HHS Public Access

Author manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2020 December 11.

Published in final edited form as:

Med Image Comput Comput Assist Interv. 2020 October ; 12265: 25–35.

doi:10.1007/978-3-030-59722-1_3.

Automated Measurements of Key Morphological Features of Human Embryos for IVF

B. D. Leahy[1,2,*], **W.-D. Jang**[1,*], **H. Y. Yang**[3,*], **R. Struyven**[1], **D. Wei**[1], **Z. Sun**[1], **K. R. Lee**[2], **C. Royston**[2], **L. Cam**[2], **Y. Kalma**[4], **F. Azem**[4], **D. Ben-Yosef**[4], **H. Pfister**[1], **D. Needleman**[1,2]

[1]School of Engineering and Applied Sciences,Harvard University, Cambridge MA 02138, USA

[2]Department of Molecular and Cellular Biology,Harvard University, Cambridge MA 02138, USA

[3]Harvard Graduate Program in Biophysics, Harvard University, Cambridge MA 02138, USA

[4]Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

Abstract

A major challenge in clinical In-Vitro Fertilization (IVF) is selecting the highest quality embryo to transfer to the patient in the hopes of achieving a pregnancy. Time-lapse microscopy provides clinicians with a wealth of information for selecting embryos. However, the resulting movies of embryos are currently analyzed manually, which is time consuming and subjective. Here, we automate feature extraction of time-lapse microscopy of human embryos with a machine-learning pipeline of five convolutional neural networks (CNNs). Our pipeline consists of (1) semantic segmentation of the regions of the embryo, (2) regression predictions of fragment severity, (3) classification of the developmental stage, and object instance segmentation of (4) cells and (5) pronuclei. Our approach greatly speeds up the measurement of quantitative, biologically relevant features that may aid in embryo selection.

Keywords

Deep Learning; Human Embryos; In-Vitro Fertilization

1 Introduction

One in six couples worldwide suffer from infertility [7]. Many of those couples seek to conceive via In-Vitro Fertilization (IVF). In IVF, a patient is stimulated to produce multiple oocytes. The oocytes are retrieved, fertilized, and the resulting embryos are cultured *in vitro*. Some of these are then transferred to the mother's uterus in the hopes of achieving a pregnancy; the remaining viable embryos are cryopreserved for future treatments. While transferring multiple embryos to the mother increases the potential for success, it also increases the potential for multiple pregnancies, which are strongly associated with increased maternal morbidity and offspring morbidity and mortality [25]. Thus, it is highly

bleahy@seas.harvard.edu.

*These authors contributed equally to this work.

desirable to transfer only one embryo, to produce only one healthy child [29]. This requires clinicians to select the best embryos for transfer, which remains challenging [27].

The current standard of care is to select embryos primarily based on their morphology, by examining them under a microscope. In a typical embryo, after fertilization the two pronuclei, which contain the father's and mother's DNA, move together and migrate to the center of the embryo. The embryo undergoes a series of cell divisions, during the "cleavage stage." Four days after fertilization, the embryo compacts and the cells firmly adhere to each other, at which time it is referred to as a compact "morula." On the fifth day, the embryo forms a "blastocyst," consisting of an outer layer of cells (the trophectoderm) enclosing a smaller mass (the inner-cell mass). On the sixth day, the blastocyst expands and hatches out of the zona pellucida (the thin eggshell that surrounds the embryo) [10]. Clinicians score embryos by manually measuring features such as cell number, cell shape, cell symmetry, the presence of cell fragments, and blastocyst appearance [10], usually at discrete time points. Recently, many clinics have started to use time-lapse microscopy systems that continuously record movies of embryos without disturbing their culture conditions [30,9,3]. However, these videos are typically analyzed manually, which is time-consuming and subjective.

Previous researchers have trained convolutional neural networks (CNNs) to directly predict embryo quality, using either single images or time-lapse videos [26,32]. However, interpretability is vital for clinicians to make informed decisions on embryo selection, and an algorithm that directly predicts embryo quality from images is not interpretable. Worse, since external factors such as patient age [12] and body-mass index [5] also affect the success of an embryo transfer, an algorithm trained to predict embryo quality may instead learn a representation of confounding variables, which may change as IVF practices or demographics change. Some researchers have instead trained CNNs to extract a few identifiable features, such as blastocyst size [18], blastocyst grade [20,11,19], cell boundaries [28], or the number of cells when there are 4 or fewer [17,21]. While extracting identifiable features obviates any problems with interpretability, these works leave out many key features that are believed to be important for embryo quality. Moreover, these methods are not fully automated, requiring the input images to be manually annotated as in the cleavage or blastocyst stage.

Here, we automate measurements of five key morphokinetic features of embryos in IVF by creating a unified pipeline of five CNNs. We work closely with clinicians to choose features relevant for clinical IVF: segmentation of the zona pellucida (Fig. 1a), grading the degree of fragmentation (Fig. 1b), classification of the developmental stage from 1-cell to blastocyst (Fig. 1c), object instance segmentation of cells in the cleavage stage (Fig. 1d), and object instance segmentation of pronuclei before the first cell division (Fig. 1e). With the exception of zona pellucida segmentation, all these features are used for embryo selection [1,27,2,24]; we segment the zona pellucida both to improve the other networks and because zona properties occasionally inform other IVF procedures [6]. The five CNNs work together in a unified pipeline, combining results to improve performance over individual CNNs trained per task by several percent.

2 Dataset

We train the CNNs using data from the Embryoscope®, the most widely-used system for IVF with standardized, time-lapse microscopy [9]. Embryoscope® images are grayscale, and taken using Hoffman Modulation Contrast microscopy [16], in which the intensity roughly corresponds to refractive index gradients. In our dataset, the Embryoscope® takes an image every 20 minutes at 7 focal planes, usually at 15 μm increments. The recorded images provide views of the embryo with different amounts of defocus; they do not provide 3D information. The embryos are recorded for 3 – 5 days, corresponding to 200 – 350 images at each focal plane (*i.e.*, 1400 – 2450 images per embryo), although embryos are occasionally removed from the incubation system for clinical procedures. To train the CNNs, we curate a dataset with detailed, frame-by-frame labels for each task.

3 Our Pipeline

For each time-lapse video, we measure 5 morphokinetic features using 5 networks:

Zona Pellucida Segmentation:

We first perform semantic segmentation to identify regions of the embryo, segmenting the image into four regions: pixels outside the well, inside the well, the zona pellucida, and the space inside the zona pellucida (the perivitelline space and embryo; Figure 2, left). We segment the images by using a fully-convolutional network (FCN [23]; based on Resnet101 [15]) to predict a per-pixel class probability for each pixel in the image. We train the FCN with images chosen from 203 embryos at 3,618 time points; we use neither a separate validation set nor early stopping for the zona segmentation.

The zona pellucida segmentation network in our pipeline takes the full 500×500 pixel image as input. We use the segmentation result to crop the images to 328×328, centered around the embryo, as input for the other networks.

Fragmentation Scoring:

With the cropped image from the zona pellucida segmentation, we score the embryo's degree of fragmentation using a regression CNN (InceptionV3 [31]). The network takes a single-focus image as input and predicts a fragmentation score of 0 (0% fragments), 1 (<10%), 2 (10–20%), or 3 (≥20%), following clinical practice. We train the network to minimize the L^1 loss on cleavage-stage images of 989 embryos at 16,315 times, each labeled with an integer score from 0–3; we use a validation set of 205 embryos labeled at 3,416 times for early stopping [13]. For each time point in the movie we analyze, we run the CNN on the three middle focal planes and take the average as the final score (Figure 3, left).

Counting and identifying cells in fragmented embryos is difficult, inhibiting the labeling of train or test data for these embryos. Moreover, since high fragmentation is strongly correlated with low embryo viability [1], in standard clinical practice highly fragmented embryos are frequently discarded. Thus, we only train the rest of the networks on embryos with fragmentation less than 2.

Stage Classification:

For low fragmentation embryos, we classify the embryo's developmental stage over time using a classification CNN (ResNeXt101 [33]). The classifier takes the three middle focal planes as input and predicts a 13-element vector of class probabilities, with 9 classes for cleavage-stage embryos (one each for 1–8 cells and one for ≥ 9 cells) and one class each for morula (M), blastocyst (B), empty wells (E), and degenerate embryos (D; Figure 4, left). To account for inaccuracies in the training data labels, we trained the classifier with a soft loss function modified from the standard cross-entropy loss

$$\log(p(\ell \mid m)) = \log\left(\sum_t p(\ell \mid t)p(t \mid m)\right), \quad (1)$$

where t is the true stage of an image, ℓ the (possibly incorrect) label, and m the model's prediction. We measured $p(\ell|t)$ by labeling 23,950 images in triplicate and using a majority vote to estimate the true label t of each image. This soft-loss differs from the regularized loss in [31] by differentially weighting classes; for instance, $p(\ell = \text{1-cell}|t = \text{1-cell}) = 0.996$ whereas $p(\ell = \text{6-cell}|t = \text{6-cell}) = 0.907$. Using the measured $p(\ell|t)$, we then trained the network with 341 embryos labeled at 111,107 times, along with a validation set of 73 embryos labeled at 23,381 times for early stopping [13]. Finally, we apply dynamic programming [4] to the predicted probabilities to find the most-likely non-decreasing trajectory, ignoring images labeled as empty or degenerate (Figure 4, center).

Cell Object Instance Segmentation:

For the images identified by the stage classifier as having 1–8 cells, we next perform object instance segmentation on each cell in the image. We train a network with the Mask R-CNN architecture [14] and a ResNet50 backbone [15], using 102 embryos labeled at 16,284 times with 8 or fewer cells; we also use a validation set of 31 embryos labeled at 4,487 times for early stopping [13]. Our instance segmentation model takes as input a single-focus image cropped from the zona segmentation and resized to 500×500. The segmentation model then predicts a bounding box, mask, and confidence score for each detected cell candidate (Figure 5, left). Both the ground-truth labels and the predicted masks overlap significantly when the embryo has 2–8 cells (Figure 5, center). We produce a final prediction by running our segmentation model on the three central focal planes; we merge candidates found across focal planes by using the one with the highest confidence score.

Pronucleus Object Instance Segmentation:

Finally, in the images identified as 1-cell by the stage classifier, we detect the presence of pronuclei. To do so, we train another object instance segmentation network with the Mask R-CNN architecture [14] and a ResNet50 backbone [15]. We use a training set of 151 embryos labeled at 9,250 times during the 1-cell stage, with a validation set of 33 embryos labeled at 1,982 times for early stopping [13]. Pronuclei are only visible during a portion of the 1-cell stage; correspondingly, about 38% of the training images contain 0, 6% contain 1, and 54% contain 2 pronuclei. The pronuclei detector takes as input a single image, cropped from the zona pellucida segmentation and resized to 500×500, and it predicts a bounding box, mask, and confidence score for each detected candidate (Figure 6, left). We run the

pronuclei detector on the three middle focal planes and merge candidates by using the one with the highest confidence score.

4 Results

We now evaluate our pipeline's performance, demonstrating the effect of each design choice in the models with ablation studies.

Zona Pellucida Segmentation:

Our zona pellucida network nearly optimally segments the test set images, taken from 36 embryos at 576 times. The FCN correctly labels image pixels 96.7% of the time, with per-class accuracies between 93–99% (Figure 2, right). The misclassified pixels arise mostly at region boundaries, roughly corresponding to the few-pixel human labeling inprecision at region boundaries.

Fragmentation Scoring:

The network predicts a score with a mean-absolute deviation of 0.45 from the test labels on the fragmentation test set of 216 embryos labeled at 3,652 times (Figure 3, right). When distinguishing between low(< 1.5) and high- (≥ 1.5) fragmentation, the network and the test labels agree 88.9% of the time. Our network outperforms a baseline InceptionV3 by 1.9%; focus averaging and cropping to a region-of-interest each provide a 1–1.5% boost to the accuracy (Table 1).

We suspect that much of the fragmentation network's error comes from imprecise human labeling of the train and test sets, due to difficulties in distinguishing fragments from small cells and due to grouping the continuous fragmentation score into discrete bins. To evaluate the human labeling accuracy, two annotators label the fragmentation test set in duplicate and compare their results. The two annotators have a mean-absolute deviation of 0.37 and are 88.9% consistent in distinguishing low- from high- fragmentation embryos. Thus, the fragmentation CNN performs nearly optimally in light of the labeling inaccuracies.

Stage Classification:

The stage classifier predicts the developmental stage with a 87.9% accuracy on the test set, consisting of 73 embryos labeled at 23,850 times (Figure 4, right). The network's accuracy is high but lower than the human labeling accuracy on the test set (94.6%). The network outperforms a baseline ResNeXt101 by 6.7%; both the soft-loss and the dynamic programming each improve the predictions by 2% (Table 1). The stage classifier struggles when there are between 5 and 8 cells (66.9% accuracy for these classes). In contrast, the stage classifier does exceedingly well on images with 1-cell (99.9%), 2-cells (98.7%), empty wells (99.4%), or blastocysts (98.0%; Figure 4, right). Despite measuring significantly more developmental stages, our stage classifier outperforms previous cell counting networks developed for human embryos [17,21].

Cell Object Instance Segmentation:

We measure the accuracy of the cell instance segmentation network using mean-average precision (mAP) [22], a standard metric for object instance segmentation tasks. Our network predicts cell masks with a mAP of 0.737 on the test set, consisting of 31 embryos labeled at 4,953 times. The model identifies cells with a precision of 82.8% and a recall of 88.4%, similar to results from other work on fewer images [28]. For correctly-identified candidates, the predicted cell area is within 17% of the true cell area 90% of the time (Figure 5, right); much of this error arises when cells strongly overlap late in the cleavage stage. Cropping to a region-of-interest provides a marginal improvement to the network's accuracy (Table 1).

Pronucleus Object Instance Segmentation:

The pronuclei segmentation network predicts masks with a mAP of 0.680 on the test set of 33 embryos labeled at 2,090 times. The network identifies pronuclei with a precision of 81.4% and a recall of 88.2%. Much of the false positive detections are from vacuoles inside the 1-cell embryo, which look similar to pronuclei. For correctly-identified candidates, the predicted pronuclei area is within 16% of the true pronuclei area 90% of the time (Figure 6, right). The pronuclei network's mAP outperforms that of a baseline Mask-RCNN by 0.03; averaging across focal planes and cropping to a region-of-interest each improves the mAP by 0.01 (Table 1).

5 Conclusions

Our unified pipeline greatly speeds up the measurement of embryos: running all five networks on a 300-image, five-day movie takes 6 minutes on a GTX Titan X. In the future, we plan to make this pipeline even faster by combining all five networks with multi-task learning [8]. Since we measure many of the key morphological features used in clinical IVF, our unified pipeline has the potential to reduce the time to grade embryos without sacrificing interpretability. Equally as important, the automatic, high-quality data produced by our pipeline will enable better retrospective chart studies for IVF, improving IVF by informing better clinical practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge M. Venturas and P. Maeder-York for help validating labels and approaches. This work was funded in part by NIH grant 5U54CA225088 and NSF Grant NCS-FO 1835231, by the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard (award number 1764269), and by the Harvard Quantitative Biology Initiative. DJN and DBY also acknowledge generous support from the Perelson family, which made this work possible.

References

1. Alikani M, Cohen J, Tomkin G, Garrisi GJ, Mack C, Scott RT: Human embryo fragmentation in vitro and its implications for pregnancy and implantation. Fertility and sterility 71(5), 836–842 (1999) [PubMed: 10231042]

2. Amir H, Barbash-Hazan S, Kalma Y, Frumkin T, Malcov M, Samara N, Hasson J, Reches A, Azem F, Ben-Yosef D: Time-lapse imaging reveals delayed development of embryos carrying unbalanced chromosomal translocations. Journal of assisted reproduction and genetics 36(2), 315–324 (2019) [PubMed: 30421343]
3. Armstrong S, Bhide P, Jordan V, Pacey A, Marjoribanks J, Farquhar C: Time-lapse systems for embryo incubation and assessment in assisted reproduction. Cochrane Database of Systematic Reviews (2019)
4. Bellman R: Dynamic programming. Science 153(3731), 34–37 (1966) [PubMed: 17730601]
5. Broughton DE, Moley KH: Obesity and female infertility: potential mediators of obesity's impact. Fertility and sterility 107(4), 840–847 (2017) [PubMed: 28292619]
6. Cohen J, Alikani M, Trowbridge J, Rosenwaks Z: Implantation enhancement by selective assisted hatching using zona drilling of human embryos with poor prognosis. Human Reproduction 7(5), 685–691 (1992) [PubMed: 1639990]
7. Cui W: Mother or nothing: the agony of infertility. (2010)
8. Dai J, He K, Sun J: Instance-aware semantic segmentation via multi-task network cascades. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 3150–3158 (2016)
9. Dolinko AV, Farland L, Kaser D, Missmer S, Racowsky C: National survey on use of time-lapse imaging systems in ivf laboratories. Journal of assisted reproduction and genetics 34(9), 1167–1172 (2017) [PubMed: 28600620]
10. Elder K, Dale B: In-vitro fertilization. Cambridge University Press (2020)
11. Filho ES, Noble J, Poli M, Griffiths T, Emerson G, Wells D: A methodfor semi-automatic grading of human blastocyst microscope images. Human Reproduction 27(9), 2641–2648 (2012) [PubMed: 22736327]
12. Franasiak JM, Forman EJ, Hong KH, Werner MD, Upham KM, Treff NR, Scott RT Jr: The nature of aneuploidy with increasing age of the female partner: a review of 15,169 consecutive trophectoderm biopsies evaluated with comprehensive chromosomal screening. Fertility and sterility 101(3), 656–663 (2014) [PubMed: 24355045]
13. Goodfellow I, Bengio Y, Courville A: Deep learning. MIT press (2016)
14. He K, Gkioxari G, Dollár P, Girshick R: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision pp. 2961–2969 (2017)
15. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 770–778 (2016)
16. Hoffman R, Gross L: Modulation contrast microscope. Applied Optics 14(5), 1169–1176 (1975) [PubMed: 20154791]
17. Khan A, Gould S, Salzmann M: Deep convolutional neural networks for human embryonic cell counting. In: European Conference on Computer Vision pp. 339–348. Springer (2016)
18. Kheradmand S, Singh A, Saeedi P, Au J, Havelock J: Inner cell mass segmentation in human hmc embryo images using fully convolutional network. In: 2017 IEEE International Conference on Image Processing (ICIP) pp. 1752–1756. IEEE (2017)
19. Khosravi P, Kazemi E, Zhan Q, Malmsten JE, Toschi M, Zisimopoulos P, Sigaras A, Lavery S, Cooper LA, Hickman C, et al.: Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. NPJ digital medicine 2(1), 1–9 (2019) [PubMed: 31304351]
20. Kragh MF, Rimestad J, Berntsen J, Karstoft H: Automatic grading of human blastocysts from time-lapse imaging. Computers in biology and medicine 115, 103494 (2019)
21. Lau T, Ng N, Gingold J, Desai N, McAuley J, Lipton ZC: Embryo staging with weakly-supervised region selection and dynamically-decoded predictions. arXiv preprint arXiv:1904.04419 (2019)
22. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL: Microsoft coco: Common objects in context. In: European conference on computer vision pp. 740–755. Springer (2014)
23. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 3431–3440 (2015)

24. Nickkho-Amiry M, Horne G, Akhtar M, Mathur R, Brison D: Hydatidiform molar pregnancy following assisted reproduction. Journal of assisted reproduction and genetics 36(4), 667–671 (2019) [PubMed: 30612209]
25. Norwitz ER, Edusa V, Park JS: Maternal physiology and complications of multiple pregnancy. Seminars in perinatology 29(5), 338–348 (2005) [PubMed: 16360493]
26. Petersen BM, Boel M, Montag M, Gardner DK: Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on day 3. Human reproduction 31(10), 2231–2244 (2016) [PubMed: 27609980]
27. Racowsky C, Stern JE, Gibbons WE, Behr B, Pomeroy KO, Biggers JD: National collection of embryo morphology data into society for assisted reproductive technology clinic outcomes reporting system: associations among day 3 cell number, fragmentation and blastomere asymmetry, and live birth rate. Fertility and sterility 95(6), 1985–1989 (2011) [PubMed: 21411078]
28. Rad RM, Saeedi P, Au J, Havelock J: A hybrid approach for multiple blastomeres identification in early human embryo images. Computers in biology and medicine 101, 100–111 (2018) [PubMed: 30121495]
29. of the American Society for Reproductive Medicine, P.C.: Guidance on the limits to the number of embryos to transfer: a committee opinion. Fertility and sterility 107(4), 901 (2017) [PubMed: 28292618]
30. Rubio I, Galán A, Larreategui Z, Ayerdi F, Bellver J, Herrero J, Meseguer M: Clinical validation of embryo culture and selection by morphokinetic analysis: a randomized, controlled trial of the embryoscope. Fertility and sterility 102(5), 1287–1294 (2014) [PubMed: 25217875]
31. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 2818–2826 (2016)
32. Tran D, Cooke S, Illingworth P, Gardner D: Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. Human Reproduction 34(6), 1011–1018 (2019) [PubMed: 31111884]
33. Xie S, Girshick R, Dollár P, Tu Z, He K: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1492–1500 (2017)

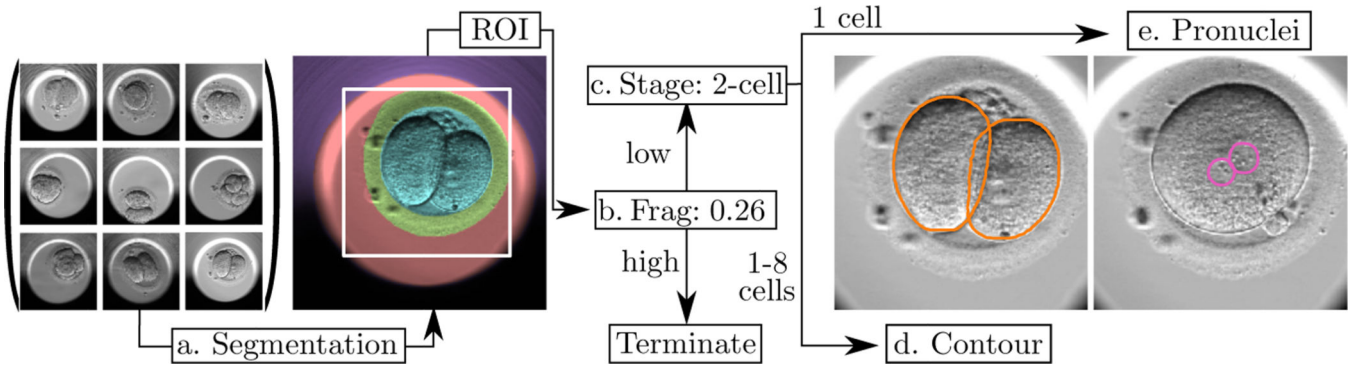

Fig. 1.

Instead of performing one task, our unified pipeline extracts multiple features from embryos. We first segment the image to locate the embryo (panel a), colored according to segmentation. The segmentation provides a region-of-interest (ROI, white box) for the other 4 networks, starting with embryo fragmentation (b); the image shown has a predicted fragmentation score of 0.26. If the embryo's fragmentation score is less than 1.5, we classify the developmental stage (c); this image is classified as a 2-cell embryo. We then detect cells in cleavage stage embryos (orange contours in d) and pronuclei in 1-cell embryos (magenta contours in e).

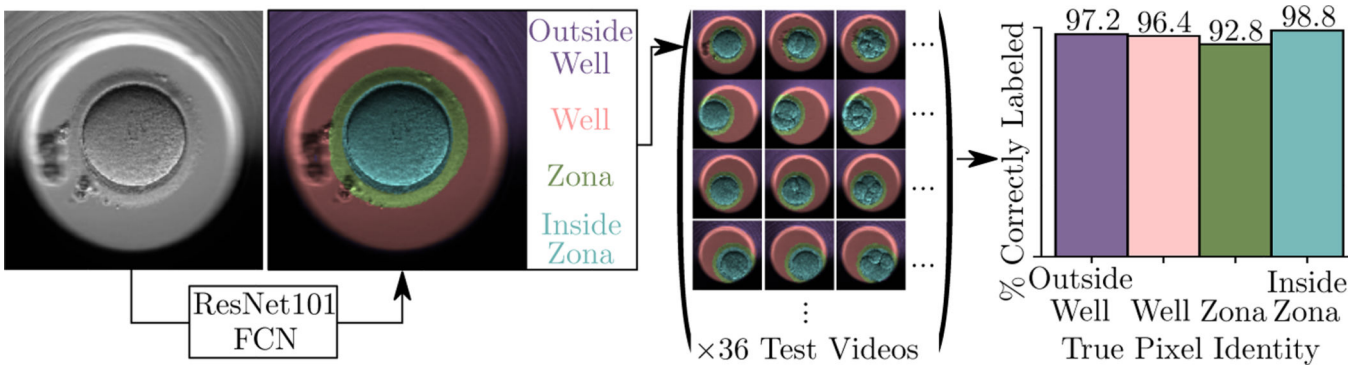

Fig. 2.

The zona pellucida network (ResNet101 FCN) performs semantic segmentation on the input image, predicting four class probabilities for each pixel (colored as purple: outside well, pink: inside well, green: zona pellucida, cyan: inside zona). Middle: 12 representative segmentations of 3 embryos from the test set. Right: the per-pixel accuracies of the segmentation on each class in the test set.

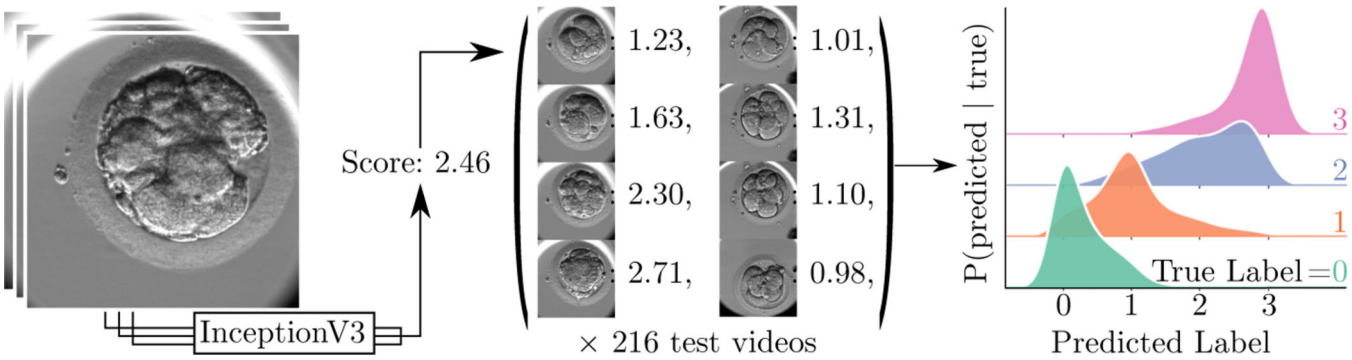

Fig. 3.

Left: The fragmentation network (InceptionV3 architecture) scores embryos with a real number from 0 – 3; the image at left is scored as a fragmentation of 2.46. Center: 8 representative fragmentation scores on the test set, shown as image: score pairs. Right: The distribution of the network's prediction given the ground-truth label on the test set. The green distribution corresponds to images with a ground-truth label of 0; orange those labeled as 1; blue, 2; pink, 3.

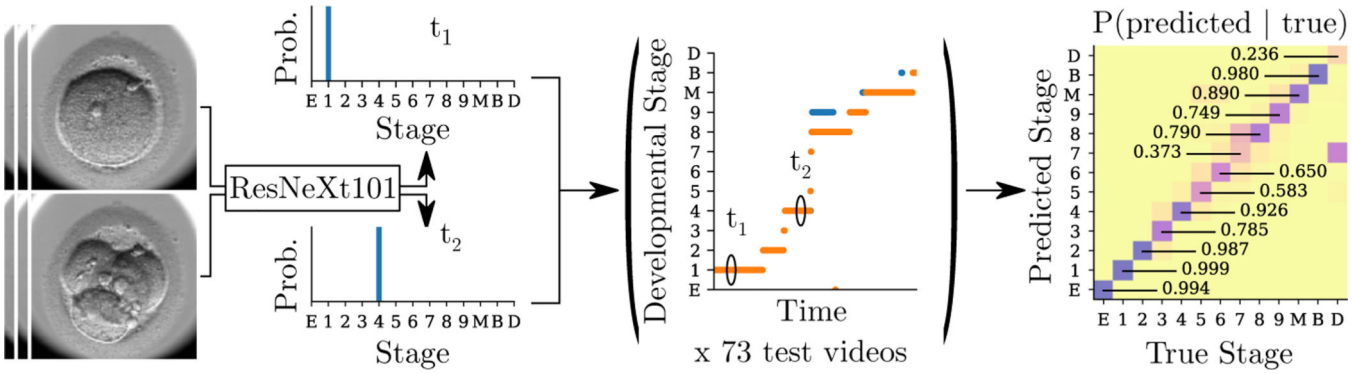

Fig. 4.

Left: The stage classification CNN (ResNeXt101) predicts a per-class probability for each image; the two bar plots show the predicted probabilities for the two images. Center: We use dynamic programming to find the most-likely non-decreasing trajectory (orange); the circled times t_1 and t_2 correspond to the predictions at left. Right: The distribution of predictions given the true labels, measured on the test set.

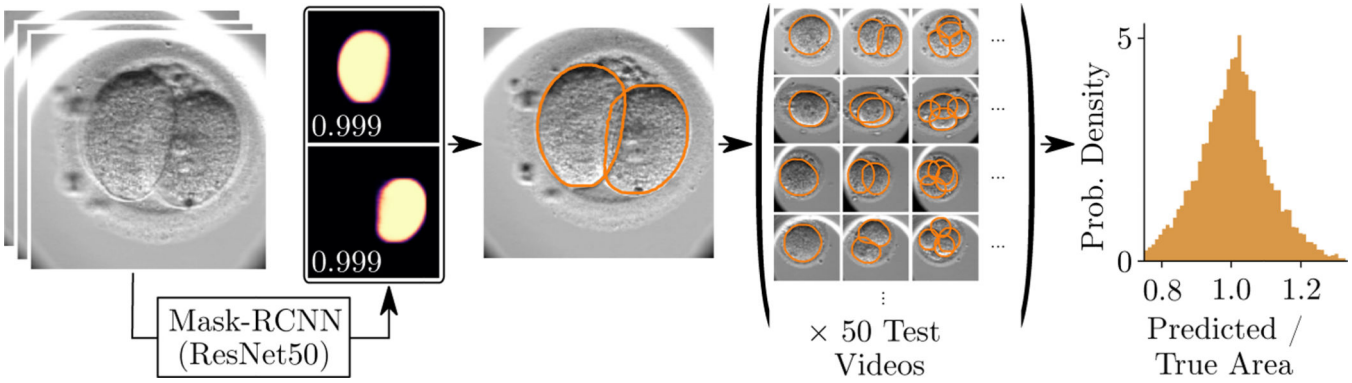

Fig. 5.

The cell detection network (Mask-RCNN, ResNet50 backbone) takes an image (left) and proposes candidates as a combined object mask and confidence score from 0–1 (second from left). Center: The boundaries of the object mask represented as the cell's contours (orange, center). Second from right: 12 cell instance segmentations for 4 embryos from the test set (shown as orange contours overlaid on the original image). Right: Histogram of the ratio of predicted to true areas for correctly identified cells in the test set.

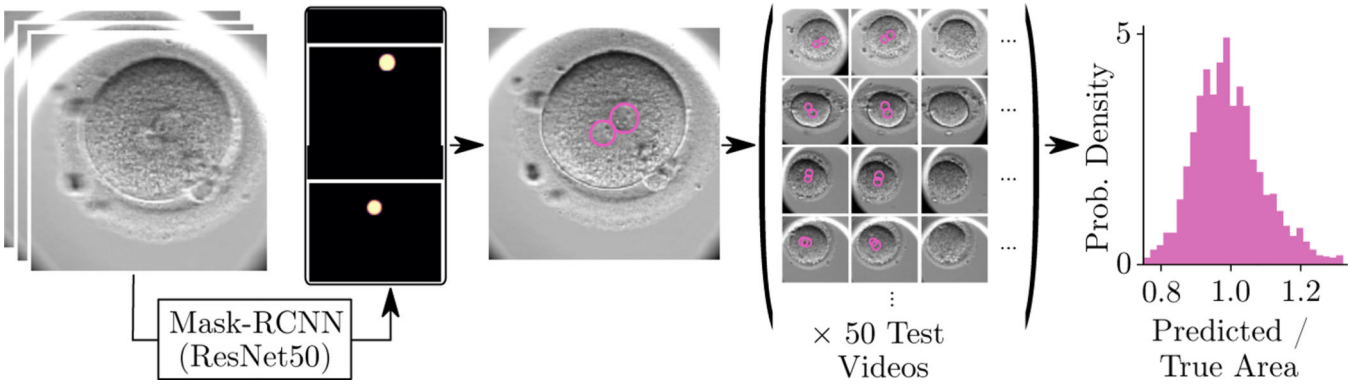

Fig. 6.

The pronuclei detection network (Mask-RCNN, ResNet50 backbone) takes an image (left) and proposes candidates as a combined object mask and confidence score from 0–1 (second from left). Center: The boundaries of the object mask represented as the pronuclei contours (magenta, center). Second from right: 12 pronuclei instance segmentations for 4 embryos from the test set (shown as magenta contours overlaid on the original image); the rightmost images illustrate true negatives after the pronuclei have faded. Right: Histogram of the ratio of predicted to true areas for correctly identified pronuclei in the test set.

Table 1.

Effect of design choices for the 4 more difficult tasks, illustrated by removing one modification to the network at a time. Test-set scores are in percent correctly classified (stage, fragmentation) and mean-average precision (blastomere, pronuclei). The best scores are boldfaced. The last row shows the test set scores using all the training data but no input from other networks and no modifications to the network.

Setting	Fragmentation (%)	Stage (%)	Blastomere (mAP)	Pronuclei (mAP)
Full Setting	**88.9**	**87.8**	0.737	**0.680**
Single Focus	87.8	84.8	**0.739**	0.668
No ROI from Zona	87.4	84.9	0.733	0.666
Using 50% Training Data	87.7	85.3	0.718	0.656
No Soft Loss	–	85.3	–	–
No Dynamic Programming	–	86.0	–	–
Single-Task Baselines	87.0 [31]	81.1 [33]	0.737 [14]	0.647 [14]