Improved Breast Mass Segmentation in Mammograms with Conditional Residual U-Net

Heyi Li[1(✉)], Dongdong Chen[1], William H. Nailon[2], Mike E. Davies[1], and David Laurenson[1]

[1] Institute for Digital Communications, University of Edinburgh, Edinburgh, UK
{heyi.li,d.chen,mike.davies,dave.laurenson}@ed.ac.uk
[2] Oncology Physics Department, Edinburgh Cancer Centre, Western General Hospital, Edinburgh, UK
bill.nailon@luht.scot.nhs.uk

Abstract. We explore the use of deep learning for breast mass segmentation in mammograms. By integrating the merits of residual learning and probabilistic graphical modelling with standard U-Net, we propose a new deep network, Conditional Residual U-Net (CRU-Net), to improve the U-Net segmentation performance. Benefiting from the advantage of probabilistic graphical modelling in the pixel-level labelling, and the structure insights of a deep residual network in the feature extraction, the CRU-Net provides excellent mass segmentation performance. Evaluations based on INbreast and DDSM-BCRP datasets demonstrate that the CRU-Net achieves the best mass segmentation performance compared to the state-of-art methodologies. Moreover, neither tedious pre-processing nor post-processing techniques are not required in our algorithm.

Keywords: Mammogram mass segmentation · Structured prediction Deep residual learning

1 Introduction

Breast cancer is the most frequently diagnosed cancer among women across the globe. Among all types of breast abnormalities, breast masses are the most common but also the most challenging to detect and segment, due to variations in their size and shape and low signal-to-noise ratio [6]. An irregular or spiculated margin is the most important feature in indicating a cancer. The more irregular the shape of a mass, the more likely the lesion is malignant [12]. Oliver *et al.* demonstrated in their review paper that mass segmentation provides detailed morphological features with precise outlines of masses, and plays a crucial role in a subsequent cancerous classification task [12].

The main roadblock faced by mass segmentation algorithms is the insufficient volume of contour delineated data, which directly leads to inadequate

accuracy [4]. The U-Net [13], as a Convolutional Neural Network (CNN) based segmentation algorithm, is shown to perform well with limited training data by interlacing multi-resolution information. However, the CNN segmentation algorithms including the U-Net are limited by the weak consistency of predicted pixel labels over homogeneous regions. To improve the labelling consistency and completeness, probabilistic graphical models [5] have been applied for mass segmentation, including Structured Support Vector Machine (SSVM) [7] and Conditional Random Field (CRF) [6] as a post-processing technique. To train the CRF integrated network in an end-to-end way, the CRF with the mean-field inference is realised as a recurrent neural network [14]. This is applied on mass segmentation [15], and achieved the state-of-art mass segmentation performance. Another limitation of CNN segmentation algorithms is that as the depth of the CNNs increase for better performing deep features, they may suffer from the gradient vanishing and exploding problems, which are likely to hinder the convergence [8]. Deep residual learning is shown to address this issue by mapping layers with residuals explicitly instead of mapping the deep network directly [8].

In this work, the CRU-Net is proposed to precisely segment breast masses with small-sample-sized mammographic datasets. Our main contributions include: (1) the first neural network based segmentation algorithm that considers both pixel-level labelling consistency and efficient training via integrating the U-Net with CRF and deep residual learning; (2) the first deep learning mass segmentation algorithm, which does not require any pre-processing or post-processing techniques; (3) the CRU-Net achieves the best mass segmentation performance on the two most commonly used mammographic datasets when compared to other related methodologies.

2   Methodology

The proposed algorithm CRU-Net is schematically shown in Fig. 1. The inputs are mammogram regions of interest (ROIs) that contain masses and the outputs
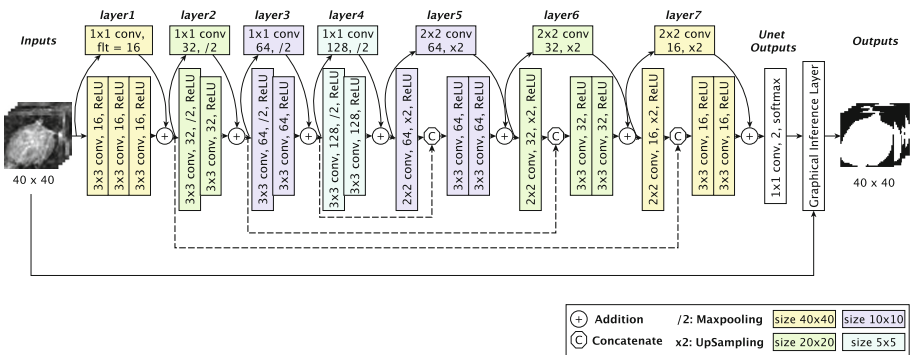

Fig. 1. Proposed CRU-Net Structure

are the predicted binary images. In this section, a detailed description of applied methods is introduced: our U-Net with residual learning, followed by the pixel-level labelling with graphical inference.

2.1 U-Net with Residual Learning

The U-Net is shown to perform well with a limited volume of training data for segmentation problems in medical imaging [13], which suits our situation. However, the gradient vanishing and explosion problem, which hinders the convergence, is not considered in the U-Net. We integrate residual learning into the U-Net to precisely segment breast masses over a small sample size training data. Assuming $\boldsymbol{x} : \Omega \rightarrow \mathbb{R}$ (Ω represents the image lattice) as an ROI and $\boldsymbol{y} : \Omega \rightarrow \{0,1\}$ as the corresponding binary labelling image (0 denotes background pixels and 1 for the mass pixels), the training set can be represented by $\mathcal{D} = \{(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})\}_{n \in \{1,\ldots,N\}}$.

The U-Net comprises of a contractive downsampling and expansive upsampling path with skip connections between the two parts, which makes use of standard convolutional layers. The output of mth layer with input $\boldsymbol{x}^{(n)}$ at pixel (i, j) is formulated as follows:

$$y_{i,j}^{(n,m)} = h_{ks}(\{\boldsymbol{x}_{s_i+\delta_i, s_j+\delta_j}\}_{0 \leq \delta_i, \delta_j \leq k}) \tag{1}$$

where k represents for kernel size, s for stride or maxpooling factor, and h_{ks} is the layer operator including convolution, maxpooling and the ReLU activation function.

Then we integrate the residual learning into the U-Net, which solves the applied U-Net network mapping $\mathcal{H}(\boldsymbol{x})$ with:

$$\mathcal{F}(\boldsymbol{x}) := \mathcal{H}(\boldsymbol{x}) - W * \boldsymbol{x} \tag{2}$$

thus casting the original mapping into $\mathcal{F}(\boldsymbol{x}) + W * \boldsymbol{x}$, where W is a convolution kernel and linearly projects x to match $\mathcal{F}(\boldsymbol{x})$'s dimensions as Fig. 1. As the U-Net layers resize the image, residuals are linearly projected either with 1×1 kernel convolutional layer along with maxpooling or upsampling and 2×2 convolution to match dimensions. The detailed residual connections of layer 2 and layer 6 are described in Fig. 2. These layers are shown as examples as all residual layers have analogous structure. In the final stage, a 1×1 convolutional layer with softmax activation creates a pixel-wise probabilistic map of two classes (background and masses). The residual U-Net loss energy for each output during training is defined with categorical cross-entropy. Mathematically,

$$f = -\sum_{i,j} \log P\big(y_{i,j}^{(n)} \mid \boldsymbol{x}^{(n)}; \boldsymbol{\theta}\big) \tag{3}$$

where P is the residual U-Net output probability distribution at position (i, j) given the input ROI $\boldsymbol{x}^{(n)}$ and parameters $\boldsymbol{\theta}$.
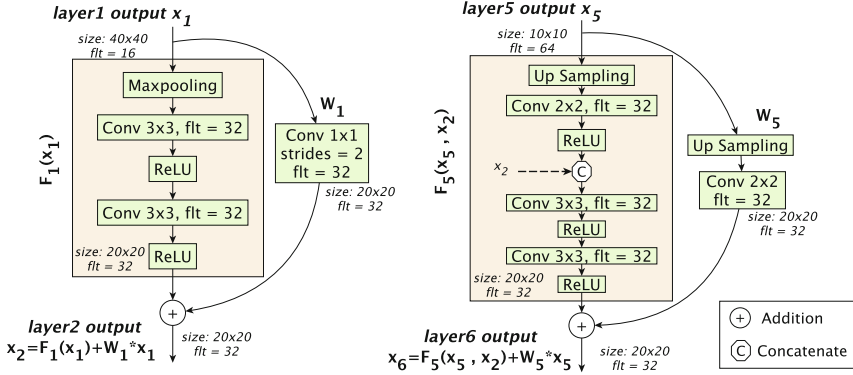
Fig. 2. Residual Learning illustration for layer2 and layer6. Other layers are equivalent to this example but with different parameters.

Note that the standard U-Net is designed for images of size 572×572. Here we modify the standard U-Net to adapt mammographic ROIs (40×40) with zero-padding for downsampling and upsampling. Residual short-cut additions are calculated in each layer. After that, feature maps are concatenated as: layer 1 with layer 7, layer 2 with layer 6, layer 3 with layer 5 as shown in Fig. 1. Both original ROIs and U-Net Outputs are then fed into the graphical inference layer.

2.2   Graphical Inference

Graphical models are recently applied on mammograms for mass segmentation. Among them, CRF incorporates the label consistency with similar pixels and provide sharp boundary and fine-grained segmentation. Mean field iterations are applied as the inference method to realise the CRF as a stack of RNN layers [14,15]. The cost function for CRF (g) can be defined as follows:

$$g = A(\boldsymbol{x}^{(n)}) - \exp \big(\sum_{i,j \in V} P\big(y_{i,j}^{(n)}\big) + \sum_{p,q \in E} \phi(y_p^{(n)}, y_q^{(n)} \mid \boldsymbol{x}^{(n)}) \big) \tag{4}$$

where A is the partition function, P is the unary function which is calculated on the residual U-Net output, and ϕ is the pair-wise potential function which is defined with the label compatibility $\mu(y_p^{(n)}, y_q^{(n)})$ for position p and q [14], Gaussian kernels k_G^1, k_G^2 and corresponding weights $\omega_G^{(1)}$, $\omega_G^{(2)}$ [10] as $\phi(y_p^{(n)}, y_q^{(n)} \mid \boldsymbol{x}^{(n)}) = \mu(y_p^{(n)}, y_q^{(n)})\big(\omega_G^{(1)} k_G^{(1)}(\boldsymbol{x}^{(n)}) + \omega_G^{(2)} k_G^{(2)}(\boldsymbol{x}^{(n)})\big)$ [6,15].

Finally, by integrating (3) and (4) the total loss energy in the CRU-Net for each input $\boldsymbol{x}^{(n)}$ is defined as:

$$\ell = (1 - \lambda)f + \lambda \cdot g(f, \boldsymbol{x}^{(n)}) \tag{5}$$

where $\lambda \in [0, 1]$ is a trade-off factor, which is empirically chosen as 0.67. And the whole CRU-Net is trained by backpropagation.

3   Experiments

3.1   Datasets

The proposed method is evaluated on two publicly available datasets INbreast [11] and DDSM-BCRP [9]. INBreast is a full-field digital mammographic dataset ($70\,\mu$ m pixel resolution), which is annotated by a specialist with lesion type and detailed contours for each mass. 116 accurately annotated masses are contained with mass size ranging from $15\,mm^2$ to $3689\,mm^2$. The DDSM-BCRP [9] database is selected from the Digital Database for Screening Mammography (DDSM) database, which contains digitized film screen mammograms (43.5 microns resolution) with corresponding pixel-wise ground truth provided by radiologists.

To compare the proposed methods with other related algorithms, we use the same dataset division and ROIs extraction as [6,7,15], in which ROIs are manually located and extracted with rectangular bounding boxes and then resized into 40×40 pixels using bicubic interpolation [6]. In work [6,7,15], extracted ROIs are pre-processed with the Ball and Bruce technique [1], which our algorithms do not require. The INbreast dataset is divided into 58 training and 58 test ROIs; The DDSM-BCRP is divided into 87 training and 87 test ROIs [6]. The training data is augmented by horizontal flip, vertical flip, and both horizontal and vertical flip.

3.2   Experiment Configurations

In this paper, each component of the CRU-Net is experimented, including $\lambda = 0, 1, 0.67$ and the CRU-Net without residual learning (CRU-Net, No R). In the CRU-Net, convolutions are first computed with kernel size 3×3, which are then followed by a skip to compute the residual as shown in Fig. 1. The feature maps in each downsampling layer are of size 16, 32, 64, and 128 respectively, while the ROIs spatial dimensions are 40×40, 20×20, 10×10 and 5×5. To avoid over-fitting, dropout layers are involved with 50% dropout rate. The resolution of two datasets are different, with the DDSM's much higher than the INbreast's. To address this, the convolutional kernel size for DDSM is chosen as 7×7 by experimental grid search. All other hyper parameters are identical. The whole CRU-Net is optimized by the Stochastic Gradient Descent algorithm with the Adam update rule.

3.3   Performance and Discussion

All state-of-art methods and the CRU-Net' performances are shown in the Table 1, where [15] are reproduced, results of [2,3,6,7] are from their papers. Table 1 shows that our proposed algorithm performs better than other published algorithms on both data sets. In INBreast, the best Dice Index (DI) 93.66% is obtained with CRU-Net, No R ($\lambda = 0.67$) and a similar DI 93.32% is achieved by its residual learning; while in DDSM-BCRP, all state-of-art algorithm performs

similarly and the best DI 91.43% is obtained by CRU-Net ($\lambda = 0$). The CRU-Net performs worse on DDSM-BCRP than INbreast, which is because of its worse data quality. To better understand the dice coefficients distribution in test sets,

Table 1. Mass segmentation performance (DI, %) of the CRU-Net and several state-of-art methods on test sets. λ is the trade off loss factor as (5).

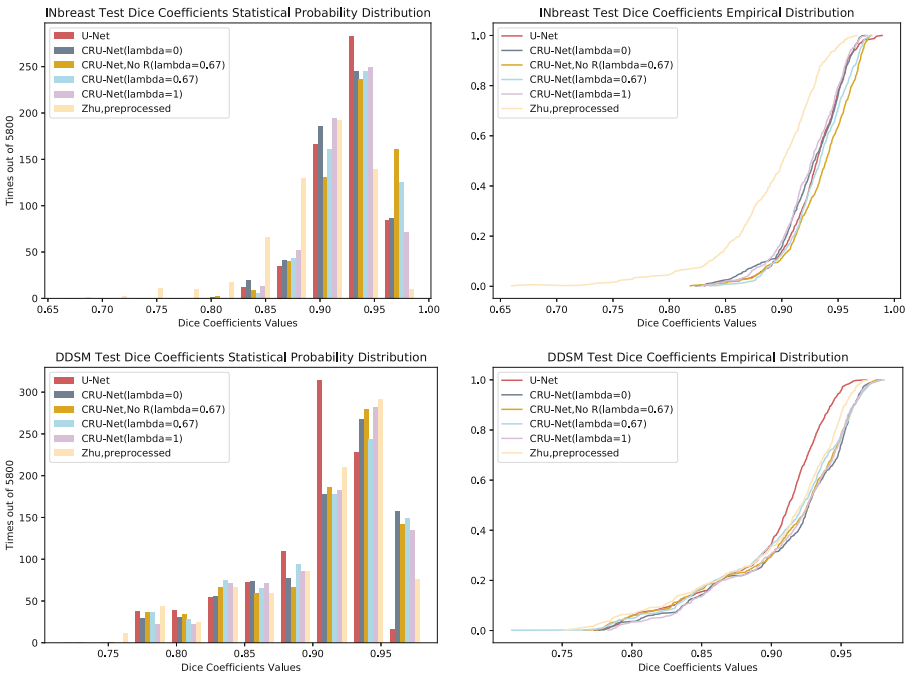| Methodology	INbreast	DDSM-BCRP	Residual	Preprocess	Postprocess
Cardoso *et al.* [3]	88	-	-	-	-
Beller *et al.* [2]	-	70	-	-	-
Dhungel *et al.* [7]	88	87	×	✓	✓
Dhungel *et al.* [6]	90	90	×	✓	✓
Zhu *et al.* [15]	89.36 ± 0.37	90.62 ± 0.16	×	✓	×
U-Net	92.99 ± 0.23	90.08 ± 0.62	×	×	×
CRU-Net ($\lambda = 0$)	92.72 ± 0.09	$\mathbf{91.43 \pm 0.02}$	✓	×	×
CRU-Net ($\lambda = 1$)	92.60 ± 0.24	91.41 ± 0.02	✓	×	×
CRU-Net, No R ($\lambda = 0.67$)	$\mathbf{93.66 \pm 0.10}$	91.14 ± 0.09	×	×	×
CRU-Net ($\lambda = 0.67$)	93.32 ± 0.12	90.95 ± 0.26	✓	×	×

Fig. 3. Test Dice Coefficients Distribution of INbreast Dataset. The first row shows the distribution of INbreast dataset and the second row shows DDSM's. The left figures depict the histogram of test dice coefficients and the rights show the sampled cumulative distribution.

Fig. 3 shows the histogram of dice coefficients and sampled cumulative distribution of two datasets. In those figures we can observe that the CRU-Net achieves a higher proportion of cases with DI > 95%. In addition, all algorithms follow a similar distribution, but Zhu's algorithm has a bigger tail than others on the INbreast data. To visually compare the performances, example contours from the CRU-Net ($\lambda = 0.67$) and Zhu's algorithms are shown in Fig. 4. It depicts that while achieving a similar DI value to Zhu's method, the CRU-Net obtains a less noisy boundary. To examine the tail in Zhu's DIs histogram (DI $\leq$ 81%), Fig. 5 compares the contours of the hard cases, which suggests that the proposed CRU-Net provides better contours for irregular shape masses with less noisy boundaries.
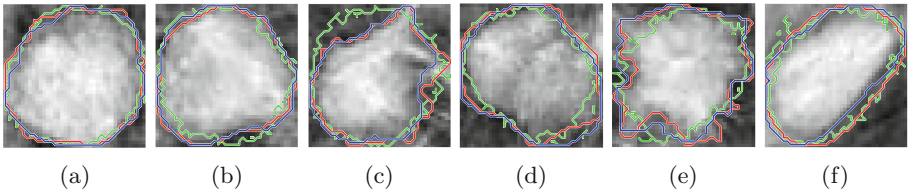

(a)   (b)   (c)   (d)   (e)   (f)

Fig. 4. Visualized comparison of segmentation results (DI > 81%) between CRU-Net and Zhu's work. Red lines denote the radiologist's contour, blue lines are the CRU-Net's results ($\lambda = 0.67$), and green lines denote Zhu's method results. (Color figure online)
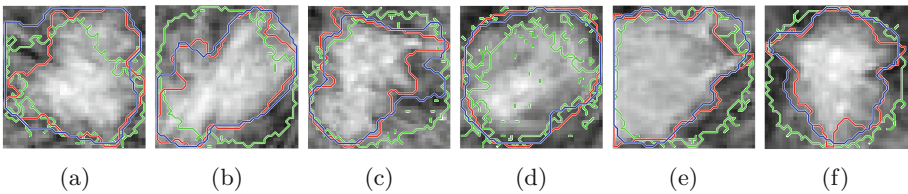

(a)   (b)   (c)   (d)   (e)   (f)

Fig. 5. Visualized comparison of segmentation results between CRU-Net ($\lambda = 0.67$) method and Zhu's work on the 5 hardest cases, when Zhu's DI $\leq$ 81%. Red lines denote the radiologist's contour, blue lines are the CRU-Net's results, and green lines are from Zhu's method. From (a) to (f), Zhu's DIs are: 70.16%, 73.47%, 76.11%, 72.95%, 80.36% and 79.98%. The CRU-Net's corresponding DIs are: 87.51%, 92.43%, 88.52%, 95.01%, 93.50% and 91.33%. (Color figure online)

4   Conclusions

In summary, we propose the CRU-Net to improve the standard U-Net segmentation performance via incorporating the advantages of probabilistic graphic models and deep residual learning. The CRU-Net algorithm does not require any tedious preprocessing or postprocessing techniques. It outperforms published

state-of-art methods on INbreast and DDSM-BCRP with best DIs as 93.66% and 91.14% respectively. In addition, it achieves higher segmentation accuracy when the applied database is of higher quality. The CRU-Net provides similar contour shapes (even for hard cases) to the radiologist with less noisy boundary, which plays a vital role in subsequent cancerous diagnosis.

References

1. Ball, J.E., Bruce, L.M.: Digital mammographic computer aided diagnosis (CAD) using adaptive level set segmentation. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2007, pp. 4973–4978. IEEE (2007)

2. Beller, M., Stotzka, R., Müller, T.O., Gemmeke, H.: An example-based system to support the segmentation of stellate lesions. In: Meinzer, H.P., Handels, H., Horsch, A., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2005. Springer, Heidelberg (2005). https://doi.org/10.1007/3-540-26431-0_97

3. Cardoso, J.S., Domingues, I., Oliveira, H.P.: Closed shortest path in the original coordinates with an application to breast cancer. Int. J. Pattern Recogn. Artif. Intell. 29(01), 1555002 (2015)

4. Carneiro, G., Zheng, Y., Xing, F., Yang, L.: Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis. In: Lu, L., Zheng, Y., Carneiro, G., Yang, L. (eds.) Deep Learning and Convolutional Neural Networks for Medical Image Computing. ACVPR, pp. 11–32. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-42999-1_2

5. Chen, D., Lv, J., Yi, Z.: Graph regularized restricted Boltzmann machine. IEEE Trans. Neural Netw. Learn. Syst. 29(6), 2651–2659 (2018)

6. Dhungel, N., Carneiro, G., Bradley, A.P.: Deep learning and structured prediction for the segmentation of mass in Mammograms. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 605–612. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_74

7. Dhungel, N., Carneiro, G., Bradley, A.P.: Deep structured learning for mass segmentation from mammograms. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 2950–2954. IEEE (2015)

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

9. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.: The digital database for screening mammography. In: Digital Mammography, pp. 431–434 (2000)

10. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Advances in Neural Information Processing Systems, pp. 109–117 (2011)

11. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. Acad. Radiol. 19(2), 236–248 (2012)

12. Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E.R., Zwiggelaar, R.: A review of automatic mass detection and segmentation in mammographic images. Med. Image Anal. 14(2), 87–110 (2010)

13. 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
14. Zheng, S., et al.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1529–1537 (2015)
15. Zhu, W., Xiang, X., Tran, T.D., Hager, G.D., Xie, X.: Adversarial deep structured nets for mass segmentation from Mammograms. arXiv preprint arXiv:1710.09288 (2017)