



End-to-End Ovarian Structures Segmentation

Diego S. Wanderley^{1,2}(✉), Catarina B. Carvalho², Ana Domingues²,
Carla Peixoto³, Duarte Pignatelli^{3,4}, Jorge Beires³, Jorge Silva^{1,2},
and Aurélio Campilho^{1,2}

¹ Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
`diego.wanderley@fe.up.pt`

² INESC TEC, Porto, Portugal

³ Centro Hospitalar de São João, Porto, Portugal

⁴ Faculdade de Medicina da Universidade do Porto, Porto, Portugal

Abstract. The segmentation and characterization of the ovarian structures are important tasks in gynecological and reproductive medicine. Ultrasound imaging is typically used for the medical diagnosis within this field but the understanding of the images can be difficult due to their characteristics. Furthermore, the complexity of ultrasound data may lead to a heavy image processing, which makes the application of classical methods of computer vision difficult. This work presents the first supervised fully convolutional neural network (fCNN) for the automatic segmentation of ovarian structures in B-mode ultrasound images. Due to the small dataset available, only 57 images were used for training. In order to overcome this limitation, several regularization techniques were used and are discussed in this paper. The experiments show the ability of the fCNN to learn features to distinguish ovarian structures, achieving a Dice similarity coefficient (DSC) of 0.855 for the segmentation of the stroma and a DSC of 0.955 for the follicles. When compared with a semi-automatic commercial application for follicle segmentation, the proposed fCNN achieved an average improvement of 19%.

Keywords: Ovarian structures segmentation · Ultrasound imaging · Convolutional neural network (CNN)

1 Introduction

Diseases of the female reproductive system can cause pain and discomfort, hormonal dysfunctions, infertility and even death, representing around 16% of all the cancers diagnosed in women worldwide and affecting annually more than 1.85 million women [10]. Due to the difficulty of distinguishing between benign and malignant tumors and to a high interobserver variability, ovarian malignant tumors result in a 68% fatality rate in the European Union [10]. A better characterization of the ovarian structures can have an important role in the early

detection of pathologies (e.g ovarian cyst, polycystic ovarian syndrome or even ovarian cancer), while it can also help the monitoring of follicle growth and distribution, important features for assisted reproductive treatments.

The brightness mode (B-mode) ultrasound imaging is commonly used in the gynecological clinical practice because it allows a visualization of the ovary and its structures. In B-mode images, follicles are represented as hypo-echogenic elliptical structures, while the stroma of the ovary exhibits a slight variation in texture relative to its surrounding tissue and has partially hyper-echogenic boundaries. Figure 1 shows an example of a gynecological B-mode image containing an ovary with three follicles, and their manual segmentation.

Image segmentation methods are often used to automatically extract objects from images, reducing the time of analysis and also diagnostic errors. However, ultrasound image segmentation is not easy due to the presence of several image artifacts and noise [6]. According to the latest review in follicles detection [7], the methods used to segment ovarian structures can only detect and measure large follicles. To the best of our knowledge, the segmentation of the stroma has not received enough attention, being only used to reduce the search space for follicle detection [1].

Neural network techniques have been achieving impressive results in visual recognition systems. Among them, fully convolutional neural networks (fCNN) are specially good at learning image features from training data and have proved to be a powerful tool for segmentation of biomedical images [8]. The herein presented research aims to explore the use of fCNNs for the segmentation of the ovarian structures, namely stroma and follicles, in a single process.

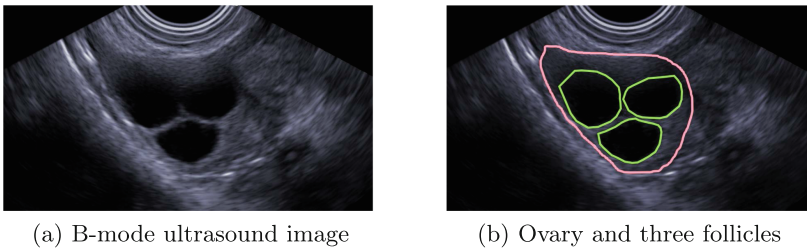


Fig. 1. B-mode image of an ovary and three follicles, and their segmentation.

2 Methodology

This section presents the methods implemented in this work to segment the ovarian structures in B-mode images. In the following subsections, the proposed system, its fCNN architecture and loss functions used are detailed.

2.1 Architecture

An overview of the proposed system is shown in Fig. 2. Switches S_1 and S_2 can be triggered to change the input data of the network and the tasks to be trained, respectively. These changes can work as regularization of the fCNN.

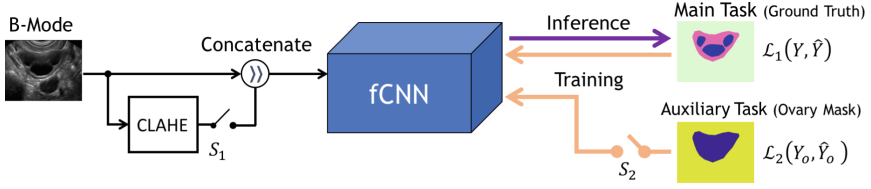


Fig. 2. Overview of the proposed system with multiple tasks.

When switch S_1 is turned on, the B-mode image is preprocessed by a contrast limited adaptive histogram equalization (CLAHE) [11] in order to enhance local contrast and improve the visualization of the ovarian structures. Both CLAHE and original images can be used as input data, as represented in Fig. 2, left side.

For the training, the switch S_2 can be used to activate the multi-task learning, which consists of using the same network to simultaneously solve multiple tasks. In this work, a mask of the ovary is used as ground truth of an auxiliary task in order to prevent the network to classify elements outside the ovary as follicles or from classifying pixels inside of the ovary as background. The auxiliary task acts as regularization during the training of the network [9], and can help the network to focus the attention on difficult cases [3].

The fCNN architecture used in this work (Fig. 3) is based on the U-net [8]. This architecture consists of a downsampling stream (left side) followed by a symmetric upsampling stream (right side). Data from downsampling stream are skip connected to the corresponding layer in the upsampling stream. The convolutional layers are followed by a batch normalization layer and ReLU activation layer; also a dropout layer is inserted between them, when pooling or concatenating operations are performed. The last layer is a 1×1 convolution followed by a softmax, which produces a pixel-wise discrete probability distribution of the three classes of interest (follicle, stroma of the ovary or background).

2.2 Loss Function

The proposed loss function can be decomposed into the main and the auxiliary tasks. Also, weight maps can be applied as regularization, in order to penalize wrong classifications. The details of each step are explained below.

The average Dice Similarity Coefficient (DSC) of each class, as proposed in [5], is the main component of the loss function. The average DSC can be defined as $\overline{\text{DSC}}(Y, \hat{Y}) = 0.5[\text{DSC}(Y_f, \hat{Y}_f) + \text{DSC}(Y_s, \hat{Y}_s)]$, where Y represents the predictions and \hat{Y} represents the ground truth (GT); the indexes f and s represent the follicles and the stroma, respectively. The background was not considered in the loss function because it is the largest region in the image and, so, the results can be heavily influenced by it.

In addition, two weight maps were computed to be applied with the DSC. The first one (W_f) intends to penalize wrong classifications between nearby follicles, as in $w(x)$, defined by U-net [8]. The value of $W_f(i)$ is calculated using

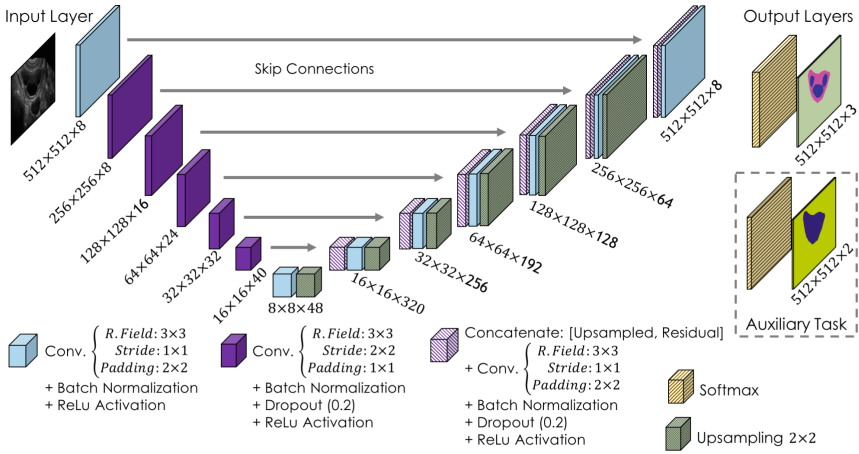


Fig. 3. Architecture of the implemented fCNN.

the distance between the i th pixel and the borders of the two nearest follicles. The second one (W_o) is applied to penalize false detections of ovarian structures in the background region, and is defined for each pixel as:

$$W_o(i) = \begin{cases} 0 & \text{if } i \text{ is inside the ovary} \\ 1 & \text{if } \Delta_o(i) > \ln(10)\sigma^2 \\ 0.1 \cdot \exp\left(\frac{\Delta_o(i)}{\sigma^2}\right) & \text{otherwise,} \end{cases} \quad (1)$$

where $\Delta_o(i)$ is the distance from pixel i to the nearest pixel of the ovary, and σ is a constant that controls the distribution of the weights around the ovary.

Then, the loss function of the main task is computed as:

$$\mathcal{L}_1(Y, \hat{Y}) = 1 - \lambda_1 \overline{\text{DSC}}(Y, \hat{Y}) + \lambda_2 \frac{\sum_i Y_f(i) W_f(i)}{\sum_i Y_f(i)} + \lambda_3 \frac{\sum_i Y_s(i) W_o(i)}{\sum_i Y_s(i)}, \quad (2)$$

where $\lambda_{1,2,3} \in \mathbb{R}^+$ are constants used to adjust the influence of each weight map.

The loss function of the auxiliary task is given by:

$$\mathcal{L}_2(Y_o, \hat{Y}_o) = 1 - \text{DSC}(Y_o, \hat{Y}_o), \quad (3)$$

where \hat{Y}_o is the GT mask of the ovary and Y_o is the predicted ovary.

Finally, the total loss function is defined as:

$$\mathcal{L}(Y, \hat{Y}) = \alpha_1 \mathcal{L}_1(Y, \hat{Y}) + \alpha_2 \mathcal{L}_2(Y_o, \hat{Y}_o) \quad \forall \alpha_{1,2} \in \mathbb{R}^+ | \alpha_1 + \alpha_2 = 1, \quad (4)$$

where $\alpha_{1,2}$ are constants used to adjust the influence of each component.

2.3 Implementation Steps

The proposed system was evaluated using six variations. The input data of the fCNN (Fig. 2) is changed by the switch S_1 ; and the switch S_2 controls the multi-task learning. When S_2 is “off”, the Eq. (4) is written with $\alpha_1 = 1$ and $\alpha_2 = 0$,

otherwise $\alpha_1 = 0.75$ and $\alpha_2 = 0.25$. Finally, the values of λ_2 and λ_3 , in Eq. (2), determine if the weight maps are added or not to the loss function. For all these experiments $\lambda_2 = \lambda_3 \in [0, 1]$, and $\lambda_1 = 1$. The values of α , λ and σ were defined empirically and were not changed during each train.

All original B-mode images were converted to gray-scale and cropped to 512×512 pixel. Aside from the batch normalization layers, no regularization or normalization to zero mean and unit variance were applied to the input images. To increase the training set, a data augmentation process using random linear transformations such as rotation, translation, flip and zoom was applied in each iteration of the training. Each iteration was performed with a batch of 4 images.

The network was trained using Adam (Adaptive Moment Estimation) optimizer [4] with an initial learning rate of 10^{-2} . In this state-of-the-art stochastic optimization method, there is a learning rate for each weight of the network, and the learning rates are adapted during the training. To reduce the probability of overtraining, an early stopping callback is set to stop the training if the improvement of the validation loss is less than 10^{-3} , during 50 epochs. This work was implemented in Python 2.7 using Keras 1.2.2 framework with TensorFlow 1.0.0 as backend.

3 Results

This section presents the dataset, the evaluation methodologies and the obtained results.

3.1 Dataset

The dataset consists of 87 B-mode images. Each image contains one ovary of a woman in childbearing age with at least one follicle. The images were acquired with an Ultrasonix SonixTouch Q+. During acquisition, the medical doctor performed semi-automatic segmentations using the Ultrasonix Auto Follicle segmentation (AF) tool [2]. It must be noted that not all of the follicles were segmented by the doctor, leading to, for instance one ovary with 4 follicles and only one semi-automatic segmentation. Posteriorly, a medical expert manually segmented each ovary and each follicle to produce the GT. The images were randomly divided as: 57 for training, 15 for validation and 15 for testing.

3.2 Evaluation

The quantitative evaluation of the results was divided into two different validation methodologies. First, the DSC between the GT and the predicted segmentations, obtained by the different trained networks, are presented. Secondly, a single follicle evaluation (SFE) was performed and then compared with the AF segmentation, mentioned in Sect. 3.1.

The motivation for SFE lays on the fact that a GT mask may have more segmented follicles than the ones annotated by the doctor using the AF tool

during the acquisition. For example, while the GT of the test set has 44 follicles manually segmented, only 25 follicles were annotated with the AF tool. The SFE verifies if a follicle segmented by the AF has a corresponding follicle in the GT and in the fCNN segmentations. Then, for each follicle present in the AF data, the DSC of GT vs AF and GT vs fCNN are computed.

In Fig. 4 two scenarios of SFE are illustrated. Figure 4(a) represents a SFE with a larger overlay while Fig. 4(b) represents an incorrect segmentation. In this case, the fCNN and the AF segmented a large single follicle which merged the existing two follicles into one, leading to an inaccurate detection.

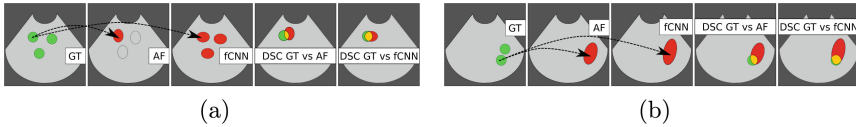


Fig. 4. Illustration of SFE for GT (green) vs AF/fCNN (red); yellow represents the agreement with GT. The correspondences between the GT and the predictions are represented by the arrows. (a) Successful case where a single follicle in GT corresponds to a single follicle segmented by AF and fCNN; (b) an incorrect segmentation since two follicles were merged. (Color figure online)

3.3 Results

The overall DSC results for the six trained networks are shown in Table 1. The fCNNs #1 and #4 show the best overall DSC for the follicles and the stroma. In Fig. 5, four examples of the segmentation performed by the developed fCNNs are shown. The highest DSC achieved for follicles was 0.955, with the fCNN #1 – Fig. 5(a), and for the stroma was 0.855, with the fCNN #3 – Fig. 5(b). Also, a standard case and the image with the worst segmentations are presented in Fig. 5(c) and (d), respectively.

Table 1. Overview of DSC for the predicted segmentations of the fCNN trained.

Architecture				Follicle DSC			Stroma of ovary DSC		
#	S_1	S_2	W. Maps	Median	Mean	Std.	Median	Mean	Std.
1	Off	Off	$\lambda_{2,3} = 0$	0.839	0.757	0.207	0.695	0.657	0.138
2	Off	On	$\lambda_{2,3} = 0$	0.844	0.691	0.331	0.651	0.525	0.251
3	Off	On	$\lambda_{2,3} = 1$	0.779	0.739	0.221	0.678	0.624	0.195
4	On	Off	$\lambda_{2,3} = 0$	0.826	0.784	0.197	0.687	0.677	0.085
5	On	On	$\lambda_{2,3} = 0$	0.809	0.765	0.172	0.612	0.632	0.088
6	On	On	$\lambda_{2,3} = 1$	0.783	0.763	0.111	0.623	0.619	0.107

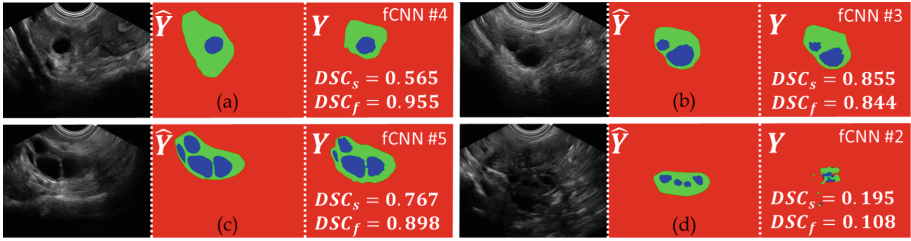


Fig. 5. Examples of segmentation results: (a) the best follicle DSC, (b) the best stroma DSC, (c) a standard case, (d) the worst image.

In a qualitative analysis, the application of multi-task learning prevented follicles for being classified outside the ovary. This approach obtained a fast convergence in training and the smallest validation errors. However, for three test images with low contrast – e.g. Fig. 5(d), the ovarian structures were poorly or not detected, which impaired the overall results. The application of CLAHE improved the results and the use of the weight map W_o solved the problem of false positive ovaries. However, weight map W_f did not significantly reduce misclassification of the pixels between too close follicles; in addition, it produced the wrong classification of the outer boundary of the follicles as background.

The results of the SFE are presented in Table 2. The AF was overcome by all architectures except the fCNN #2. The best overall results for the SFE were obtained with the simplest architecture. Although the CLAHE improved the contrast in boundary regions, the SFE did not improve when CLAHE was used.

Table 2. Overview of DSC for single follicle evaluation (SFE).

Method	Median	Mean	Std.	Max.
fCNN #1	0.823	0.735	0.206	0.961
fCNN #2	0.575	0.556	0.321	0.972
fCNN #3	0.817	0.723	0.237	0.952
fCNN #4	0.755	0.659	0.299	0.972
fCNN #5	0.775	0.691	0.237	0.971
fCNN #6	0.773	0.729	0.176	0.927
Auto Follicle (AF)	0.712	0.615	0.271	0.948

4 Conclusions

In this paper, the first supervised fCNN for the segmentation of the stroma and follicles of ovaries in B-mode images, in an end-to-end fashion, was presented. Despite being trained with a small dataset, the developed method does not depend on heavy preprocessing or post-processing strategies. The visual results

show that a fCNN can learn features that allow to distinguish the ovarian structures in B-mode images. This functionality could allow a better characterization of the overlooked stroma region. Also, the proposed method proved to be more accurate than a commercialized semi-automatic method for follicle segmentation.

Despite presenting slightly better results in the validation set, the proposed regularization techniques show worse overall DSC results for the test set, when comparing with the simplest fCNNs (#1 and #4). This may have been caused by the increasing of the complexity of the segmentation task and by the overwhelming of the data information by the regularization terms. An improvement of the proposed regularizations should be investigated to yield better results.

For future steps of this investigation, the proposed fCNN will be extended to a deeper architecture, increasing the number of learnable features. Due to the scarcity of data, a k-fold cross-validation should be applied to better evaluate the consistency of each architecture. Also, a more efficient loss function will be elaborated in order to force the network to learn the boundaries of the follicles. Finally, the increasing of the dataset is fundamental to improve the variability of the training set.

Acknowledgments. This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project “POCI-01-0145-FEDER-006961”, and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia as part of project “UID/EEA/50 014/2013”.

References

1. Chen, T., Zhang, W., Good, S., Zhou, K.S., Comaniciu, D.: Automatic ovarian follicle quantification from 3D ultrasound data using global/local context with database guided segmentation. In: IEEE 12th ICCV, pp. 795–802 (2009)
2. Eskandari, H., Azar, R.Z., Pendziwol, L.: Ovarian follicle segmentation in ultrasound images. US Patent 9,679,375 (2017)
3. Ferreira, F.T., Sousa, P., Galdran, A., Sousa, M.R., Campilho, A.: End-to-end supervised lung lobe segmentation. In: 2018 IJCNN, pp. 1–8 (2018)
4. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR (2014). <http://arxiv.org/abs/1412.6980>
5. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 4th International Conference on 3DV, pp. 565–571 (2016)
6. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: a survey. IEEE Trans. Med. Imaging **25**(8), 987–1010 (2006)
7. Potočník, B., Cigale, B., Zazula, D.: Computerized detection and recognition of follicles in ovarian ultrasound images: a review. Med. Biol. Eng. Comput. **50**(12), 1201–1212 (2012)
8. 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

9. Ruder, S.: An overview of multi-task learning in deep neural networks. arXiv (2017). <http://arxiv.org/1706.05098>
10. Stewart, B.W., Wild, C.P.: IARC World Cancer Report 2014. International Agency for Research on Cancer, Lyon (2014)
11. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: Heckbert, P.S. (ed.) Graphics Gems IV, pp. 474–485. Academic Press Professional Inc, USA (1994)