Boundary Regularized Convolutional Neural Network for Layer Parsing of Breast Anatomy in Automated Whole Breast Ultrasound

Cheng Bian[1], Ran Lee[1], Yi-Hong Chou[2], and Jie-Zhi Cheng[3(✉)]

[1] School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China
[2] Department of Radiology, Taipei Veterans General Hospital, Taipei 11217, Taiwan
[3] Department of Electrical Engineering, Chang Gung University, Taoyuan 33302, Taiwan
jzcheng@ntu.edu.tw

Abstract. A boundary regularized deep convolutional encoder-decoder network (ConvEDNet) is developed in this study to address the difficult anatomical layer parsing problem in the noisy Automated Whole Breast Ultrasound (AWBUS) images. To achieve better network initialization, a two-stage adaptive domain transfer (2DT) is employed to land the VGG-16 encoder on the AWBUS domain with the bridge of network training for AWBUS edge detector. The knowledge transferred encoder is denoted as VGG-USEdge. To further augment the training of ConvEDNet, a deep boundary supervision (DBS) strategy is introduced to regularize the feature learning for better robustness to speckle noise and shadowing effect. We argue that simply counting on the image context cue, which can be learnt with the guidance of label maps, may not be sufficient to deal with the intrinsic noisy property of ultrasound images. With the regularization of boundary cue, the segmentation learning can be boosted. The efficacy of the proposed 2DT-DBS ConvEDNet is corroborated with the extensive comparison to the state-of-the-art deep learning segmentation methods. The segmentation results may assist the clinical image reading, particularly for junior medical doctors and residents and help to reduce false-positive findings from a computer-aided detection scheme.

Keywords: Segmentation · Ultrasound · Breast · Deep learning

1 Introduction

Automated whole breast ultrasound (AWBUS) is a new medical imaging modality approved by FDA in 2012. The AWBUS technology can automatically depict the whole anatomy of breast in a 3D volume, and hence enable the chance of offline thorough image reading. However, the advantage of the volumetric imaging may also introduce more workload for radiologists. Even for a senior radiologist, the reading of a AWBUS volume can take tens of minutes to reach confident diagnostic workup, due to the large image information and difficulty of interpretation of ultrasound images. Accordingly, the manpower shortage of radiologists can be possibly expected with the popularization of this new imaging technology. To improve the reading efficiency, an automatic

M. Descoteaux et al. (Eds.): MICCAI 2017, Part III, LNCS 10435, pp. 259–266, 2017.
DOI: 10.1007/978-3-319-66179-7_30

segmentation method is proposed in this study to parse the AWBUS images into the breast anatomic layers of subcutaneous fat, breast parenchyma, pectolis muscles and chest wall. The layer decomposition is shown in Fig. 1. The layer parsing of breast anatomy in the AWBU images can assist the clinical image reading for less-experienced radiologists and residents. Meanwhile, the breast density, which is an important biomarker for cancer risk [1, 13], can be easily computed with the parsing results. In the context of computer-aided detection, the layer parsing may also help to exclude false-positive detections [2].

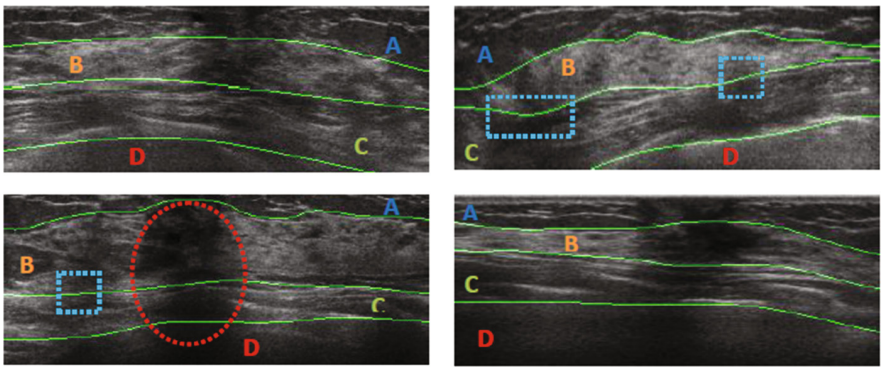

Fig. 1. Illustration of anatomical layers in AWBUS. (A), (B), (C) and (D) indicate layers of subcutaneous fat, breast parenchyma, muscle and chest wall, respectively. The green lines are septa boundaries of layers. The red dotted circle indicate a significant shadowing effect, whereas the blue dotted rectangles suggest regions that are difficult for layer differentiation.

Referring to Fig. 1, the segmentation task for layer parsing in AWBUS images can be very challenging. It shall not only deal with the intrinsic low quality properties of ultrasound images like the speckle noise and shadowing effect, but also tackle the distribution overlapping problem of echogenicity among different breast anatomical layers. The low image quality and echogenicity overlapping problems may also lead to ill-defined septa boundaries in places between the consecutive layers, see Fig. 1, and hence render layer parser task more problematic. On the other hand, the appearance and morphology of breast anatomic layers can vary significantly from person to person. For people with less breast density, the fat layer can be thicker, whereas the AWBUS images that depict dense breasts may have larger breast parenchyma. Therefore, the issue of high inter-subject variability for the anatomical structures should also be well considered in the design of layer parsing algorithm.

In the literature, most works focused on the segmentation of breast lesions in ultrasound images [14, 15]. To our best knowledge, this is less related work for breast anatomy parsing in AWBUS images. In this study, we propose to leverage the deep convolutional encoder-decoder network, ConvEDNet for short [3–6], with the **2-stages domain transfer (2DT)** and **deep boundary supervision (DBS)**, i.e., deep supervision [7] with boundary cue, techniques for layer parsing in the AWBUS images. The

ConvEDNet [3–6] can perform end-to-end training for the semantic segmentation, and is typically constituted with the convolutional (encoder) and deconvolutional (decoder) paths, which learn useful object features and restore the object geometry and morphology at the resolution of the input images, respectively. The learning of ConvEDNet is majorly based on the image context cues with the guidance of object labels [3–6]. However, as discussed earlier, simply usage of the image context cues may not be sufficient to address the issues of low image quality, ill-defined septa boundaries, etc. Accordingly, we further incorporate the boundary cue drawn by radiologists with auxiliary learning techniques of 2DT and DBS to boost the training of ConvEDNet. The cues of boundary and image context are complementary to each other, and can be synergized to achieve promising image analysis performance [8]. The details of 2DT and DBS will be elaborated latter.

The proposed 2DT-DBS ConvEDNet is extensively compared with the state-of-the-art ConvEDNets, i.e., FCN [3], DeconvNet [4], SegNet [5] and U-Net [6]. We also perform the ablation experiments to illustrate the effectiveness of the implementation of the 2DT and DBS techniques. One related deep learning method [8] which also fuses the image context and boundary cues for the training of FCN in the multi-task paradigm is implemented for comparison. Specifically, in [8] the object labeling and boundary delineation are treated as two tasks to co-train the FCN. Our formulation on the other hand adopts the deep supervision strategy to augment the feature learning in the encoder path with the auxiliary boundary cue that encodes the object geometry and morphology. With the extensive experimental comparison, it will be shown that the proposed 2DT-DBS ConvEDNet can outperform other baseline methods for the layer parsing of breast anatomy in AWBUS.

2 Method

The architecture of our network is illustrated in Fig. 2. The mainstream architecture is a ConvEDNet. The segmentation is based on 2D AWBUS images in this study. The data annotation and the learning/testing of layer parsing methods are performed independently on the sagittal and axial 2D views of AWBUS images, and the final annotation and layer parsing results are reached by averaging the three septa boundaries of the four layers from the two corresponding boundaries of the two 2D views.

Since the ultrasound data are relatively noisy, the segmentation capability of the encoder-decoder path in ConvEDNet may not be sufficient to address the challenging issues in our problem. In this study, we employ the 2DT and DBS to augment the network training. As shown in Fig. 2, our network is equipped with five auxiliary side networks to impart the boundary knowledge to regularize the feature learning.

The computational breast anatomy decomposition in the AWBUS images is formulated as a pixel classification problem with four classes of subcutaneous fat, breast parenchyma, pectolis muscle, and chest wall. Given the annotated label map set, C, and the original 2D AWBUS image set, X, the training of the ConvEDNet tries to seek proper neural parameters, W_c, with the minimization of the loss function:

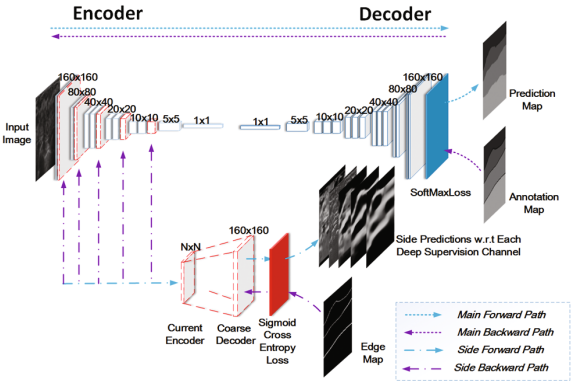

Fig. 2. Our boundary regularized ConvEDNet architecture. For the encoder layer size of each side network, N is the size of the connecting layer of mainstream encoder.

$$\mathcal{L}(C,X;W_c) = \mathcal{L}_c(C,X;W_c) + \|W_c\|_2, \tag{1}$$

where $\mathcal{L}_c(\cdot)$ is the cross entropy function [12], and $\|\cdot\|_2$ is the L_2 norm for regularization. The minimum of the loss function (1) can be sought by stochastic gradient descend for the end-to-end learning of segmentation.

2.1 Our Mainstream ConvEDNet (MConvEDNet)

Similar to [4], the encoder of the MConvEDNet is composed of VGG-16 [9] net with removal of the last classification layer, see Fig. 2. We change the kernel size of conv6 and deconv6 as 5 to fit our data. The unpooling layers at the decoder are paired with the max-pooling layers of the encoder. The locations of maximum activations at the max-pool layers are memorized with switch variables to assist the unpooling process.

2.2 Two-Stages Domain Transfer (2DT)

Since the cost for collection and annotation of medical images is relatively expensive, the common approach to attain good performance with the deep learning technique is to initialize network with parameters learnt from natural images [10]. However, considering that the domains of natural and AWBUS images are quite different, we propose to engage the knowledge transfer of model parameters in two stages. Specifically, the first stage of domain transfer is carried out to employ the VGG-16 [9] as the encoder followed by a decoder with single deconvolutional layer for the anatomical edge detection in AWBUS images. The learning of the edge detector is guided by the boundary maps, where the three septa boundaries of the four layers are drawn. To boost the learning of edge detection, deep supervision with the boundary maps is also implemented by the same 5 auxiliary side networks shown in Fig. 2. This type of edge detector network is also called Holistically-nested Edge Detector (HED) net [11]. The training for the AWBUS edge detector will land VGG-16 encoder into the AWBUS domain to be

familiar with the presence of speckle noise and shadowing effect. Similar to [11], the AWBUS edge detection is formulated as a 2-class differentiation with edge label as 1 whereas non-edge label as 0. The learnt encoder for the AWBUS edge detector is denoted as VGG-USEdge. The tasks of anatomic edge detection and layer parsing may relate to each but remain different. Therefore, the VGG-USEdge encoder may provide more useful prior knowledge than VGG-16. Accordingly, the VGG-USEdge is applied to initialize the encoder network of our MConvEDNet.

2.3 Deep Boundary Supervision (DBS)

As can be found in Fig. 2, the MConvEDNet is relatively deep and hence the gradient vanishing issue can possibly occur in the network training. Meanwhile, the learning process can also be thwarted with the difficult issues discussed earlier. To further boost the learning process, the deep supervision strategy is employed. Here, we introduce the cue of layer boundaries with the deep supervision strategy to improve the learning. To further illustrate the efficacy of boundary cue, we further implement two comparison options. The first option is the deep supervision with label map cue on MConvEDNet. The second one is to perform the DBS on both encoder and decoder, which totally have 10 auxiliary side networks. It will be shown that the pure deep DBS can better boost the segmentation than the other deep supervision strategies.

The DBS is realized by adding auxiliary side networks to the endings of 5 layers in the encoder of MConvEDNet. The auxiliary side networks are shallow and simply constitutes of coupled single convolutional and deconvolutional layers, see Fig. 2. Given the neural parameters, W_e^p, of an auxiliary side network p, $1 \le p \le Q$; Q is the total number of convolutional layers at the encoder, and the edge map set of layer boundaries, E, the learning of the end-to-end segmentation with the DBS can be realized by the minimization of the reformulated loss function of

$$\mathcal{L}\left(C, X; W_c\right) = \mathcal{L}_c\left(C, X; W_c\right) + \left\|W_c\right\|_2 + \sum\nolimits_p^Q \mathcal{L}_e\left(E, X; W_e^p\right), \tag{2}$$

where $\mathcal{L}_e(\cdot)$ is the class-balanced cross entropy function for the auxiliary side networks that considers the non-balance issue between edge and non-edge classes [11]. With the minimization of cost function, the encoder network be equipped with the capability to drive the prediction masks of mainstream network as close to the manual label maps as possible, and keep the output edge maps of the side networks not deviating from the annotated edge maps significantly. For the comparing implementation of deep supervision with label map cue, we can simply replace the training map set E with C. The deep supervision with both label and edge map cues need two parallel side networks which consider training map sets of C and E, respectively.

2.4 Implementation Details

The learning rate of the mainstream network is initialized as 0.01, while the weight decay and momentums parameters are set as 0.0005 and 0.9, respectively. For the auxiliary

networks of deep supervision, the learning rates are 10^{-6}, where the parameters of weight decay and momentum are the same as those of MConvEDNet. No dropout is implemented but the batch normalization is adopted. The architectures of auxiliary side networks for the edge detector net and MConvEDNet are the same for simplicity, but with different random initialization on network parameters. Our method is developed based on the Caffe environment [12].

3 Dataset and Annotation

The AWBUS data were collected from Taipei Veterans General Hospital, Taipei, Taiwan, with the approval of their institutional review board (IRB). 16 AWBUS volumes acquired from 16 subjects are involved in this study. The subject ages range from 30 to 62. The non-human dark regions of all AWBUS images are excluded and leaving all image contents with the size of 160×160. The annotation for the boundaries of the breast anatomical layers in the AWBUS images was performed by a radiologist with 5 years of experience in breast ultrasound. The annotated data were further reviewed by a very senior radiologist with experience of medical ultrasound more than 30 years to ensure the correctness of the annotation. Each AWBUS volume contains around 170–200 2D images and the overall number of 2D images is 3134.

4 Experiments and Results

The evaluation of the AWBUS image segmentation is based on leave-one-out cross validation (LOO-CV). The basic unit of LOO-CV is an **AWBUS volume** but not a 2D image. Two assessment metrics, intersection over union (IoU) [4] and curve distance (CD) [15], are adopted for the quantitative evaluation between the computerized segmentation results and manual annotations. The CD is the averaged absolute distance between two comparing lines. The state-of-the-art ConvEDNets of FCN, DeconvNet, SegNet and U-Net are also implemented as baseline methods for comparison. Meanwhile, the multi-task method [8], denoted as "Multitask" which fuses image context and boundary cues is also implemented for comparison. The combinational options of 2DT and DBS are also implemented to show the effect of each technique on our problem. As discussed in Sect. 2.3, to illustrate efficacy of DBS, the implementation of DBS on encoder and decoder (FullyDBS) and deep supervision with label map (DLS) are also performed. To show the effectiveness of 2DT, we also implement the random parameter initialization (RandI) for the DBS+ConvEDNet.

Table 1 reports the mean $\pm$ standard deviation statistics of the CD and IoU metrics for the segmentation results of each implementation over the LOO-CV scheme. Specifically, the segmentation performances w.r.t. the three septa boundaries in-between layers (CD) and four anatomic layers (IoU) are listed in the columns of Table 1. The layers A, B, C and D represents fat, breast parenchyma, pectolis muscles and chest wall, respectively. The lines 1, 2 and 3 are the septa boundaries w.r.t. the layer pairs of "A/B", "B/C", and "C/D". It is worth noting that the our MConvEDNet is based on the

DeconvNet [4]. To give the visual comparison, the segmentation results of all methods involved in this study are listed in Fig. 3.

Table 1. Segmentation performances of different methods. "Main" represents our mainstream ConvEDNet (MConvEDNet). It is worth noting that the encoders of "DeconvNet" and "DBS +Main" are initialized with VGG-16.

Metrics	CD (pixel)			IoU (%)			
Method	Line1	Line2	Line3	A	B	C	D
FCN [3]	8.0 ± 4.8	9.2 ± 4.9	10.6 ± 7.7	74.2 ± 11.5	50.2 ± 20.2	62.8 ± 13.7	72.5 ± 16.6
SegNet [5]	7.8 ± 7.0	11.6 ± 6.8	13.2 ± 8.9	75.3 ± 14.3	50.9 ± 18.6	54.7 ± 18.6	67.1 ± 17.7
U-Net [6]	6.36 ± 5.9	9.98 ± 6.3	11.9 ± 8.4	76.7 ± 13.2	53.8 ± 17.9	57.4 ± 17.5	68.3 ± 18.5
DeconvNet [4]	4.7 ± 4.1	6.6 ± 4.9	9.2 ± 7.9	82.8 ± 9.4	67.0 ± 16.8	69.3 ± 15.6	74.7 ± 16.6
DLS+Main	5.1 ± 4.3	6.7 ± 4.3	10.0 ± 7.5	81.5 ± 10.4	65.7 ± 15.9	67.8 ± 13.0	73.9 ± 16.5
FullyDBS+Main	5.9 ± 5.2	7.5 ± 4.6	10.4 ± 7.9	79.7 ± 10.7	61.8 ± 17.9	65.9 ± 14.0	72.3 ± 16.4
DBS+Main	4.2 ± 3.6	5.9 ± 4.1	9.1 ± 7.2	84.4 ± 8.3	69.3 ± 15.5	70.2 ± 12.9	75.5 ± 16.1
DBS+Main+2DT	**3.9 ± 3.6**	**5.6 ± 3.9**	**8.3 ± 7.0**	**86.8 ± 7.9**	**72.2 ± 14.6**	**72.4 ± 12.6**	**76.1 ± 15.9**
DBS+Main+RndI	10.5 ± 7.6	13.7 ± 7.1	14.6 ± 9.7	69.1 ± 14.4	60.6 ± 17.5	50.8 ± 19.1	64.5 ± 18.7
Multitask [8]	4.9 ± 4.1	6.9 ± 4.3	9.6 ± 7.4	82.2 ± 9.2	64.9 ± 16.6	67.4 ± 13.5	74.4 ± 16.1

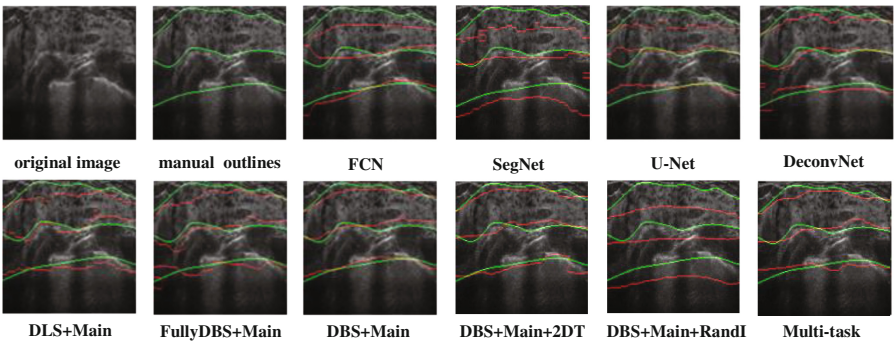

Fig. 3. Visual Comparison for the layer parsing results from different implementations. The layer boundaries of computerized results are drawn with red color, whereas the manual outlines of radiologists are colored in green.

5 Discussion and Conclusion

As can be observed from Fig. 3 and Table 1, the FCN segmentation results are relatively not stable. Some regions are obviously mislabeled. Therefore, the FCN may have less robustness to the low ultrasound image quality. On the other hand, the DeconvNet is relatively more suitable for our problem, because of deep decoding path. The SegNet results appear worse than the results of FCN in the muscle (C) layer and septa boundary between muscle and chest wall layers. It thus suggests that the fix of feature map with the same size may not help on our problem. The results of U-Net are in-between the results of SegNet and FCN, though the skip connection strategy is adopted in U-Net to alleviate the gradient vanishing problem. Therefore, the feature learning is relatively

difficult even with the skip connections between the encoder and decoder correspondences. Accordingly, the incorporation of boundary cue may help to improve the ultrasound image segmentation.

It can be found in Table 1 that the best segmentation performance can be achieved by our method "DBS+Main+2DT" with both IoU and CD metrics. Therefore, it may suggest that our 2DT-DBS ConvEDNet may have better capability to withstand speckle noise, shadowing and other challenges shown in the introduction section. Based on the extensive comparisons with other baseline implementations, the efficacy of the 2DT-DBS ConvEDNet on the layer parsing problem can be corroborated.

Acknowledgement. This work was supported by the National Natural Science Funds of China (No. 61501305), the Shenzhen Basic Research Project (No. JCYJ20150525092940982), the Natural Science Foundation of SZU (No. 2016089).

References

1. McCormack, V.A., dos Santos Silva, I.: Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol. Biomark. Prev. **15**, 1159–1169 (2006)
2. Tan, T., et al.: Chest wall segmentation in automated 3D breast ultrasound scans. Med. Image Anal. **17**, 1273–1281 (2013)
3. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: CVPR 2015, pp. 3431–3440 (2015)
4. Noh, H., et al.: Learning deconvolution network for semantic segmentation. In: ICCV 2015, pp. 1520–1528 (2015)
5. Badrinarayanan, V.: Segnet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. arXiv preprint arXiv:1505.07293 (2015)
6. 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). doi:10.1007/978-3-319-24574-4_28
7. Lee, C.-Y., et al.: Deeply-Supervised nets. In: AISTATS, 2015, June 2015
8. Chen, H., et al.: DCAN: deep contour-aware networks for accurate gland segmentation. In: CVPR 2016, pp. 2487–2496 (2016)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
10. Shin, H.-C., et al.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE TMI **35**, 1285–1298 (2016)
11. Xie, S., Tu, Z.: Holistically-nested edge detection. In: ICCV 2015, pp. 1395–1403 (2015)
12. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: ACM MM 2014, pp. 675–678 (2014)
13. Gubern-Merida, A., et al.: Breast segmentation and density estimation in breast MRI: A fully automatic framework. IEEE JBHI **19**, 349–357 (2015)
14. Huang, Q., et al.: Breast ultrasound image segmentation: a survey. IJCARS **12**, 1–5 (2017)
15. Cheng, J.-Z., et al.: ACCOMP: augmented cell competition algorithm for breast lesion demarcation in sonography. Med. Phys. **37**(12), 6240–6252 (2010)