

INNER CELL MASS SEGMENTATION IN HUMAN HMC EMBRYO IMAGES USING FULLY CONVOLUTIONAL NETWORK

S. Kheradmand, A. Singh, P. Saeedi, J. Au**, J. Havelock***

School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada.

* Signal Processing Group, Department of Engineering, University of Cambridge, U.K.

** Pacific Center for Reproductive Medicine, Burnaby, BC, Canada.

ABSTRACT

The success of In-Vitro Fertilization (IVF) greatly relies on the quality of the Inner Cell Mass (ICM) obtained at day 5 of embryo development. Unfortunately, ICM segmentation is difficult due to its shape variability and unconstrained profile. This paper proposes a two-stage pipeline that first uses a preprocessing step to remove artifacts from the input images which are then used by the Fully Convolutional Networks (FCN) to produce the ICM segmentation. The paper also proposes a novel data augmentation technique, specific to this application. The pre-processing step is shown to accelerate the learning of the FCN while data augmentation avoids overfitting and lead to better generalization. The performance of the proposed pipeline is evaluated on pixel classification accuracy and Jaccard index, on a dataset of 8460 images augmented from 235 images. The proposed method outperforms the state-of-art by about 28% on Jaccard index.

Index Terms— IVF, Blastocyst Segmentation, ICM, Fully Convolutional Networks, Embryo Quality Assessment.

1. INTRODUCTION

In-Vitro Fertilization (IVF) is a five-day process in which an egg is fertilized with a sperm in a laboratory setting to produce a blastocyst. Embryologists manually analyze numerous blastocysts to select the ones with the highest viability. To increase the chance of pregnancy, generally, more than one blastocyst is transferred to the uterus, which may result in multiple births. Hence, selecting just one embryo with the high viability is desired. Gardner [1] introduced a system to measure viability based on 3 factors: blastocyst developmental stage, the quality of Inner Cell Mass (ICM) and Trophectoderm (TE) as shown in Fig. 1. The ICM has been considered the most important predictor of pregnancy and live births because it contains the cells that will give rise to the fetus [2].

Although numerous attempts have been made in the past to automate embryo analysis [3], [4], [5], [6], only two attempts has been successful in segmenting the ICM. Santos et al. [7] proposed a level-set based semi-automatic method for grading blastocyst by extracting ICM. Unfortunately, no

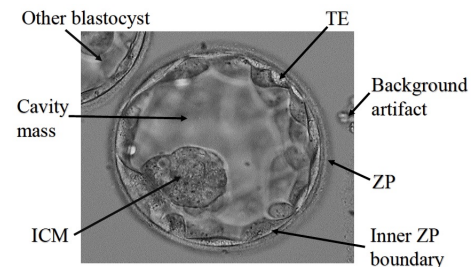


Fig. 1. Illustration shows the components of a blastocyst. The background artifact and partial blastocyst are also marked.

quantitative results representing the outcome of the ICM segmentation are provided. Kheradmand et al. [8], segmented ICM by extracting image patch features that were used by a neural network to perform the required segmentation. However, the reported segmentation quality of the ICM was less than 50%. Since ICM is an important factor in determining the embryo viability, there is a strong need for a method to perform ICM segmentation with high accuracy and quality.

Deep Convolutional Neural Networks (DCNNs) have made great advances at numerous classification [9] and segmentation [10] tasks in computer vision and speech applications over the past few years. Fully Convolutional Networks (FCNs) has recently become popular for medical image segmentation applications. Milletari et al. [11] proposed a fully convolutional neural networks for volumetric medical image segmentation while Wang et al. [12] used them for MRI image segmentation. Xie et al. [13] used FCNs to perform automated microscopic cell counting. However, these networks produce state-of-the-art results only for large datasets and tend to overfit [14] for applications with smaller datasets which is generally the case with medical imaging.

In this paper, a two-stage ICM segmentation pipeline is proposed that first uses the pre-processing method on the input image to remove the background artifacts and partial blastocysts, as shown in Fig. 1. The pre-processed image is then used by the FCN network to learn the invariant and discriminatory features of the ICM region. The removal of the background artifacts improves the learning of the FCN. Since the

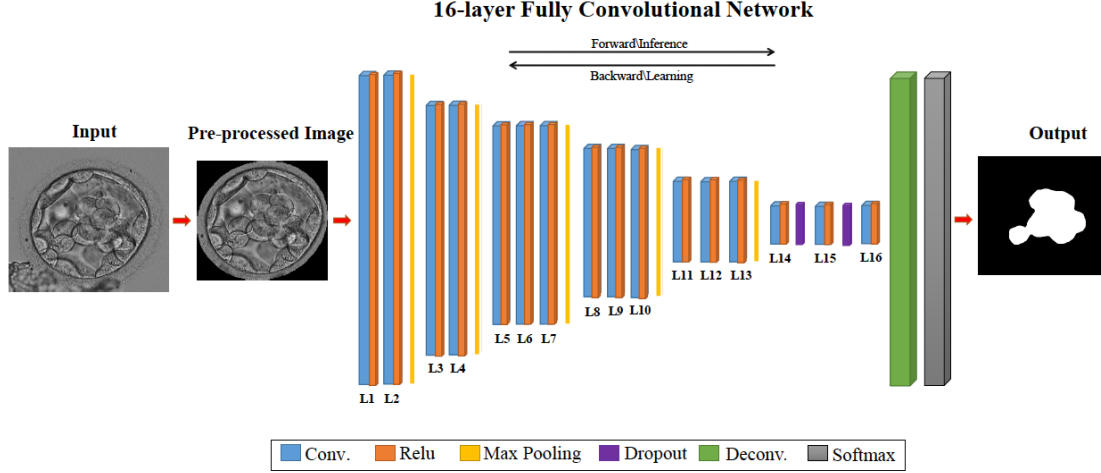


Fig. 2. The input image is preprocessed and then given as input to the 16-layer FCN that learns invariant and discriminatory features of the ICM to produce the desired segmentation. Convolutional layers are labeled, and their details are provided in Table 1.

amount of data present in this application is not sufficient to train an FCN from scratch, a pre-trained network on ImageNet is used (found in [15]). The paper also proposes a data augmentation technique, specific to this application that helps to avoid overfitting of the network and further results in better generalization. The performance is evaluated on the: (i) pixel classification accuracy, and (ii) Jaccard index on the augmented dataset.

Section 2 details the proposed ICM segmentation pipeline. Section 3 presents the proposed augmentation technique with experimental results while Section 4 draws conclusions.

2. ICM SEGMENTATION PIPELINE

This section presents the proposed two-stage ICM segmentation pipeline. The first part details the pre-processing procedure while the second part presents the FCN architecture used for segmentation. The pipeline is shown in Fig. 2.

2.1. Pre-processing

This section details the automatic pre-processing step [5] to remove the background artifacts and partial blastocysts present in the input image by removing pixels outside the ZP boundary. In order to detect the inner ZP, the center of the blastocyst is first detected by fitting a circular model to the edges of the image. To detect the edges that belong to the inner ZP boundary, the circular model is divided to θ ($\theta = 6$) orientations from the center. In each orientation, strong edges with the prominent direction computed using phase congruency are retained while the rest are removed. The retained edges are used to construct the inner ZP boundary (Fig. 3(a)). To ensure that all ICM, TE, and Cavity pixels are inside the detected inner ZP boundary, a disk shape morphological structuring element of size k ($k = 18$) is used to dilate the detected inner ZP boundary. Outside of the dilated inner ZP

is removed after which the image is cropped such that the dilated inner ZP fits the image (Fig. 3(b)).

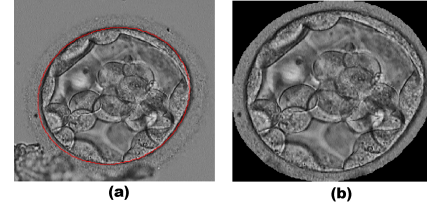


Fig. 3. Pre-processing step. (a) Detected Inner ZP boundary (b) Pre-processed image after removing the background artifacts.

2.2. Fully Convolutional Networks

Fully Convolutional Networks (FCNs) proposed by Long et al. [10] are visual models used to produce pixel-wise image segmentation. These networks use the fractional convolution layers to learn filters for upsampling in order to map or connect coarse outputs to the dense pixel space, to produce per-pixel labels. FCN architecture with 32-pixel stride is used for ICM segmentation as shown in Fig. 2. This architecture consists of 16 convolutional layers, 5 max-pooling layers of size 2×2 , 15 Relu layers, 2 dropout and one fractional (de-convolutional) layer. The filters of the FCN network are initialized to a 16-layer VGG network [16], pre-trained on the ImageNet dataset (found in [15]). The details of the convolutional layers for the FCN network are presented in Table. 1.

Table 1. FCN Architecture [10]: A layer with b number of filters of size $a \times a$ is denoted as: a,b. Lx-Ly denotes that all the layers from Lx to Ly have the same filter size and number of filters.

Layer	L1 - L2	L3 - L4	L5 - L7	L8 - L13	L14	L15	L16
size	3, 64	3, 128	3, 256	3, 512	7, 4096	1, 4096	1, 2

3. EXPERIMENTS

The performance of the proposed ICM segmentation pipeline is demonstrated on an augmented dataset of 8460 images obtained from the original 235 images on (i) pixel classification accuracy and, (ii) Jaccard index. ICM segmentation is also performed using FCN on both raw and preprocessed images to demonstrate the benefits of the preprocessing step. The performance of the proposed architecture is also compared with the state-of-the-art method for ICM segmentation [8]. The dataset used for experimentation and experimental results are presented below. The section below also details the technique used to augment the original dataset.

3.1. Dataset

Data and Ground Truth Acquisition: The blastocyst images used in this experiment have been acquired using an Olympus IX71 inverted microscope with a Nomarski (DIC) optics. The software used for capturing the images is the Research Instrument (RI) Cronus 3. All the images are captured at 1.6 and a lens objective of 20. Each HMC human blastocyst image is manually labeled by an expert embryologist to produce pixel by pixel annotations (ICM or the background).

Dataset Augmentation Technique: The blastocyst dataset used in this paper contains a total of 235 images of different sizes. This dataset is augmented to 8460 images using the proposed augmentation technique, specific to this application. The increase in size of the datasets can help to the overfitting of FCN and provide better generalization. The blastocyst images are elliptical structures with ICM located inside as shown in Fig. 1. In order to increase the number of images, we first apply pre-processing step to all 235 images, then each image is rotated 36 times (10 degrees apart) from the center of the detected ZP. Since blastocysts are elliptical with no cardinal directions, generating the rotated versions will preserve the original data distribution. In without pre-processing pipeline, we rotate the original image around the center of the image. In both cases, a total number of 8460 images are generated.

3.2. FCN Training

FCN network is trained on 80% (6768) of images (10-fold cross validation) chosen randomly from the dataset of 8460 augmented images. It is made sure that if a certain image is present in the training or test set, it's rotated versions are also in the same category. Also, the images that are used for training the network with the preprocessing pipeline are also used for the without preprocessing pipeline. This ensures that a valid non-data dependent comparison is made between the two pipelines. The parameters used to train FCN are shown in Table 2. Since the train and test sets are chosen randomly, the distribution of the images on train and test sets might affect the results. To make sure the evaluation is independent of the data in training and test set, we trained the network 7 times, and each time used different training and test datasets. The results shown in Table 3 and Table 5 are the average over 7

Table 2. Parameter used by FCN while training [10]. Learning Rate (LR), Momentum (Mom.), Batch Size, and Weight Decay (WD)

LR	Mom.	Batch Size	WD
0.0001	0.9	20	0.0005

time experiments. The network training is terminated if the classification error on training and validation sets changes less than 2% of the current value after 10 epochs.

3.3. Quantitative and Qualitative Evaluation

The performance of the ICM segmentation pipeline, evaluated on the quantitative and qualitative results, is presented for both the with and without pre-processing method pipelines.

3.3.1. Quantitative Results

The FCN pipeline trained with and without the pre-processing method is used to segment the ICM region from the blastocyst images. The performance of both the pipelines is evaluated on: (i) Pixel classification accuracy and, (ii) Jaccard Index. Pixel classification accuracy represents the pixels correctly identified for the background as well as ICM region while Jaccard index measures the true pixels identified only for the ICM region. Hence, Jaccard index is the important measure compared to the pixel classification accuracy as it represents the performance of the proposed pipeline to segment the ICM. Both of the metrics are defined below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

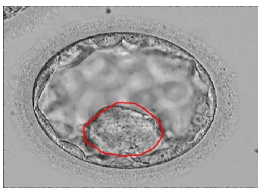
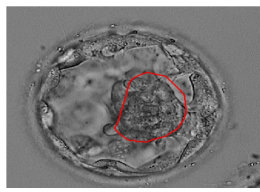


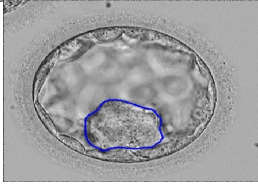
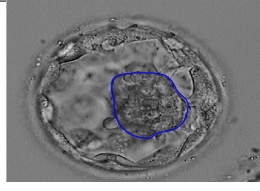

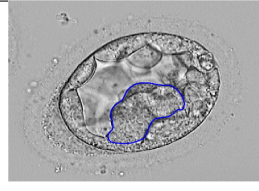
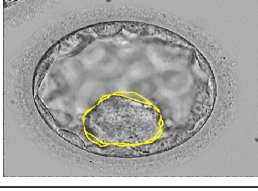
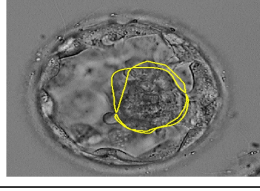
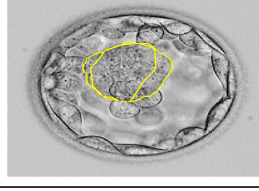
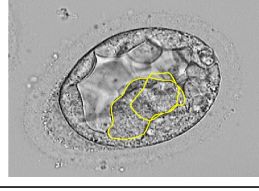
$$Jaccard\ Index = \frac{TP}{TP + FN + FP} \quad (2)$$

TP represents true positives (correctly extracted ICM pixels), FP presents the false positives (incorrectly extracted ICM pixels), FN measures the false negatives (missed ICM pixels) and TN shows true negatives (correctly extracted background pixels).

Table 3 presents the ICM segmentation results on the training and test datasets for both FCN pipelines trained with and without the pre-processing method. As observed from the table, the ability of the pipeline to identify true pixels for the whole blastocyst image (accuracy) is almost same for both pipelines (with and without pre-processing). However, the correct pixels of the ICM region are identified with more than 25% Jaccard index measure for the with pre-processing method approach. In addition, the pre-processing step removes the visual artifacts and partial blastocyst from the image which accelerates the learning of the FCN. It can be observed from the Table 3 that the FCN with preprocessing converges 15 epochs earlier than the FCN pipeline without pre-processing.

In order to observe the effect of data augmentation, the FCN network is trained directly on the original dataset of 235

Table 4. Qualitative results of the network. (First row) The detected ICM region in red. (Second row) The ICM region in GT, provided manually by experts, in blue. (Third row) The difference in detection in yellow. (Last row) Jaccard index (Jac.).

	Input Image 1	Input Image 2	Input Image 3	Input Image 4
Proposed Method				
Ground truth				
Difference				
Jac.	85.9	77.0	68.5	42.4

images (188 for training and 47 for test). The result, as shown in Table 3, prove that the data augmentation helps in a better generalization to test data and avoids overfitting to the training data.

Table 3. Accuracy (%) and Jaccard index (%) for ICM segmentation with and without pre-processing step on Training and Test sets. Number of epochs required for convergence (NC) is also shown. Aug. stands for Augmentation.

	Train set		Test set		
Pre-Processing?	Accuracy	Jaccard	Accuracy	Jaccard	NC
Yes + Aug.	96.8	78.4	95.6	76.5	16
No + Aug.	95.9	60.2	95.1	54.4	31
No Aug.	98.4	52.4	97.4	30.8	–

3.3.2. Qualitative Results

The superior performance of the FCN pipeline trained with the pre-processing step is shown in Table 3. Table 4 presents the segmentation results on four different ICM images obtained using the FCN pipeline which is learned with the pre-processing method. The extracted ICM boundaries are overlaid on the input images. We show the best, the average, and the worst results. We think that such information may provide some perspective for the readers of this paper. In this table, the first row shows the visual presentation of the output ICM region generated by the FCN pipeline. The second row shows the ICM ground truth determined by our expert embryologists. The third row depicts the difference between the output of our algorithm and the ground truth. Finally, the fourth row

lists the Jaccard index in percentage for each image.

3.4. Comparison with the state-of-the-art

Table 5 compares the result of the proposed pipeline with pre-processing to the work proposed in [8]. It can be observed from the table that the proposed method outperforms the state-of-the-art on both accuracy and Jaccard index. It is important to note that the value of Jaccard index is mathematically calculated using the Precision (56.4%) and Recall (75.6%) values stated in [8]. Also, the dataset size is increased from 193 (130 for training and 63 for test) in [8] to 235 in this paper.

Table 5. Comparison of the results based on accuracy (%) and Jaccard index (%) with the state of the art.

	Proposed method	method at [1]
Jaccard Index	76.5	48.0
Accuracy	95.6	93.0

4. CONCLUSION

In this paper, we propose a two-stage pipeline that uses a pre-processing step with FCN to produce ICM segmentation. The paper also proposed a data augmentation technique. The pre-processing step results in accelerated learning of the FCN while the data augmentation method avoids overfitting of the network and provides better generalization. The proposed method was able to outperform the state-of-the-art by a significant margin on the Jaccard index.

5. REFERENCES

- [1] D.K. Gardner, J. Stevens, C.B. Sheehan, and W.B. Schoolcraft, "Analysis of blastocyst morphology," *Human preimplantation embryo selection*, pp. 79–87, 2007.
- [2] C. Lagalla, M. Barberi, G. Orlando, R. Sciajno, M. Bonu, and A. Borini, "A quantitative approach to blastocyst quality evaluation: morphometric analysis and related ivf outcomes," *Journal of assisted reproduction and genetics*, vol. 32, no. 5, pp. 705–712, 2015.
- [3] A. Singh, J. Au, P. Saeedi, and J. Havelock, "Automatic segmentation of trophectoderm in microscopic images of human blastocysts," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 382–393, 2015.
- [4] A. Khan, S. Gould, and M. Salzmann, "Segmentation of developing human embryo in time-lapse microscopy," *Biomedical Imaging (ISBI), IEEE 13th International Symposium on*, pp. 930–934, 2016.
- [5] D. Yee, P. Saeedi, and J. Havelock, "An automatic model-based approach for measuring the zona pellucida thickness in day five human blastocysts," *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, pp. 877–880, 2013.
- [6] J.A. Noble, D. Wells, et al., "A review on automatic analysis of human embryo microscope images," *The open biomedical engineering journal*, vol. 4, no. 1, 2010.
- [7] E. Santos Filho, J.A. Noble, M. Poli, T. Griffiths, G. Emerson, and D. Wells, "A method for semi-automatic grading of human blastocyst microscope images," *Human Reproduction*, vol. 27, no. 9, pp. 2641–2648, 2012.
- [8] S. Kheradmand, P. Saeedi, and I. Bajic, "Human blastocyst segmentation using neural network," *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–4, 2016.
- [9] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CCVPR)*, pp. 3431–3440, 2015.
- [11] F. Milletari, N. Navab, and S.A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *3D Vision (3DV), Fourth International Conference on*, pp. 565–571, 2016.
- [12] Y. Wang, Z. Sun, C. Liu, W. Peng, and J. Zhang, "Mri image segmentation by fully convolutional networks," *IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 1697–1702, 2016.
- [13] W. Xie, J.A. Noble, and A. Zisserman, "Microscopy cell counting with fully convolutional regression networks," *MICCAI 1st Workshop on Deep Learning in Medical Image Analysis*, 2015.
- [14] R. Mao, H. Zhu, L. Zhang, and A. Chen, "A new method to assist small data set neural network learning," *Proceeding of the sixth International Conference on Intelligent Systems Design and Applications*, pp. 17–22, 2006.
- [15] "<https://github.com/bvlc/caffe/wiki/model-zoo>," .
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.