ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

A Deep Learning approach for oocytes segmentation and analysis

Paolo Andreini[1] *, Niccolò Pancino[2], Filippo Costanti[2], Gabriele Eusepi[3], and Barbara Toniella Corradini[2] *

1 - University of Siena, Department of Information Engineering and Mathematics, Siena, Italy

2 - University of Florence, Department of Information Engineering, Florence, Italy

3 - SILOG Sistemi Logici s.r.l., Siena, Italy

Abstract. Medical Assisted Procreation (MAP) has seen a sharp increase in demand over the past decade, due to a variety of reasons, including genetic factors, health conditions altered by stress and pollution, as well as delayed pregnancy and age-related loss of fertility. The success of MAP techniques is strongly correlated to the dexterity of a human operator, who is asked to classify and select healthy oocytes to fertilize and return to the uterus. This work describes a deep learning approach to the segmentation of oocyte images, to support operators in their selection, to improve the success probability of MAP.

1 Introduction

Medical Assisted Procreation (MAP) has seen an increasing demand over the years: in Italy, MAP demand registered a 10% increase between 2013 and 2019 [1]; an increase in terms of thousands of MAP cycles was registered in Europe (75k to 900k) and in Asia (60k to 550k) between 1993 and 2017 [2]. In this work, we focus on second level in-vitro fertilization (IVF), a set of pharmacological therapies, surgical and laboratory procedures which consist in placing spermatozoa and oocytes on a plate with an ideal culture medium. If fertilization is successful, the embryo is obtained, which can be transferred back to the uterus for gestation. There are two different IVF techniques, FIVET and ICSI: both collect the oocytes and perform fertilization in the laboratory, then transfer back the highest quality embryos. The main difference between the two is the way the fertilization is stimulated: in FIVET (Fertilization In Vitro with Embryo Transfer), once the oocytes have been taken, they are placed on a plate with the spermatozoa, to facilitate fertilization; in ICSI (IntraCytoplasmic Sperm Injection), which is used in case of spermatozoa with low motility, spermatozoa are directly injected into the oocyte. It is worth noting that the success of both FIVET and ICSI depends, among the other genetic factors, on the capability of the human operator to perform a good selection of embryos. The present work aims at showing how a semantic segmentation approach based on Deep Learning (DL) can be used to support the operators, to improve the success probability of the ICSI procedure.

*Barbara Toniella Corradini and Paolo Andreini contribute equally to the work.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

1.1 Motivations

ICSI was the most used IVF technique in 2019 (85.9% of the cases) and it showed a success probability between 50% and 80% for fertilization. A key factor for the success of ICSI is the capability of the operator to select high quality oocytes. The quality of an oocyte is low if it shows abnormalities, such as: smooth endoplasmic reticulum (SER), vacuoles, granular cytoplasm and inclusions. Among the others, the detection of inclusions is an important step in determining the quality of the oocytes. In this work, we focus on the characterization of inclusions in oocyte images, to support embryologists in the use of ICSI. In particular, with the aim of both inferring the presence of inclusions and locating them, a semantic segmentation approach was used. The main reasons to use DL as a support tool for human operators can be summarized as follows.

- Explainability: image semantic segmentation can be described as a pixel level image classification, which provides the capability to localize the object of interest into the image. This makes the human operator interpretation of the system output easier.
- Consistency: while the interpretation by the human expert can be significantly variable, experiments performed by an AI system can be repeated until a reliable result is obtained, so that operator errors can be prevented.

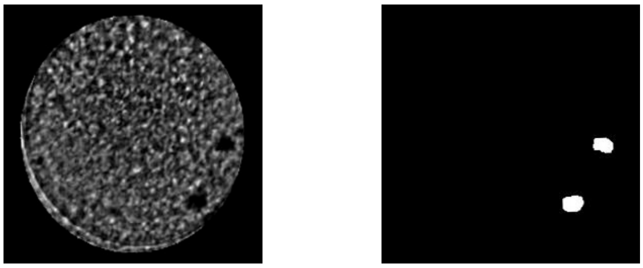
1.2 Segmentation with Deep Learning

Image segmentation is a Computer Vision task that aims to group parts of an image which can be considered homogeneous based on some visual features. In this work, in particular, we focus on semantic segmentation (i.e., classification of all the pixels of an image into semantic classes). Application fields of image segmentation are numerous: among the others, medical image analysis (e.g., skin lesion classification, agar plate, chest X-ray and brain NMR analysis, etc.) [3][4], autonomous driving (e.g., pedestrian detection) [5][6], video surveillance and scene text segmentation [7].
Various algorithms for image segmentation have been developed in the literature, which can be categorized as: threshold based [8], edge based, region based [9] and clustering based [10].
Recently, due to the success of Deep Learning (DL) models in a wide range of vision applications, there has been a substantial amount of work aimed at developing image segmentation approaches using DL models.

1.3 Structure of the paper

The paper is structured as follows: in Section 2, the data preprocessing and the neural network architecture used in the experiments are introduced; in Section 3, the experimental setup and results are presented; finally, Section 4 collects the conclusion of this work and proposes further developments and challenges.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1.
Available from http://www.i6doc.com/en/.



(a) (b)

Fig. 1: Sample image from the training set in (a); the corresponding mask in (b)

2 Material and Methods

2.1 Dataset

The dataset was initially composed by 40x inverted images captured by using a Leica Micromanipulator with Nikon camera (Fig. 1a), provided with respective masks highlighting inclusions (Fig. 1b).

Split	W/ Inclusion	Resolution
Training Original	18	256×256
Training Augmented	1000	256×256
Training Synthetic	1000	256×256
Test Original	8	256×256

Table 1: Dataset organization

Table 1 shows the numerosity of the dataset, organized as follows: Training Original (i.e., the set of microscope images); Training Augmented (i.e., images from the training set augmented by randomly flipping and rotating); Training Synthetic (i.e., the set of synthetic images augmented by generation as described in Section 3.1); Test Original (i.e., the microscope image set used as a test set).

2.2 Neural network model

The Segmentation Multiscale Attention Network (SMANet) [11] is a DL architecture for image segmentation, formed by three main components: a ResNet encoder, a multiscale attention module, and a convolutional decoder.
The SMANet encoder uses a ResNet50 architecture [12] with dilated convolution (i.e. atrous convolution [13]) that allows to perform feature extraction of the input image. The convolutional encoder is attached to an attention module providing pixel-wise attention for the features extracted in the encoding step, and returning two attention maps. The attention mechanism learns to focus on regions containing the desired object. The attention maps are then pixel-wise multiplied with the CNN encoded representation, previously passed through a 1

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

by 1 convolutional layer to conveniently reduce the feature map dimensions. The SMANet attention module, which outputs a pixel-wise prediction, is followed by a two level decoder to recover small details at a higher resolution.

3 Experiments

3.1 Experimental setup

In order to successfully train a deep segmentation network, a large amount of supervised data is usually required. Nonetheless, data deficiency is a typical drawback of DL biomedical applications, due to privacy issues and to the inherent difficulty on obtaining supervised data. In this work, we focus on the use of data augmentation, based on a DL based image generation technique, to face the lack of a large set of annotated data. In particular, following the procedure employed in [14] and [15], we used an image-to-image translation approach to transform semantic label maps in fake oocyte images, that can be used to enlarge our small training set. The label maps are created by randomly placing from one to five circular spots with variable diameter (from 2 to 7 pixels). We advocate that this closely resembles the true distribution of the inclusions in the oocyte which, based on the opinion of experts, is totally random. The generated images, following the proposed approach, can be used as an effective type of data augmentation. The experimental setup can be divided in the following phases:

- Training on real images, in which SMANet was trained on real images augmented by performing standard augmentation operations, i.e. flips and rotations;
- Training on synthetic images, in which synthetic images obtained by the aforementioned approach were used as training set;
- Fine-tuning, in which a model pre-trained on synthetic images was fine-tuned on real images.

Experiments run on a machine equipped with Intel Core i7-9800x CPU, Nvidia GeForce GTX1080Ti GPU and 32GB of RAM.

3.2 Results

The experimental results are reported in Table 2. As a first outcome, it is worth noting that synthetic images can be effectively used as data augmentation. Indeed, better results are achieved by training on synthetic images rather than using real data, demonstrating the high quality of the generated images. Moreover, the results can be further improved by fine-tuning the model, pre-trained on the syntethic data, on the real images.

Trained on	Accuracy	Mean IoU
Real	0.9985	0.6415
Synthetic	0.9974	0.6937
Real + Synthetic	0.9988	0.7220

Table 2: Segmentation performance with real images as test set (IoU stands for Intersection over Union and quantifies the percentage of overlapping between the targets and the predictions, whereas the accuracy is the percentage of correctly predicted pixels)

4 Conclusions and next developments

As far as we know, this is the first work that addresses the problem of semantic segmentation of oocyte images based on DL techniques, despite the great interest on this kind of investigation in the medical field. Nonetheless, few remarks on possible improvements can be made: as a matter of further research, the collection of original images will be enlarged, which is expected to greatly improve the performance of the network. Furthermore, other parameters having a strong influence on the evaluation of oocytes may be considered in the segmentation task: to this aim, more detailed segmentation maps will be drawn, taking into account other characteristics and abnormalities of oocytes.
As a consequence, this work can be considered as a first step towards a more complex tool, capable to characterise oocytes in a complete manner and give the human operator a quality score, obtained through an objective and reproducible procedure.

Acknowledgements

Special thanks to A.G.I medica (AGI) for providing the oocite image dataset and to SILOG Sistemi Logici s.r.l. for being a key contributor to the production of this article.

References

[1] Ministero della Salute. Relazione del ministro della salute al parlamento sullo stato di attuazione della legge contenente norme in materia di procreazione medicalmente assistita (legge 19 febbraio 2004, n. 40, articolo 15), 2021.

[2] GD. Adamson, J. de Mouzon, G. Chambers, F. Zegers-Hochschild, O. Ishihara, M. Banker, S. Dyer, and M. Kupka. International committee for monitoring assisted reproductive technology: world report on assisted reproductive technology. 2017.

[3] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

[4] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli. Image generation by GAN and style transfer for agar plate image segmentation. *Computer methods and programs in biomedicine*, 184:105268, 2020.

[5] S. Ghosh, A. Pal, S. Jaiswal, K. C. Santosh, N. Das, and M. Nasipuri. SegFast-V2: Semantic image segmentation with less parameters in deep learning for autonomous driving. *International Journal of Machine Learning and Cybernetics*, 10(11):3145–3154, 2019.

[6] M. Treml et al. Speeding up semantic segmentation for autonomous driving. In *Proceedings of the 29th Conference on Neural Information Processing Systems*, 2016.

[7] S. Bonechi, P. Andreini, M. Bianchini, and F. Scarselli. COCO_TS dataset: pixel-level annotations based on weak supervision for scene text segmentation. In *International Conference on Artificial Neural Networks*, pages 238–250. Springer, 2019.

[8] K. Bhargavi and S. Jyothi. A survey on threshold based segmentation technique in image processing. *International Journal of Innovative Research and Development*, 3(12):234–239, 2014.

[9] S. K. Ahammad and V. Rajesh. Image processing based segmentation techniques for spinal cord in MRI. *Indian Journal of Public Health Research & Development*, 9(6), 2018.

[10] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang. Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm. *Computers and Electronics in Agriculture*, 165:104962, 2019.

[11] S. Bonechi, M. Bianchini, F. Scarselli, and P. Andreini. Weak supervision for generating pixel-level annotations in scene text segmentation. *Pattern Recognition Letters*, 138:1–7, 2020.

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[13] G. Papandreou, I. Kokkinos, and P.-A. Savalle. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *arXiv preprint arXiv:1412.0296*, 2014.

[14] P. Andreini, G. Ciano, S. Bonechi, C. Graziani, V. Lachi, A. Mecocci, A. Sodi, F. Scarselli, and M. Bianchini. A two-stage GAN for high-resolution retinal image generation and segmentation. *Electronics*, 11(1):60, 2021.

[15] G. Ciano, P. Andreini, T. Mazzierli, M. Bianchini, and F. Scarselli. A multi-stage GAN for multi-organ chest X-ray image generation and segmentation. *Mathematics*, 9(22), 2021.