Predicting the Success of Blastocyst Implantation from Morphokinetic Parameters Estimated through CNNs and Sum of Absolute Differences

Julio Silva-Rodríguez*, Adrián Colomer*, Marcos Meseguer†, Valery Naranjo*
*Instituto de Investigación e Innovación en Bioingeniería, I3B, Universitat Politècnica de València, Valencia, Spain
Email: adcogra, vnaranjo@i3b.upv.es
†Instituto Valenciano de Infertilidad, Universidad de Valencia, Valencia, Spain
Email: marcos.meseguer@uv.es

***Abstract*—The process of In Vitro Fertilization deals nowadays with the challenge of selecting viable embryos with the highest probability of success in the implantation. In this topic, we present a computer-vision-based system to analyze the videos related to days of embryo development which automatically extracts morphokinetic features and estimates the success of implantation. A robust algorithm to detect the embryo in the culture image is proposed to avoid artifacts. Then, the ability of Convolutional Neural Networks (CNNs) for predicting the number of cells per frame is novelty combined with the Sum of Absolute Differences (SAD) signal in charge of capturing the amount of intensity changes during the whole video. With this hybrid proposal, we obtain an average accuracy of 93% in the detection of the number of cells per image, resulting in a precise and robust estimation of the morphokinetic parameters. With those features, we train a predictive model based on Random Forest classifier able to estimate the success in the implantation of a blastocyst with more than 60% of precision.**

***Index Terms*—Embryo, IVF, morphokinetic parameters, implantation, image processing, machine learning, deep learning.**

I. Introduction

Nowadays, infertility is one of the main problems of reproductive health in the current society. According to [1], in 2010 around 48.5 million couples were affected by infertility problems. In Vitro Fertilization (IVF) is one of the main techniques to face this problem. However, this process is characterized by a low pregnancy rate, propitiating the implantation of more than one blastocyst that could lead to multiple pregnancies [2]. In order to mitigate the number of undesired multiple births, the quality of the implanted blastocyst is under research in recent years [3]. Time-lapse embryo imaging during the in vitro culture provides images of the embryo each 5-25 minutes. These images allow analyzing the morphology during its different stages without affecting the culture conditions with the aim of selecting the best blastocysts to be transferred [4]. Currently, the challenge resides in obtaining relevant features which correlate with the success in the implantation. Computer-vision-based systems are vital to accomplish this objective because of their ability to deal with huge amounts of data and remove the high level of subjectivity and intra-observer variability.

Many morphokinetic parameters that correlate with embryo implantation have been proposed in the literature, such as the polar body extrusion [5], pro-nuclei quality [6], grade of fragmentation in the cells [7] or the zona pellucida thickness [8]. However, the most relevant features related to embryo quality are the timing of the first cell divisions [9]. Those parameters are defined as we illustrate in Fig. 1. In particular, the timestamps of embryo division from 2 to 5 cells are represented as $t2$, $t3$, $t4$, $t5$ respectively. Additionally, the temporal intervals related to the mitosis are defined as a linear combination of cell division instants, being $cc2 = t3 - t2$ and $cc3 = t5 - t3$. The correlation of those morphokinetic features (related to the timing in the cellular division) with the success in embryo implantation has been analyzed in [9]–[11]. In particular, the authors of [9] present a hierarchical model based on the fact that embryos presenting timing division levels far from the population average values have less probability of implantation. It is important to remark that in the most of the state-of-the-art approaches involving the computation of morphokinetic features, the creation of automatic predictive models for embryo implantation is not addressed [12]–[15].

This work presents an artificial-intelligence-based system that takes as an input the time-lapse videos of the embryo culture and automatically estimates the morphokinetic features detailed in [9] to predict the success of the implantation. We present a robust algorithm to locate the embryo in the culture image avoiding artifacts in the next morphokinetic feature-extraction stage. The feature-extraction step represents the main contribution of this paper. We propose a hybrid approach based on the combination of the predicted number of cells per frame using Convolutional Neural Networks (CNNs) and the temporal information of the Sum of Absolute Differences (SAD). This proposed hybrid methodology outperforms the most relevant state-of-the-art methods of automatic estimation of morphokinetic parameters. In the second step and making use of the obtained features, we train a Random Forest classifier with the aim of predicting the success in the implantation. To the best of our knowledge, this is the first time that a fully-automated system (without combining image-based features with manually morphological classification) to characterise the

978-9-0827-9703-9/19/$31.00

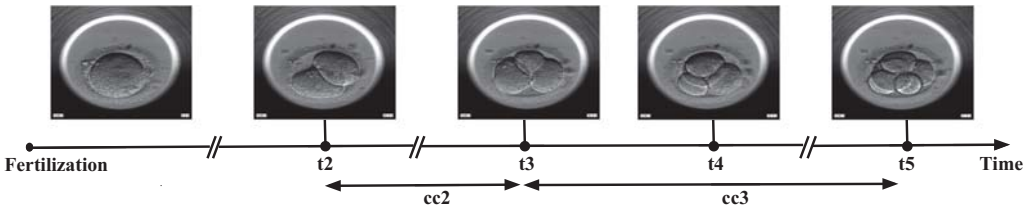


Fig. 1. Morphokinetic parameters related to first embryo cleavage.

embryo quality and estimate the embryo success in the embryo implantation is presented.

II. Materials and methods

A. Materials

The database used in this work is composed of 263 time-lapse videos in which the monitoring of the embryo development during the first days is carried out by means of the EmbryoSlide® system in the Valencian Institute of Infertility (IVI). The acquired videos provide a gray-level vision of the embryo, registering one image of 500×500 pixels every 15 minutes during 4 days. The embryos belong to IVF processes where only one embryo was transferred. For each embryo of the database, a label containing if it was a successful birth (liveborn) or unsuccessful implantation (non-implantation) is available. For the validation of the morphokinetic features, every frame of each video was manually annotated according to the number of cells in the image (i.e. 1 cell, 2 cells, 3 or 4 cells, 5 or more cells). Note that the database is composed of around 70000 images, where the aforementioned classes are distributed in 25%, 20%, 20% and 35% respectively.

B. Preprocessing

Due to the acquisition procedure of embryo culture videos, some gel artifacts and sperm are present in the raw dataset. The first step of the proposed methodology focuses on the automatic extraction of a region of interest (ROI) centered on the blastocyst with the aim of avoiding noise in the following steps of the algorithm. Previous works developed the detection of the ROI removing the known background of the microscope device, and thresholding the image to obtain the biggest element centroid [13]. This method could be sensitive to the number of artifacts and changes in the illumination, that could affect the threshold applied. In this work, a robust algorithm for embryo location in the culture is developed. This is based on the changes produced in the cytoplasm of the embryo between consecutive frames, while the artifacts in the image remain constant. Let $v[x, y, t]$ be a grayscale embryo video, with x and y the spatial coordinates and t the time instant. The following steps are defined to automatically obtain a ROI centered on the embryo.

1) Obtain the absolute difference between one frame (Fig. 2 (a)) and the following. The result of this operation is a image with speckle noise inside the embryo (see Fig. 2 (b)).

$$v_d[x, y] = abs(v[x, y, t_i] - v[x, y, t_{i+1}]) \quad (1)$$

2) Smooth the speckle image with a convolutional gaussian filter $G[x, y]$ (Fig. 2 (c)).

$$v_{ds}[x, y] = v_d[x, y] \circledast G[x, y] \quad (2)$$

3) Enhance the center of the embryo, applying to $v_{ds}[x, y]$ an erode operation with a circular element $ee[x, y]$ of radious $ree = 70$ pix. (Fig. 2 (d)).

$$v_{dse}[x, y] = (v_{ds} \ominus ee)[x, y] \quad (3)$$

4) Detection of the embryo center coordinates $[c_x c_y]$, being those the points where the projections of the image in x and y axis (see Fig. 2 (e)) get the maximum intensity.

$$c_x = argmax(\sum_y v_{dse}[x, y]) \quad (4)$$

$$c_y = argmax(\sum_x v_{dse}[x, y]) \quad (5)$$

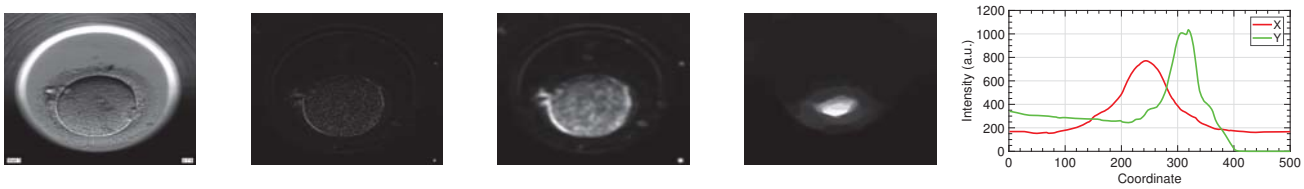
Once the center of the embryo is obtained, a rectangular window around it is cropped, whose size is constant for all the samples.

C. A hybrid method to estimate morphokinetic parameters in embryos

Once the embryo is located and extracted from the image, the next step is to detect the cell division timings. For that purpose, we propose a novel method built as a combination of two different approaches. The first one is based on the estimation of the amount of intensity changes produced in each frame, considering that the intensity changes produced by a cellular division in consecutive frames are higher than those produced in regular frames in which the embryo remains constant. The intensity changes of consecutive video frames can be computed following the Sum of Absolute Differences (SAD) defined as follows:

$$SAD[t_i] = abs(\sum_x \sum_y (v[x, y, t_i] - v[x, y, t_{i+1}])) \quad (6)$$

The SAD signal provides a value per frame, and it registers peaks in the cellular division moments as shown in Fig. 3



(a) (b) (c) (d) (e)

Fig. 2. Algorithm for embryo location in the culture image. (a): One frame $v[x, y, t_i]$. (b): absolute difference image between (a) and the consecutive frame, $v_d[x, y]$. (c): $v_{ds}[x, y]$, image (b) after applying a gaussian filter to smooth it. (d): $v_{dse}[x, y]$, results of applying a morfological filter of erosion over (c). (e): projections in x and y of (d).

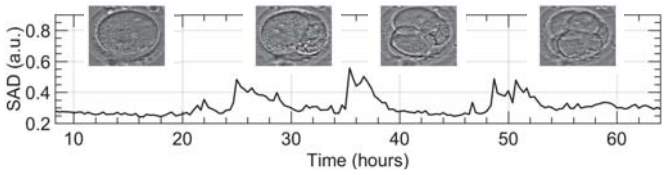


Fig. 3. Signal SAD and cellular divisions.

The second approach focuses on building a classifier able to predict the number of cells in each frame. Let $v[x, y, t_i]$ be an embryo grayscale image at instant i, the objective is to predict the label l_i, corresponding to the different target classes defined as $l \in L$ {1 cell, 2 cells, 3-4 cells, 5+ cells}. Note that the images with 3 and 4 embryos are joined in the same class. Due to the low frame rate used in the acquisition process, we have available an insufficient number of frames to discriminate between 3 and 4 cells. Then, for all the frames in one video, the signal $N[t]$, i.e. the number of cells in each instant, is obtained. This approach is developed following a strategy based on automatic feature learning through Convolutional Neural Networks (CNNs). The proposed architecture (presented in Table I) is built upon the CNN proposed in [13] and it is trained using the cropped embryo images as input during 15 epochs in which the cross-entropy function is minimized by means of the Stochastic Gradient Descent (SGD) optimizer defining a learning rate $lr = 0.005$ and a batch size $bs = 32$.

In the inference stage, the CNN gives, for each image, a probability of membership to each one of the possible classes. Of course, the final label for each video frame is obtained by the class which provides the highest probability. The CNN does not take into account biological constrictions to take its prediction. For example, its decision does not contemplate that the number of cells cannot decrease along time. In this work, we provide this information to the predictive model by means of Conditional Random Field (CRF) [15]. This mechanism consists of minimizing the energy of the cost function defined as follows:

$$E(N[t]) = \sum_{t=1}^{t=T} \psi_t^U(N_t) + \sum_{t=1}^{t=T-1} \psi_t^P(N_t, N_{t+1}) \qquad (7)$$

Being $\psi_t^U(N_t)$ the unitary term, defined as the inverse logarithm of the probability given by the CNN in the number of cells, and $\psi_t^P(N_t, N_{t+1})$ the pair term, that gives infinite energy if $N_t <= N_{t+1}$, and energy zero otherwise. The process is implemented iteratively.

TABLE I
ARCHITECTURE OF THE CNN USED.

N	Layer	Size
0	Input	128 x 128
1	Conv	24 x 11 x 11
2	Conv	64 x 5 x 5
3	Conv	96 x 3 x 3
4	Conv	96 x 3 x 3
5	Conv	64 x 3 x 3
6	FC	512
7	FC	512
8	FC	512
9	FC	4
10	Softmax	64 x 3 x 3
11	Classification	64 x 3 x 3

Conv: Convolutional + Relu
Fc: Fully-Connected + Relu
Max-Pooling 3 x 3 applied after Conv 1, 2 and 5

The signal $N[t]$ is constrained by the image acquisition system that captures images 2D of a 3D object. This produces that, after one cellular division, one cell could occlude others, being impossible to detect them. This is a problem for the predictive models which provide $N[t]$. However, the $SAD[t]$ signal should register a peak at the cell division instant taking into advantage the information of the previous and current frames. Then, both signals have complementary information and, as a novelty in this work, they are combined to improve the accuracy and robustness of the system. In order to compute the morphokinetic features, the transitions registered by the $N[t]$ signal are analyzed (e.g. the temporal instants where $N[t]$ changes from one to two cells is assigned as $t2$ parameter). If one of the transitions is lost, the complementary information captured by the $SAD[t]$ signal is introduced. In particular, the timestamp in which the $SAD[t]$ signal presents a prominent peak in the expected area of the lost transition is now selected to calculate the morphokinetic parameter to be estimated.

D. Non-implantation / Liveborn prediction

Once the morphokinetic features are obtained, a machine-learning model is trained to predict the success in the embryo implantation. The labels used are liveborn if it is successful

and non-implantation otherwise. The extensive study of the morphokinetic features developed in [9] shows the non-linear correlation of those with implantation success. Taking this fact into account, a Random Forest model is trained as predictor. To the best of the authors' knowledge, this is the first time that a non-linear classifier is trained to predict the success in the implantation from the previously estimated timings of first cellular divisions.

III. Results

A. Evaluation of the morphokinetic parameters estimation

In order to validate the proposed methodology, it is followed a 5 fold cross-validation strategy performed on the 80% of the videos of the database, while the remaining 20% of the videos are used to test the predictive models. Regarding the estimation of the morphokinetic parameters, the prediction of the number of cells per frame is evaluated. Note that the accuracy per class is the figure of merit used to quantify the performance of each model. The results obtained for each combination of the methods presented in II are shown in Table II. In particular, the raw predictions of the CNN are compared with the predictions obtained when the CNN output is combined with the information of the SAD signal. Additionally, the influence of the post-processing CRF method is studied. Note that a comparison with the work proposed in [13] is also reported although it is not a direct comparison due to the inexistence of public databases to establish fair analogies.

TABLE II
Precision in the cell number detection

-	Accuracy(%)				
Method	**1 cell**	**2 cells**	**3-4 cells**	**5+ cells**	**Avg.**
Validation set					
CNN	93.99	82.03	70.68	76.97	80.93
CNN+CRF	95.70	88.20	74.45	85.68	86.01
CNN+CRF+SAD	**98.11**	**91.60**	**82.39**	**97.34**	**92.36**
Khan et al. [13]	100.00	99.47	82.60	90.54	93.15
Test set					
CNN+CRF+SAD	**99.56**	**87.83**	**83.46**	**95.12**	**91.49**

CNN: Convolutional Neural Network ; CRF: Conditional Random Field
SAD: Sum of Absolute Differences

The CNN approach to predict the number of cells obtains an average accuracy of 80%, while the CRF post-processing improves the average accuracy around 6%. As main novelty, in the present work, a combination of the SAD signal with the CNN + CRF model is proposed, to improve those cases in which cells are covered by other cells or artifacts. When SAD information is included the results improve in a 6% in average, obtaining a multi-class accuracy of 92%. Remark that the precision obtained when predicting +5 cells outperforms the most relevant state-of-art works (remember about the indirect character of the comparison). The results obtained in the testing database are also presented in Table II. A similar performance of the system can be observed. This fact reflects the absence of overfitting in the trained predictive models and the robustness in the estimation of the number of cells per frame that allows an accurate calculation of the morphokinetic parameters.

B. Validation of the non-implantation/liveborn predictive model

The prediction of the success in embryo implantation is done with the morphokinetic parameters presented in I following the conclusions of [9] about the possibility to improve pregnancy rates with that information. The results about the prediction of the non-implantation (NI) or liveborn (LB) of the embryo are presented for each one of the 5 cross-validation groups separately. Specifically, the accuracy per class and the average values are reported in Table III.

TABLE III
Accuracy of LB/NI prediction in blastocyst transfer

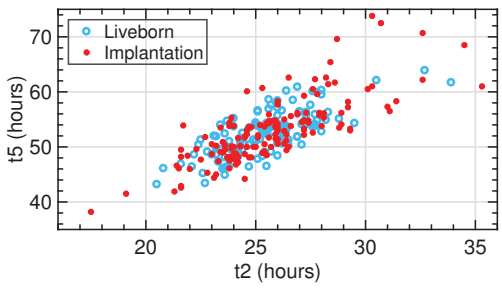
-	Accuracy(%)		
Group	**NI**	**LB**	**Avg.**
CV-I	43.33	**66.66**	54.99
CV-II	**70.83**	38.88	54.85
CV-III	56.52	61.11	58.81
CV-IV	62.50	42.30	52.40
CV-V	50.00	33.33	41.66
CV-Avg.	56.63	48.46	52.54
Test	73.33	46.70	**60.00**

CV: Cross-Validation
NI: Non-Implantation ; LB: Liveborn

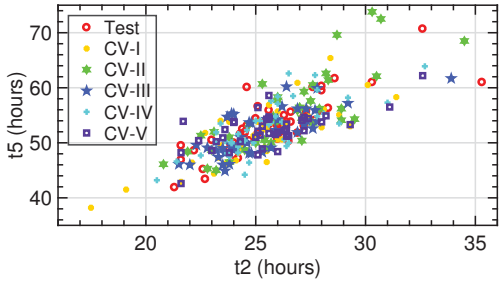
The average precision for both classes in the training set is 52%. This could be interpreted as a random classifier between both classes, without prediction ability. However, analyzing each cross-validation fold separately, some of them present a higher precision in the classification, mainly for the non-implantation class (i.e. CV-II, CV-III or CV-IV). The performance of the LB/NI predictive model on the test set should be carefully analyzed. A wide improvement on the accuracy for the non-implantation class (73.33%) is registered giving place to an average precision for both classes of 60%.

In order to understand the low results obtained during the cross-validation strategy, and study the ability of the morphokinetic features to estimate the implantation of the embryo, Fig. 4 shows two of the features, concretely $t2$ against $t5$. In particular in Fig. 4(a), we represent $t2$ against $t5$ according to the liveborn/non-implantation, while in Fig. 4.(b) $t2$ versus $t5$ is depicted according to the training and test group, where each sample was randomly assigned.

In Fig. 4(a) it can be observed how the majority of the samples are scattered around the average value of $t2$ and $t5$, independently to the labelled class (i.e. liveborn and non-implantation). The main differences between classes can be shown in the extreme values, those with high and low $t2$ and $t5$ parameters. Note that, in these extreme cases there exist more samples with non-implantation than liveborn. The differences in the results obtained between the different cross-validation groups and the test subset can be explained paying attention to the distribution of the samples in the groups. Fig. 4(b) shows that in the groups CV-I and CV-V there are not many



(a)



(b)

Fig. 4. Distribution of parameter $t2$ (in hours) against $t5$ (in hours). In (a) those parameters are represented according to the embryo labels (i.e. liveborn/non-implantation), while in (b) the representation is performed according to the groups in cross-validation and test. CV: cross-validation.

samples with high values of $t2$ and $t5$ features and this could propitiate a drop in the precision for the non-implantation class. By contrast, in the test group, there are few samples with high values in $t2$ and $t5$ and the final predictive model is trained using all the cross-validation samples, ensuring extreme cases in the training set. This fact suggests that the results obtained on the test subset represent better the quality of the morphokinetic features to predict the success in the implantation. Increasing the amount of samples of the database, a more scattered random distribution of all possible values $t2$ and $t5$ along the groups would propitiate more consistent results in the cross-validation strategy.

IV. Conclusions

In this paper we present a system to predict the success in the implantation of the blastocyst from time-lapse videos of the in vitro culture, using morphokinetic parameters related to the timing in the first cell divisions. The computer-vision-based system covers different steps. First, a robust pre-processing method to detect the embryo in the image. Then the morphokinetic parameters are accurately calculated, improving the state-of-the-art results, with a novel combination of CNNs to predict the number of cells per frame and the SAD signal that estimates the amount of intensity changes along the time-lapse video. Finally, and for the first time in this field, a Random Forests model is trained to automatically predict the success in the implantation of the blastocysts from the morphokinetic parameters previously estimated.

In future works, the whole information of the embryo videos should be taken into account instead of only considering the discrete information given by the morphokinetic parameters. To accomplish this objective and estimate the success in the implantation, a hybrid approach composed by CNNs and recurrent neural neworks (RNN) will be proposed. Additionally, more time-lapse videos will be acquired enlarging the blastocyst database.

Acknowledgement

This work has been founded by the Ministry of Economy, Industry and Competitiveness under the SICAP project (DPI2016-77869-C2-1-R). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

References

[1] Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA, "National, regional, and global trends in infertility prevalence since 1990: A systematic analysis of 277 health surveys", 2012.

[2] De Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, Sccaravelli G, Smeenk J, Vidakovic S and Gooddsens V, "ART in Europe, 2014: results generated from European registries by ESHRE", Human Reproduction, vol. 33, 2018.

[3] Casper R, Haas J, Hsieh TB, Bassil R ans Mehta C, "Recent advances in in vitro fertilization [version 1; referees: 2 approved]", F1000Research 2017, 6(F1000 Faculty Rev):1616, 2017.

[4] Aparicio-Ruiz B, Basile N, Perez S, Bronet F, Remohi J and Meseguer M, "Automatic time-lapse instrument is superior to single-point morphology observation for selecting viable embryos: retrospective study in oocyte donation", Fertility and Sterility, vol. 106, 2016.

[5] Payne D, Flaherty SP, Barry MF and Matthews CD, "Preliminary observations on polar body extrusion and pronuclear formation in human oocytes using time-lapse video cinematography", Human Reproduction, vol. 12, 1997.

[6] Azzarello A, Hoest T and Mikkelsen AL, "The impact of pronuclei morphology and dynamicity on live birth outcome after time-lapse culture", Human Reproduction, vol. 27, 2012.

[7] Santos Filho E, Noble JA and Wells D, "A review on automatic analysis of human embryo microscope images", The Open Biomedical Engineering Journal, 2010.

[8] Coello A, Meseguer M, Galn A, Alegre L, Remohi J and Cobo A, "Analysis of the morphological dynamics of blastocysts after vitrification/warming: defining new predictive variables of implantation", Fertiliy and Sterility, 2017.

[9] Meseguer M, Harrero J, Tejera A, Hilligsoe KM, BirgerRamsing N and Remohi J, "The use ef morphokinetics as a predictor of embryo implantation", Human Reproduction, 2011.

[10] Hlinka D, Kalatova B, Uhrinova I, Dolinska S, Rutarova J, Rezacova J, Lazarovska S, Dudas M,"Time-Lapse Cleavage Rating Predicts Human Embryo Viability". Physiological Research, 2012.

[11] Milewski R, Ajduk A, "Time-lapse imaging of cleavage divisions in embryo quality assesment", Society for Reproduction and Fertility, vol. 154, 2017.

[12] Morales D, Bengoetxea E and Larraaga P, "Automatic segmentation of zona pellucida in human embryo images applying an active contour model", Proceedings of the Medical Imgae Understanding and Analysis, 2008.

[13] Khan A, Gould S, Salzmann M, "Deep Convolutional Neural Networks for Human Embryonic Cell Counting", Computer Vision ECCV 2016 Workshops. Lecture Notes in Computer Science, vol. 9913, 2016.

[14] Khan A, Gould S, Salzmann M, "Automated monitoring of human embryonic cells up to 5-cell stage in time-lapse microscopy images", 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015.

[15] Khan A, Gould S, Salzmann M, "A Linear Chain Markov Model for Detection and Localization of Cells in Early Stage Embryo Development", 2015 IEEE Winter Conference on Applications of Computer Vision, 2015.