Human Embryo Cell Centroid Localization and Counting in Time-Lapse Sequences

Lisette Lockhart*, Parvaneh Saeedi*, Jason Au†, Jon Havelock†

* Simon Fraser University
† Pacific Centre for Reproductive Medicine

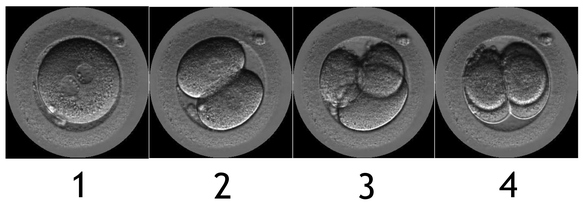
Motivation

- More than 30,000 In Vitro Fertilization (IVF) treatments performed annually in Canada [1]
  - Low success rate (~3.5 embryo transfers per pregnancy [1])
- Transferring highest quality embryos will improve likelihood of implantation
  - Monitor embryo during *in vitro* development with time-lapse imaging
- Knowing when embryonic cells divide is indicative of embryo quality [2],[3]
  - Annotation is time-consuming and subjective
  - Automate cell centroid localization to measure cell stage quickly and objectively

Related Work

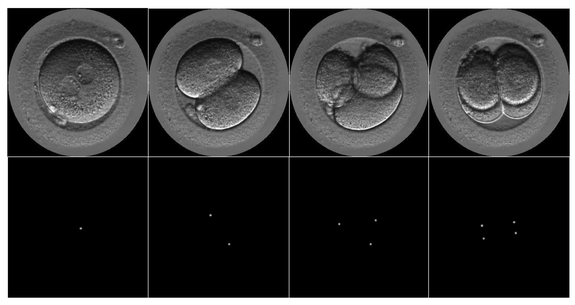
Cell Counting in Images via Classification

- Minimal annotations required (cell stage onset frame)
- Uses Convolutional Neural Network (CNN) classification models
- Cannot capture cell orientation or movement of cells throughout sequence



Cell Counting in Images via Localization

- More annotations needed (cell centroid coordinates)
- Achieved using CNN segmentation models with structured regression output layer
- Captures more information about cells enabling further assessment



Methodology

Structured Regression Network

Fully convolutional regression network
- ResNet-18 feature encoder
- Progressive Upsampling Convolution

- Weighted mean squared error
  - Address severely imbalanced foreground/background pixels
    - m,n: pixel height,width
    - y: ground truth regression mask
    - ŷ: predicted regression mask
    - α_0 and α_1: adjustable weighting parameters

$$\mathcal{L}(y,\hat{y}) = \sum_{n=1}^{N}\sum_{m=1}^{M} \frac{((y_{m,n}-\hat{y}_{m,n})^2 \cdot ((\alpha_0 \cdot \frac{y_{m,n}}{\max y}) + \alpha_1))}{M \cdot N}$$

Temporal Context Prior

- Cell centroid location and cell count relies considerably on previous frame
  - Add centroid regression mask from previous frame to provide context (Multi-Input I)
- Centroid mask from previous frame has no indication of cell movement
  - Add optical flow diagram between subsequent frames to provide context (Multi-Input II)
- Encode context as attention with squeeze-excitation
  - Add attention modules to layers with most channels

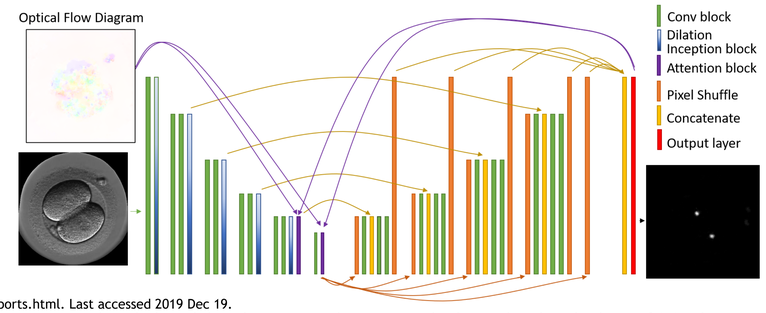
Sampling Procedure

- Very little movement between most frames
- Avoids grouping together similar samples

Algorithm 1: Training with predicted outputs from previous frame

Input: sequence frames $\mathbb{X}_1$, predicted centroid masks from previous frame $\hat{\mathbb{X}}_2$
Output: centroid masks $\mathbb{Y}$, predicted centroid masks $\hat{\mathbb{Y}}$
while *loss not plateaued* **do**
1. Train on $([\mathbf{x}_1^{(i)}, \hat{\mathbf{x}}_2^{(i)}], \mathbf{y}^{(i)}), i \in \{2,3,...,N\}$ on all sequences in training set for one epoch
2. Predict on $([\mathbf{x}_1^{(i)}, \hat{\mathbf{x}}_2^{(i)}], \mathbf{y}^{(i)}), i \in \{1,2,...,(N-1)\}$
3. Store $\hat{\mathbf{y}}^{(i)}, i \in \{1,2,...,(N-1)\}$ as $\hat{\mathbf{x}}_2^{(i)}, i \in \{2,3,...,N\}$

Network Architecture



Experimental Results

- 108 human embryo time-lapse sequences from 1-4 cell stage
  - Dot-annotated cell centroids
- Training, validation, and test sets randomly selected as 70%/15%/15% of sequences
  - 5-fold cross-validation
- Adam optimizer with initial learning rate 3×10^{-5}

Cell Centroid Localization

Model	Distance to nearest centroid (in pixels)				
	1-cell	*2-cell*	*3-cell*	*4-cell*	*Total*
U-Net [29]	2.88	4.25	4.72	4.43	4.24
Cell-Net [21]	2.97	4.14	4.94	4.68	4.38
Multi-Input I (Proposed)	**2.51**	3.98	4.73	4.28	4.05
Multi-Input II (Proposed)	2.57	**3.95**	**4.35**	**4.20**	**3.98**

- Detection: < 5 pixels from nearest ground truth centroid
- Near Miss: ≥ 5 and < 8 pixels from nearest ground truth centroid
- Total Miss: ≥ 8 pixels from nearest ground truth centroid

Model	Cell detection rate (in %)		
	Detection	*Near Miss*	*Total Miss*
U-Net [29]	80.0	11.7	8.3
Cell-Net [21]	77.1	11.9	11.0
Multi-Input I (Proposed)	80.1	**11.0**	8.9
Multi-Input II (Proposed)	**80.9**	11.3	**7.8**

Cell Counting

$$Cell\ Stage\ Acc. = \frac{TP_i + TN_i}{\sum_{s=1}^{S} N_s}, i \in \{1,2,3,4\},$$

$$Total\ Acc. = \frac{\sum_{i=1}^{4} TP_i}{\sum_{s=1}^{S} N_s},$$

Model	Cell Stage Prediction Accuracy (in %)				
	1-cell	*2-cell*	*3-cell*	*4-cell*	*Total*
U-Net [29]	92.8	67.4	61.6	**78.4**	77.7
Cell-Net [21]	96.2	**81.8**	67.5	62.3	77.5
Multi-Input I (Proposed)	**97.7**	78.8	**69.2**	68.6	79.3
Multi-Input II (Proposed)	95.7	74.7	69.0	75.8	**80.2**

Conclusions & Future Work

- ✓ Structured regression suitable for cell centroid localization for cell counting in embryo sequences
- ✓ Foreground/background pixel imbalance relaxed using temporal relationship in embryonic cell development
- ✓ Training strategy samples diverse batches of data for network gradient updates
- ➢ Extend to entire 5-day in vitro embryo development sequence

[1] Canadian Fertility & Andrology Society: Canadian Assisted Reproductive Technologies Register Plus (CARTR Plus), https://cfas.ca/cartrannual- reports.html. Last accessed 2019 Dec 19.
[2] Jacobs, C., Nicolielo, M., Erberelli, R., Mendez, F., Fanelli, M., Cremonesi, L., Aiello, B., and Lorenzon, A.R.: Correlation between morphokinetic parameters and standard morphological assessment: what can we predict from early embryo development? A time-lapse-based experiment with 2085 blastocysts. JBRA Assisted Reproduction (2020)
[3] Basile, N., Vime, P., Florensa, M., Aparicio Ruiz, B., Garcia Velasco, J. A., Remohi, J., and Meseguer, M.: The use of morphokinetics as a predictor of implantation: A multicentric study to define and validate an algorithm for embryo selection. Human Reproduction 30 (2), pp. 276-283 (2014)