistent learning on our MPSEgNet model to alleviate the inaccurate segmentation caused by noisy human annotations (labels). As shown in Fig. 2, the consistent learning framework includes two modules: a teacher network and a student network. Considering that breast tumors occupy only a small portion of an MR image, we first crop one central view and multiple adjacent views from the MR image containing entire tumor regions. When feeding the center view into the teacher network, the prediction should be identical with the predictions obtained from the student network by inputting adjacent views. Thus, the predictions of the student network can be used as an extra supervision to drive the teacher network to output consistent segmentation free of noise.

Specifically, one center view X and 8 adjacent views $X_1, X_2 \dots X_8$ of size of 352×352 are generated from one MRI slice, where the latter are positioned top, bottom, left, right, upper left, lower left, upper right, lower right, and lower right, respectively, 10 pixels away from the center. In each batch, the teacher and student networks output the predictions P and $P_1, P_2 \dots P_8$ given views X and $X_1, X_2 \dots X_8$, respectively. Then we take the average over $P_1, P_2 \dots P_8$ to get the pseudo label \bar{P} , whose entropy β was used to evaluate the uncertainty and determine the probability of correct pixel.

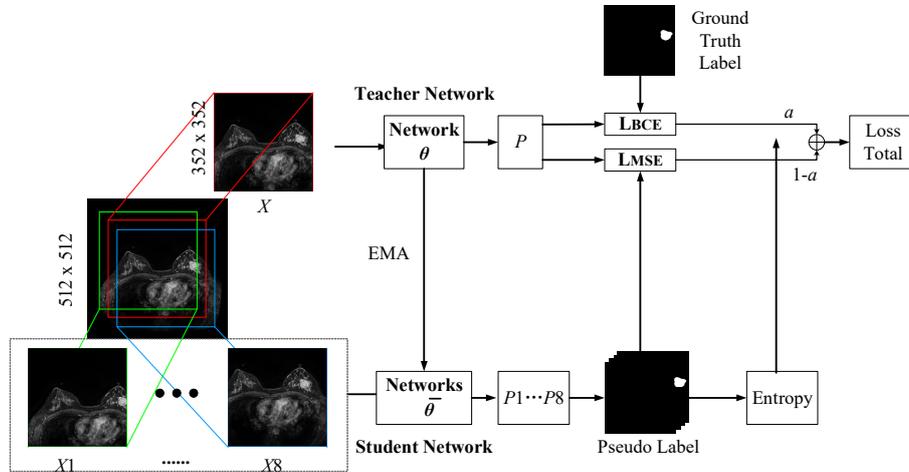


Figure 2. Consistent learning framework for our model, where X is the center view of slice, and X_1 - X_8 are the adjacent views around the center view.

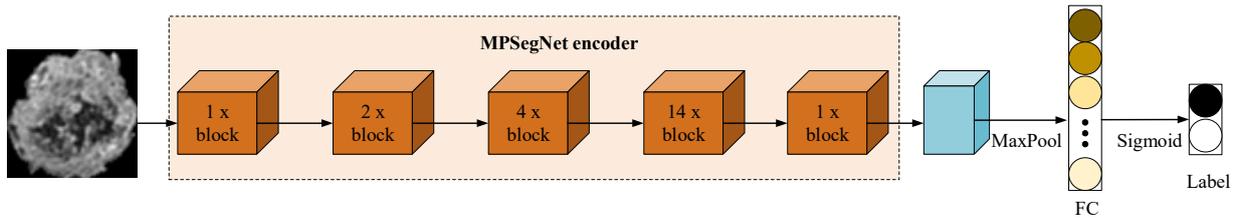


Figure 3. Prediction model for SLN Metastasis of breast cancer patients, which takes the MPSEgNet encoder as the backbone.

$$\beta = E[-\bar{P} \cdot \log \bar{P}] \quad (1)$$

where β represents the uncertainty estimate of each pixel, and $E[\cdot]$ the expectation operator. Note that consistent predictions among adjacent views output a low uncertainty, while inconsistent predictions are likely to have a high uncertainty.

Finally, we optimize the teacher network by combining the BCE loss and mean squared error (MSE) loss:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{BCE} + (1 - \alpha) \cdot \mathcal{L}_{MSE} \quad (2)$$

where \mathcal{L}_{BCE} is the BCE loss between the ground truth label and the prediction P of the teacher network, while \mathcal{L}_{MSE} is the MSE between the prediction P and the pseudo label \bar{P} of the student network. The consistency coefficient $\alpha = e^{-\beta}$ controls the balance between two losses. High uncertainty has a small α (tending to zero) that drives the model to focus on the MSE constraint between two predictions P and \bar{P} . In contrary, low uncertainty has a large α (tending to one), meaning that the model will focus on the BCE constraint from the original label.

The teacher network parameterized by θ can be updated by \mathcal{L}_{total} in Eq. (2). The student network parameterized by $\bar{\theta}$ is optimized by the exponential moving average (EMA) of θ as $\bar{\theta} = \gamma\theta + (1 - \gamma)\bar{\theta}$, where γ is a smoothing coefficient.

C. Prediction of SLN Metastasis

Given segmented breast tumors, we perform the binary classification between breast cancer patients with and without SLN metastasis. We aim to investigate the relationships between the predictive performance and segmentations under different consistencies. As shown in Fig. 3, this classification model shares the same backbone as the MPSEgNet encoder. We also use the parameters in segmentation as the pre-training of the classification model, because there are similar features between two tasks more or less. The backbone is followed by max-pooling and a fully connected layer. We feed segmented breast tumors into the classification model to generate the metastatic prediction results. Of note, each segmentation is obtained by averaging the prediction over all 9 views (one center view and 8 adjacent views). Thus, the impact of segmentation under different thresholds or consistencies on the prediction of SLN metastasis can be fully investigated.

D. Implementation

All models are coded using pytorch 1.8.0 with python 3.7 on a server with one GTX 3060 GPU. For tumor segmentation, all images are resized into 288×288 . The batch size and maximum training epoch are set as 2 and 150, respectively. The initial smoothing coefficient γ is 0.99 and is updated by $\gamma = \min((1 - 1/(\text{epoch} + 1)), \gamma)$. For SLN metastasis prediction, all tumor images are resized into 48×48 . The batch size and maximum training epoch are set as 4 and 50, respectively. We

use AdamW optimize in both tasks, with initial learning rates of 3×10^{-4} and 1×10^{-3} for tumor segmentation and SLN metastasis prediction, respectively. The cosine annealing strategy is employed to adjust the learning rate with the minimum learning rate and half cycle being 1×10^{-8} and 150/50 (segmentation/prediction), respectively.

III. EXPERIMENTAL RESULTS

A. Dataset

We enrolled 205 breast cancer patients in this study, including 81 and 124 with and without SLN metastasis, respectively. All patients underwent MRI examination by a 3.0 T system (Signa HDxt; GE Medical Systems) and the post-contrast sequences were acquired after the injection of 0.2 mmol/kg body weight of gadolinium-DTPA at a rate of 2.0 mL/s followed by 20 mL saline solution. We manual annotated 205 post-contrast MRIs of breast tumors (3426 slices) using the Medical Imaging Interaction Toolkit (v2015.5.0). Then, we divided the all the slices at a ratio of $\sim 8:1:1$ into a training set ($n=157, 2722$ slices), a validation set ($n=25, 375$ slices) and a testing set ($n=23, 375$ slices), which are used for the training, validation and testing of our developed model, respectively.

B. Comparison with State-of-the-Arts

We compared our model to state-of-the-art segmentation methods, including UNet [10], FCN8s [7], AUNet [11] and ResUNet [12]. As shown in Table 1, our MPSEgNet achieved an average Dice coefficient of 0.8004, better than all the comparative methods. In particular, we improved the Dice coefficient by 0.0614, 0.0635, 0.0425 and 0.0240 compared to UNet, FCN8s, AUNet and ResUNet, respectively. In addition, our MPSEgNet has the highest sensitivity of 93.91%, indicating the advantage of this model in addressing under-segmentation.

TABLE I. PERFORMANCE COMPARISON BETWEEN OUR MODEL AND PREVIOUS MODELS IN INDEPENDENT TESTING SET

Method	Dice	Sensitivity (%)	Specificity (%)
UNet	0.7422	84.90	99.79
FCN8s	0.7401	82.01	99.80
AUNet	0.7611	87.93	99.79
ResUNet	0.7796	85.88	99.83
MPSEgNet	0.8004	93.91	99.77

C. Evaluation of Consistent Learning

Table 2 gives the results of our model after performing consistent learning. At low thresholds (0.1-0.4), our model tends to over-segment breast tumors of interest as indicated by high sensitivities (94.77-97.15%) and low specificities (99.75-99.77%). In contrary, high thresholds (0.6-0.9) lead to under-segmentation of our model, manifesting as decreased sensitivities (87.45-92.77%) and increased specificities

(99.78-99.80%). The intermediate threshold of 0.5 corresponds to the result of MPSEgNet in Table 1. Of note, the highest Dice coefficient of 0.8213 is obtained when the threshold equals 0.8, a relatively high consistency among different views. This suggests that the consistent and slight under-segmentation is beneficial to overall segmentation performance for breast tumors.

TABLE II. SEGMENTATION PERFORMANCE UNDER DIFFERENT CONSISTENT THRESHOLDS IN INDEPENDENT TESTING SET

Threshold	Dice	Sensitivity (%)	Specificity (%)
0.1	0.7517	97.15	99.75
0.2	0.7717	96.30	99.76
0.3	0.7845	95.62	99.76
0.4	0.7927	94.77	99.77
0.5	0.8004	93.91	99.77
0.6	0.8089	92.77	99.79
0.7	0.8173	91.53	99.78
0.8	0.8213	90.23	99.79
0.9	0.8206	87.45	99.80

D. Prediction of SLN Metastasis

The predictive results for SLN metastasis under different consistent thresholds are shown in Table 3. It is unexpected that the highest Dice coefficient of 0.8213 (at 0.8 threshold) generates only suboptimal predictive performance with the area under the curve (AUC) and accuracy of 0.6920 and 73.91%, respectively. The optimal predictive performance is obtained when the threshold equals 0.3-0.4, achieving an AUC and accuracy of 0.7321-0.7634 and 73.21-78.26%, respectively. We notice that corresponding Dice coefficients are only 0.7845-0.7927, *i.e.* the interval of over-segmentation. Overall, the optimal segmentation may not be directly related to the accurate prediction of SLN metastasis, which is noteworthy in future breast cancer diagnosis.

TABLE III. PREDICTION PERFORMANCE FOR SLN METASTASIS UNDER DIFFERENT SEGMENTATIONS BY DIFFERENT CONSISTENT THRESHOLDS.

Threshold	Dice	AUC	Accuracy
0.1	0.7517	0.6380	60.87
0.2	0.7717	0.6294	65.22
0.3	0.7845	0.7634	78.26
0.4	0.7927	0.7321	73.21
0.5	0.8004	0.6607	69.57
0.6	0.8089	0.5893	65.22
0.7	0.8173	0.6607	69.57
0.8	0.8213	0.6920	73.91
0.9	0.8206	0.5893	65.92

IV. CONCLUSION

We have developed a consistent learning-based breast tumor segmentation model named MPSEgNet and evaluated its application in SLN metastasis prediction. First, we proposed a new architecture based on RepVGG and CBAM that can learn the multi-scale features of breast tumors. Then, we used a consistent learning framework to solve the interference of noise labels on the model to obtain robust segmentation results. Finally, we evaluated the MPSEgNet for predicting SLN metastasis under different consistent thresholds. The results showed the superiority of our method over other state-of-the-art methods. A high consistency among multiple views can boost the segmentation performance of our MPSEgNet model during consistent learning. However, the optimal segmentation does not produce the best SLN

metastatic prediction results, which implies that the dependence of classification upon segmentation needs to be elaborately investigated further. Overall, this study facilitates more accurate segmentation of breast tumors and provides an initial analysis between tumor segmentation and prediction of SLN metastasis, which has potential significance for the precise medical care of breast cancer patients.

REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2018," *Ca-a Cancer Journal for Clinicians*, vol. 68, pp. 7-30, Jan-Feb 2018.
- [2] C. Nos, C. Harding-MacKean, P. Freneaux, A. Trie, M. C. Falcou, X. Sastre-Garau, *et al.*, "Prediction of tumour involvement in remaining axillary lymph nodes when the sentinel node in a woman with breast cancer contains metastases," *British Journal of Surgery*, vol. 90, pp. 1354-1360, Nov 2003.
- [3] Y. Dong, Q. Feng, W. Yang, Z. Lu, C. Deng, L. Zhang, *et al.*, "Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI," *European Radiology*, vol. 28, pp. 582-591, Feb 2018.
- [4] L. Han, Y. Zhu, Z. Liu, T. Yu, C. He, W. Jiang, *et al.*, "Radiomic nomogram for prediction of axillary lymph node metastasis in breast cancer," *European Radiology*, vol. 29, pp. 3820-3829, Jul 2019.
- [5] C. Liu, J. Ding, K. Spuhler, Y. Gao, M. S. Sosa, M. Moriarty, *et al.*, "Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI," *Journal of Magnetic Resonance Imaging*, vol. 49, pp. 131-140, Jan 2019.
- [6] R. Chai, H. Ma, M. Xu, D. Arefan, X. Cui, Y. Liu, *et al.*, "Differentiating axillary lymph node metastasis in invasive breast cancer patients: A comparison of radiomic signatures from multiparametric breast MR sequences," *Journal of Magnetic Resonance Imaging*, vol. 50, pp. 1125-1132, Oct 2019.
- [7] J. Long, E. Shelhamer, T. Darrell, and Ieee, "Fully Convolutional Networks for Semantic Segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 3431-3440.
- [8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, GERMANY, 2015, pp. 234-241.
- [9] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, *et al.*, "AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Physics in Medicine & Biology*, vol. 65, pp. 055005, Feb 2020.
- [10] Y. Yan, P. H. Conze, E. Decenciere, M. Lamard, G. Quellec, B. Cochener, *et al.*, "Cascaded multi-scale convolutional encoder-decoders for breast mass segmentation in high-resolution mammograms," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, GERMANY, 2019, pp. 6738-6741.
- [11] Gao, Y., *et al.* Dense Encoder-Decoder Network based on Two-Level Context Enhanced Residual Attention Mechanism for Segmentation of Breast Tumors in Magnetic Resonance Image. in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2019. San Diego, CA.
- [12] Fang, Z., *et al.*, Combining a Fully Convolutional Network and an Active Contour Model for Automatic 2D Breast Tumor Segmentation from Ultrasound Images. *Journal Of Medical Imaging And Health Informatics*, 2019. 9(7): p. 1510-1515.
- [13] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding and J. Sun, "RepVGG: Making VGG-style ConvNets Great Again," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 13728-13737.
- [14] Woo, S., Park, J., Lee, JY., Kweon, I.S. (2018). CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*, vol 11211. Springer, Cham.
- [15] Shi, J., Wu, J. (2021). Distilling Effective Supervision for Robust Medical Image Segmentation with Noisy Labels. In: *et al.* *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. MICCAI 2021. Lecture Notes in Computer Science*, vol 12901. Springer,