Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 542–549

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

Ovarian Structures Detection using Convolutional Neural Networks

Diego Wanderley[a,*], Carlos Ferreira[a,b], Aurélio Campilho[a,b], Jorge Silva[a,b]

[a]*Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal*
[b]*Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto, Portugal*

Abstract

The detection of ovarian structures from ultrasound images is an important task in gynecological and reproductive medicine. An automatic detection system of ovarian structures can work as a second opinion for less experienced physicians or complex ultrasound interpretations. This work presents a study of three popular CNN-based object detectors applied to the detection of healthy ovarian structures, namely ovary and follicles, in B-mode ultrasound images. The Faster R-CNN presented the best results, with a precision of 95.5% and a recall of 94.7% for both classes, being able to detect all the ovaries correctly. The RetinaNet showed competitive results, exceeding 90% of precision and recall. Despite being very fast and suitable for real-time applications, YOLOv3 was ineffective in detecting ovaries and had the worst results detecting follicles. We also compare CNN results with classical computer vision methods presented in the ovarian follicle detection literature.

Keywords: follicle detection; ovary detection; ovarian structures, gynecological ultrasound; convolutional neural networks.

* Corresponding author.
E-mail address: dswanderley@gmail.com

10.1016/j.procs.2021.12.047

1. Introduction

Ultrasound is an imaging modality based on the emission of high-frequency sound waves and the capture of the returned echoes, which allows the evaluation of soft human body tissue in a non-invasive manner. Due to the fact of being safe, easy to use, and low-cost, ultrasound imaging became a relevant tool in gynaecology, being the ultrasound brightness mode (B-mode) the principal medical imaging modality applied to the inspection of internal organs, such as the ovary. A B-mode image of an ovary allows the identification of the ovarian stroma, follicles, cysts, and tumours.

Object detection consists of locating and identifying certain objects in an image according to their class. The location can be represented by a rough mark around the object or a pixel-wise classification. The object detection task can be used for the computerised analysis of medical images to detect anatomical structures under analysis. For example, the automatic detection of follicles assists in follicular monitoring during infertility treatment. However, the automatic detection of anatomic structures from ultrasound images has been shown complex due to the signal artifacts such as attenuation, speckle, shadows, and signal dropout.

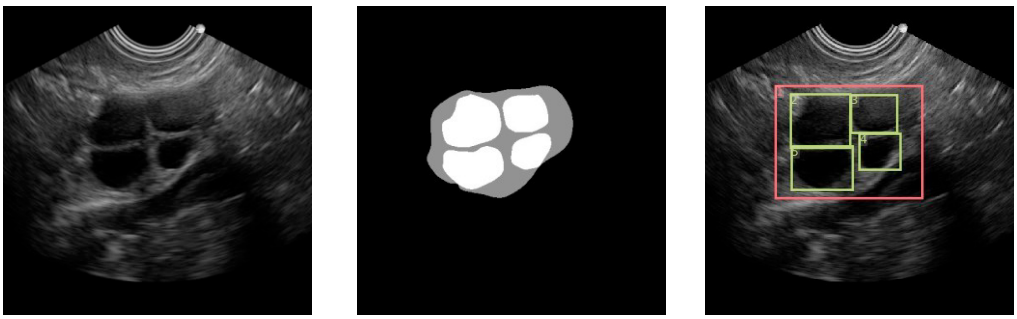
The literature presents some computerised methods for the locating of ovarian follicles in B-mode images, mainly using segmentation tasks. The published works using classical computer vision on 2-D B-mode image only allow locating large, or ready to ovulate follicles used for assisted reproduction procedures [1]. Convolutional Neural Networks (CNN) are started to be used for both the detection and segmentation of ovarian structures in 2-D images.

A three-step algorithm based on 2D region growing was proposed in [2] for follicle detection. The first step is a segmentation of the follicles to produce coarse estimations. The second step applies a region growing algorithm to each estimation to produce accurate delimitation of the follicles candidates. The third step uses prior knowledge about the follicles (e.g. shape and size) to classify the detected regions as follicle or non-follicle. According to [1], this method was the most mature for automatic follicle detection from 2D images using classical computer vision methods, having been tested on 50 images with a follicle recognition rate (FRR) of 78%.

A similar approach was applied in [3], where some geometric features of the follicle such as the area, the ratio of the major axis length to minor axis length, the compactness, the extent, and the centroid location, were used to recognize ovarian follicles from transvaginal ultrasound images. The reference data were extracted from a training set of 16 images. The FRR was 85% for a test set of 20 images

A cellular automaton combined with neural networks (NN) was used for follicle segmentation in [4] and subsequent detection of follicle instances. The results had an FRR of 60% and a follicle misidentification rate (FMR) of 30%. The misidentification rate was reduced to 14% using a support-vector machine (SVM) to optimise the cellular neural networks templates [5], and 28 images were used in the validation.

In [6], an active contour algorithm was applied to the B-mode images and 15 features were extracted from the follicle candidates. Then, a particle swarm optimization (PSO) was employed to select the optimal features to be used as input of a multilayer perceptron neural network. The goal of this network is to evaluate if the candidate is a follicle or a non-follicle. The authors reported an accuracy of 98.3% but there is not much information about the validation and the testing sets.



(a) B-Mode (b) Manual segmentation (c) Bounding box

Fig. 1: Example of an image with one ovary and four follicles.

A fully convolutional neural network (FCN) was presented in [7] to perform semantic segmentation of the ovarian stroma and follicles in B-mode images. This work also presented some variations with pre-processing and multi-task learning. A private dataset was used consisting of 57 images for training, 15 for validation and 15 for testing. The segmentations were assessed by the Dice similarity coefficient (DSC), with a mean value of 0.78 for follicles and 0.68 for stroma.

In [8], two methods were used to test a public dataset of 3D ovarian volumes. The first baseline applies an FCN to segment the follicles in each slide of the three orthogonal planar cross-sections, and then combines the results. The second baseline is a method based on Directional 3D Wavelet Transform. The authors concluded that the dataset contains enough data for training a CNN without overfitting.

The works presented here use a segmentation step to later locate each follicle. The main problem with the location using segmentation in 2-D images is often the prediction masks for anatomical structures appear connected, which can lead to an error in the results if the masks are not correctly separated. The herein presented research aims to explore the state-of-the-art CNNs for object detection applied for the detection of the ovarian structures, namely ovary and follicles in ultrasound B-mode images. Such an approach does not need a segmentation step and can also be used as a delimiter for segmentation systems. In addition, the CNN results are compared with classical computer vision methods results.

2. Methodology

2.1. Dataset

Our dataset consists of 100 ultrasound B-mode images, scan converted, with 8-bit grayscale resolution and 512 × 512-pixel size. These images were acquired with an Ultrasonix SonixTouch Q+, which performs the beam-formed radio-frequency data conversion to B-mode with a specific algorithm. The frame rate is dependent on several factors such as the operating mode, the probe model, and the number of focal points. For capturing the images of this dataset, the maximum frame rate was 40 frames per second (fps).

The images were randomly divided as 64 for training, 18 for validation and 18 for testing. The images in the testing set contained 57 follicles. Aside from the batch normalization layers, no regularization or normalization to zero mean and unit variance was applied to the input images. Each image contains one ovary of a woman in childbearing age with at least one follicle. One medical expert manually segmented each ovary and each follicle. The bounding boxes of the segments were used as ground truth for the detection. Figure 1 presents an example of this process.

No image pre-processing was applied between the capture and the CNN inputs. The use of pre-processing before an FCN input is discussed in [7], and although it improves punctual results, there was no overall gain.

2.2. CNN Detectors

The object detection task can be divided into location and classification. Two-stage detectors perform these tasks sequentially, and one-stage detectors perform them in parallel. We tested three popular CNN-based detectors followed by non-maximum suppression (NMS). A two-stage, the **Faster R-CNN** [9], and two one-stage, the **RetinaNet** [10] and the **YOLOv3** [11].

These three networks make use of anchor boxes to improve the speed and efficiency of the object location. Each anchor box is associated to a mini batch across the feature maps of the network output. The network predicts the probability of an object existing inside an anchor, based on intersection-over-union (IoU) between the anchor and the object location.

The Faster R-CNN is a state-of-the-art object detector of the Region Based Convolutional Neural Networks

(R-CNN) family. After the backbone, the Faster R-CNN has a Region Proposal Network (RPN) to perform the object location, followed by the classification head. To perform a multi-scale detection, the Faster R-CNN detector can be connected to a Feature Pyramid Network (FPN) [12]. The FPN consists of bottom-up and top-down pathways, connected by lateral connections, that merge the features from the backbone layers with the up-sampling stream, allowing the detection at all levels of the FPN.

The RetinaNet is a popular one-stage object detector based on the FPN architecture that uses a detection head with two subnets at the output, one for bounding boxes regression and the other for classification. The RetinaNet is trained using the novel focal loss [10] function, which consists of a modified cross-entropy loss to reduce the impact of the class imbalance problem, caused especially bye the background class.

Both Faster R-CNN and RetinaNet use a ResNet [13] backbone. In order to test the influence of the backbone on the detection results, we trained the RetinaNet with different variations of ResNet. We also use a multi-branch variation of ResNet, known as ResNeXt [14].

YOLOv3 [11] is an one-stage multi-scale object detector that allows real-time predictions. The backbone is a CNN with 53 convolutional layers, named Darknet-53, an extension of the original Darknet [15] that has only 7 convolutional layers. In addition to the original implementation, we use two more variations of YOLOv3: the YOLOv3-SPP, that extends the Darknet-53 with a spatial pyramid pooling (SPP) [16]; and the YOLOv3-Tiny that applies the detection layers on only two levels of the original Darknet backbone.

We defined the Faster R-CNN with a ResNet-50 and FPN backbone as our baseline since the Faster R-CNN is the state-of-the-art network in object detection, and the only two-stage network used in this research. For the one-stage networks, we vary the backbones or the number of anchors to measure the influence of these changes on the results. The reason for varying several backbones is to verify the efficiency of the networks of a stage with different depths. With such data, it is possible to establish in the future the minimum architecture required for use on embedded platforms and real-time applications that have an acceptable efficiency.

2.3. Validation and Metrics

The Average Precision (AP) is an estimate of the area under the curve of the Precision vs Recall graph. In this work, the AP is obtained by interpolating the precision at each level, as explained in [17]. The trained models were then evaluated with the testing set using the mean Average Precision (mAP), the Precision, the Recall, and the F1 score, considering both follicle and ovary classes. Furthermore, we consider true-positive, the detections whose bounding boxes have an IoU with the ground truth greater than 0.5.

During the training, the validation was performed by measuring the mAP of both classes (ovary and follicle). The final version of each model was selected at the epoch with the best validation mAP.

3. Results

The tests were performed on a desktop with a NVIDIA GeForce RTX 2060 SUPER (8G VRAM) GPU and an AMD Ryzen 5 3500X CPU. Table 1 presents the models tested their training losses, their validation mAP and their inference average time. Since the cost functions are different for each network family, the loss function values vary greatly. For example, the RetinaNet losses are never greater than 1, whereas the YOLO losses start with values higher than 100. A better comparison can be made using the validation set mAP. Faster R-CNN presented the best training and validation results.

Regarding inference time, YOLOv3 Tiny is the fastest model (17 fps), and Faster R-CNN the slowest (8 fps), as expected. The increase of the backbone depth on RetinaNet impacts up to 50% more processing time, decreasing from 12 fps with ResNet-18 to 8 fps with ResNeXt-101.

Table 2 presents the models average results for each class detection. YOLOv3 is a lightweight architecture, focusing on detection speed, which sacrifices detection assertiveness. For this reason, YOLOv3 models performed worse than the other two architectures. The results for the follicles are similar to the general ones, as there is a strong class imbalance since an image contains only one ovary but may contain several follicles.

The Faster R-CNN confirmed its state-of-the-art object detector condition, achieving the best mAP (94.6%), Recall (94.7%), and F1 Score (95.1%) values for all classes. Faster R-CNN model failed to detect 6 follicles and detected 6 false positive follicles. The RetinaNet architecture presented competitive results with Faster R-CNN, with the deeper models achieving mAP values greater than 90%. The lightest RetinaNet models outperformed YOLOv3 testing results by a large advantage. The RetinaNet with ResNet-18 is only 8% slower than YOLOv3 but exceeds the mAP by 37.5 points. RetinaNet with ResNet-18 presented the highest Precision (96.7%).

Table 1. Models definitions, their training losses, validation mAP and inference average time.

Model				Training Loss	Validation mAP	Inference Avg. Time
#	Detector	Backbone	Anchors			
FRN	Faster R-CNN	ResNet-50	9	0.286	97.7%	125 ms
RN1	RetinaNet	ResNet-18	9	0.688	87.3%	82 ms
RN2	RetinaNet	ResNet-34	9	0.595	90.4%	89 ms
RN3	RetinaNet	ResNet-50	9	0.716	87.8%	92 ms
RN4	RetinaNet	ResNet-101	9	0.519	89.3%	104 ms
RN5	RetinaNet	ResNeXt-50	9	0.690	92.2%	96 ms
RN6	RetinaNet	ResNeXt-101	9	0.401	92.8%	121 ms
Y1	YOLOv3	Darknet-53	6	9.311	42.6%	76 ms
Y2	YOLOv3	Darknet-53	9	7.360	59.8%	76 ms
Y3	YOLOv3-SPP	Darknet-53	6	7.537	59.6%	79 ms
Y4	YOLOv3-SPP	Darknet-53	9	8.340	64.7%	102 ms
Y5	YOLOv3 Tiny	Darknet	4	5.368	54.2%	57 ms
Y6	YOLOv3 Tiny	Darknet	6	4.748	65.8%	66 ms

The RetinaNet with deeper backbones models achieved higher IoU results and produced fewer false negatives in comparison with the RetinaNet models with shallow backbones. For example, the RetinaNet with ResNeXt-101 produced 7 false negatives against 15 from ResNet-18 and 11 from ResNeXt-50. However, the depth of the backbone did not show a relevant impact on the detection of false positive follicles.

In general, the YOLO models presented many false positive follicles. Such results suggest that more selective thresholds need to be applied to the object's confidence values, according to the different classes, requiring higher values for the detection of follicles and lower values for the detection of ovaries.

Regarding the detection of the ovaries, Faster R-CNN detected all ovaries, and RetinaNet failed only once with ResNet-18. Despite registering a 100% Precision for all models, the YOLOv3 architecture was not successful to locate ovaries, reaching only 50% of detection in two models.

Figure 2 show some results from Faster R-CNN, RetinaNet with ResNet-18, RetinaNet with ResNext-101, and YOLOv3-SPP with 6 anchors, compared with the ground truth. The objects indicated in the ground truth are numbered from top to bottom and from left to right. The first row represents the Faster R-CNN best result, while the second row represents its worst result. The third and the fourth rows contain two random images. Table 3 presents the IoU for all true positive detections illustrated in Figure 2, where the objects are identified by their classes (O for ovary or F for follicle) followed by their numbers in the ground truth images.

Table 2. Average results for follicles, ovaries and both classes detections, on all tested models.

Model	Follicles				Ovaries				Overall			
#	Precision	Recall	F_1	mAP	Precision	Recall	F_1	mAP	Precision	Recall	F_1	mAP
FRN	91%	89%	90%	89%	100%	100%	100%	100%	96%	95%	95%	95%
RN1	93%	74%	82%	72%	100%	94%	97%	94%	97%	84%	90%	83%
RN2	90%	79%	84%	78%	100%	100%	100%	100%	95%	90%	92%	89%
RN3	87%	72%	79%	70%	100%	100%	100%	100%	94%	86%	86%	85%
RN4	91%	86%	88%	85%	100%	100%	100%	100%	95%	93%	94%	92%
RN5	90%	81%	85%	80%	100%	100%	100%	100%	95%	90%	93%	90%
RN6	88%	88%	88%	85%	100%	100%	100%	100%	94%	94%	94%	93%
Y1	86%	75%	80%	67%	100%	11%	20%	20%	93%	43%	50%	39%
Y2	58%	91%	71%	64%	100%	28%	44%	44%	79%	60%	57%	46%
Y3	82%	86%	84%	75%	100%	44%	62%	62%	91%	65%	73%	60%
Y4	63%	92%	74%	62%	100%	50%	67%	67%	81%	71%	71%	56%
Y5	74%	81%	77%	63%	100%	50%	67%	67%	87%	65%	72%	56%
Y6	69%	84%	76%	63%	100%	40%	56%	56%	84%	62%	66%	51%

Table 3. IoU of true positives detections with ground truth in Figure 2.

Image	Model	IoU with the ground truth						
		O1	F2	F3	F4	F5	F6	F7
1	FRN	0.94	0.84	0.79	0.93	0.83	0.90	0.87
	RN1	0.82	0.69	0.00	0.70	0.61	0.93	0.00
	RN6	0.81	0.76	0.68	0.73	0.76	0.75	0.00
	Y3	0.61	0.79	0.00	0.59	0.68	0.67	0.52
2	FRN	0.50	0.86	0.00	0.80	0.62	-	-
	RN1	0.50	0.00	0.66	0.00	0.00	-	-
	RN6	0.53	0.74	0.69	0.68	0.00	-	-
	Y3	0.00	0.73	0.71	0.63	0.50	-	-
3	FRN	0.93	0.86	0.62	0.62	0.92	-	-
	RN1	0.80	0.69	0.70	0.70	0.88	-	-
	RN6	0.81	0.83	0.79	0.79	0.71	-	-
	Y3	0.00	0.64	0.58	0.58	0.82	-	-
4	FRN	0.74	0.86	0.50	0.94	0.90	-	-
	RN1	0.82	0.83	0.00	0.90	0.76	-	-
	RN6	0.80	0.83	0.51	0.73	0.85	-	-
	Y3	0.34	0.71	0.00	0.76	0.71	-	-

Table 4 shows some published results of follicle detection on ovarian 2D ultrasound, without using deep learning, sorted by the year of publication, compared with some CNN methods presented in this work. It is important to highlight that all these methods were validated using their private datasets, which prevents the comparison of their results. The metrics used for comparison in Table 4 are Precision and Recall, however, in [6], the Recall value is not presented, instead, the authors informed a Specificity of 96.8%. Furthermore, some publications did not identify (NI) the number of images used in the training, validation or testing subsets.

Although one classical computer vision method has the best Precision result and another the best Recall result, the results of the CNNs are promising, as they present high and balanced results of both Precision and Recall values. Faster R-CNN and RetinaNet with deeper backbones have low false-negative rates, which account for high Recall values, when compared to most classical computer vision methods. However, since the datasets are different, an effective conclusion cannot be made.

Table 4. Comparison results for follicle detection from different methods using different datasets.

Method	Year	Images (follicles)			Precision	Recall
		Train	Val.	Test		
Three-step algorithm [2]	2002	NI	NI	50 (302)	71%	78%
Cellular automaton + NN [4]	2004	7	NI	50 (302)	70%	60%
Cellular automaton + SVM [5]	2006	4	NI	28 (168)	86%	58%
Geometric features [3]	2009	16	NI	20 (48)	98%	85%
PSO + NN [6]	2017	20	NI	NI	-	100%
FRN		64	18	18 (57)	91%	89%
RN1		64	18	18 (57)	93%	74%
RN5		64	18	18 (57)	90%	81%
RN6		64	18	18 (57)	88%	88%
Y3		64	18	18 (57)	82%	86%

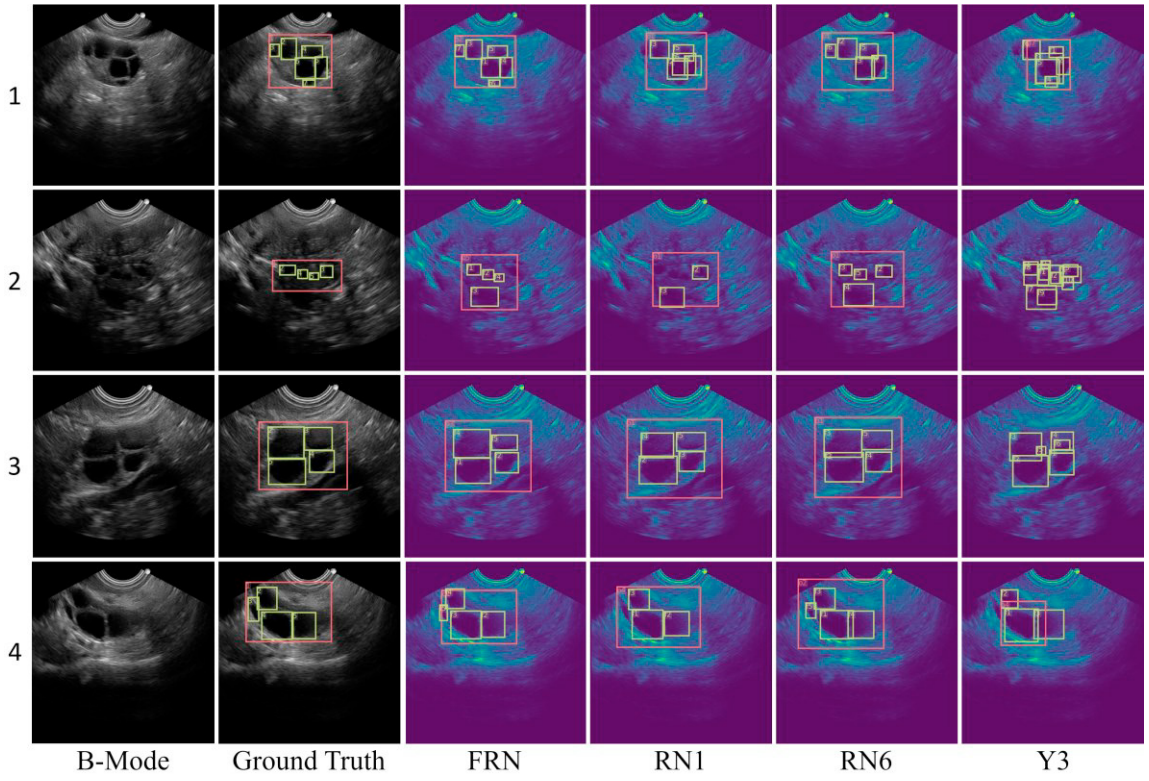


Fig. 2: Samples of detection results from four model compared to Ground Truth.

4. Conclusions

In this work, we trained, in a supervised end-to-end fashion, and evaluated the Faster R-CNN, RetinaNet, and YOLOv3 models for detection of ovarian structures, namely ovary and follicles, in B-mode ultrasound images. The tested models do not demand pre-processing steps, and only an NMS step is used to reduce the number of false positive detections. The results show that CNN models for object detection can detect healthy ovarian structures without overfitting.

The Faster R-CNN model presented the best results for both follicle and ovary classes. The RetinaNet proved to be a reliable one-stage detector, being able to improve the AP of the follicles when the depth of the backbone is increased. Both models were able to detect ovaries in the testing set images. In contrast, the YOLOv3 models failed to detect ovaries and detected multiple false positives follicles. Despite being the fastest, YOLOv3 is not recommended for this task. Furthermore, all tested CNNs present limitations for use with a high frame rate, using the tested hardware.

For each model, we used the same threshold values to ignore low confidence detections and to determine true positives. The future steps of this research include the selection of optimal thresholds per class to define the true positives for each model, and the use of segmentation masks in training to improve accuracy. In addition, increasing the data set for training is essential for the production of more general models.

Acknowledgements

This work was funded by National Funds through the FCT - Portuguese Foundation for Science and Technology within the scope of project NORTE-01-0145-FEDER000016 and UIDB/50014/2020. Carlos Ferreira is funded by the FCT grant contract SFRH/BD/146437/2019.

References

[1] B. Potočnik, B. Cigale, D. Zazula. (2012) "Computerized detection and recognition of follicles in ovarian ultrasound images: a review." Medical & Biological Engineering & Computing **50 (12)**: 1201–1212.

[2] B. Potočnik, D. Zazula. (2002) "Automated analysis of a sequence of ovarian ultrasound images. part I: segmentation of single 2d images." *Image and Vision Computing* **20 (3)**: 217–225.

[3] P. S. Hiremath, J. R. Tegnoor. (2009) "Recognition of follicles in ultrasound images of ovaries using geometric features." *2009 International Conference on Biomedical and Pharmaceutical Engineering*: 1–8.

[4] B. Cigale, D. Zazula. (2004) "Segmentation of ovarian ultrasound images using cellular neural networks." *International Journal of Pattern Recognition and Artificial Intelligence* **18 (04)**: 563–581.

[5] B. Cigale, M. Lenic, D. Zazula. (2006) "Segmentation of ovarian ultrasound images using cellular neural networks trained by support vector machines", in B. Gabrys, R. J. Howlett, L. C. Jain (eds.), *Knowledge-Based and Intelligent Information and Engineering Systems. Lecture Notes in Computer Science,* **4253***, Springer, Berlin, Heidelberg*: 515–522.

[6] O. R. Isah, A. D. Usman, A. M. S. Tekanyi. (2017) "A hybrid model of pso algorithm and artificial neural network for automatic follicle classification." *International Journal Bioautomation* **21 (1)**.

[7] D. S. Wanderley, C. B. Carvalho, A. Domingues, C. Peixoto, D. Pignatelli, J. Beires, J. Silva, A. Campilho. (2019) "End-to-end ovarian structures segmentation", in R. Vera-Rodriguez, J. Fierrez, A. Morales (eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Springer International Publishing, Cham*: 681–689.

[8] B. Potočnik, J. Munda, M. Reljič, K. Rakić, J. Knez, V. Vlaisavljević, G. Sedej, B. Cigale, A. Holobar, D. Zazula. (2020) "Public database for validation of follicle detection algorithms on 3d ultrasound images of ovaries." *Computer Methods and Programs in Biomedicine* **196**: 105621.

[9] S. Ren, K. He, R. Girshick, J. Sun. (2017) "Faster R-CNN: Towards real-time object detection with region proposal networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39 (6)**: 1137–1149.

[10] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár. (2017) "Focal loss for dense object detection." *2017 IEEE International Conference on Computer Vision (ICCV)*: 2999–3007.

[11] J. Redmon, A. Farhadi. (2018) "Yolov3: An incremental improvement." *CoRR arXiv* **1804.02767**. URL http://arxiv.org/abs/1804.02767

[12] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie. (2017) "Feature pyramid networks for object detection." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 936–944.

[13] K. He, X. Zhang, S. Ren, J. Sun. (2016) "Deep residual learning for image recognition." *2016 Conference on Computer Vision and Pattern Recognition (CVPR):* 770–778.

[14] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He. (2017) "Aggregated residual transformations for deep neural networks." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 5987–5995.

[15] J. Redmon, S. K. Divvala, R. B. Girshick, A. Farhadi. (2015) "You only look once: Unified, real-time object detection." *CoRR arXiv* **1506.02640**. URL http://arxiv.org/abs/1506.02640

[16] Z. Huang, J. Wang. (2019) "DC-SPP-YOLO: dense connection and spatial pyramid pooling based YOLO for object detection." *CoRR arXiv* **1903.08589**. URL http://arxiv.org/abs/1903.08589

[17] R. Padilla, S. L. Netto, E. A. B. da Silva. (2020) "A survey on performance metrics for object-detection algorithms." *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*: 237–242.