A DEEP NEURAL CNN MODEL WITH CRF FOR BREAST MASS SEGMENTATION IN MAMMOGRAMS

Ridhi Arora
Department of Computer Science & Engineering
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand
India

Balasubramanian Raman
Department of Computer Science & Engineering
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand
India

***Abstract*—Malignancy in women's breast is known to be the second most common form of cancer. Early detection can help diagnose the disease effectively, but it continues to grow manifolds due to reasons unknown. Therefore, to aid radiologists in the effective treatment of breast cancer, an end-to-end deep learning-based architecture for ROI-based breast mass segmentation is proposed. The architecture involving Residual connections and Group convolution in U-Net (RGU-Net), contains encoder and decoder blocks with different resolutions and feature sizes. The architecture captures multi-level features from the encoder-decoder architecture using the residual connections and group convolution. Moreover, to improve the field-of-view of the filters, atrous convolutions are added. Later, for better visualization, the predicted masks are labelled (structured learning) using a conditional random field (CRF) to analyse the mass boundaries explicitly. A publicly available INBreast dataset is used to validate the method, which is augmented to produce robust results. The experimental results produced from the proposed approach outperformed the conventional mass segmentation algorithms, demonstrating its effectiveness.**

***Index Terms*—Breast Mass, Feature processing, Semantic Segmentation, Residual Mapping, Group Convolution, Image labeling**

I. Introduction

The global trend in breast cancer cases and related deaths has significantly increased in recent times. As per the statistics, $1,762,450$ new cancer cases are diagnosed in the United States [1] with future estimation of $62,930$ new cases of breast cancer and $95,830$ new cases of skin cancer in females. This has made it a major public health concern with high morbidity and mortality rates [2]. At the same time, if this is diagnosed in early stages, it can be cured effectively. Among all types of breast anomalies, breast mass segmentation is the most common but an encumbered form of segmentation. It involves extreme care due to the irregular and spiculated shape of masses, heterogeneously distributed in the background. The more sporadic the shape of a mass, the more it becomes sure to have a malignant lesion [3]. Many works on the lesion segmentation and classification have been performed to aid experts in the adjoining field [3]–[6]. Ronneberger *et al.* introduced the first of its kind, a segmentation network, U-Net, which helps in automatically segmenting the images (of variable size) and attained reasonable results for variable cell structures [7]. Reviews on the biomedical 2D and 3D image segmentation have been performed by many researchers [8], [9]. A Work by Yang *et al.* [10] on medical image segmentation utilized the GAN structures and segmented (3D) computed tomography (CT) liver images. A work on MRI head and neck tumors using modified U-Net by Zhao *et al.* uses dilated convolution to extract multi-level features from the acquired dataset [11]. DeepLab [12], a work by Liang-Chieh *et al.* investigated PASCAL VOC-2012 dataset using atrous convolutions, atrous spatial pyramid pooling (ASPP) and DCNNs with probabilistic graphical models to efficiently predict object boundaries. Another work by Rouhi *et al.* for breast tumour segmentation and classification is introduced, incorporating Artificial Neural Network (ANN) and Cellular Neural Network (CeNN) [13].

We propose an intuitive way of semantic mass segmentation from ROI-based breast mass images to continue the technological advancements for mass segmentation. The proposed model, RGU-Net, incorporates the following advantages: (a) incorporating recurrent layer-like path for feature re-use and combination (residual connections), (b) incorporating more features from the encoder layers using feature projection capability of group convolution layer, (c) feature preservation with the involvement of convolution operation with stride 2 for down-sampling, and (d) applying CRF for image labeling (structured learning).

II. Methodology

A. Network architecture

In Figure 1, a schematic representation of the proposed model, RGU-Net, is presented. The figure's left part is the compression (encoder) path where the image features are extracted. In contrast, the right path depicting the expansion (decoder) path decompressing the features until the original size of the image is obtained.

The Encoder: The left part of the network contains different convolution blocks (conv blocks) that operate at different

resolutions. Each block includes two sub-blocks of convolution layers followed by Batch Normalization (BN) and a rectified linear unit (ReLU) layer. Each sub-block in a block is connected through a group convolution layer to extract spatial-dimension features. Unlike the approach used in [7], residual mappings are also incorporated in every block. Every stage's input is skipped and added (element-wise addition) with the output of the last ReLU layer of that block.

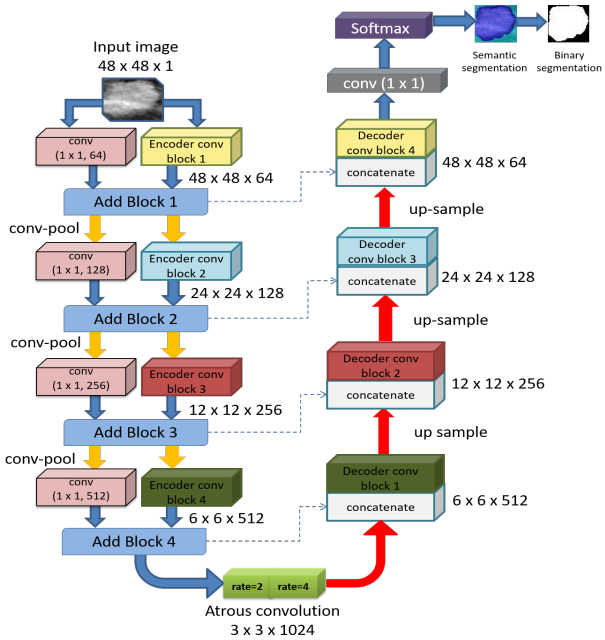

Fig. 1. Illustration of feature re-use and combination ability of the residual connection and group convolution in the proposed framework.

- Convolution layers in each sub-block have the kernel size of 3×3. The group convolution in the form of channel-wise separable convolution with filter size 7×7 and 1 filter per group of channels is used. To enhance feature extraction and extract minute details, residual identity connections with 1×1 convolution operation are added (Figure 2 (right)) in each block.

$$y = F(x, W_i) + x \tag{1}$$

  In equation 1, W_i represent a set of i weight matrices (W_1 in the example) occurring in the layers of the residual (skipped) layers. The "identity shortcuts" are referring to performing the element-wise addition of x (input to that corresponding block) with the output of the residual connections.
- In the added residual identity connections, the input and the output number of channels are the same. Identity connections between the sub-blocks of a block blend the features across channels and allow the networks to have more depth and avoid degradation.
- For preserving the inter-block features, the max-pool operation is replaced by a convolution operation with a kernel of 4×4 and a stride of 2 (conv-pool). As every subsequent layer extracts features by considering only the 2×2 image patches, the resulting feature maps are halved. This strategy of applying stride 2 in convolution layer works similar to the pooling layers that helps preserve some more image features.

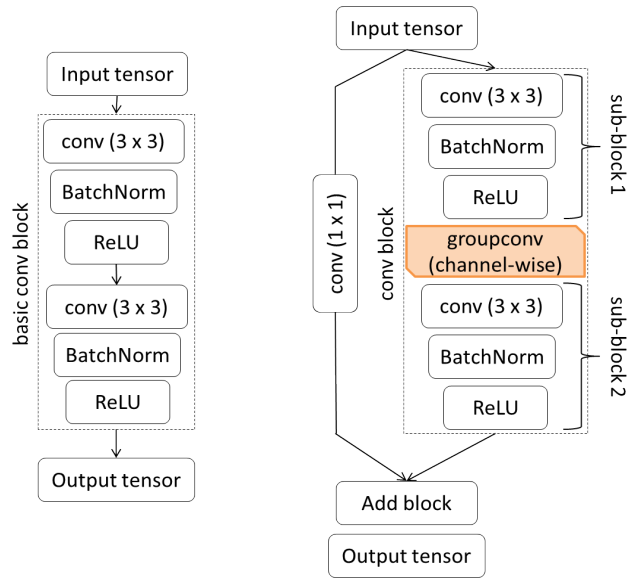

Fig. 2. Two types of convolution blocks, basic convolution block (left), and the proposed RGU-Net (right).

The Deepest part: Down-sampling operations in the left part of the network continuously reduce the size of the input signals and increase the receptive field of the network. Therefore, to allow wider field-of-view and extraction of context semantic information, two atrous convolution layers (as in [11]) with kernel size 3×3 and a dilation factor of 2 and 4, respectively are added. They visibly enhance the segmentation of images by precisely locating the contour boundaries.

The Decoder: The right path of the network extends the spatial support to the low-level features with the high-resolution localization features and combines them to output a 2-channel segmentation map.

- Similar to the encoder, group convolution layer in the form of channel-wise separable convolution is applied in every convolution block with kernel size 7×7 and 1 filter per group of channels. A de-convolution operation is applied after every convolution block to increase the size of the inputs.
- Contextual information from the left path is concatenated with the high-resolution feature information of the right path in the network. They are as depicted in the form of horizontal skip connections in the network in Figure 1. They allow for the concatenation of the contextual features extracted from the compression path with the localization features of the upsampling path to improve the quality of the final segmentation.
- Lastly, a 1×1 convolution is applied to produce output feature maps of the same size as that of the input. Then they are converted to probabilities for pixel-wise

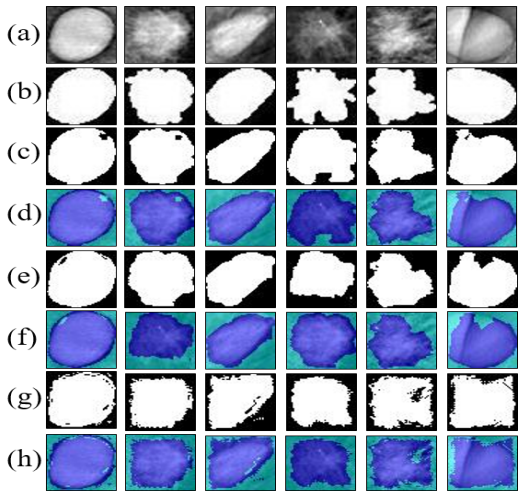

Fig. 3. Ablation studies on the proposed model (on epoch 65), where (a) original image, (b) binary labels (given), (c) prediction mask using ADAM optimizer, (d) semantic prediction mask using ADAM optimizer, (e) prediction mask using RMSPROP optimizer, (f) semantic prediction mask using RMSPROP, (g) prediction mask using SGDM optimizer, (h) semantic prediction mask using SGDM optimizer.

TABLE I
INTROSPECTION OF THE PROPOSED MODEL ON DIFFERENT OPTIMIZERS TO EXAMINE THE ROBUSTNESS OF THE APPROACH USED. ADAM: USING OPTIMIZER ADAM ON RGU-NET, RMSPROP: USING OPTIMIZER RMSPROP ON RGU-NET, SGDM: USING OPTIMIZER SGDM ON RGU-NET.

Epochs	Optimizer	Jac (in %)	Acc (in %)	Sen (in %)	Spe (in %)	Precision (in %)
65	ADAM	86.3	89.4	91.3	84.1	94
	RMSPROP	85.3	88.6	91.6	81.1	92.6
	SGDM	79.4	83.5	88.1	71.8	88.9
75	ADAM	86	89.1	91.5	82.7	93.4
	RMSPROP	85	88.3	91.1	81	92.6
	SGDM	77.2	82.1	89.4	66.7	84.9
85	ADAM	85.1	88.5	91.6	80.6	92.4
	RMSPROP	85.7	89.1	93.6	79.1	91
	SGDM	75.7	80.8	88.5	64.6	83.9

classification into foreground and background regions with 0 reserved for background and 1 for the foreground (mass), by applying *softmax* activations.

B. Conditional Random Field (CRF) for prediction labeling

Segmentation map labeling is a critical phenomenon as it requires adequate knowledge about the mass boundaries in the image. Correct labeling can provide the right picture, which later can be used for obtaining accuracy levels. One such method of labeling, CRF, is used in this study. It is a structured form of learning which is well suited for segmentation tasks [14]. CRF architecture incorporates hierarchical connectivity where an output function y, as $y = (y_1, ..., y_P)$, dependent on input sequence x, is defined as $F(y, x)$. The prediction for all the output variables are therefore calculated using $f(x) = \text{argmax}_y F(y, x)$. Typically, the proper factorization of $F(y, x)$ with respect to multivariate output y leads to a graphical model for which argmax can be maximized and is calculated using inference methods efficiently. In this work, CRF [15] is therefore used as a post-processing techniques for correctly annotating the segmentation maps produced from the proposed model. It uses pairwise edge potentials defined using the linear combination of Gaussian and bilateral kernels. The first kernel is the appearance kernel which ensures that the adjacent pixels with similar colour should be placed in one class. In contrast, bilateral kernel fills the place for inference on the RGB pixel values of the image. For this work, the default kernel value is used.

III. EXPERIMENTS AND RESULTS

A. Dataset and augmentation

The proposed model is validated on a publicly available mass segmentation dataset: INBreast dataset [16]. It is a mammographic mass analysis dataset containing high-quality mammogram images and provides accurate and precise lesion contours. The dataset includes 116 images, which in itself have 58 images reserved for training and rest 58 for testing. Similar data split is used in the comparison process. We further augment the dataset for effective training. Augmentations of horizontal flipping, vertical flipping, and horizontal and vertical flipping are applied. This makes the dataset 4 times of the original dataset. The images' size is changed from 40×40 to 48×48 to fit to the model specification and effectively extract the required features.

B. Implementation and Training

The segmentation network contains repeated application of encoder and decoder blocks in the downsampling and upsampling paths which are trained on 65 epochs for effective training. The experiments on the augmented image dataset are conducted on a workstation with Intel® Xeon(R) Gold 5120 CPU @ 2.20GHz×56 with 93.1 GB RAM on Ubuntu 18.04.2 LTS operating system with NVIDIA Quadro P5000 with 16GB

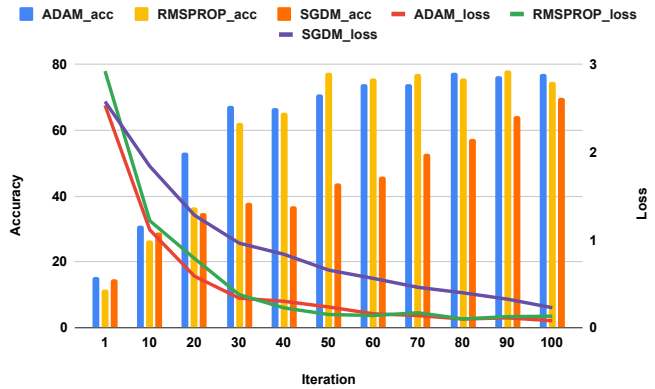

Fig. 4. Comparison of the proposed approach with itself for accuracy and loss values on the application of ADAM, RMSprop and SGDM optimizers on the proposed framework (till 100 iterations).

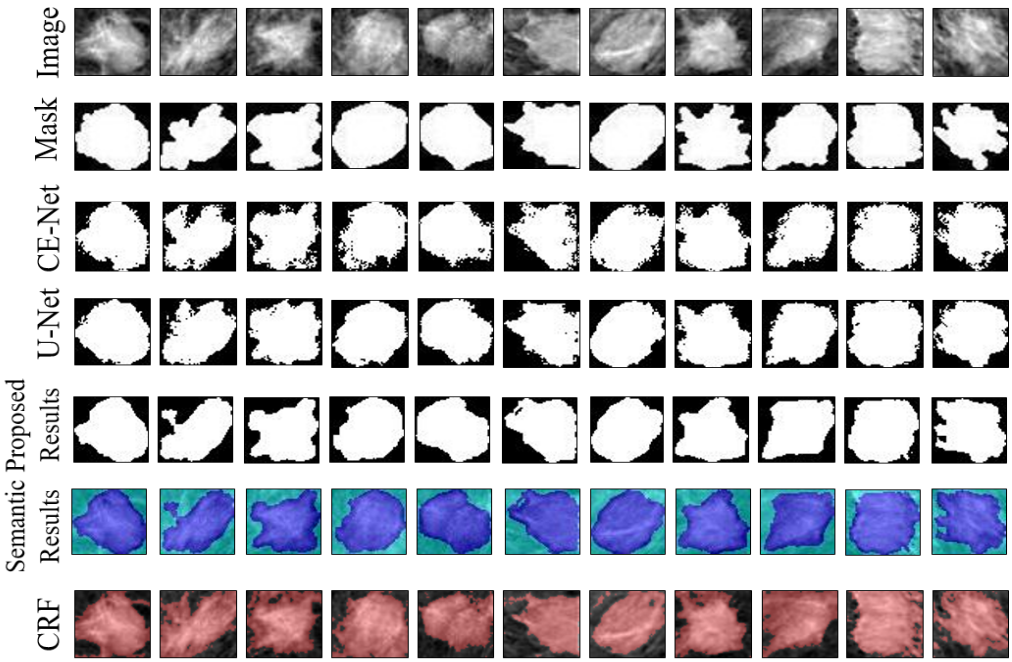

Fig. 5. Visualization of the segmentation results produced using CE-Net, U-Net and the proposed approach, RGU-Net. Blue region: are the semantic segmentation results obtained from the RGU-Net and Red region: is the CRF labeling on the produced binary segmentation results in row 7.

Graphics. The programming language of MATLAB R2019a is used. Hyper-parameter values are set for the implementation of the proposed model framework. For optimization, adaptive moment estimation (ADAM) [17] is used with cross-entropy loss at the pixel-classification layer with a learning rate of $1e-3$. After performing the ablation study with the model, the batch size is set to 16. A dropout of 0.5 is applied at the end of the encoder path and the start of the decoder path.

C. Results and Discussion

To better understand the proposed model, we have used Jaccard index (Jac), dice co-efficient (Dice), pixel accuracy (Acc), sensitivity (Sen), specificity (Spe), and Precision as the performance measures.

1) Ablation Studies: The residual identity connection with group convolution in the network has made the model pristine in finding the boundaries of the lesion. Ablation experiments are therefore performed on the proposed framework using the augmented BUS image dataset. Experimental results on the proposed model on the use of different optimizers with different epochs are tested and are tabulated in Table I. To further display the segmentation efficiency, Figure 3 shows the visual results for the ablation experiments. Figure 3 (b) displays the ground truth of the original input image of Figure 3 (a) with Figure 3 (c) and (d) depicting the predicted segmentation mask and their semantic segmentation masks by using ADAM optimizer. Similarly, Figure 3 (e) and (f) depicts the predicted segmentation mask and their semantic segmentation masks by using the root mean square gradient propagation (RMSPROP) optimizer. Figure 3 (g) and (h) shows the predicted segmentation mask and their semantic segmentation masks by using Stochastic gradient descent with momentum (SGDM) optimizer. Every column of the figure depicts a different input BUS image. It can be inferred that the results produced by ADAM optimizer with the proposed methods has outdone the other optimizers. However, the DDSM ROI-based image dataset [1] has more irregular ROIs when compared with the INBreast dataset. Therefore, when the proposed model is tested on DDSM dataset, it yielded

TABLE II
COMPARISON OF PROPOSED METHOD WITH STATE-OF-THE-ART METHODOLOGIES.

Method	**Dice (in %)**
DeepLab [12]	67.7
Deep Structure Learning [18]	88
FCN [19]	89.4
Structure Learning + CNN [20]	90
CE-Net [21]	91
Multi-FCN-CRF with Adversarial Training [19]	90.7
U-Net	92
RGU-Net (with SGDM)	88.5
RGU-Net (with RMSprop)	92.1
RGU-Net (with ADAM)	**92.6**

[1]https://drive.google.com/a/uci.edu/file/d/0B-7-8LLwONIZU0l2N3hXdU96Y2M/view?usp=sharing

all round size mass boundaries which are incorrect when compared to the given binary labels of the input images.

(i) Qualitative Analysis: Figure 4 shows the graph for the comparison of accuracy and loss among the application of different optimizers on the proposed approach. For better understanding, results for first 100 iterations (till 8 epochs) are shown which helps us to understand the performance of RGU-Net better. It is inferred that the loss value continues to decrease with the increasing iterations. Figure 5 manifests the segmentation results produced by the proposed framework. As it can be seen, the proposed method has produced better and comparable lesion contours of the masses with the heterogeneous boundaries.

(ii) Comparison with the state-of-the-art methodologies: To evaluate the segmentation performance of the proposed model, we validated Deep structured learning [18], FCN [19], Structure Learning + CNN [20], CE-Net [21], Multi-FCN-CRF with Adversarial Training [19] models on the same dataset under similar environment. DeepLab [12] results are obtained by predicting the test results from the previously trained DeepLab models. U-Net is the model with encoder depth 4 but without residual mapping among the convolution blocks. RGU-Net (with RMSprop) uses RMSprop optimizer with the proposed model. SGDM was also applied on the model, called RGU-Net (with SGDM). The comparative results are tabulated in Table II. It is to note that the proposed model, RGU-Net (with ADAM), outdoes the state-of-the-art methods for the acquired breast mass image dataset.

IV. Conclusion

This paper has proposed a semantic segmentation architecture, RGU-Net, for 2D BUS ROI breast mass image dataset. Proposed framework involves the ability of feature fusion using the advantages of group convolution and residual mappings in the convolution layers (conv layers) placed in the form of encoder-decoder fashion. CRF-based labeling on the produced segmentation maps has helped to analyze the mass boundaries accurately. Experiments on the proposed framework demonstrate that the method can accurately locate mass contours and perform pixel classification while retaining the contextual information. Publicly available INBreast dataset is used, which on introspection (ablation study) shows the model can perform well under different training strategies. Besides, the model is generic, which can be used for various biomedical image segmentation tasks.

Acknowledgments

The authors would like to thank Ministry of Human Resource Development (MHRD), Government of India, India for their constant support to carry out this research work.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[2] Y. Iwamoto, S. Kaucher, E. Lorenz, T. Bärnighausen, and V. Winkler, "Development of breast cancer mortality considering the implementation of mammography screening programs–a comparison of western european countries," *BMC public health*, vol. 19, no. 1, pp. 1–9, 2019.

[3] M. Jiang, S. Zhang, Y. Zheng, and D. N. Metaxas, "Mammographic mass segmentation with online learned shape and appearance priors," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 35–43.

[4] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig *et al.*, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," *Expert Systems with Applications*, vol. 139, p. 112855, 2020.

[5] M. Wodzinski and A. Skalski, "Resection-based demons regularization for breast tumor bed propagation," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 2018, pp. 3–12.

[6] N. Antropova, B. Huynh, and M. Giger, "Performance comparison of deep learning and segmentation-based radiomic methods in the task of distinguishing benign and malignant breast lesions on dce-mri," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. International Society for Optics and Photonics, 2017, p. 101341G.

[7] 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.

[8] R. Chaieb and K. Kalti, "Feature subset selection for classification of malignant and benign breast masses in digital mammography," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 803–829, 2019.

[9] D. Yi, R. L. Sawyer, D. Cohn III, J. Dunnmon, C. Lam, X. Xiao, and D. Rubin, "Optimizing and visualizing deep learning for benign/malignant classification in breast tumors," *arXiv preprint arXiv:1705.06362*, 2017.

[10] D. Yang, D. Xu, S. K. Zhou, B. Georgescu, M. Chen, S. Grbic, D. Metaxas, and D. Comaniciu, "Automatic liver segmentation using an adversarial image-to-image network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 507–515.

[11] B. Zhao, J. Soraghan, G. Di Caterina, and D. Grose, "Segmentation of head and neck tumours using modified u-net," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–4.

[12] 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.

[13] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and cnn segmentation," *Expert Systems with Applications*, vol. 42, no. 3, pp. 990–1002, 2015.

[14] I. Batatia, "A deep learning method with crf for instance segmentation of metal-organic frameworks in scanning electron microscopy images," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 625–629.

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

[16] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations (ICLR)*, 2015.

[18] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep structured learning for mass segmentation from mammograms," in *2015 IEEE international conference on image processing (ICIP)*. IEEE, 2015, pp. 2950–2954.

[19] W. Zhu, X. Xiang, T. D. Tran, G. D. Hager, and X. Xie, "Adversarial deep structured nets for mass segmentation from mammograms," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 847–850.

[20] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 605–612.

[21] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.