Automating Assessment of Human Embryo Images and Time-Lapse Sequences for IVF Treatment

by

Lisette Lockhart

B.Eng., University of Victoria, 2018

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Applied Science

in the
School of Engineering Science
Faculty of Applied Sciences

© Lisette Lockhart 2021

SIMON FRASER UNIVERSITY
Summer 2021

Copyright in this work is held by the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Declaration of Committee

Name: **Lisette Lockhart**

Degree: **Master of Applied Science**

Thesis title: **Automating Assessment of Human Embryo Images and Time-Lapse Sequences for IVF Treatment**

Committee: **Chair:** Ljiljana Trajkovic
Professor, Engineering Science

Parvaneh Saeedi
Supervisor
Associate Professor, Engineering Science

Andrew Rawicz
Committee Member
Professor, School of Engineering Science

Marinko Sarunic
Examiner
Professor, School of Engineering Science

Abstract

As the number of couples using In Vitro Fertilization (IVF) treatment to give birth increases, so too does the need for robust tools to assist embryologists in selecting the highest quality embryos for implantation. Quality scores assigned to embryonic structures are critical markers for predicting implantation potential of human blastocyst-stage embryos. Timing at which embryos reach certain cell and development stages in vitro also provides valuable information about their development progress and potential to become a positive pregnancy. The current workflow of grading blastocysts by visual assessment is susceptible to subjectivity between embryologists. Visually verifying when embryo cell stage increases is tedious and confirming onset of later development stages is also prone to subjective assessment. This thesis proposes methods to automate embryo image and time-lapse sequence assessment to provide objective evaluation of blastocyst structure quality, cell counting, and timing of development stages.

Keywords: Medical image analysis; deep learning; embryology

Acknowledgements

I would like to thank my Supervisor, Parvaneh Saeedi, for the opportunity to work on this fascinating project and for the guidance and support throughout my degree. Thanks to my committee members for guidance and making my degree possible. Thanks to all the friends who have supported me, especially Sorour and Reza for their companionship in the lab and on the court. Thanks to Alexandra for sparking my interest in medical image analysis and for the continued encouragement. Finally, a big thanks to my brother, Brandon, and my parents for their ongoing encouragement and support.

Table of Contents

List of Tables

List of Figures

Acronyms

BE	Blastocyst Expansion.
CAM	Class Activation Map.
CNN	Convolutional Neural Network.
HMC	Hoffman Modulation Contrast.
ICM	Inner Cell Mass.
IVF	In Vitro Fertilization.
LSTM	Long Short-Term Memory.
MAE	Mean Absolute Error.
MSE	Mean Squared Error.
NLL	Negative Log Likelihood.
PCRM	Pacific Centre for Reproductive Medicine.
PVS	Perivitelline Space.
ReLU	Rectified Linear Unit.
SD	Standard Deviation.
TE	Trophectoderm.
ZP	Zona Pellucida.

Chapter 1

Introduction

1.1 IVF Treatment

In vitro fertilization (IVF) treatment is a commonly performed medical procedure for couples suffering from infertility. In 2016, 31,274 IVF treatment cycles were performed in Canada [72]. Despite its popularity, the treatment still has a relatively low success rate, with only 29.4% embryo transfers leading to a live birth [72]. Since it takes nearly 3.5 embryo transfers to successfully give birth, there is a large financial and emotional burden imposed on couples attempting to conceive using IVF treatment.

In IVF treatment, eggs and sperm are extracted from patients and combined in a lab environment. Multiple fertilized eggs (i.e. embryos) are developed outside the patient's body, and the most promising embryo(s), according to morphological characteristics, is transferred into the patient's uterus. Multiple embryos (2-3) may be transferred at once to increase the likelihood of success. This, in turn, increases the chances of multiple pregnancy, which poses additional health risks to both the patient and fetuses/babies.

To increase the success rate of IVF treatment, the mechanisms and effects of different incubation conditions affecting embryo development are highly sought after. Preimplantation genetic screening (PGS), performed by embryo biopsy and genetic testing, is a great indicator for negative implantation outcomes. However, it is not a good indicator for positive implantation outcome. It is also performed infrequently due to added treatment cost [72]. Morphological assessment of human embryos is still used in the majority of IVF cycles. This thesis focuses on two key aspects of embryo assessment:

1. morphology (i.e. structural appearance) of embryo components using single images, and

2. morphokinetics (i.e. timing of changes) of embryo development using time-lapse sequences.

These new approaches set the future path for developing intelligent systems that can assist embryologists with assessing embryo quality in IVF treatment and understanding which embryos have the highest likelihood of leading to a successful pregnancy.

1.2 Embryo Development In Vitro

During IVF treatment, a female's ovaries are hyper-stimulated to produce multiple eggs for external fertilization. The fertilized eggs (embryos) are incubated in controlled environmental conditions until they reach the blastocyst stage (usually on the 5th day post-fertilization). The blastocyst with the best development progress is selected for implantation into the patient's uterus. Many embryos are developed in every IVF cycle to increase the likelihood of getting high quality candidates. Embryos are evaluated throughout the incubation process and scored at multiple stages to assess their quality as they develop. The development process is shown in Fig. 1.1.

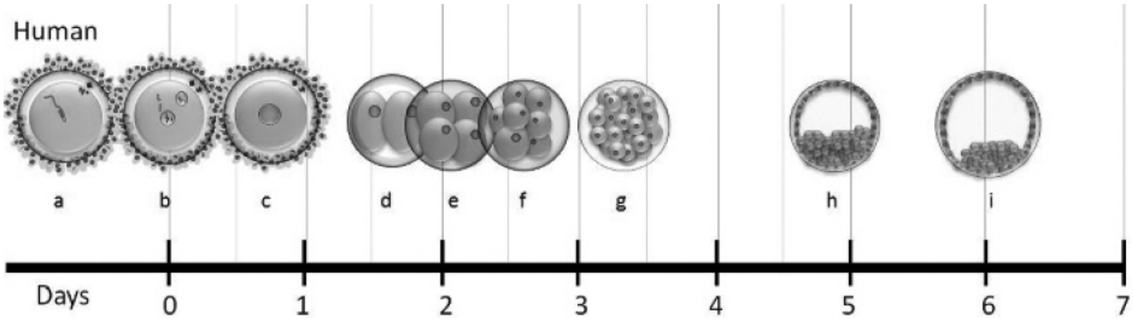


Figure 1.1: Timeline of blastocyst development (pronuclear formation is considered day 0): fertilization (a), pronuclear formation (b), zygote (c), 2-cell embryo (d), 4-cell embryo (e), 8-cell embryo (f), morula (g), early blastocyst (h), and late blastocyst (i) [61].

When a fertilized egg divides into two cells, it enters the cleavage stage of its development. The cells of a normal two-cell embryo will later divide and create a four-cell embryo. Each cell in the four-cell embryo will divide again to form an eight-cell embryo, etc. Embryologists have investigated some attributes at this stage that may indicate which embryos will advance further. These attributes include similarity in size, little or no fragmentation, division time and synchronicity. Healthy embryos have a fairly strict rate of progression starting at the time of their fertilization. Some studies suggest that embryos with optimal cleavage speed defined by: 2 cells at 25-27h, 4 or 5 cells at 44-46h, and at least 7 cells at 66-68h post insemination might have higher implantation and/or clinical pregnancy rates. Slow progression in cell division usually is an indicator of a low quality embryo [11].

A protective glycoprotein layer called zona pellucida (ZP) encapsulates the inner embryos structures and ensures close cell proximity between blastomeres as the embryo develops. The ZP has several functions including restricting penetration of other sperm cells once an egg is fertilized, protection of pre-implantation embryos, and preventing premature implantation.

As the embryo develops, denser clusters of blastomeres (i.e. embryonic cells) form. In normal embryonic development, and embryo reaches morula stage at day 4. This is characterized by increased blastomere adhesion (i.e. cells becoming more attached). Blastomeres then develop into 2 structures: trophectoderm (TE) - a mass of cell lining the inner ZP which later develops into the placenta, and inner cell mass (ICM) - a mass of cells near the centre of the embryo that later develops into the fetus. Concurrently, the embryo absorbs fluid from the surrounding media to create the blastocoel - a fluid-filled cavity providing structural support and nutrients. Typical blastocyst structures are shown in Fig. 1.2.

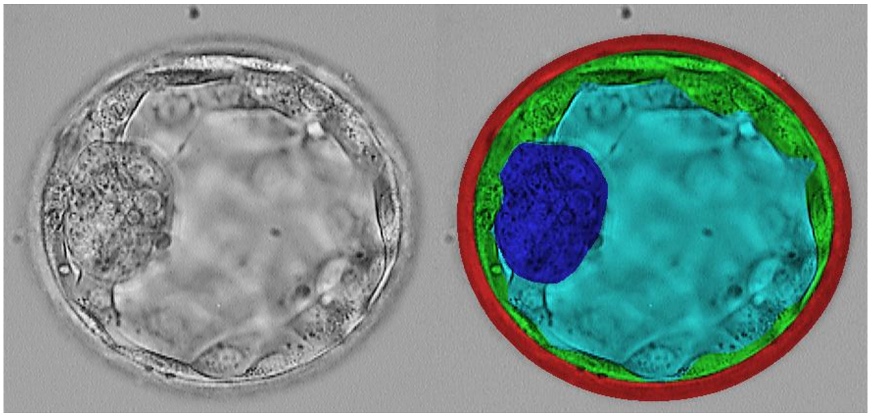


Figure 1.2: Blastocyst ZP (red), TE (green), ICM (dark blue), and blastocoel (light blue).

An embryo typically reaches blastocyst stage at day 5. By this point, the trophoblast (outer membrane), ICM, and blastocoel are clearly visible. The ICM is well-defined and its many cells are compact. There are few fragments from cleavage stage in the blastocoel or perivitelline space (PVS - space between inner ZP and outer TE). As the blastocyst develops, the blastocoel volume becomes > 50% of the embryo volume and the ZP becomes thinner.

1.3 Embryo Imaging

Assessing embryo quality is limited by image acquisition capability: a) Dyes are often added to specimens in microscopy images to accentuate structures of interest. These dyes cannot be used on embryos as they could cause adverse affects for development. b) High resolution microscopy devices that can potentially damage the embryo must also be excluded. Therefore, optical imaging methods remain the only safe imaging technique. Since transparent specimens like embryos do not absorb light, the imaging technique must exploit differences in refractive index between embryo structures and also surrounding medium.

Phase Contrast microscopy [2] accentuates the difference between light passing around and light that is diffracted by a specimen. The light passing through a specimen has the same amplitude of light passing around it, by lags by approximately 1/4 wavelength. This phase difference varies by specimen structure, and is used to construct an image. However, it suffers from a characteristic halo artifact at object boundaries (bright outer areas and dark inner areas), which makes it difficult to distinguish boundaries between objects. Images taken with phase contrast microscopy also have a flat 2D appearance. Hoffman Modulation Contrast (HMC) microscopy in Fig. 1.3 also relies on differences in the phase portion of light passing through specimens to construct an image. It has a light amplitude filter (shown in Fig. 1.4) that produces different shades based on specimen surface structure. This technique eliminates the halo artifact at object boundaries and gives objects a 3D appearance. HMC microscopy has therefore become the most commonly used imaging technique for embryos.

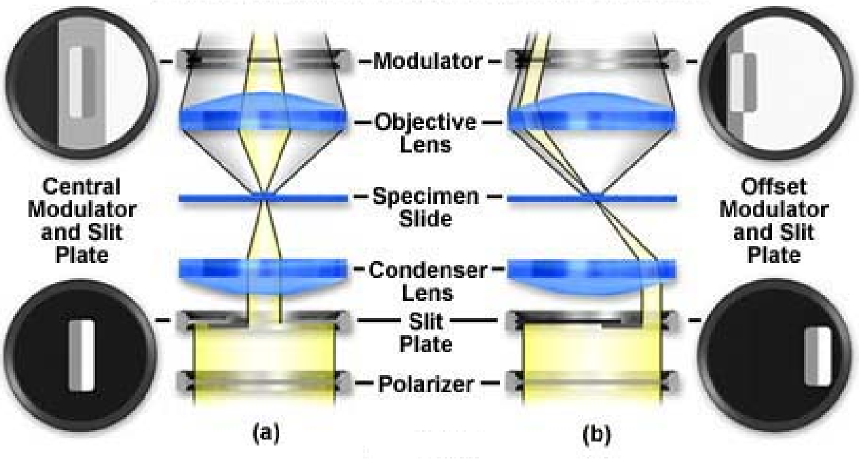


Figure 1.3: Embryo structures with changing optical densities diffract light into a dark, gray, or bright zone with HMC microscopy. Offsetting the modulator and slit plate increases resolution [1].

Embryo images taken with HMC microscopes suffer from object depth ambiguity and weak borders along structures. Peaks can be confused for valleys when viewing object edges, making it difficult to visually assess embryos. Unlike most other medical imaging techniques, the background pixels of HMC microscopy embryo images are a mid-range (gray) value instead of black. Image processing algorithms for locating embryo structure boundaries must rely on increasing and decreasing pixel intensity, instead of one or the other.

Embryos used to be developed in incubators where they had to be transferred for imaging. Now, many clinics are adopting incubators with time-lapse imaging systems that have a built-in microscope for continual embryo monitoring. Embryo images taken from a traditional incubator and time-lapse imaging incubator HMC microscope are shown in Fig. 1.5. Some work has shown embryos developed in these time-lapse incubators have greater pregnancy outcome [46]. Continual embryo imaging throughout development enables another

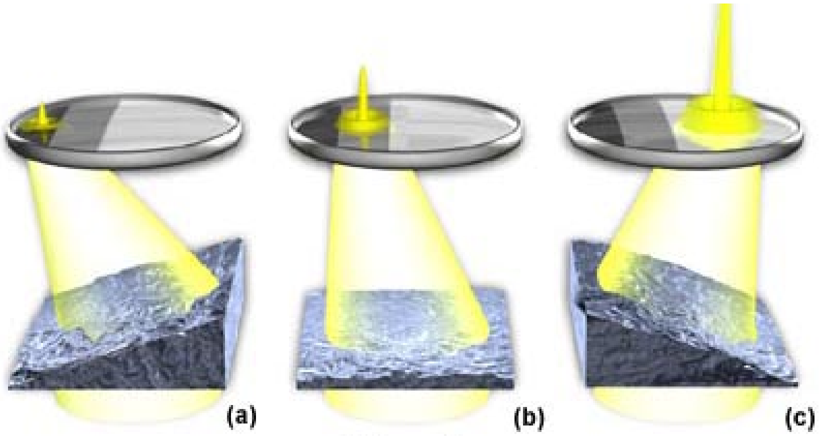


Figure 1.4: Imaging Phase Gradients for negative gradients (a), zero gradient (b), and positive gradients (c) [1].

set of measurable parameters that can be correlated with embryo quality. These additional morphokinetic (time-based) parameters can be used to assess embryo quality with little added cost.

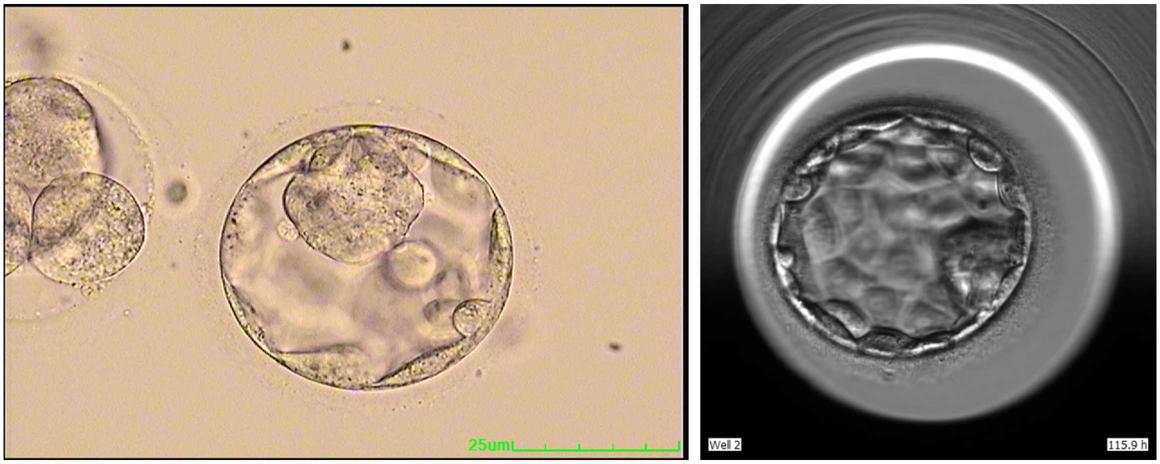


Figure 1.5: Microscopic embryo image from traditional incubator (left) and time-lapse imaging incubator (right).

The embryo images used in this thesis were collected at the Pacific Centre for Reproductive Medicine (PCRM) in Burnaby, BC. The images and time-lapse sequences were taken at a single focal plane.

In traditional incubators, removing the embryo for imaging could potentially damage the embryo. Embryos are imaged only as often as necessary, and are not repeated if the focal plane or focus is off. Acquiring 3 images of each blastocyst focused on different structures was explored previously but no improvement was found. Therefore only a single blastocyst image focused on all the structures was taken for each sample.

In time-lapse sequences, the focal plane is set across the sequence and cannot easily be changed to get a better view of certain structures. Each sequence is 380 frames with 15-minute acquisition interval. Although acquiring images at multiple focal planes could provide a better view of different structures, it significantly increases the memory and processing required at the clinic.

1.4 Embryo Quality Assessment

While live birth is the ultimate measure of success in IVF treatment, several biological, procedural, and environmental factors other than embryo quality can prevent such outcome. Therefore, implantation and clinical pregnancy outcomes can be better suited for evaluating embryo quality. Implantation outcome is confirmed with a biochemical pregnancy test following the embryo transfer. A clinical pregnancy is confirmed by ultrasound detection of a gestational sac or fetal heartbeat during gestation.

Non-invasive embryo imaging is used to assess embryo morphology (visual attributes) for quality markers correlated to higher pregnancy outcome. With more widespread availability of incubators containing time-lapse imaging systems, monitoring changes in embryo development enables use of morphokinetic (time-based) parameters for providing insight into its implantation potential.

1.4.1 Blastocyst Morphology

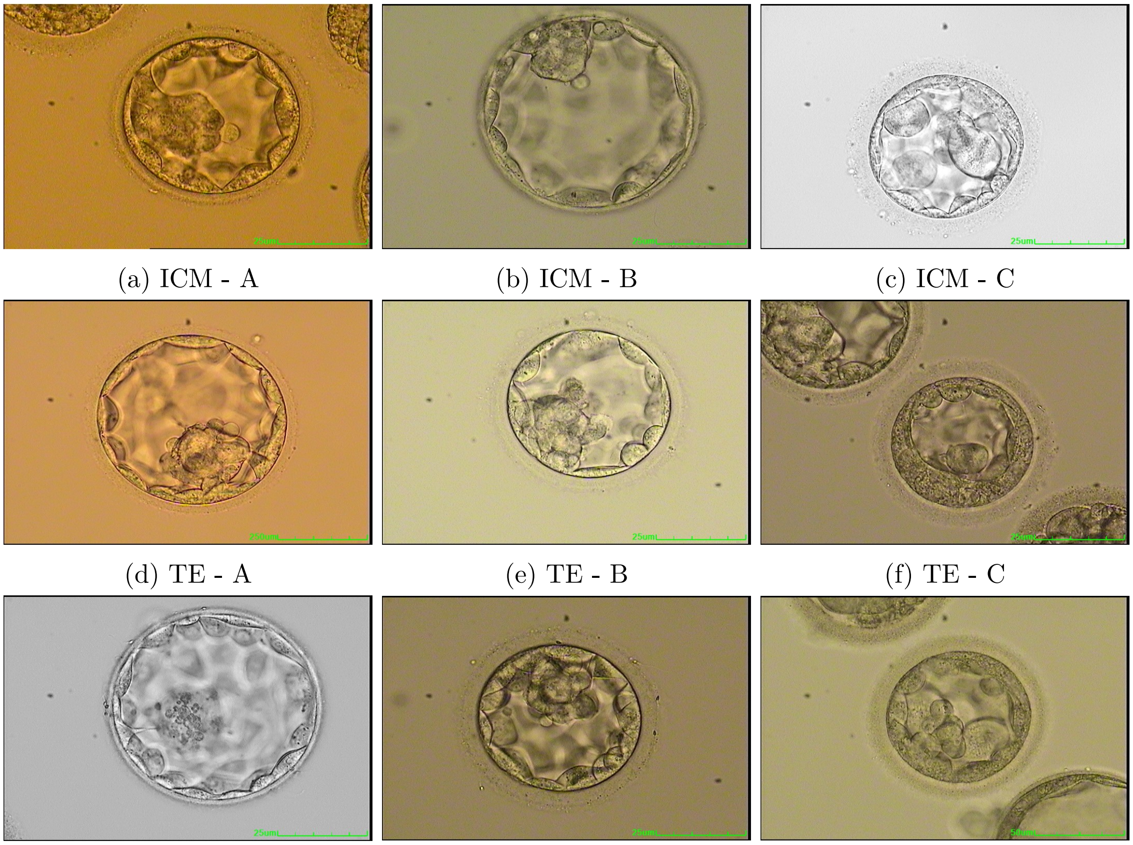
The most common method for ranking viability of blastocysts (i.e. embryos around day-5 post-fertilization) for transfer is using non-invasive visual inspection based on morphological characteristics [21]. However, this is a challenging task as living embryos can only be imaged using non-invasive techniques. Assessing blastocyst quality from images is therefore susceptible to subjectivity. There are blastocyst morphology factors that have shown to correlate with higher implantation potential [22], [39], which were used to develop the Gardner grading system [24].

Three main structures are used to assess blastocyst quality according to the Gardner grading system, summarized in Table 1.1. ICM grade varies with amount of cells and how densely they clump together. TE grade assesses the tissue lining of the ZP for amount of cells and whether they form a cohesive ring. Blastocyst expansion (BE) indicates metabolic competency and is evaluated based on the ZP thickness and blastocoel volume [24]. Examples of each grade are shown in Fig. 1.6.

Higher blastocyst morphology score was shown to increase implantation and ongoing pregnancy rates [23]. Having more high quality blastocysts in an embryo transfer cycle led to higher likelihood of ongoing pregnancy and live birth [81]. However, grading criteria are vague and embryologists show inconsistency due to subjectivity and expert level when performing morphological embryo grading [12]. Automatic grading is therefore desirable

Table 1.1: Blastocyst grading description according to Gardner grading system [24]. Each grade is listed from highest to lowest quality.

	ICM Grade
A	Numerous and tightly packed cells
B	Several and loosely packed cells
C	Few cells
	TE Grade
A	Many cells organized in epithelium
B	Several cells organized in loose epithelium
C	Few cells
	BE Grade
4	Blastocoel volume larger than early embryo, ZP is thin
3	Blastocoel fills the blastocyst, ZP is thick
2	Blastocoel fills greater than half of the blastocyst



(a) ICM - A

(b) ICM - B

(c) ICM - C

(d) TE - A

(e) TE - B

(f) TE - C

(g) BE - 4

(h) BE - 3

(i) BE - 2

Figure 1.6: Examples of blastocyst image with different ICM grades (top), TE grades (middle), and BE (bottom).

for generating unbiased, consistent decisions about blastocyst quality to help embryologists make the best possible transfers to lead to positive implantation and ultimately live birth.

In addition to the challenging biological variation present in embryos of each grade, automating blastocyst grading can be challenging due to image artifacts. Certain embryo structures can be more or less in focus based on the microscope image acquisition. Images can contain parts of multiple embryos that neighbour the blastocyst of interest, and the surrounding embryo culture media can contain embryo fragments.

1.4.2 Embryo Morphokinetics

The timing of early cell stage onset and duration can give insight into whether an embryo will develop into a fully formed blastocyst as well as its implantation potential. The onset of later morula and blastocyst stages can be used to monitor embryo development and predict implantation potential.

Cell Cleavage Timing and Stage Duration

An embryo starts with one cell (a.k.a. zygote stage) that divides repeatedly throughout incubation in vitro. Individual cells (i.e. blastomeres) can be identified and counted in early stages, and used to track embryo development. The beginning and duration of each embryonic cell stage provide valuable insight into embryo quality, including 2-cell stage onset [30], 2-cell stage duration [6], [45], [13], 3-cell stage onset [6], 3-cell stage duration [45], [13], 4-cell stage onset [30], and 5-cell stage onset [6], [45], [13]. Detecting at which frame these cell cleavage events occur can be associated with treatment time to track development progress and rank embryos for implantation potential.

Monitoring the embryo through in vitro development to determine morphokinetic parameters is tedious, and susceptible to differences of opinion between embryologists. Automating embryo development monitoring for objective morphokinetic analysis has therefore become of interest to embryologists.

Embryonic cell stage images pose multiple challenges for analysis. Embryos can have irregular-shaped cells, cells with occluded boundaries, severely overlapped cells, and cell fragmentation (see Fig. 1.7). In some cases, two cells of similar size and shape overlap by nearly 100% of their area, so they can be easily mistaken as one cell. Small cell fragments can occlude cell boundaries, while other fragments have similar size, shape, and appearance as actual cells.

Analyzing time-lapse sequences requires more computing power as there are more frames to process and can have some additional issues with image quality. Consecutive frames can have very little change in embryo development (e.g. most frames between cell cleavage) or significant movement of cells (e.g. during cell cleavage). Some changes (like 3-cell stage) cannot always be captured in the 15 minute acquisition interval. Frames are occasionally captured while a cell is almost finished dividing, sharing more similar appearance to frames

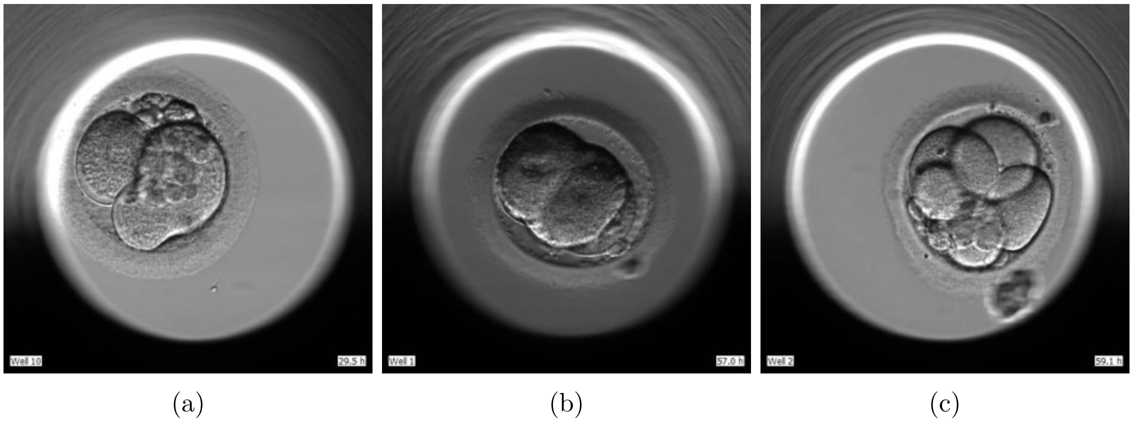


Figure 1.7: Embryos with irregular-shaped cell (a), severe cell overlap at 4-cell stage (b), and large cell-like fragment (c) from time-lapse sequence frames.

of the next higher cell stage. Cleavage frames have motion artifacts that blur cell boundaries. Occasionally, a cell begins to divide or fragment, then is reabsorbed in a later frame.

1.4.3 Morula and Blastocyst Stage Onset

Onset of certain embryo development stages has shown to be correlated with blastocyst quality and likelihood of implantation, including morula [16, 48, 60] and blastocyst [16, 30] stages. Examples of embryo sequence frames at pre-morula (a.k.a. cleavage), morula, and blatocyst stages are shown in Fig. 1.8. Detecting at which frame these development stage onsets occur can be used to predict whether an embryo shows normal development progress and its likelihood of leading to positive pregnancy outcome.

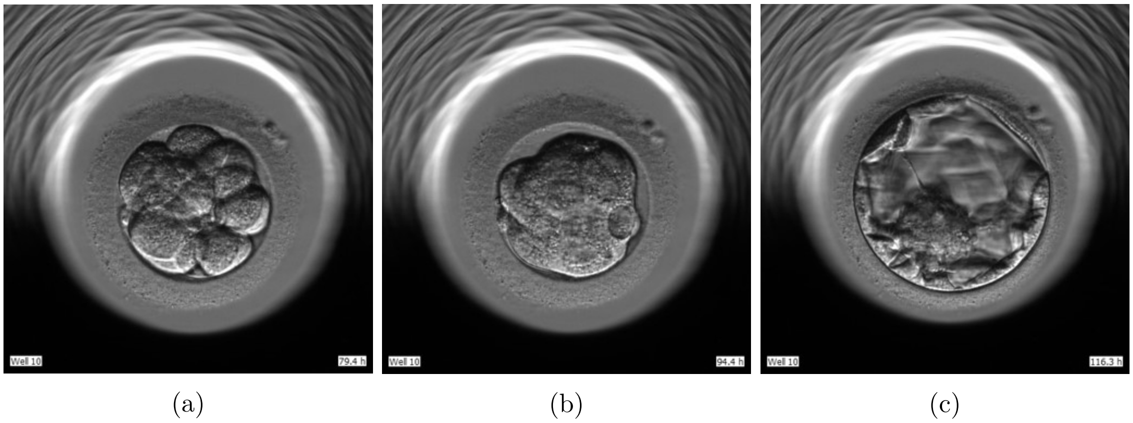
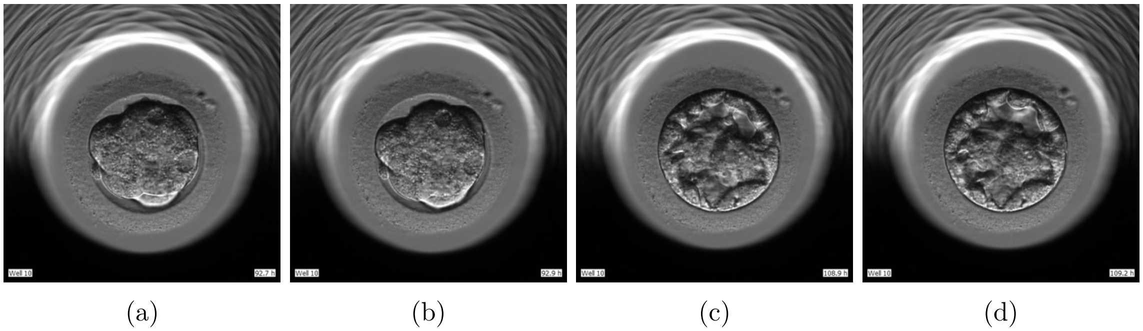


Figure 1.8: Typical cleavage (pre-morula) (a), morula (b), and blastocyst (c) stage embryo time-lapse sequence frames.

Automating development stage detection of day 4-5 embryos is difficult due to small variation between stages, as shown in Fig. 1.9. Predicting onset of morula stage is even more challenging because cell adhesion of an embryo throughout morula stage often increases and decreases, sometimes reverting in appearance to cleavage stage. Similarly, blastocyst stage embryos can contract such that they appear more like a cleavage or morula stage embryos.



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

Figure 1.9: Example of frames directly before (a) and after (b) morula stage onset (92.9 hr post-fertilization), and before (c) and after (d) blastocyst stage onset (109.2 hr post-fertilization). Differences are small between consecutive frames at stage onsets.

1.5 Automating Embryo Quality Assessment

Automating embryo quality assessment can provide robust, objective, and standardized predictions of embryo quality and potential pregnancy outcome. Using traditional image processing techniques to extract high-level information from embryos is challenging because the images are grayscale with weak structure borders. Artifacts from overlapping objects, motion, or neighbouring structures can complicate analysis. Differences between embryos and image acquisition adds further variation to account for. Deep learning methods can be trained on examples of embryo images with desired outputs to perform automated assessment while learning to overcome these challenges.

Deep learning techniques are at the forefront of many image analysis tasks, including classifying images into classes, detecting objects of interest, and segmenting specific structures. These techniques rely on having sufficient examples of images and their desired outputs to learn through an iterative process of how to perform the desired task.

In particular, convolutional neural networks (CNNs) [38], [74], [69], [27] were hugely successful at the ImageNet Large Scale Visual Recognition Challenge [63] and have since been used extensively in visual recognition tasks. An example CNN architecture (VGG16) is shown in Fig. 1.10. They contain a series of 2D convolution filters, non-linear activations (e.g. rectified linear unit or ReLU), and spatial pooling/downsampling operations to encode spatial features from the input image. 1D fully connected layers use these featuresto assign probabilities of the input image belonging to different classes. All layers have randomly initialized parameters (i.e. weights) that are adjusted by backpropagating an output error between predicted and manually annotated outputs.

For classification, it is straightforward to leverage information from CNNs pre-trained on a large-scale database (e.g. ImageNet) and transfer the knowledge to a different dataset

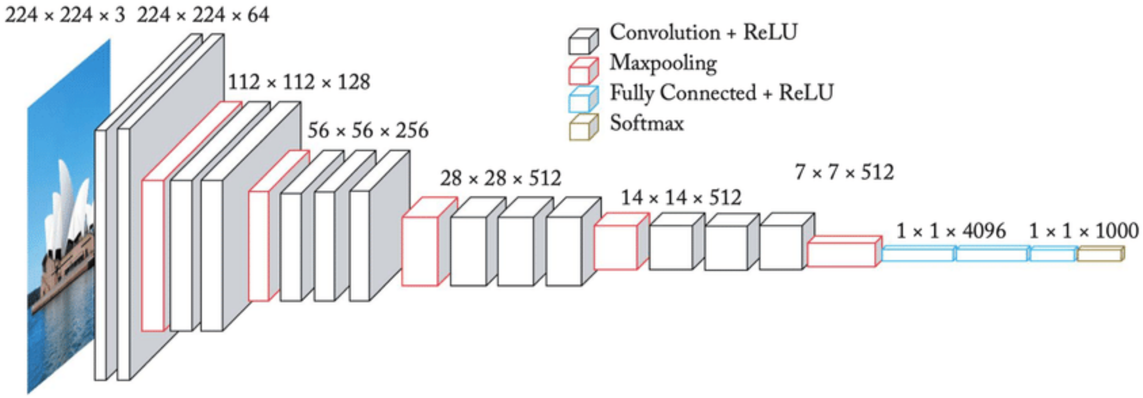


Figure 1.10: VGG16 architecture consisting of convolution, spatial pooling (downsampling), and fully connected layers with non-linear ReLU activation [8]. Numbers indicate *height* × *width* × *depth* of the spatial features at each layer of the network.

(e.g. medical image classification). Model weights can be fine-tuned for a medical image classification task with a shorter training period (fine-tuning) on the desired dataset. By starting with parameters already optimized for a different image classification task, training time is greatly reduced. Additionally, reasonable classification performance can often be obtained on datasets where the lack of training data prevents model convergence when training from randomly initialized weights.

CNNs can also predict outputs of the same size as the input image, which is common for image segmentation and registration tasks. These networks encode the input image into a latent space representation then decode these features into a high-resolution a map corresponding to where structures of interest are located. CNNs have been successfully trained for various medical image analysis tasks including classification via transfer learning [77], 2D image segmentation [62] and 3D volume segmentation [47], landmark localization [25], [85], and image registration [5], [4].

Although deep learning methods show promise for many medical image analysis tasks, their application for automating embryo image assessment poses several challenges. Datasets often have a small number of samples since individual clinics perform relatively few procedures per year and sharing data between facilities is restricted. Embryo image quality is limited by acquisition capability. Embryo images are a single 2D focal plane of a 3D structure, potentially missing relevant information from adjacent depths or different angles.

Despite the challenges for automated embryo image assessment, deep learning-based techniques are utilized in this thesis to extract morphology scores from blastocyst images and morphokinetic parameters from embryo time-lapse sequences. Images and time-lapse sequences were acquired and annotated at PCRM and the algorithms were developed retrospectively. Blastocyst morphology grades were predicted for ICM, TE, and BE and can be directly used to assess blastocyst quality. Cell centroid coordinates and cell counts were predicted in 1- to 4-cell time-lapse sequence frames and can be used to extract cell stage onset and duration as well as cell tracking. Morula and blastocyst stage onset frames were predicted in time-lapse sequences and can be used to assess if an embryo's development time is within normal range.

1.6 Thesis Outline

The remainder of this thesis is organized by embryo assessment task. Each task aims at extracting different morphological or morphokinetic parameters that can be used to assess embryo development progress or quality. Strategies are proposed to address issues of limited training data, dataset class imbalance, and image and annotation quality.

1.6.1 Blastocyst Grading

Quality of embryos' ICM, TE, and expansion progress measured with the Gardner grading system are morphological parameters associated with implantation potential. In Chapter 2, an image classification network is proposed to assign ICM, TE, and BE grades to blastocyst images. A multi-label multi-class classification network simultaneously performs all three grading tasks, compared to previous work that simplifies the task to binary good versus poor quality classification or uses three separate networks. The network base is pre-trained on a large image classification database then fine-tuned on blastocyst images. Severe class distribution imbalance in ICM grade is addressed by leveraging information from classification of other grades. This was the first (known) attempt at assigning scores for ICM, TE, and BE quality simultaneously by a single CNN. The grades predicted by this algorithm could be used to assess likelihood of leading to positive implantation outcome.

1.6.2 Cell Centroid Localization and Cell Counting

Onset and duration of early embryo cell stages are morphokinetic parameters that are predictive of embryonic development and implantation potential. In Chapter 3, a structured regression network is proposed to detect embryonic cell centroids in embryo time-lapse sequences. A convolutional regression network is trained on Gaussian-annotated centroid maps to localize embryonic cell centroids. Different from previous works, spatio-temporal relationship between sequence frames is incorporated through additional inputs to exploit natural cell stage development and spatial constraints during training. The detected cell centroids are counted for each frame to determine embryo cell stage. The embryonic cell centroids predicted by this algorithm could be used for tracking cell movement and cell count for each frame could be used to determine cell stage onset and duration.

1.6.3 Embryo Stage Classification and Onset Detection

Onset of morula and blastocyst development stages are morphokinetic parameters that are predictive of an embryo's implantation potential. Automating development stage detection of day 4-5 embryos is especially difficult due to small variation between stages. In Chapter 4, an image classification network is proposed to detect embryo development stage with new learning strategies that explicitly address challenges of this task. Synergic loss encourages the network to recognize and utilize stage similarities between different embryos. Short-range temporal learning incorporates chronological order to embryo sequence predictions. Image and sequence augmentations complement both approaches to increase generalization to unseen sequences. Embryo stage classification predictions across each sequence are restructured to follow monotonic non-decreasing order. The minimum index at which each stage occurs is then chosen as the stage onset. The morula and blastocyst onset times predicted by this algorithm could be used to assess embryo development progress and likelihood of leading to positive implantation outcome.

1.7 Scholarly Contributions

Throughout this program, two peer-reviewed conference papers were published. Another peer-reviewed conference papers was submitted and is currently under review. They are listed below in chronological order, corresponding to work described in Chapters 2, 3, and 4, respectively.

1. Lockhart, Lisette and Saeedi, Parvaneh and Au, Jason and Havelock, Jon. Multi-Label Classification for Automatic Human Blastocyst Grading with Severely Imbalanced Data. In *21st International Workshop on Multimedia Signal Processing*, pages 1-6. IEEE, 2019.

2. Lockhart, Lisette and Saeedi, Parvaneh and Au, Jason and Havelock, Jon. Human Embryo Cell Centroid Localization and Counting in Time-Lapse Sequences. In *25th International Conference on Pattern Recognition*, pages 8306-8311. IEEE, 2021.

3. Lockhart, Lisette and Saeedi, Parvaneh and Au, Jason and Havelock, Jon. Embryo Development Stage Onset Prediction from Time-Lapse Imaging with Synergic Loss and Temporal Learning. Submitted to *24th International Conference on Medical Image Computing & Computer-Assisted Intervention*, 2021.

Chapter 2

Blastocyst Grading

2.1 Problem Description

Assessing the development progress and quality of blastocyst ICM and TE structures, and expansion can be used to rank embryos for implantation in an IVF treatment cycle. This is currently performed by visual assessment, with grades commonly assigned according to the Gardner grading system. Manual grading is prone to subjectivity between expert embryologists due to large biological variation and interpretation of the grading criteria. The goal of this project is to develop an automated algorithm for assigning quality scores to embryo ICM, TE, and BE from single images. The algorithm must be able to distinguish biological variation between multiple grades using a small, severely imbalanced training set.

2.2 Related Work

A pre-trained deep CNN was fine-tuned to classify blastocyst images into good or poor quality with high accuracy achieved by aggregating predictions over several focal depths [35]. Blastocysts were partitioned into three categories (good, fair, and poor quality), but the middle (fair quality) class was disregarded. In practice, there can be multiple good quality blastocysts per cycle and multiple-grade blastocyst scoring can determine those with highest implantation potential. A similar method predicted blastocyst quality according to the Gardner grading system, assigning scores for ICM, TE, and BE [10]. A separate network was optimised for each grade, requiring three models and three training periods to perform the blastocyst grading.

Blastocyst TE has been semi-automatically segmented using ellipse fitting by direct least squares and variational level set algorithm [66]. TE segmentation was fully automated using Retinex algorithm and level set segmentation [70], and texture analysis with watershed segmentation [64]. The ICM component has been segmented using image processing methods including variational level set algorithm [65] and texture analysis [64], [52]. Deep CNNs have also been used to segment blastocyst ICM, including fully convolutional networks [34] and stacked dilated U-Net [55]. These later works show promise that deep learning can be extended to the small blastocyst image dataset in this thesis.

Transfer learning has been applied in many medical image classification tasks thanks to publicly available network weights from pre-training on ImageNet [63] database. Training networks with transfer learning versus random weight initialization were shown to improve trained convolution kernels for thoraco-abdominal lymph node detection and interstitial lung disease classification [68] and be more robust to amount of data available for training [76]. Transfer learning was used for skin cancer detection from skin lesion images [18] and pneumonia detection from chest x-rays [58]. Following these popular implementations, transfer learning was used for embryo image analysis with classification of blastocyst image quality [35], [10] and classification of embryonic cell stage [50]. Due to previous success of transfer learning and the limited number of blastocyst images for training, transfer learning was used as the foundation of blastocyst grading in this work.

Multi-class classification on imbalanced datasets has been approached for deep neural networks using cost sensitive learning [32], [78] and majority to minority class knowledge transfer [79]. Cost sensitive learning is suitable single-label classification, but cannot directly translate to multi-label classification problems with varying class distributions for each label and combinations of classes for each sample. Transferring class knowledge from majority to minority classes can be used for large-scale classification problems with sufficient variation between classes, but blastocyst classification typically involves small datasets with a small spatial variation between classes. Stratified sampling is used in this work to address class imbalance to ensure an appropriate amount of class sample representation during training.

2.3 Proposed Methods

To assign quality scores to blastocyst images, image pre-processing is first performed to remove artifacts and improve data quality for analysis. Baseline image classification networks are trained for single-grade blastocyst scoring. A single multi-grade blastocyst image scoring network is proposed to improve classification performance and the training pipeline. Stratified sampling is performed to ensure representation of minority classes for training and testing.

2.3.1 Image Pre-processing

Image pre-processing is used throughout this thesis to isolate the most important image information for embryo assessment (i.e. the embryo) and eliminate artifacts. This is a particularly important step for embryo image assessment tasks since datasets are small and so low quality samples cannot be omitted from or compensated for in training.

Most images in the dataset (those collected from PCRM) were 3-channel images with image height and width of 479 and 720 pixels, respectively. The remaining images collected from online sources were either 3-channel or 1-channel and their height and width varied.

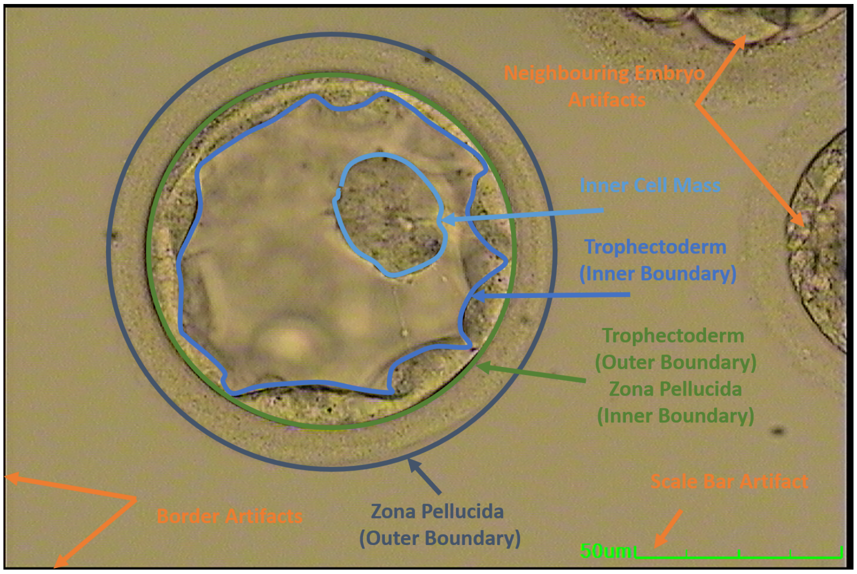

Neighbouring Embryo Artifacts
Inner Cell Mass
Trophectoderm (Inner Boundary)
Trophectoderm (Outer Boundary) Zona Pellucida (Inner Boundary)
Scale Bar Artifact
Border Artifacts
Zona Pellucida (Outer Boundary)
50um

Figure 2.1: Typical blastocyst image with labeled morphological components (ICM, TE, and ZP), and common artifacts (image borders, scale bar, and neighbouring embryos).

Every unprocessed blastocyst image (from PCRM) contained rows and columns of black pixels along their border. This was a sharp intensity change from the adjacent gray values and could influence segmentation techniques. They were therefore removed by cropping the image borders by a fixed number of pixels.

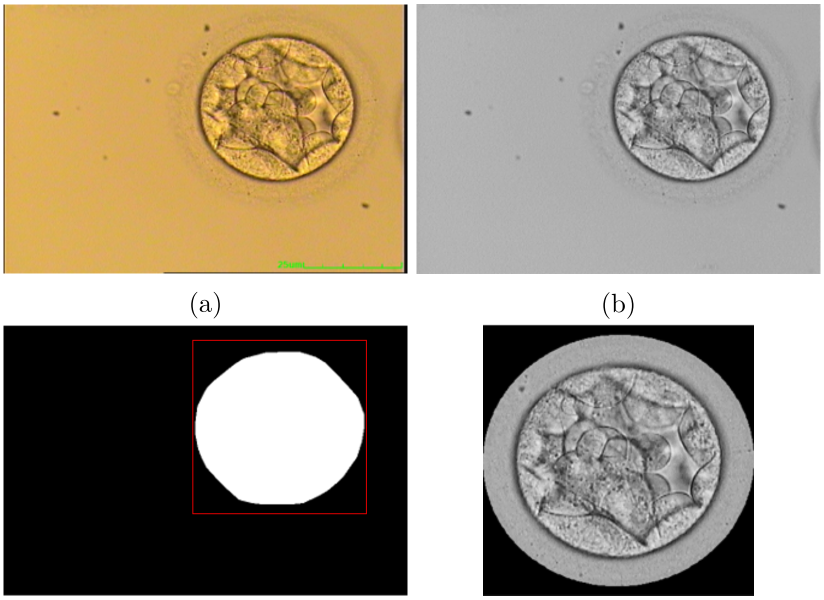
A green scale bar in the bottom right corner of each image is used by embryologists to identify the size of blastocysts. Again exhibiting a sharp intensity change compared to surrounding pixels, these extraneous pixels were replaced with more representative values. Scale bar pixels were found by scanning all the pixels in the bottom 7/8 rows and left 1/2 columns. They were identified when red and blue channels had a pixel intensity value of 0, and the green channel had a value of 255. Green scale bar pixels were replaced with the mean intensity value in a 5x5 window, excluding neighbouring green pixels. All images were converted to grayscale for consistency across the dataset.

To segment the relevant embryo image area, sobel edge filtering followed by morphological operations were applied to create a binary segmentation mask of blastocyst inner ZP area. To remove small sections that disrupted the elliptical shape along the outer contour, the mask was subtracted from its convex hull and the remaining pieces were dilated. The inverse of this mask was multiplied by the largest connected component, thereby removing sections with sharp orientation change from the elliptical contour tangent lines. The largest connected component was again isolated and filled.

The binary segmentation mask of the inner blastocyst was dilated by 20% of the total blastocyst area to ensure any ZP classified as background was included in the cropped image. An ellipse was constructed using the eccentricity, orientation, minor axis length, and major axis length of the dilated mask using Eq. (2.1), where a and b are major and minor axis lengths, respectively, h and k are shifts in the x and y directions, respectively, and α is orientation angle measured from x axis. This eliminated commonly occurring conical edges the convex hull was susceptible to.

$$1 = \frac{((x-h)\cos\alpha + (y-k)\sin\alpha)^2}{a^2} + \frac{((x-h)\sin\alpha - (y-k)\cos\alpha)^2}{b^2} \tag{2.1}$$

The elliptical mask was multiplied by the grayscale image and cropped to the smallest bounding square. Zero padding was added to the image wherever the ellipse exceeded image dimensions. A square crop shape was chosen over rectangular as image inputs are square in CNN-based image classification tasks. Pixels surrounding the blastocyst mask were set to zero (black) rather than using actual image values. This ensured all cell fragments and neighboring embryos were excluded. Image pre-processing steps are shown in Fig. 2.2.



(a)

(b)

(c)

(d)

Figure 2.2: Image pre-processing steps: original image (a), borders and scale bar removed and converted to grayscale (b), blastocyst segmentation map with red box indicating crop region (c), and final centred and cropped image (d).

Images were resized to height and width of 320 × 320 pixels when used as network input. This standardized the network input size and was found to be large enough to capture fine-grained details necessary for classification. Single-channel grayscale images were repeated to create 3-channel images to match the pre-trained CNN input dimensions.

2.3.2 Single-label Multi-Class Networks

A baseline was established using separate networks for ICM, TE, and BE grading (i.e. single category, multiple scores). Three state-of-the-art deep CNNs pre-trained on the ImageNet database were compared for this task: VGG16 [69], ResNet50 [27], and InceptionV3 [75]. VGG16 is comprised of stacks of layers each with 3×3 convolution kernels that increase in number of feature maps with increasing network depth. ResNet50 utilizes identity connections in parallel with residual blocks to facilitate information travel through the network. InceptionV3 is a set of stacked modules comprised of different sized convolution filters in parallel to incorporate additional context at each stage. The amount of trainable weight parameters excluding final fully connected layers is 14.7 million, 23.6 million, and 21.8 million for VGG16, ResNet50, and InceptionV3, respectively.

Pre-trained weights in convolution layers form the base of each network. Final feature maps, obtained by passing the input images through the base, are pooled to a scalar value for each channel by assigning the maximum value of each feature. These pooled nodes are connected to a 32-node ReLU-activated fully connected layer with 50% dropout [73] rate. In each training iteration (forward pass of input data and backward pass of output error), 16 nodes are randomly chosen to be disconnected from the network. At test time, the 32 nodes' weights are halved since there are twice as many nodes. This strategy prevents the network to rely on relationships between nodes and avoid learning weights too specific to the training data.

The 32-node layer is connected to a 3-node softmax-activated output layer, corresponding to three quality scores in a blastocyst grade. Only the last three convolution layers were fine-tuned so the network would avoid learning combinations of features specific to training data. When training the larger ResNet50 and InceptionV3 model bases, there were no layers left as trainable since they were much more prone to overfitting.

2.3.3 Multi-label Multi-Class Network

Training separate networks for the different blastocyst grading tasks was inconvenient as models were trained individually and experimental results and analysis were computed separately. There was also redundancy as features were encoded from the exact same set of images multiple times. Ideally, a single network should perform all three grading tasks simultaneously to make better use of encoded network features, reduce training time, and streamline experimental analysis.

A single network, shown in Fig. 2.3, was proposed to classify all grades concurrently (i.e. multiple categories with multiple scores). Similar to the baseline network, pre-trained VGG16 convolution and pooling layers encoded image features. However, a multi-branch classifier used the pooled feature maps to predict quality scores for all grades in one pass. Each branch had its own set of layers and could therefore be optimized separately. Each classifier branch was similar to the single-label network fully connected classifier, except an 8-node layer was added between the 32-node and output layers. This facilitated a more specific mapping of the shared convolutional features to the individual grade outputs.

Each output layer had an individual error function with respect to the ground truth annotation for its respective grade. Parameter weight updates in the fully connected portion were based solely on the error backpropagated from their respective blastocyst grade label. Parameter weights from the trainable portion of the convolution layers were updated as average backpropagated error from all three labels. This was also a form of regularization to convolution layer weights to prevent overfitting to the training data.

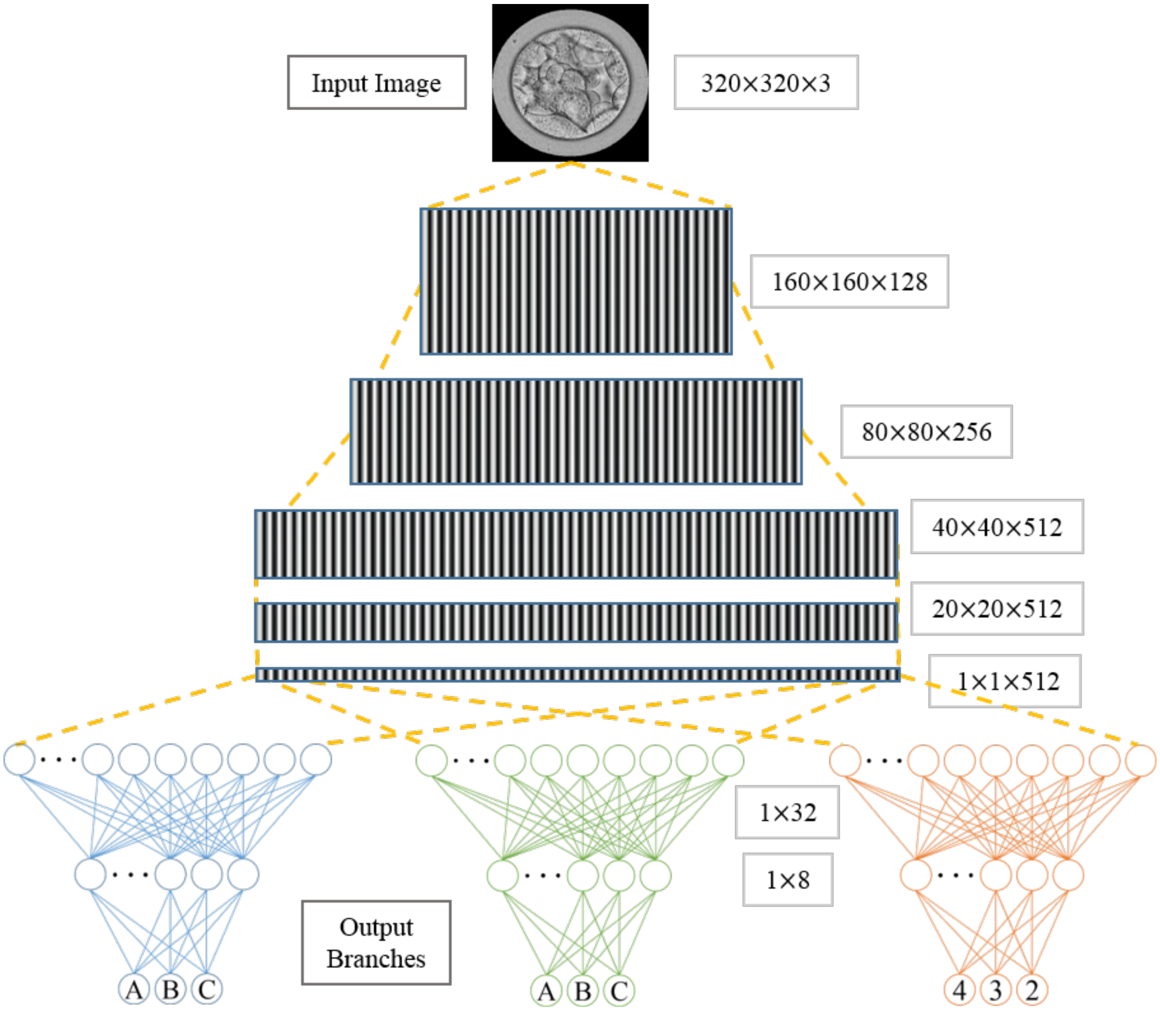


Figure 2.3: Blastocyst grading network diagram with size and number of convolution layer feature maps and fully connected layer nodes. Input images are fed through a series of convolution and pooling layers shared between grading labels. Extracted feature maps are used by separate fully connected classifier branches to assign ICM, TE, and ZP grades simultaneously.

Training a single multi-label classification network required approximately $\frac{1}{3}$ the number of network parameters and training time compared to three single-label classification baseline networks.

2.3.4 Network Training Details

All networks were trained using categorical cross-entropy loss with scalar grading labels converted to one-hot encoded binary arrays. RMSprop optimizer [15] was used to update parameters using an initial learning rate of 1×10^{-6}, scheduled to decrease if no improvement to 0.3× its current value. Training was done for 500 epochs with early stopping.

During training, random data augmentations were performed each epoch to add sample variation. Images were rotated between angle of 0-360°, shifted horizontally and vertically between image width and height of 0-15%, zoomed between image size of 0-15%, and linearly mapped by shear transformation between image size of 0-5%. These augmentations add variation to each sample to artificially increase the amount of data for better model generalisation. During training and testing, each sample was subtracted by its mean and divided by its standard deviation to reduce the effect of illumination differences.

2.3.5 Stratified Sampling

Ideally, classification datasets should have the same amount of samples per class during training to ensure equal likelihood of each sample belonging to any class. Unfortunately, ICM grade exhibits severe class distribution imbalance, with only 1.0% of image samples belonging to the minority class. Classification networks trained with imbalanced data receive more samples from the majority class and often use class probability rather than data features to make predictions. It is possible to achieve high ICM classification performance by simply predicting all samples as majority class A.

To ensure that samples from the minority ICM class were represented throughout experiments, the data was stratified by partitioning samples according to their ICM grade. Samples in majority and middle ICM classes 'A' and 'B' were split into sample sets according to a set percentage of the total data. However, the 7 available samples in minority ICM class 'C' were fixed to 3, 2, and 2 for training, validation, and test sets, respectively. Samples assigned to each set were chosen randomly by data shuffling before each experiment. Although this technique was not required for the single-label baseline models performing TE and ZP grading, the same stratification of ICM sample distribution was used to compare results with less bias across models.

2.4 Experimental Results

MATLAB R2018b was used for blastocyst grading image pre-processing. Deep neural networks were trained using Keras 2.2.4 framework with TensorFlow 1.11 backend.

2.4.1 Dataset

Images for automatic blastocyst grading were collected and labeled by two embryologists at PCRM. Of the 704 total images, 674 were collected at from the clinic between 2012 and 2018, and the remaining 30 images were gathered from online sources. Images were acquired using an Olympus IX71 inverted microscope. The sample distribution for each grade is shown in Table 2.1. Since the images collected from PCRM were from embryos that were implanted, the majority of blastocyst images had high ICM and TE grades and very few had ICM or TE grade C.

Table 2.1: Blastocyst grade distribution for 704 images. Each image has 3 grades. Majority classes are in bold.

ICM		***TE***		***BE***	
A	**507 (72.0%)**	A	**382 (54.3%)**	4	248 (35.2%)
B	190 (27.0%)	B	268 (38.1%)	3	**300 (42.6%)**
C	7 (1.0%)	C	54 (7.6%)	2	156 (22.2%)

2.4.2 Setup

The classification performance was measured using 3-fold cross-validation. The dataset was randomly split into training (70%), validation (15%), and test (15%) sets within the stratified sampling constraints. Training and evaluation was performed three times per experiment with a different combination of samples across the sets in each fold.

Confusion matrices show how images were classified, with correct predictions along the diagonal. Accuracy, precision, and recall averaged over each grade label show overall performance. Accuracy was calculated by globally dividing true positive predictions in all classes by the total number of samples, thereby weighing all predictions identically. Precision and recall were computed for each class individually, then each were averaged each across the three classes. The equation for accuracy is given in Eq. (2.2), and equations for precision and recall are given in Eq. (2.3). TP_i, FP_i, and FN_i are the numbers of true positive, false positive, and false negative predictions, respectively, belonging to class i, and $N_{samples}$ is the total number of samples in the set.

$$\text{Accuracy} = \frac{\sum_{i=1}^{3} TP_i}{N_{samples}} \tag{2.2}$$

$$\text{Precision} = \frac{1}{3}\sum_{i=1}^{3}\frac{TP_i}{TP_i + FP_i} \qquad \text{Recall} = \frac{1}{3}\sum_{i=1}^{3}\frac{TP_i}{TP_i + FN_i} \tag{2.3}$$

2.4.3 Quantitative Results

Confusion matrices of outputs from baseline and proposed networks are presented in Table 2.2. The proposed multi-label classification network with VGG16 base had the most correctly predicted samples in all BE grade classes and two TE grade classes. It was the only network to correctly assign TE grade C class samples. VGG16 baseline network classified all samples to the ICM majority class, and had the most correctly predicted samples in TE grade B class. ResNet50 was the only network to correctly classify an ICM grade C class sample and also correctly classified the most ICM grade B class samples. In all networks, majority class samples were correctly classified more often and minority class samples of ICM and TE grades were almost entirely misclassified.

Table 2.2: Confusion matrices of 318 images across 3 test folds. Correctly predicted samples are in blue and highest number of correctly predicted samples per class are bold.

	Model	Actual Classes										
		Label	***ICM***			***TE***			***BE***			
		Class	A	B	C	A	B	C	Class	4	3	2
Predicted Classes	ResNet50 [27] 3×Single-Label	A	116	50	2	121	82	20	4	18	58	44
		B	83	**24**	3	40	28	9	3	11	21	35
		C	29	10	**1**	12	6	0	2	33	61	37
	InceptionV3 [75] 3×Single-Label	A	203	79	6	89	93	10	4	6	7	11
		B	18	2	0	60	53	13	3	25	60	54
		C	7	3	0	0	0	0	2	36	76	43
	VGG16 [69] 3×Single-Label	A	**228**	84	6	103	46	2	4	29	11	0
		B	0	0	0	66	**78**	23	3	45	103	18
		C	0	0	0	0	0	0	2	0	24	88
	VGG16 - Multi-Label (Proposed)	A	214	63	4	**151**	57	0	4	**57**	9	0
		B	14	21	2	21	59	19	3	17	**113**	11
		C	0	0	0	0	7	**4**	2	0	21	**90**

Overall classification accuracy, precision, and recall are summarised in Table 2.3. Pretrained ResNet50 and InceptionV3 baseline models did not adapt well to the task of blastocyst grading. Scoring higher accuracy in ICM and TE grading than BE grading suggests that performance was improved by assigning more samples to the majority class, which also caused lower precision and recall scores.

For ICM grading, single-label VGG16 simply assigned all samples to the majority class for all three folds, indicating the data distribution was the primary prediction factor. However, TE grading accuracy was higher than assigning all samples to the majority class. Grading performance of VGG16 was considerably higher on average compared to ResNet50 and InceptionV3 networks, and it was therefore used in multi-label classification.

Table 2.3: Blastocyst grading performance averaged across grade classes. Accuracy, precision, and recall are expressed in percentage with standard deviation (SD) across three folds.

Model	***ICM***			***TE***			***BE***		
	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
ResNet50 [27] 3×Single-Label	44.3 ±26.9	37.7 ±27.2	32.0 ±17.4	46.9 ±14.8	30.2 ±27.6	31.4 ±35.5	23.9 ±8.1	24.9 ±8.7	25.3 ±9.0
InceptionV3 [75] 3×Single-Label	64.5 ±12.6	26.8 ±38.1	30.5 ±50.7	44.6 ±3.0	29.5 ±25.6	32.0 ±30.1	34.3 ±7.2	32.0 ±9.8	30.2 ±18.5
VGG16 [69] 3×Single-Label	71.7 ±0.0	23.9 ±41.4	33.3 ±57.7	56.9 ±11.5	38.3 ±34.9	41.3 ±35.8	69.2 ±12.0	71.0 ±8.4	65.6 ±23.3
VGG16 - Multi-Label (Proposed)	**73.9** ±2.0	**44.3** ±39.6	**39.6** ±48.6	**67.3** ±4.7	**56.1** ±18.4	**51.0** ±35.3	**81.8** ±3.9	**82.5** ±3.4	**81.7** ±6.5

The proposed multi-label classification network achieved the greatest blastocyst grading performance for all metrics. With imbalanced data in ICM and TE grades, accuracy was significantly higher than precision and recall, indicating sample distribution influenced predictions. However, accuracy for these grades were higher than assigning all samples to the majority class (71.7% and 54.7% accuracy, respectively), demonstrating blastocyst image features were learned to distinguish between classes.

Precision and recall scores in respective grades were similar, showing a proportional amount of misclassifications in minority and majority classes. Higher BE grading metrics show better classification performance was achieved with less bias towards a particular class when sample distribution is more balanced. Standard deviations for ICM and TE grades show the train/test split was highly influential in performance of minority classes.

Comparison to Related Work

In [35], the problem of blastocyst grading was greatly simplified by assigning blastocysts into good, fair, or poor quality classes then disregarding the middle (fair) class. They fine-tuned an InceptionV1 classification network to assign blastocyst images (with 7 focal planes) as good or poor quality and reported classification accuracy of 96.94% for a single fold. Their study included 10,148 embryos with 50,392 images (14× the number of embryos available for this thesis), with some poor quality embryos omitted to achieve a balanced number of good and poor quality embryos.

In [10], separate networks were trained for blastocyst grading using Gardner grading system. They fine-tuned three ResNet50 classification networks to assign ICM, TE, and BE grades to blastocyst images. They reported classification accuracy of 89.63%, 82.84%, and 95.73% for ICM, TE, and BE, respectively, for a single fold. Their study included 16,201 embryos with 171,239 images (23× the number of embryos available for this thesis), with imbalanced distribution of samples across scores.

The classification accuracy achieved in this thesis was 73.9%, 67.3%, and 81.8% for ICM, TE, and BE, respectively, averaged across 3 folds. Although these results are significantly lower than in [10], this can be largely attributed to having only 704 embryos (704 blastocyst images) available for fine-tuning the classifier and cross-validation. The methods proposed in this thesis performed fine-grained blastocyst grading to enable comprehensive ranking of blastocysts while also addressing the redundancy of and data distribution imbalance with fine-tuning three networks to assign different grades to the same images.

2.4.4 Visual Results

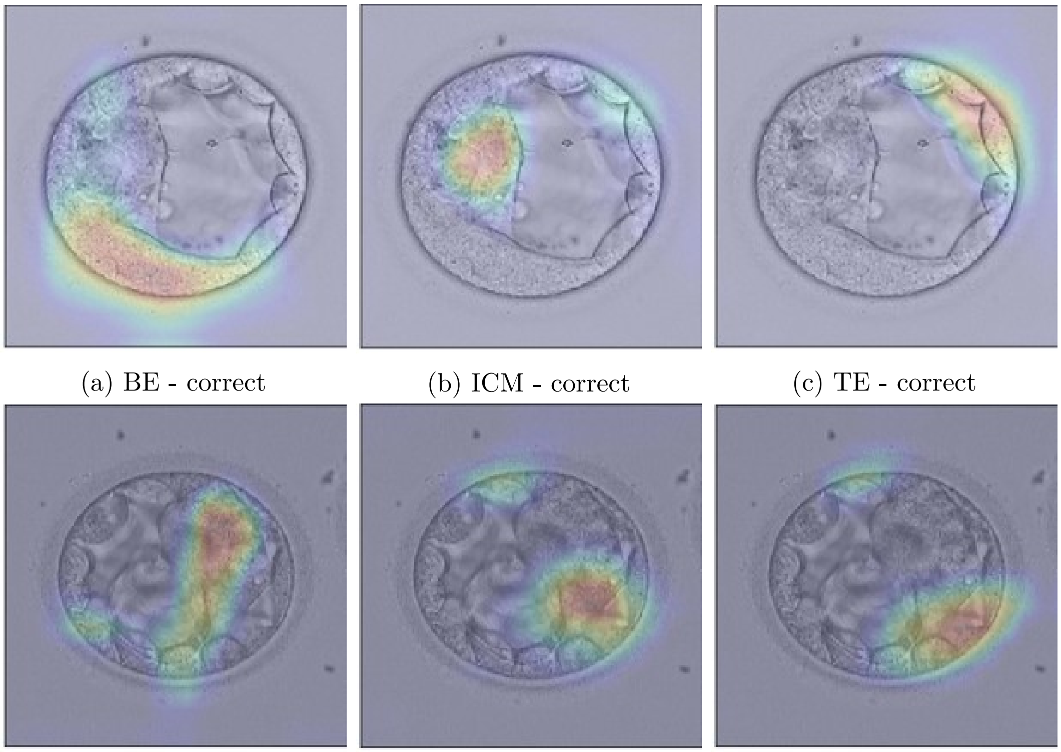
Class Activation Maps (CAMs) computed from the network weights show the region(s) in the input image that contributed towards the final grade prediction. They are computed by tracing the predicted class node activation back through the network, resulting in a continuous-valued single channel 2D map with the same height and width as the input image. These maps can be plotted with a colormap and overlaid on the input image, highlighting how important each pixel was for classification. CAMs are overlaid on input images from ICM, TE, and BE grading in Fig. 2.4, where red pixels indicate highest contribution towards grading prediction and no colour represents no contribution towards grading prediction.

Correct classifications occurred when the corresponding anatomical structure was the primary focus of the network (i.e. ICM region was red for ICM grading, TE region was red for TE grading, and ZP or blastocoel region was red for BE grading). Incorrect classifications occurred when a different structure or small part of a structure not representative of the whole contributed to the final prediction. Even though images were fed through the same feature extractor, the grading branches were specialized enough to focus on different regions of the image for classification.

2.5 Conclusions

An image classification network was trained for assigning scores to single blastocyst images for ICM and TE quality, and BE. Classification performance of ICM, TE, and BE grading was improved by combining the three labels into separate network output branches with a shared feature encoder. Training time and memory footprint were reduced by using a single network for analysis. As most model parameters were shared for all grades, the bias inherent to skewed data distribution was reduced by jointly optimizing the three grading tasks.

This algorithm can be used to automatically predict the ICM, TE, and BE grade of blastocysts according to the Gardner grading system. These scores are morphological parameters that could be used to assess embryo quality in IVF treatment.



(a) BE - correct

(b) ICM - correct

(c) TE - correct

(d) BE - incorrect

(e) ICM - incorrect

(f) TE - incorrect

Figure 2.4: Input images with overlaid CAMs are shown for correct (top row) and incorrect (bottom row) examples of ICM, TE, and BE grading from respective network branches.

Chapter 3

Cell Centroid Localization and Cell Counting

3.1 Problem Description

Ability to measure at what time (hours posts-fertilization) embryos reach cells stages 2-5 gives insight into their development progress and implantation potential. Time-lapse imaging systems built into incubators enable determination cell stage by visual assessment to the nearest 15-minute increment. Visually assessing cell stage onset and duration is tedious and can be prone to error due to biological variation that impedes visibility of cells. The goal of this project is to develop an automated algorithm for localizing cell centroids in early embryo time-lapse sequences and extracting the cell count for each frame. The algorithm must detect cells with varying shapes and levels of boundary occlusion and overlap.

3.2 Related Work

Automated cell counting has been performed via image classification and structured regression methods. These families of methods are compared in Fig. 3.1. While classification approaches require only image-level annotations, they fail to capture the variation in cell orientations and mechanics of cell movement throughout development. With pixel-level annotations, a structured regression approach localizes cell centroids and can be used to monitor cell movement and orientation.

Semi-automatic embryonic cell detection was performed in [26] and [54], providing valuable information on cell size, shape, and symmetry as well as cell stage duration and cleavage times. In both these works, human annotations (cell centroid locations in [26] and number of cells in [54]) were required as input to generate cell boundaries. Automatically locating cell centroids could fully automate multi-instance cell segmentation when combined with either of these two cell detection approaches. Automatic ellipse-fitting approaches for blastomere detection were developed in [71] up to 4-cell stage and [33] up to 8-cell stage.

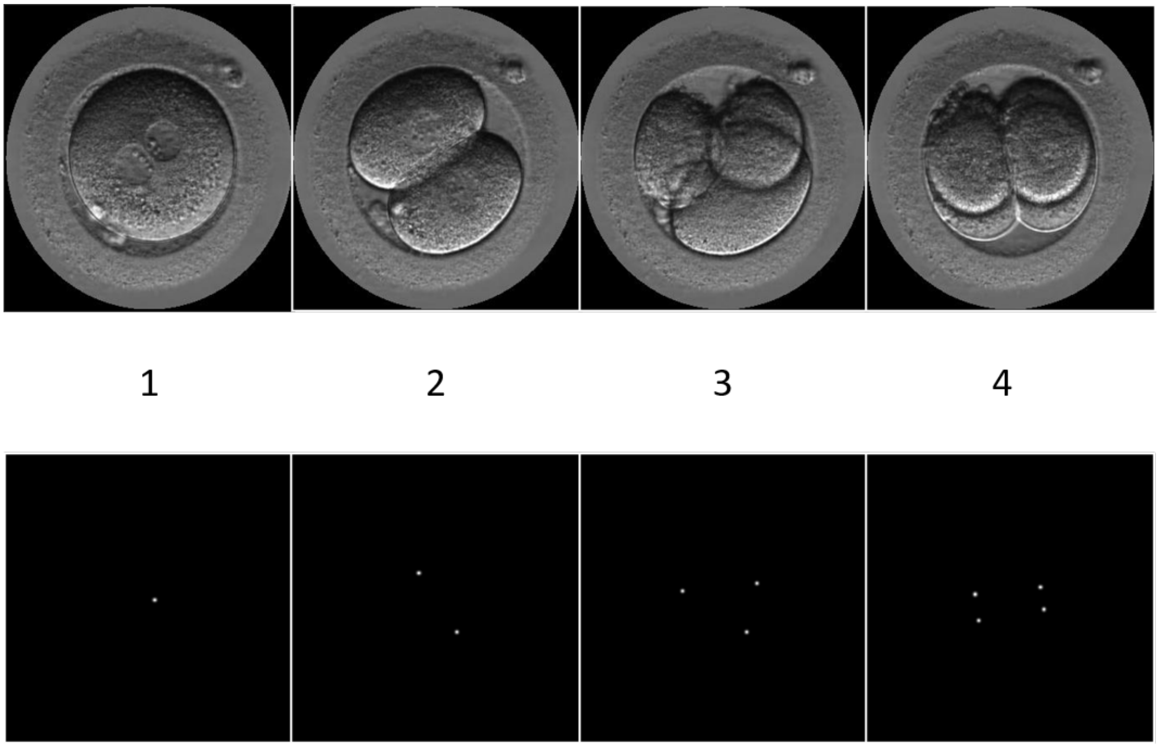


Figure 3.1: Cell counting in embryo time-lapse sequences via cell classification (middle row) and cell centroid localization (bottom row).

Embryonic cell stage was first classified using a CNN in dark-field microscopy image sequences [31]. More recently, embryonic cell stage has been classified in HMC microscopy time-lapse sequences. Temporal fusion combined extracted features of multiple frames for single-frame cell stage classification [50]. A single or multiple frames were used as input for multi-frame cell stage classification with temporal ensembling [43]. Confidence-based majority voting between two single-frame CNN classifiers improved cell stage classification over a single CNN [17]. Embryos were localized by combining Haar feature-based cascade classifier and their radiating lines, then used as input for cell stage classification [59]. A two-stage network was trained for simultaneous embryo localization and embryonic cell stage classification with weakly-supervised segmentation [40]. Most recently, a global optimization method was used to restructure predicted embryo cell stage classification [44]. While classification-based embryo staging provides useful information for cleavage time and stage duration, more insight into embryo quality can be gained by localizing each cell in a given frame.

In blood cell counting tasks, cells have been represented in microscopy images as density maps centred at their centroids. Semantic segmentation networks were converted to regression networks by changing the final network layer to have continuous output values conducive to density maps. Counting is performed as the summation of detected cell centroids from thresholded density maps [82], [51], [80], [3]. A similar regression network separated overlapping cells using their density maps for nuclei segmentation in histopathology images [49].

Embryonic cell counting via centroid localization using a convolutional regression network was introduced in [53] and extended in [56]. This network utilized a ResNet50 feature encoder with Atrous Spatial Pyramid Pooling [9] to generate a rich set of low resolution features. Encoded features were reconstructed into a high resolution density map by Progressive Upsampling Convolution. The cell centroid localization method was validated on a small dataset of selected frames with clear cell distinction and lacked challenging frames near cell cleavages that occur frequently in time-lapse sequences. Cell centroid masks were computed by Gaussian-fitting to cell segmentation masks. The cell centroid annotation shape, size, and orientation varied proportionally with each cell, providing extra information in the loss during cell centroid localization training.

Explicitly adding spatial context through attention blocks [28] has proven successful in several visual recognition tasks, including classification [28], semantic segmentation [20], and super-resolution [88]. In convolution layers with many channels, attention blocks guide contribution of local features by weighting each channel by a scalar value determined by higher resolution features. In this work, attention blocks incorporate spatio-temporal information from current frame and secondary input images into encoder layers. Low resolution features are emphasized or muted based on their association to cell centroid locations at the previous frame and cell movement between frames.

The work in this thesis builds off the method in [56] with the following differences:

- The fully convolutional regression network architecture was scaled down in depth (number of layers) and width (number of channels).
- Dilated convolution layers were implemented in parallel instead of cascaded.
- Additional inputs (previous frame prediction and optical flow diagram between frames) were incorporated into attention blocks in the regression network.
- A data sampling strategy was developed for the additional inputs.
- Cell centroid masks were generated from dot-annotations instead of cell segmentation masks.

3.3 Proposed Methods

To detect embryonic cell centroids, image pre-processing is first performed to remove highly salient embryo well regions and improve data quality for analysis. A fully convolutional regression network is trained to localize cell centroids from density map representations of cell centroid pixel coordinates. Temporal relationship between adjacent sequence frames is encoded by adding the previous frame's predicted centroid map and optical flow diagram pre-computed between consecutive frames as network input. These additional inputs are incorporated into squeeze-excitation blocks, providing attention for convolution layer channels that should be most relevant for analyzing the current frame.

3.3.1 Image Pre-processing

Embryo time-lapse sequence frames are a single-channel images with height and width of 500 pixels. An example embryo time-lapse sequence frame from the dataset is shown in Fig. 3.2. Every embryo time-lapse sequence contains a circular well where the embryo is located. There is also a light-saturated patch along the upper-half of the inner perimeter of the well.

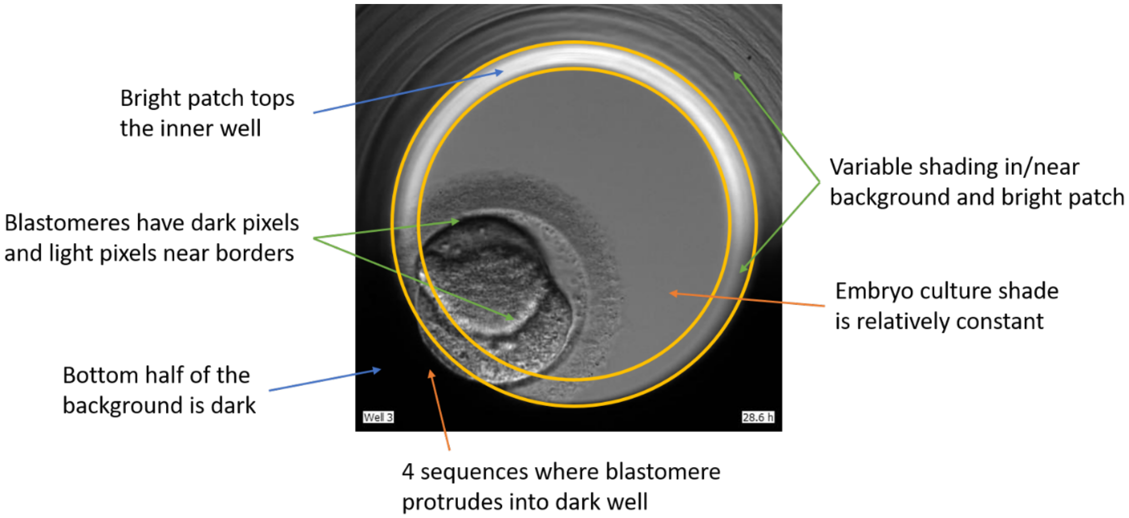

Bright patch tops the inner well
Blastomeres have dark pixels and light pixels near borders
Bottom half of the background is dark
Variable shading in/near background and bright patch
Embryo culture shade is relatively constant
4 sequences where blastomere protrudes into dark well
Wel 3
28.6 h

Figure 3.2: An embryo sequence frame with observations used to develop the image pre-processing pipeline.

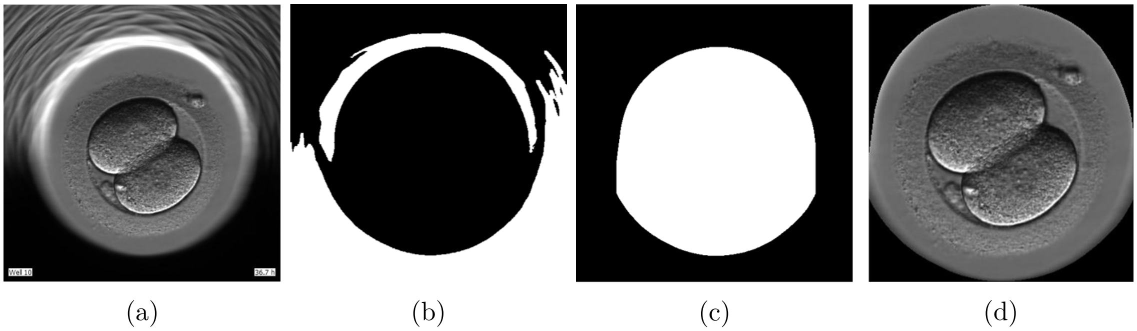
With pre-processing, the embryo and culture media were isolated and the remaining pixels were cropped out (i.e. set to zero). The cell centroid localization algorithm could therefore process only the relevant pixels and ignore potentially unfair cues like the timestamp in the bottom-right corner. The sequence frame pre-processing steps are shown in Fig. 3.3 and were as follows:

1. The 2 largest connected components of brightest and darkest pixels (corresponding to upper and lower well boundaries) were isolated. Pixel intensity values were thresholded and connected components were refined with morphological operations.

2. Connected components were joined along the minimum distance line between components on the left and right sides. The top component bottom perimeter was linked to the bottom component top perimeter using geometric constraints.

3. The embryo frame was multiplied element-wise to the segmentation mask, thereby eliminating image patterns from the microscope as well as well number and timestamp.

Lastly, the microscope well containing the embryo was centred by cropping to its smallest bounding rectangle.



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

Figure 3.3: Image pre-processing steps for cell centroid localization: original image (a), largest bright and dark components (b), embryo area mask (c), and final centred image (d).

Even with image pre-processing, the task of cell centroid localization remains challenging due to irregular-shaped cells, cells with occluded boundaries, severely overlapped cells, and large cell-like fragments.

3.3.2 Fully Convolutional Regression Network

Convolutional structured regression networks encode input images into high-channel low resolution features through convolution operations and non-linear activation. These features are decoded by an inversely proportional set of upsampling operations to construct a high resolution output map. A single linearly-activated output layer assigns pixel values that correspond to centroid proximity. The trained network predicts cell centroid location probabilities, which are thresholded for cell counting.

The embryonic cell centroid localization regression network utilizes a ResNet18 [27] feature encoder. This is a shallower and narrower encoder than that used in [56], making it quicker to train on the larger time-lapse sequence dataset and less prone to overfitting to the many nearly identical frames between cell cleavages.

Since cells naturally change (decrease) in size throughout embryo development, dilated convolution inception modules [57] are added to encode features at multiple resolutions without adding more parameters. These modules contained convolution layers with dilation rates of 1×, 2×, 4×, and 8×, each with $\frac{1}{4}$ the number of filters of the previous layer. The dilated convolutions were performed in parallel instead of in cascade as in [56] due to better performance seen in [57].

Progressive Upsampling Convolution [56] transforms features from low to high resolution with explicit emphasize on each network feature size. Multiple pairs of sub-pixel convolution - pixel shuffle upsampling operations of inversely proportional upsampling rates are performed in parallel. Pixel shuffling [67] increases feature resolution while minimising information loss of upsampling. The set of high resolution feature maps create a continuous-valued output density map for structured regression.

As cell centroid annotations are sparse binary maps, they are smoothed with a Gaussian kernel to increase the amount of non-background pixels. This increases informative feedback to the network weights during training. These density maps encode background and cell centroid as low and high pixel values, respectively. The Gaussian kernel size was chosen to provide maximum nonzero pixels while maintaining separation between cells in the image.

Previously used loss functions for cell counting regression - L2 loss [82], [80], [49] and L1 loss [51], [3] - are ill-suited for embryo cell segmentation where there are few cells (i.e. non-zero regions) in the image. When the output masks are dominated by the background class, the network tends to optimize for the trivial solution of predicting all pixels as background. To address the imbalanced ratio of background pixels, spatially-weighted MSE [56] was used to penalize incorrect predictions near cell centroids more harshly. The loss, for pixel coordinates (m, n), is

$$\mathcal{L}(y, \hat{y}) = \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{((y_{m,n} - \hat{y}_{m,n})^2 \cdot ((\alpha_0 \cdot \frac{y_{m,n}}{\max y}) + \alpha_1))}{M \cdot N} \tag{3.1}$$

where M is the image height, N is the image width, y is the ground truth smoothed centroid map, $\hat{y}$ is the predicted centroid map, and α_0 and α_1 control the weighting of each pixel according to their proximity to the nearest ground truth centroid.

3.3.3 Temporal Context Prior

Embryo cell stage is a monotonically non-decreasing phenomenon, which restricts the cell count at any frame. Spatial location of cells are constrained by their enclosure within the zona pellucida (a membrane separating inner embryo from outer environment), and embryo position in the incubator microwell. Knowing an embryo's cell stage and centroid location(s) at the previous frame is therefore highly relevant to analyze the current frame. Predicted cell centroid maps at the previous frame $\hat{\mathbb{X}}_2$ were added as a secondary input and incorporated into encoder attention blocks for cell centroid localization. Compared to [28] where attention blocks take only features from previous residual block, this module also utilizes global features from $\hat{\mathbb{X}}_2$. Features are concatenated in the squeeze operation. This method is referred to as Multi-Input I.

Iterating through batches of sequences in parallel will group together samples of similar cell stage and background appearance. This could bias network gradient updates towards ideal kernel weights specific to each batch, and not the best weights for the entire dataset. Diverse batches of images from different cell stages and sequences were selected by sampling image and predicted map pairs in random order. This was possible by storing the entire training set of predicted maps after each epoch. This secondary input training strategy is shown in Algorithm 1. The strategy was similarly applied to validation set sequences. For testing, predictions were performed sequentially one frame at a time, using each prediction as secondary input to the following frame of each sequence.

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
    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
    Predict on $([\mathbf{x}_1^{(i)}, \hat{\mathbf{x}}_2^{(i)}], \mathbf{y}^{(i)}), i \in \{1, 2, ..., (N-1)\}$
    Store $\hat{\mathbf{y}}^{(i)}, i \in \{1, 2, ..., (N-1)\}$ as $\hat{\mathbf{x}}_2^{(i)}, i \in \{2, 3, ..., N\}$
```

Cell centroid localization maps can have the same values whether they precede frames with no cell movement or full cell division. They can be misleading as a prediction prior without indication of cell movement between frames. Optical flow diagrams can contextualize cell movement between frames by approximating cell velocity from changes in brightness patterns. Object velocity is encoded in 2-channel images with same height and width as the sequence frames. The 1st and 2nd channels represent movement in the horizontal and vertical directions, respectively. An examples of optical flow diagrams through cell division is shown in Fig. 3.4 in RGB image format.

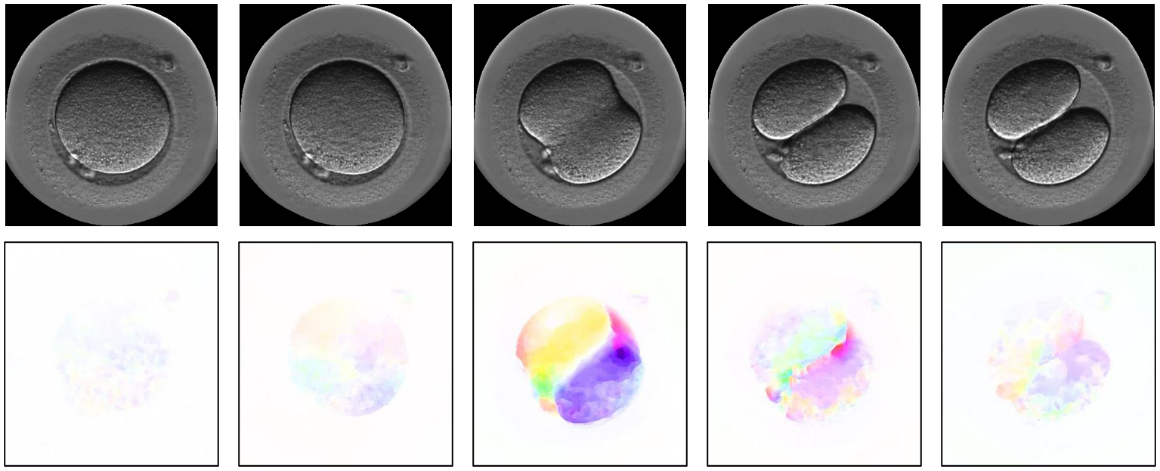


Figure 3.4: Optical flow diagrams (converted to RGB images) between consecutive above-left and above sequence frames. Black borders were added to flow diagrams for clarity. White pixels indicate no movement and bright pixels indicate large movement between frames.

Optical flow diagrams were added as another secondary input to indicate temporal relationship to the previous frame predicted mask. These inputs were calculated using [14] between each consecutive pair of frames using TV-L^1 algorithm [83] prior to training. The flow magnitude was clipped to the range [-20, 20]. Flow magnitude normalization was omitted so the network could utilize large changes in cell movement. Global temporal features from optical flow diagrams were also concatenated in the squeeze operation of encoder attention blocks. This is referred to as Multi-Input II.

Cell centroid location predictions at the previous frame and optical flow diagrams were incorporated into channel attention blocks. Local spatial features from the current frame, global features from previous frame predictions, and temporal context of optical flow diagrams were squeezed into scalar channel descriptors to emphasize relevant features. Network training procedure is unchanged by adding attention blocks. The network architecture for Multi-Input II is shown in Fig. 3.5. Multi-Input I has the same network architecture without optical flow input.

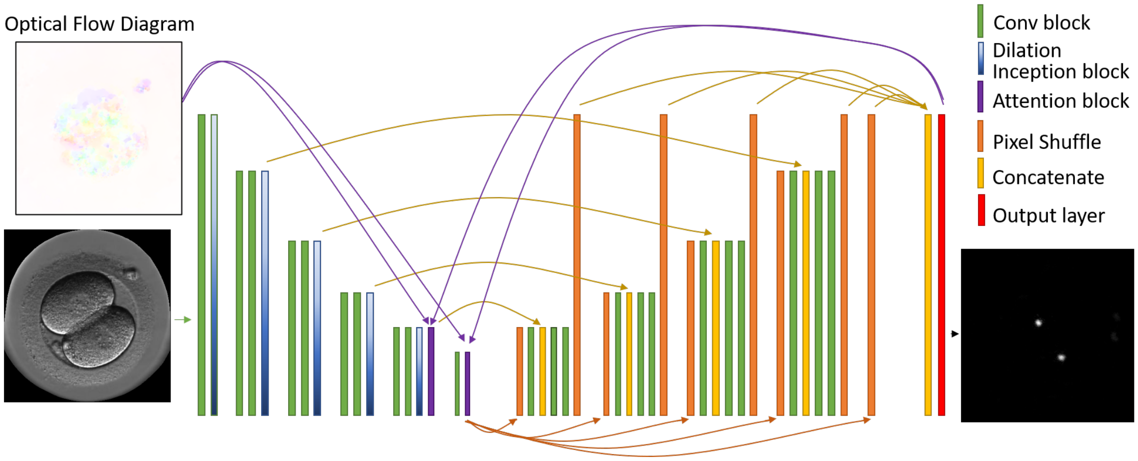


Figure 3.5: Proposed network diagram for Multi-Input II. Spatial features are extracted in a single encoder branch and are decoded by multiple parallel branches. Secondary inputs are incorporated into high-channel encoder attention blocks.

3.3.4 Network Training Details

The baseline state-of-the-art medical image semantic segmentation network, U-Net [62], was converted to a regression network by replacing the output layer with linearly-activated convolution as performed for cell counting in [80]. In the proposed network, attention block locations were chosen empirically on high-channel convolution layers (encoder modules 5 and 6) where the number of channels was large enough to incorporate previous frame context without sacrificing feature encoding capacity. The proposed networks were trained with Adam optimizer [36] with initial learning rate of 3×10^{-5}. The learning rate was reduced by 0.3 after 15 epochs of non-decreasing loss to a minimum rate of 1×10^{-7}. Training was terminated after plateau of non-decreasing loss of 30 epochs. All networks were trained with spatially-weighted mean squared error (MSE) with coefficients α_0 and α_1 set to 2 and 0.15, respectively, for all experiments.

3.3.5 Cell Counting

Cell centroid regions were found by applying Otsu thresholding to the predicted cell centroid maps. Connected components in the binary thresholded map were analyzed and any components with major axis length greater than double minor axis length were divided into two regions. These were assumed to be two closely located cells and were separated at midway along the region's morphological skeleton (i.e. controlled erosion resulting in a single pixel representation of the shape). This process is shown in Fig. 3.6. The centre pixel of each connected component was considered cell centroid. Connected components were summed for cell count.

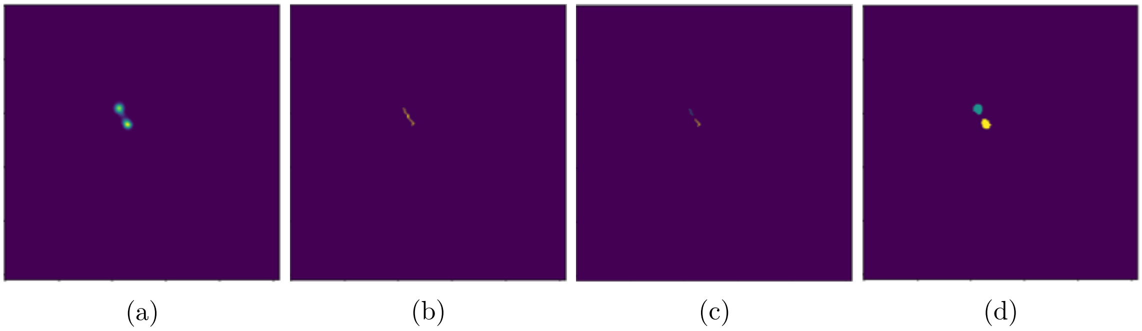

(a) (b) (c) (d)

Figure 3.6: Process for dividing closely cell centroids: unprocessed centroid localization regression map prediction (a), skeleton of Otsu thresholded regression map (b), divided skeleton (c), post-processed centroid prediction (d).

3.4 Experiments

Embryo time-lapse sequence frames were pre-processed using MATLAB R2018b. Deep neural networks were trained using Keras 2.2.4 framework with TensorFlow 1.11 backend.

3.4.1 Cell Centroid Localization

Dataset

The dataset contains 108 human embryo time-lapse sequences at 1-4 cell stage collected at the Pacific Centre for Reproductive Medicine from 2017-2018. A frame was captured every 15 minutes beginning from the start of pronuclear phase to end of 4-cell stage. Cell centroids were manually annotated following cell stage annotations by expert embryologists. 36,035 cell centroids were annotated, with cell count distribution across dataset and sequences given in Fig. 3.7 and Fig. 3.8. 3-cell stage in embryo development is much shorter on average than other stages and was not captured in the imaging interval for some sequences.

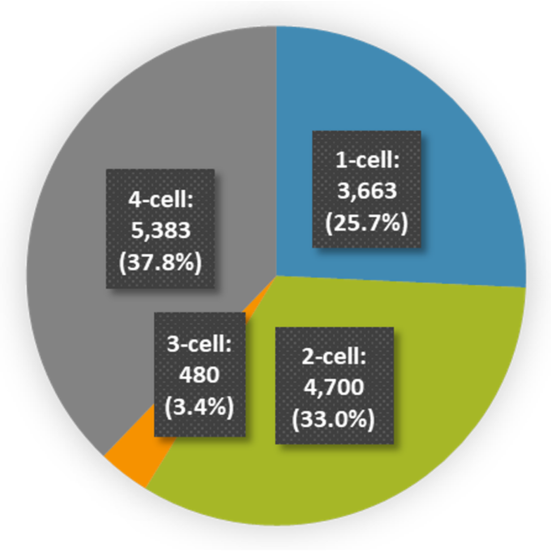


Figure 3.7: Cell count distribution across the dataset. The number of frames at each cell stage is imbalanced, with significant minority of samples at 3-cell stage.

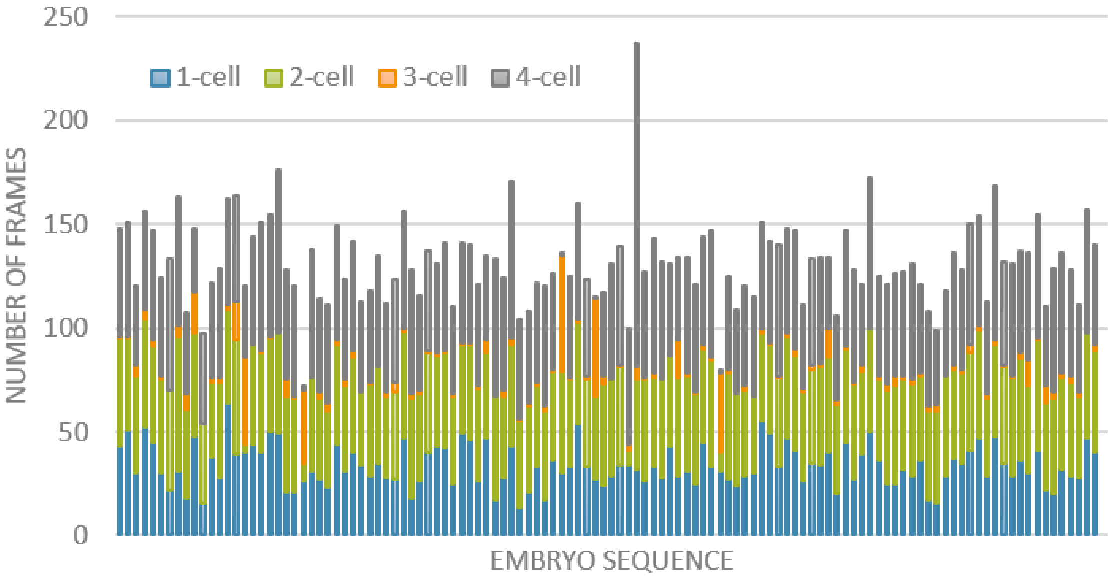


Figure 3.8: Cell count distribution for each embryo sequence. The number of frames at each cell stage varies greatly between sequences.

Cell Centroid Annotation

Unlike [56], no cell segmentation masks were available to generate cell centroid annotations. Since the number of time-lapse sequence images was much larger than the number of single frames in [56], only cell centroids were manually annotated to reduce annotation time. To annotate cell centroids, a image annotation tool (labelImg [42]) was modified for annotating landmarks in embryo time-lapse sequences. This python graphical user interface (GUI) uses PyQt5 library to assign image-level labels for classification or bounding boxes for object detection. Modifications for embryo cell centroid localization included:

- Changing bounding box annotations and outputs to a single point with centroid number classes
- Changing XML annotation file output format to point coordinates and cell count
- Automatically copying the previous frame's centroids when moving to the next frame (if unannotated)
- Adding gridlines to precisely locate cell centroids

The interface automatically saves x,y coordinates to XML file for each frame, automatically loads any previously saved landmarks from XML files, and enables moving forward and backward between frames with keyboard shortcuts.

All 36,035 embryo time-lapse sequence frames at 1-4 cell stage were annotated using this image annotation tool. Since cells have small amount of motion between most non-cleavage frames, copying centroid annotations from the previous frame when moving to the next frame greatly reduced annotation time.

Centroid coordinates were converted from the GUI XML output files to binary 2D masks, where centroid pixels were set to one. Since the number of background pixels vastly outweighed the number of foreground pixels, the binary centroid maps were converted to density maps to increase the ratio of foreground to background pixels during training. Otherwise, the network could get nearly perfect segmentation performance by the trivial solution of predicting every pixel as background. Convolutional filtering with a Gaussian kernel of $\sigma = 1.5$ was applied to the dot centroid maps and saved prior to training.

Setup

Training, validation, and test sets were established by randomly selecting 70%, 15%, and 15% of sequences, respectively, so that all frames in a sequence were contained in one set. 5-fold cross-validation was performed to reduce any bias from the selection of test set sequences. Dataset rolling and network training were repeated 5× for cross-validation. Results reported are the average across 5 test sets. Inputs and output ground truth masks were resized to 256×256 pixels. Previous mask inputs were initialised as zero matrices.

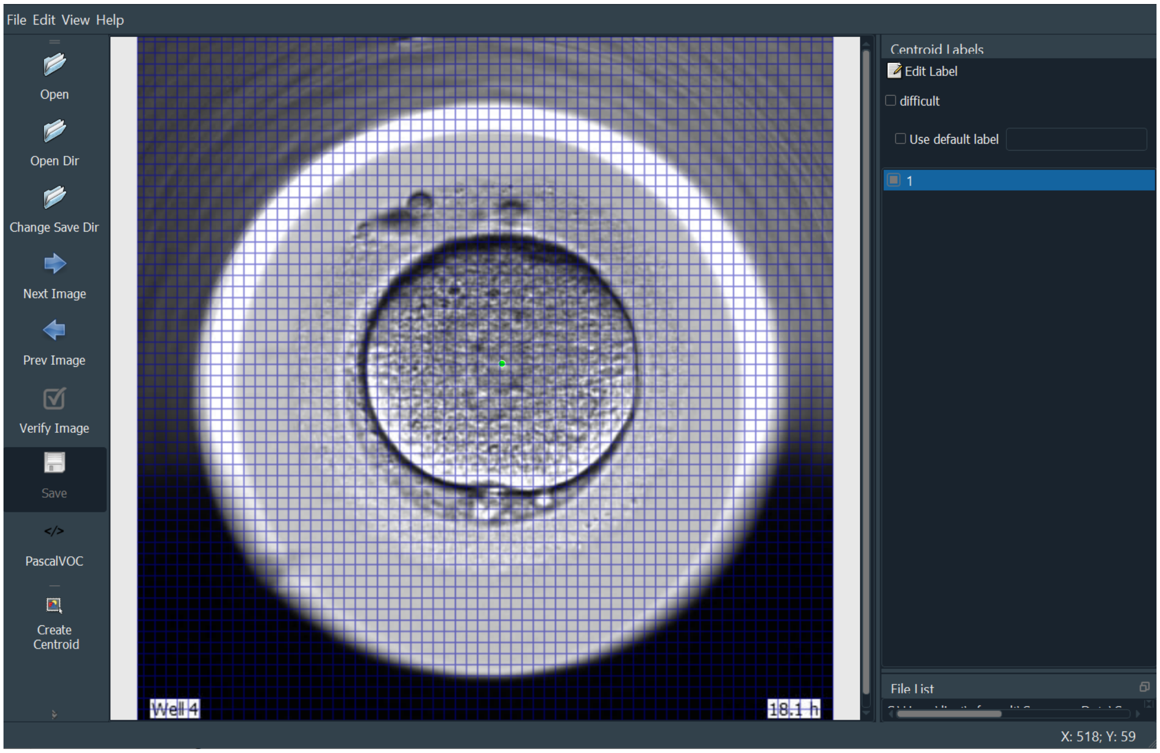


Figure 3.9: Cell centroid annotation GUI for label embryonic cell centroid pixels.

Cell centroid localization performance was first evaluated by measuring Euclidean distance (in pixels) to the nearest ground truth centroid on a per-cell basis. Lower distance error indicates predicted cell centroids are closer to their associated ground truth centroids.

Quantitative Results

Cell centroid distance error at each cell stage and sample-weighted average are presented in Table 3.1.

Table 3.1: Embryo cell centroid localization performance - distance error.

Model	Distance to nearest centroid (in pixels)				
	1-cell	***2-cell***	***3-cell***	***4-cell***	***Total***
U-Net [62]	2.88	4.25	4.72	4.43	4.24
Cell-Net [56]	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**

Cell centroid localization performance was also evaluated by measuring the rate of cell detection on a per-cell basis. Embryonic cell centroids were considered detected if they were < 5 pixels from the nearest ground truth centroid. A near miss or total miss was assigned if predicted centroids were ≥ 5 and < 8 pixels or $>= 8$ pixels, respectively, from the nearest ground truth centroid. Cell centroid detection rates at each cell stage and sample-weighted average are presented in Table 3.2.

Table 3.2: Embryo cell centroid localization performance - detection rate.

Model	Cell detection rate (in %)		
	Detection	***Near Miss***	***Total Miss***
U-Net [62]	80.0	11.7	8.3
Cell-Net [56]	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-Net, the previous state-of-the-art embryonic cell centroid localization and counting network, achieved lower scores than U-Net architecture in several categories. This was likely caused by having far more trainable parameters, and therefore overfitting to the training set. Empirically, training U-Net with unweighted MSE loss [80] could not produce Gaussian-shaped density maps, emphasising the importance of spatially-weighted loss function in [56].

Centroid localization distance error was smallest and cell counting accuracy was highest for all networks at the 1-cell stage, where samples had little or no cell overlap or fragmentation. Although 4-cell stage had the most samples, centroid localization performance was lower for some methods due to the complexity of cell orientations, cell overlap, fragmentation, and error propagated from previous frames.

Secondary network inputs reduced how far predicted cell centroids were from corresponding ground truth centroids. Adding previous frame predicted centroid masks (Multi-Input I) reduced centroid localization distance error at 1-, 2-, and 4-cell stages over baseline networks. Additionally incorporating flow diagrams (Multi-Input II) further improved distance error at 2-, 3-, and 4-cell stages. While Multi-Input I showed comparable cell detection rate to U-Net, the rate of correct detections increased and rate of total misses decreased with added cell movement context in Multi-Input II.

Visual Results

Qualitative results in Fig. 3.10 compare algorithm performance on samples with challenging textures and cell orientations. Overall, cell centroid distance to ground truth are closer in proposed methods, with fewer falsely identified cells. Since the network is predicting a Gaussian distribution for each cell centroid, predicted centroid regions smaller than the ground truth size indicate less confidence that a cell is present and those larger indicate less confidence in cell boundary.

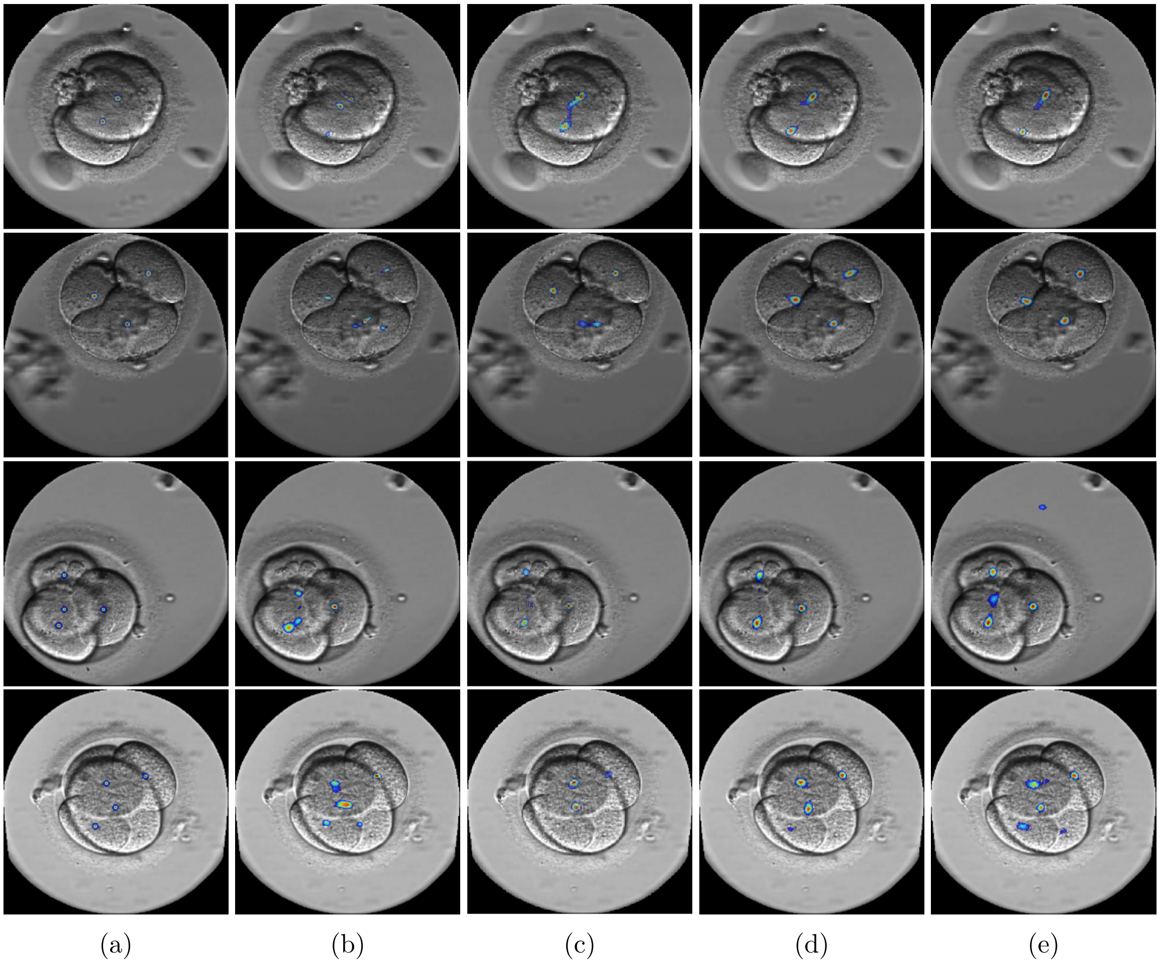


Figure 3.10: Qualitative cell centroid localization performance with ground truth (a), U-Net (b), Cell-Net (c) Multi-Input I (d), and Multi-Input II (e). Sequence frames are overlaid with predicted centroids, where red pixels indicate highest probability of cell centroid location.

Comparison to Related Work

The cell centroid localization methods in [56] and [80] were trained and tested on the dataset created for this thesis work. In blood cell counting tasks, there are many more cells in each input image, so the ratio of foreground to background pixels is more balanced. The U-Net proposed for cell centroid regression in [80] performed best at the 4-cell stage, when there was the highest amount of foreground pixels.

The embryonic cell centroid network in [56] was trained on a small dataset (176 images taken from traditional incubator) that lacked the many nearly identical frames between cleavages contained in time-lapse sequences. It was also trained on Gaussian-fitted segmentation mask annotations, which contain more information about the cell size and shape than Gaussian-filtered dot annotations used in this thesis. It therefore did not perform as well on the dataset used in this thesis.

The methods proposed herein aimed to address the challenges of training with time-lapse sequences while making use of the temporal relationship between consecutive frames.

3.4.2 Cell Counting

Setup

Since the cell centroid map predictions from cell centroid localization experiments were used for cell counting, the same dataset partitioning from Section 3.4.1 were used here. The cell count for each frame was determined from the sum of centroids in the annotation XML file.

Cell counting performance was measured with classification accuracy:

$$\begin{aligned} \textit{Cell Stage Acc.} &= \frac{TP_i + TN_i}{\sum_{s=1}^{S} N_s}, \quad i \in \{1,2,3,4\} \\ \textit{Total Acc.} &= \frac{\sum_{i=1}^{4} TP_i}{\sum_{s=1}^{S} N_s} \end{aligned} \tag{3.2}$$

where TP_i and TN_i are true positive and true negative predictions for cell stage i, respectively, S is the number of sequences in the test set, and N_s is the number of frames in sequence s.

Results

Cell counting accuracy is presented for each cell stage in Table 3.3. The total accuracy is averaged across all the samples in the test set.

The baseline U-Net tended to over-detect cells at 2- and 3-cell stage. U-Net therefore showed higher cell counting accuracy at the 4-cell stage, and lower accuracy at the 2- and 3-cell stage. Conversely, the baseline Cell-Net tended to under-detect cells at 3- and 4-cell stages. Cell-Net therefore showed higher cell counting accuracy at the 2-cell stage, and lower accuracy at the 3- and 4-cell stage. Multi-Input I had a milder tendency to under-detect

Table 3.3: Embryo cell counting performance per cell stage.

Model	Cell Stage Prediction Accuracy (in %)				
	1-cell	***2-cell***	***3-cell***	***4-cell***	***Total±SD***
U-Net [62]	92.8	67.4	61.6	**78.4**	77.7±8.6
Cell-Net [56]	96.2	**81.8**	67.5	62.3	77.5±6.8
Multi-Input I (Proposed)	**97.7**	78.8	**69.2**	68.6	79.3±6.6
Multi-Input II (Proposed)	95.7	74.7	69.0	75.8	**80.2**±3.5

cells at 4-cell stage, but scored the highest counting accuracy at 1- and 3-cell stages. While Multi-Input II did not score the highest counting accuracy at any single cell stage, its balanced performance across all stages led to highest overall accuracy.

Lower cell counting results with Cell-Net were likely due to overfitting to certain cell orientations during training. It is easier for a network with a greater number of parameters to fit better to nearly repeated frames where there is little cell movement, thereby reducing generalization to new sequences during testing. The standard deviation across folds was reduced with Multi-Input II.

Effect of Artifacts

The presence of artifacts had a significant effect on performance. The actual 4-cell stage frames that were mis-counted as an earlier cell stage by Multi-Input I were investigated in more detail. As seen in Fig. 3.11, there were a small number of sequences that contributed to the majority of incorrectly counted cells. This is an indication that the network was sensitive to image quality since sequences with artifacts had many incorrect cell counts throughout 4-cell stage.

Comparison to Related Work

Early automated embryonic cell counting was performed by blastomere detection using ellipse-fitting approaches from cell boundary annotations in [71] and [33]. The method in [71] was evaluated on only 40 images from a single source. The method in [33] was evaluated on 468 images from three sources. Their time-lapse sequence dataset portion was restricted to no more than one image per cell stage per sequence (thereby omitting multiple difficult samples occurring in the same sequence). However, they reported an average overall quality of 83% for 1-8 cell stage, higher than the overall accuracy of 80.2% for 1-4 cell stage achieved in this thesis. This shows the value of using more time-consuming cell boundary annotations over cell centroid annotations.

Classification-based cell counting approaches often show very weak performance for 3-cell stage (e.g. 55.15% accuracy in [44] and 23.91% accuracy in [31]). The small amount

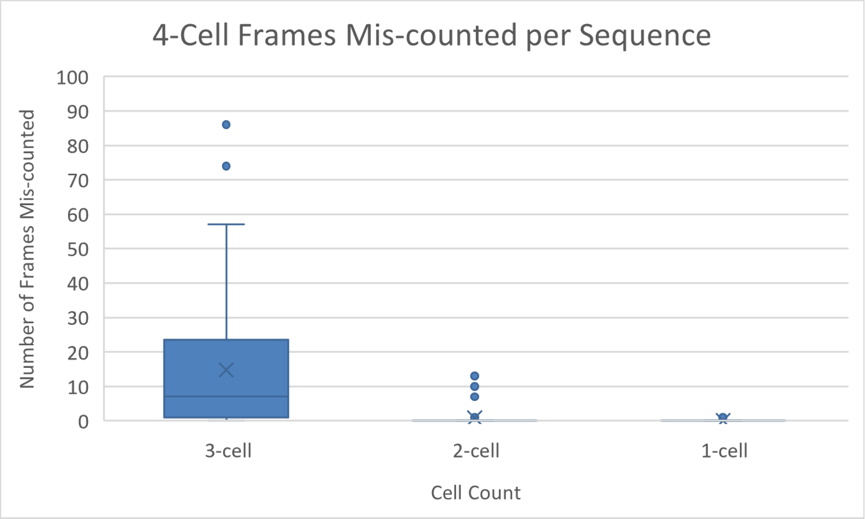


Figure 3.11: Summary of 4-cell stage frames mis-counted per sequence. The median number of frames incorrectly classified per sequence (indicated by horizontal line) was 7 frames while the mean (indicated by an x) was 14.59 frames. A small amount of sequences had a high number of mis-counted 4-cell frames.

of training samples for 3-cell stage (3.4% in this dataset) makes it extremely difficult for a classifier to properly learn discernible features unique to that stage. Since the centroid localization approach learns to detect individual cells, the performance for 3-cell stage counting (69.0% accuracy) was much higher compared to these classification approaches.

3.5 Conclusions

A structured regression network was trained to localize embryonic cell centroids in time-lapse sequences. Imbalance between the number of foreground and background pixels was addressed using the temporal relationship of cell location and movement between consecutive sequence frames. A proposed training strategy incorporated predicted cell centroid(s) at the previous frame and expected cell movement from optical flow diagrams. Cell centroid localization performance were improved by using spatio-temporal network attention for each frame. Image post-processing applied to cell centroid prediction masks extracted cell count for each frame.

This algorithm can be used to automatically predict the location of cells and cell count for each frame in embryo time-lapse sequences up to 4-cell stage. The frame at which embryos reach 2-4 cell stage and duration in frames embryos spend at 1-3 stages can be computed from the cell count predictions. Using the time-stamp on each frame, the stage onset and duration can be expressed in hours post-fertilization and compared to normal ranges for embryo development in vitro. Furthermore, predicted cell centroid locations in each time-lapse sequence frame could be used for cell tracking.

Chapter 4

Embryo Stage Classification and Onset Detection

4.1 Problem Description

Knowing at what time (hours post-fertilization) embryos reach morula and blastocyst stage gives insight into their development progress and implantation potential. Time-lapse imaging systems built into incubators enable determination morula and blastocyst onset by visual assessment to the nearest 15-minute increment. Visually assessing morula or blastocyst stage onset is prone to subjectivity since the changes between morula and blastocyst stage embryos near their onset are subtle. The goal of this project is to develop an automated algorithm for detecting morula and blastocyst stage onset. The algorithm must be sensitive enough to detect the exact onset frame while also generalizing to new sequences with different biological variability not seen during training.

4.2 Related Work

Embryo staging of embryonic cleavage times and morula through expanded blastocyst stage was first performed using traditional image processing techniques [19]. A deep learning approach was later proposed employing a fine-tuned CNN classifier took as input multiple focal planes of each time-lapse imaging frame to classify embryos as 1-9 cell, morula, or blastocyst stage [41].

CNN-based embryo stage and embryonic cell stage classification methods analyze image frames individually then perform post-processing by considering all frames in an embryo sequence. Dynamic programming used in [41], [44] is a family of methods that iterates through each sequence frame, restructuring windows of sequential predictions around each frame. Another strategy is numerical optimization for the entire sequence with negative label likelihood (NLL) loss [40], [43], [50] or mean absolute error (MAE) [44] of raw predictions or Earth Mover's Distance (EMD) of predicted stage histograms [40], [43], [50].

Medical image classification tasks facing intra-class variation and inter-class similarity have been improved using synergic loss. Facial expression recognition utilized pairwise learning between with synergic loss two identical models in [87]. A similar pairwise learning approach with synergic loss was used for brain MR and skin lesion classification in [86]. Given challenging variation between embryo stages (especially near stage onsets) synergic loss is explored in this work for embryo stage classification.

Embryo time-lapse sequences have also been analysed to perform quality scoring of blastocysts [37], [7] and to predict likelihood of implantation [7]. In these works, spatial features are extracted with CNNs then fed to a recurrent neural network (RNN) that incorporates temporal context [37] or genetic algorithm [7] to aggregate sequence information for predicting a single quality score per sequence. While the output of the network in [37] is a single score for each sequence, an RNN is adapted in this work to incorporate temporal context learning for stage classification predictions at every frame.

While image classification networks have shown to be successful at embryo early cell stage classification, there are biological challenges that reduce the efficacy of these methods for later morula and blastocyst stage onset detection. The differences between morula and blastocyst stage embryos near their onset are more subtle than the presence of a new cell. Also, development through morula and blastocyst stages is not monotonic non-decreasing like cell stage development. Cell compaction increases and decreases repeatedly through morula stage, as shown in Fig. 4.1. Blastocyst stage embryos can contract, as shown in Fig. 4.2. These challenges are explicitly addressed in this work using synergic loss and temporal learning.

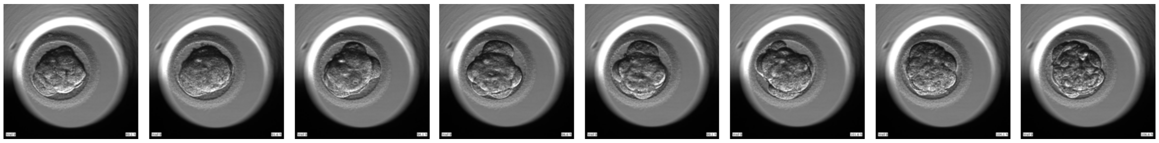


Figure 4.1: Progression of an embryo through morula stage. The adhesion or compaction of cells increases and decreases, changing the texture of the cell mass.

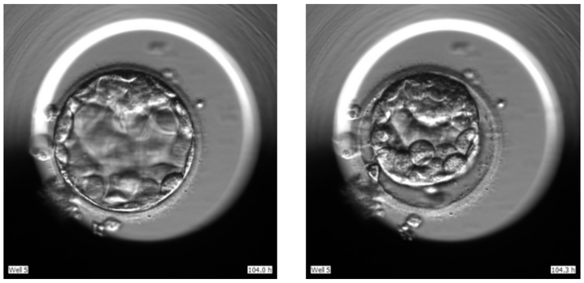


Figure 4.2: Blastocyst contraction event shown in two consecutive sequence frames. The PVS (between inner ZP and outer TE) increases drastically after contraction.

4.3 Proposed Methods

For embryo stage classification, image pre-processing techniques are used to isolate the embryo in time-lapse sequence frames. An image classification network is trained with synergic loss and temporal learning to assign each frame as either cleavage, morula, or blastocyst stage. The predicted stage classifications across each sequence are restructured to be monotonic non-decreasing. Morula and blastocyst stage onsets are inferred from the minimum index where each of these stages occurred.

4.3.1 Image Pre-processing

Although the same time-lapse sequences from cell centroid localization were used for embryo staging, the image pre-processing technique from cell centroid localization failed for later development stages when the embryo took up more of the embryo well area (see Fig. 4.3). For most experiments, the only image pre-processing step was removing (i.e. setting to zero) pixels corresponding to embryo well and time stamp located along the bottom rows of the image. These 1-channel images were 500 × 500 pixels before being resized for the network input.

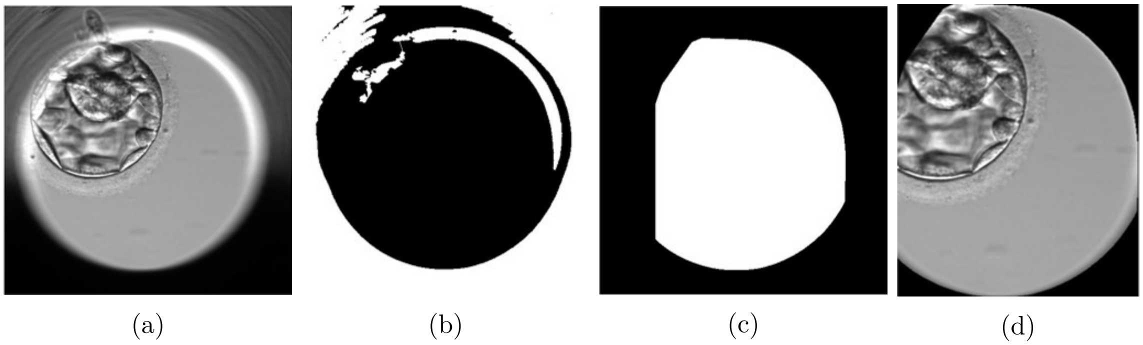


Figure 4.3: Image pre-processing from cell centroid localization failure on blastocyst image: original image with well number and time stamp removed (a), largest bright and dark components with upper well bright region not fully captured (b), embryo area mask (c), and final centred image with part of blastocyst cut-off (d).

For the experiments employing pre-processed images, a circle was cropped around the centered embryo, as shown in Fig. 4.4. The circle centroid was computed as the centroid of dilated Canny-detected edges from the inner embryo well, estimated by empirical spatial constraint. The circle radius was set to be the same across each sequence since embryo size changes throughout development and this feature can help classify development stage. Circle radius was set as 190 pixels. The resulting image size of 380 × 380 pixels was large enough to always contain the largest expanded blastocysts.

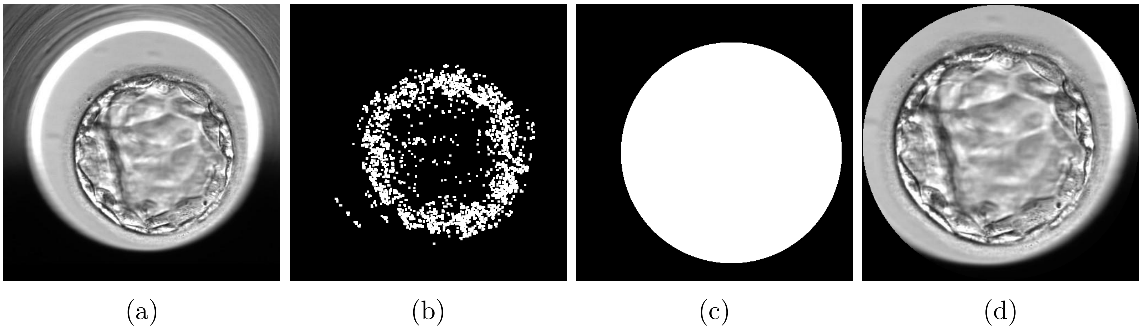

(a) (b) (c) (d)

Figure 4.4: Image pre-processing steps for embryo staging: original image with well number and time stamp removed (a), inner embryo well dilated edges (b), embryo area mask (c), and final centred image (d).

4.3.2 Classification Labels

Embryo stage onsets were manually annotated in time (hours post-fertilization). The frames associated with stage onsets were determined by cropping the image to the time stamp box in the bottom right corner of each sequence frame, as shown in Fig. 4.5. Characters were read from the time stamp crop using MATLAB's built-in optical character recognition (OCR) function. The "h" was disregarded and the remaining 4 or 5 characters were concatenated to form a number in the range of tens to hundreds with one decimal place. Whenever the annotation time did not exactly correspond to a sequence frame, the onset frame for that stage was set as the nearest frame following the annotation time.

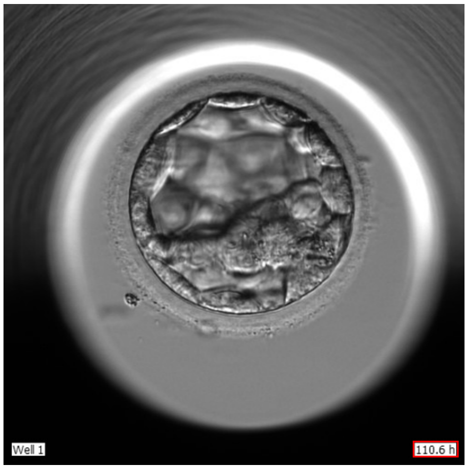


Figure 4.5: Sequence frame with time stamp shown inside the red box. The time stamp was present at the same coordinates for every frame in the dataset.

All frames following the morula and blastocyst stage onset were labeled as morula and blastocyst stage for classification, respectively. This approach requires minimal expert labeling, but is susceptible to noisy labels when cell adhesion (attachment) decreases or blastocysts contract.

4.3.3 Development Stage Classification

Timing of development stage onset is a structured regression problem, solving it using deep learning techniques is highly unsuitable due to inadequate number of embryo sequences available. However, image classification algorithms have been widely studied and shown great success in recent years. Predicting development stage onset was therefore performed by first classifying sequence frames into development stage categories then applying signal processing techniques to infer stage onset timing.

The baseline network is a pre-trained VGG16 [69] model convolutional feature extractor with only the last convolutional layer trainable during fine-tuning. This is topped with a 32-node ReLU-activated fully-connected layer and a softmax-activated output layer. This network architecture was chosen to reduce overfitting to the training set sequences. The full dataset (45,209 frames) is much smaller than the 14 million ImageNet database images used to pre-train VGG16. Since there are only 117 uniques embryos in the dataset with minor differences between output classes (stages), a CNN of VGG16's size would easily extract features that are too specific to for classifying the training set sequences without generalizing to new sequences. Using a pre-trained model enables the network to train with a much small number of images and hyperparameter search with less risk of overfitting.

Larger architectures (e.g. ResNet50) have the advantage of increased field of view, utilising more embryo area for classification. However, they are more prone to overfitting, which was a concern in this task. These newer architectures also use batch normalization [29] in the initial training - a technique that regularizes convolutional kernel weights based on the kernel's mean and variance for each batch. Microscopic embryo images share very different batch statistics than natural ImageNet images. The outputs of every convolutional layer could be less relevant since the inputs are out of initial training distribution.

4.3.4 Synergic Network

There is often more variation between frames at the same stage in different sequences than frames at different stages in the same sequence. To encourage the network to learn embryo-independent features that are similar between stages, pairwise learning was implemented using two identical baseline networks with unshared weights [86]. An additional mini-network concatenated nodes from the 2$^{\text{nd}}$ last (fully-connected) layer of each branch. This tensor was fed through a 32-node fully-connected layer followed by a single-node sigmoid-activated output layer. For inference, predictions from the two network branches were averaged. Synergic loss penalized incorrect prediction of whether the images fed to each branch were at the same stage using binary cross-entropy between the actual (y_s) and predicted ($\hat{y}_s$) stage similarity.

$$\mathcal{L}(y_s, \hat{y}_s) = -y_s log(\hat{y}_s) + (1 - y_s) log(1 - \hat{y}_s) \tag{4.1}$$

Despite the relatively small number of trainable parameters, the network continued to overfit to the training set. Since embryos show little change before and after stage onset, many image augmentations could change the stage label. To challenge the network during training without adding unwanted label noise, mixup augmentation [84] was used to blend image samples and smooth the corresponding ground truth labels by the same factor. A portion of images and labels from one synergic network branch were blended with a random selection from those fed to the other branch.

4.3.5 Temporal Learning

As mentioned, an embryo's appearance while progressing through morula and blastocyst stages can revert to that of a previous stage. Instead of sampling random images in each batch, images were analyzed in sequences to incorporate short-range temporal dependency during training. An LSTM layer was added after each synergic CNN branch output layer. Classification loss (categorical cross-entropy) was measured between the actual (y_t) and predicted ($\hat{y}_y$) stages and backpropagated both before and after the LSTM layer for each frame.

$$\mathcal{L}(y_t, \hat{y}_t) = -y_t log(\hat{y}_t) \tag{4.2}$$

The convolutional feature extractors, stage classifiers, and LSTM layers were trained together in an end-to-end manner. The starting index (between zero and batch size) for each full embryo sequence was randomly chosen every epoch to sample different batches. The final network diagram in Fig. 4.6 shows how embryo stage is predicted for each image individually with the CNN-based classifier and is then refined using temporal context in the LSTM layer.

Since each stage onset was only sampled once per sequence, only 2 batches per sequence contained stage transitions. To train on more complex sequences containing transitions, extra sequence batches near stage onsets were added. This stage onset oversampling added up to 4 extra batches per stage onset or 8 extra batches per embryo sequence. 4 batches cannot always be sampled before reaching the end of sequence for blastocyst stage, which can have as few as one frame at that stage.

For onset oversampling, the 1$^{\text{st}}$ extra sequence batch starting SI index was randomly chosen as $SI_s \in [24, 32)$ frames before stage onset SO_s, for stage s. This index was increased by 8 frames up to three times, creating up to 4 batches containing the stage transition. The extra four batches have indices $[SO_s - SI_s + 8 \times 0, SO_s - SI_s + 8 \times 0 + 32)$, $[SO_s - SI_s + 8 \times 1, SO_s - SI_s + 8 \times 1 + 32)$, $[SO_s - SI_s + 8 \times 2, SO_s - SI_s + 8 \times 2 + 32)$, $[SO_s - SI_s + 8 \times 3, SO_s - SI_s + 8 \times 3 + 32)$. An example is shown in Fig. 4.7.

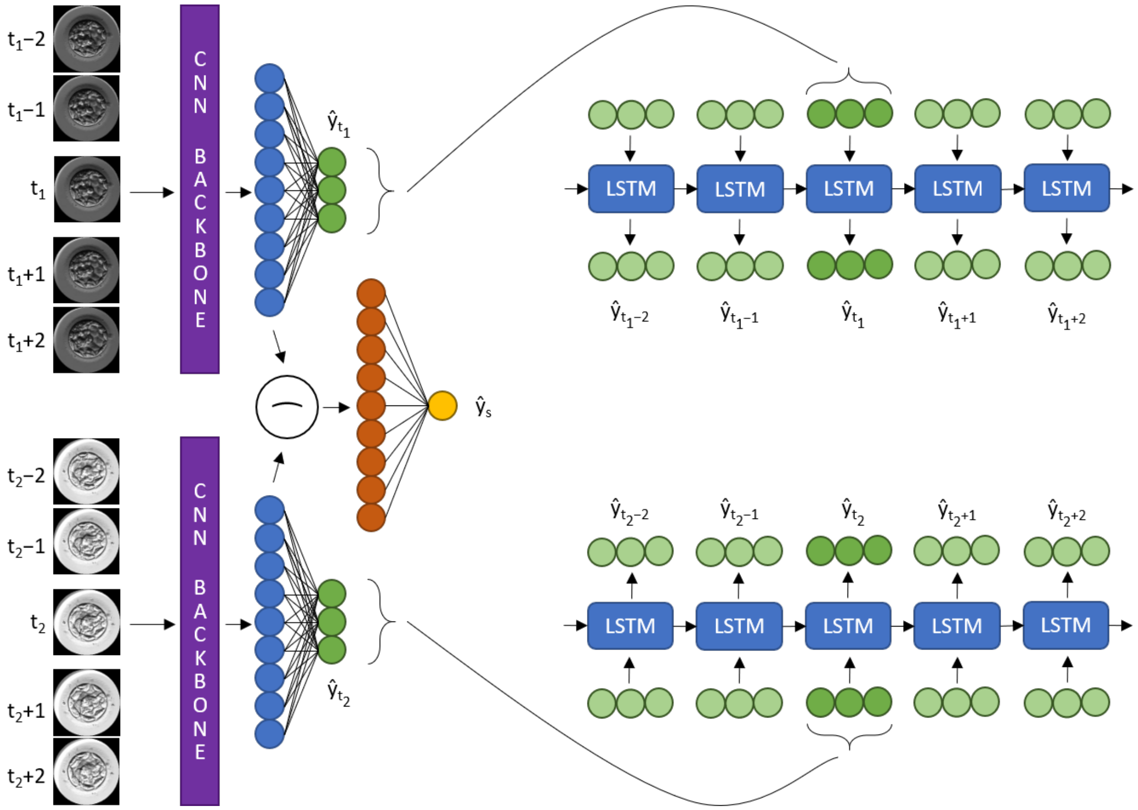
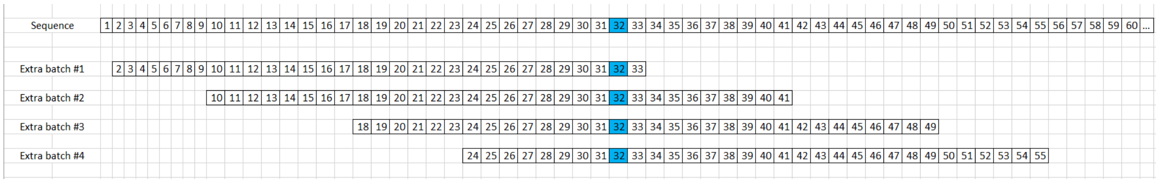


Figure 4.6: Proposed network diagram for embryo staging. Two image sequence batches are fed in parallel through separate convolutional feature extractors and then classified into stages. Staging predictions are refined with an LSTM. Classification error in Eq. (4.2) is computed at both input and output to LSTM layer. Fully-connected layers from each classifier are concatenated and used to predict whether the input image fed through each branch belong to the same stage. Synergic loss in Eq. (4.1) from this binary output is backpropagated through both classifier branches.



Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Extra batch #1		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Extra batch #2										10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Extra batch #3																		18	19	20	21	22	23	24	25	26	27	28	29	30
Extra batch #4																								24	25	26	27	28	29	30

Sequence	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	...
Extra batch #1	31	32	33																												
Extra batch #2	31	32	33	34	35	36	37	38	39	40	41																				
Extra batch #3	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49												
Extra batch #4	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55						

Figure 4.7: Example of onset oversampling with stage onset $SO_s = 32$ (blue) and extra sequence starting index $SI_s = 30$. The sequence frame indices for the 4 extra batches are shown.

4.3.6 Network Training Details

Images were resized to 320×320 pixels. Standard images augmentations were used including horizontal and vertical flipping, rotation up to 360°, horizontal and vertical translation by up to 10% of the image height/width. Network training used Adam optimizer with initial learning rate chosen empirically as 1×10^{-4} for networks containing LSTM layers and 3×10^{-5} otherwise. The learning rate was reduced on plateau of 8 epochs by a factor of 0.9 and early stopping was applied after validation loss had no longer decreased for 15 epochs.

Image and label blending with mixup augmentation used scaling factor sampled from beta distribution with $\alpha = 0.2$ for each instance. A batch consisted of two 32-frame image sequences (one for each synergic branch) for LSTM networks or 32 random frames otherwise. Image sequences for LSTM network batches were sampled by iterating through embryo sequences with stride 32. When embryo sequence length was not divisible by 32, another batch was sampled containing the final $(N - 32, N]$ frames of that sequence, where N is the number of sequence frames.

4.3.7 Embryo Stage Onset Detection

Stage onset is retrieved from the minimum sequence index at which that stage was predicted. Embryo sequences have monotonic non-decreasing development stages, though this property is not guaranteed in predictions without temporal post-processing. Predicted time-lapse sequence stage predictions can oscillate between stages, as shown in Fig. 4.8. The long-range (sequence-wide) temporal dependency of staging predictions is addressed by restructuring predicted stage labels across sequences.

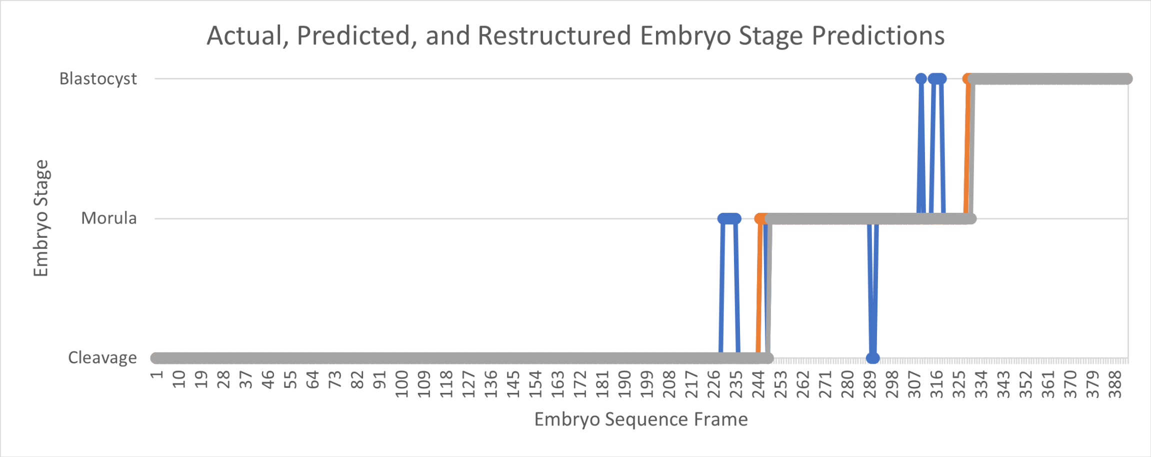


Figure 4.8: Actual (gray), predicted (blue), and restructured (orange) embryo stage predictions for a test sequence. Morula and blastocyst stage onset error were reduced from 19 to 4 frames and 21 to 2 frames, respectively, for this sequence using MAE minimization.

Stage predictions for each embryo sequence were optimized by minimizing error between unprocessed stage predictions and all possible series of monotonic non-decreasing predictions. NLL loss as used in [43,50] and MAE (Global) as used in [44] were implemented. For both optimizations, all possible sets of monotonic non-decreasing predictions are considered and that which gives the lowest (NLL or MAE) loss is chosen.

4.4 Experiments

Embryo time-lapse sequence frames were pre-processed using Python 3.6 scikit-image library. Deep neural networks were trained using Keras 2.2.4 framework with TensorFlow 1.11 backend.

4.4.1 Embryo Stage Classification

Dataset

Experiments were performed on 117 human embryo time-lapse imaging sequences collected at the PCRM from 2017-2019. Frames were acquired every 15 minutes, capturing embryo development from zygote stage (approx. 18 hours post-fertilization) to blastocyst stage (approx. 5 days post-fertilization). The morula and blastocyst stage onset times were annotated by an embryologist at PCRM. The overall image and sequence distributions for embryo staging are shown in Fig. 4.9 and Fig. 4.10, and overview of stage onset and duration is summarized in Table 4.1.

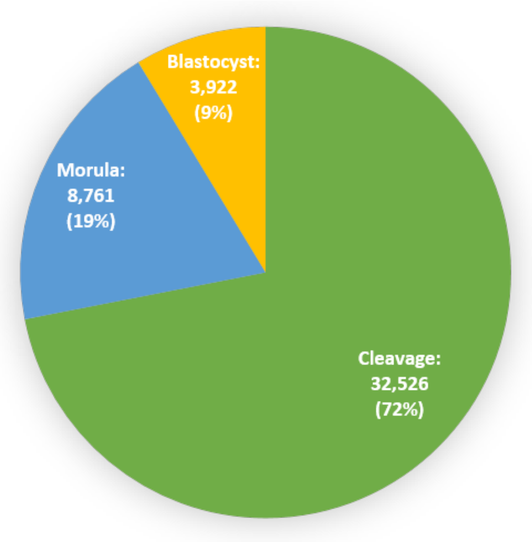


Figure 4.9: Embryo stage distribution across the dataset. Cleavage stage frames make up the majority of samples while morula and blastocyst stage have fewer samples.

Setup

Embryo sequences were randomly partitioned into training, validation, and test sets using ratio 70/15/15%, respectively. To reduce performance bias in chosen training/test splits, 5-

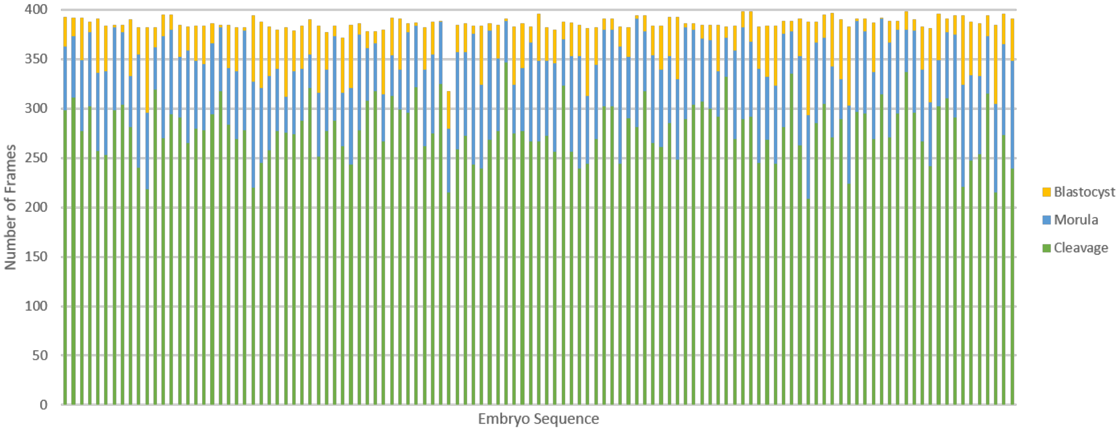


Figure 4.10: Embryo stage distribution for each sequence. The number of frames at blastocyst stage varies greatly between sequences.

Table 4.1: Dataset overview of embryo stage onset and duration.

	Stage Onset Timing (frame)		
Stage	**Min.**	**Max.**	**Average ± SD**
Cleavage	-	-	-
Morula	210	348	279.00 ± 28.82
Blastocyst	281	392	353.88 ± 24.24
	Stage Duration (frames)		
Stage	**Min.**	**Max.**	**Average ± SD**
Cleavage	209	347	278.00 ± 28.82
Morula	34	133	74.88 ± 19.71
Blastocyst	1	95	33.52 ± 21.89

fold cross-validation was performed by rolling the sequences in each set. Results presented were averaged across the 5 folds.

Embryo stage classification performance is presented for the baseline VGG16 with 3 final convolution layers fine-tuned and fully-connected classifier. These results are compared with adding of synergic learning, mixup augmentation, LSTM layer, onset oversampling, and image pre-processing. Classification performance was measured with precision and recall in Eq. (4.3), and F1-Score in Eq. (4.4), where TP_i, FP_i, and FN_i are the numbers of true positive, false positive, and false negative predictions, respectively, belonging to class i.

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \qquad \text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{4.3}$$

$$\text{F1-Score}_i = \frac{TP_i}{TP_i + \frac{1}{2}(FP_i + FN_i)} \tag{4.4}$$

Results

The quantitative results of embryo stage classification for each stage and network or training strategy modification are summarized in Table 4.2.

Table 4.2: Embryo stage classification performance.

Syn. Loss	**mixup Aug.**	**LSTM**	**Onset Overspl.**	**Image Pre-proc.**	**Stage**	**Prec.**	**Rec.**	**F1-Score ±SD**
Baseline					Cleavage	96.91	95.63	96.27±0.4
					Morula	81.17	85.37	83.22±0.8
					Blastocyst	92.05	91.38	91.71±1.7
✓					Cleavage	96.78	96.12	96.45±0.3
					Morula	83.13	84.34	83.73±1.6
					Blastocyst	91.32	**93.43**	92.36±1.8
✓	✓				Cleavage	97.15	95.89	96.52±0.4
					Morula	82.48	86.63	84.50±0.6
					Blastocyst	92.86	92.37	92.62±1.6
✓	✓	✓			Cleavage	97.01	**98.13**	**97.57**±0.6
					Morula	**87.85**	86.79	87.31±2.0
					Blastocyst	**94.86**	88.41	91.52±4.0
✓	✓	✓	✓		Cleavage	97.72	97.03	97.37±0.4
					Morula	85.68	**89.57**	87.58±1.1
					Blastocyst	94.58	90.46	92.47±2.1
✓	✓	✓	✓	✓	Cleavage	**97.85**	97.28	**97.57**±0.6
					Morula	87.12	88.85	**87.97**±1.8
					Blastocyst	92.56	92.90	**92.73**±1.4

The baseline network was particularly susceptible to incorrectly classifying embryos as morula stage and mis-classified many blastocyst stage embryos as morula stage.

Synergic learning reduced the number of embryos incorrectly classified as morula stage and increased the number of embryos correctly classified as blastocyst. Averaging predictions from the two synergic network branches slightly improved the results over each branch individually. Adding mixup augmentation increased the number of embryos correctly classified as morula stage.

Incorporating temporal learning with an LSTM layer significantly improved morula stage classification, achieving the fewest falsely predicted morula stage embryo classifications. However, the highest amount of blastocyst stage classification were incorrectly missed. Although short-range temporal classification consistency was improved, morula stage misclassifications often continued far before or after the onset frame. Adding stage onset oversampling correctly predicted more embryos as blastocyst stage. This sampling strategy compensated for the LSTM performance drop at the blastocyst stage. While correctly predicting the highest number of embryos as morula stage, this included more incorrectly predicted morula stage embryos.

Image pre-processing ensured maximal embryo area was visible as input to the network, reducing pixel information loss during image resizing. The subtle cell adhesion (texture) differences could be better extracted and used for predicting morula stage. The F1-score SD was lowest for synergic loss with mixup and highest when LSTM was added.

Choice of Architecture

Convolutional feature encoders for embryo stage classification were investigated by training different models on a single dataset fold. VGG16 and ResNet50 architectures with various trainable layers were compared. Staging F1-scores for each stage are shown in Table 4.3.

The experiment shows that pre-trained features in ResNet50 did not transfer as well as VGG16 for embryo image staging. Only when the all the ResNet50 layers (including batch normalization) were re-computed could it make use of convolutional features. Since the networks were trained on a single NVIDIA GeForce GTX 1080 Ti GPU with 11 GB memory, the fully trainable ResNet50 could only be trained with batch size of 8. The GPU memory constraints limited the use of synergic or temporal learning for larger architectures.

Setting all layers as trainable for fine-tuning gave the worst VGG16 performance as the network could overfit heavily to the training set. Freezing all layers did not allow the discriminative final convolution layers to adapt to embryo images. Having the final 3 layers trainable enabled the network to learn features relevant for embryo stage classification without overfitting severely to the training data. VGG16 with final 3 layers trainable was used for all embryo stage classification experiments.

Table 4.3: Embryo stage classification performance on single fold for architecture selection.

Model	Layers Trainable	Stage	F1-Score
VGG16	All	Cleavage	82.12
		Morula	0.0
		Blastocyst	0.0
VGG16	Final 3	Cleavage	96.70
		Morula	84.20
		Blastocyst	91.70
VGG16	None	Cleavage	92.65
		Morula	70.61
		Blastocyst	90.24
ResNet50	Final 5	Cleavage	83.30
		Morula	0.0
		Blastocyst	0.28
ResNet50	None	Cleavage	96.20
		Morula	77.10
		Blastocyst	88.32

Weak Blastocoel Segmentation

Segmenting the blastocoel is challenging due to weaker boundaries with surrounding structures, varying texture from structures underneath, and looser geometric characteristics than the ZP. Therefore previous methods for detecting blastocyst stage onset do so by segmenting the ZP and using its width for blastocyst stage classification [19]. Embryologists rely on ZP thickness and blastocoel volume for assessing blastocyst onset.

Class activation maps were used to visualize which regions in the image were most relevant for correctly classified blastocyst stage images. As seen in Fig. 4.11, the network relied mainly on blastocoel region for blastocyst stage prediction. This indicates the blastocoel could be a better structure for predicting blastocyst stage onset than ZP for image classification networks. This also demonstrates blastocyst stage classification requiring only image-level annotations could be used for weakly-supervised blastocoel segmentation.

Comparison to Related Work

In [41], a large CNN backbone (ResNeXt101) taking as input 3 focal planes for each frame was proposed for embryo stage classification. Their overall stage classification accuracy (for 13 classes after prediction restructuring) was reported as 87.9%. They also reported human labelling development stage classification accuracy (by comparing the ground truths from 2 annotators) was 94.6%. Due to the overfitting issues observed during training with larger ResNet50 classification backbone, training strategy improvements with a smaller VGG16 backbone were investigated to improve morula and blastocyst stage classification. In this

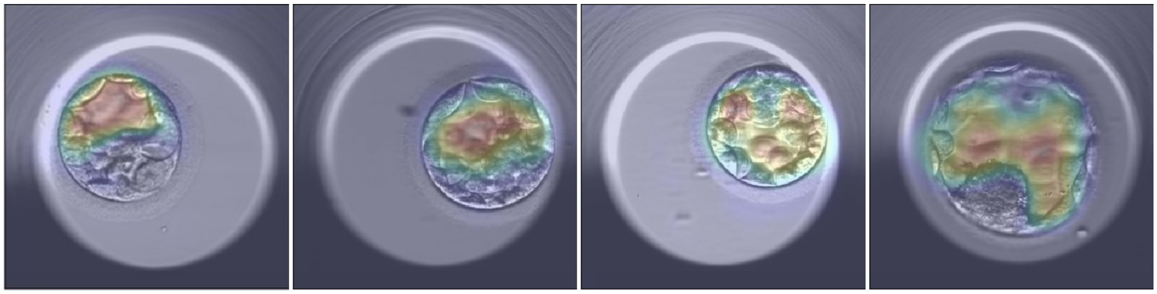


Figure 4.11: Class activation maps overlaid on embryo time-lapse images correctly classified as blastocyst stage. Red regions show that blastocoel region structure contributed the most for stage prediction.

thesis work, the overall stage classification accuracy (for 3 classes after global prediction restructuring and single focal plane) was 95.9%.

4.4.2 Embryo Stage Onset Detection

Since the raw embryo stage classification predictions from embryo stage classification experiments were used for stage onset detection, the same dataset and test set partitioning from Section 4.4.1 were used here.

Setup

Stage onset performance was measured as the mean absolute error in frames between predicted and actual stage onset in Eq. (4.5), where M is the number of embryo sequences in the test set, n is the frame number in an embryo sequence, and y_{i,n_m} and $\hat{y}_{i,n_m}$ are the true and predicted stage i cell counts for sequence m at frame n.

$$\text{MAE}_i = \frac{1}{M}\sum_{m=1}^{M} |\min_{n_m} y_{i,n_m} - \min_{n_m} \hat{y}_{i,n_m}| \tag{4.5}$$

Results

Morula and blastocyst stage onset detection results on unprocessed sequences, and NLL loss and Global MAE optimized sequences are summarized in Table 4.4.

Unprocessed sequences contained many incorrect morula stage predictions in frames far from the actual onset. Both temporal post-processing strategies significantly improved morula stage onset. Global MAE occasionally optimized to slightly better stage onsets, though both losses gave nearly identical solutions.

Minimizing NLL loss and Global MAE for prediction restructuring is optimal only when the number of mis-classifications before and after the actual onset are relatively balanced and therefore cancel out. In the experiments, incorrect stage predictions were often biased such that the majority of mis-classifications occurred before or after the actual onset.

Table 4.4: Mean absolute stage onset error with restructured predictions.

Syn. Loss	**mixup Aug.**	**LSTM**	**Onset Overspl.**	**Image Pre-proc.**	**Stage**	**Unproc.**	**NLL**	**MAE**
Baseline					Morula	46.32	13.96	13.64
					Blastocyst	5.21	5.17	5.17
✓					Morula	39.86	12.37	12.18
					Blastocyst	5.76	4.77	4.77
✓	✓				Morula	41.51	12.60	12.50
					Blastocyst	4.76	**4.32**	**4.32**
✓	✓	✓			Morula	20.23	11.98	11.98
					Blastocyst	5.67	5.12	5.12
✓	✓	✓	✓		Morula	24.40	11.94	11.94
					Blastocyst	5.04	4.47	4.47
✓	✓	✓	✓	✓	Morula	20.16	**11.04**	**11.04**
					Blastocyst	5.26	4.77	4.72

Higher classification performance therefore did not guarantee lower stage onset error after restructuring.

Comparison to Related Work

In [19], image processing techniques were used to detect embryo stage onsets. They reported stage onset error as a median of times between predicted and actual onset. However, they did not take the absolute value so both positive and negative errors were computed in the mean. They reported morula and blastocyst stage onset median error of -0.38 and -0.91 hours, respectively. In this thesis, the morula and blastocyst stage onset median errors are -1 frame (0.25 hours) and 0 frames (0 hours), respectively. This metric fails to capture the magnitude of stage onset errors, but shows the proposed algorithm is not biased towards predicting morula or blastocyst onset early or late.

4.5 Conclusions

An image classification network was trained for classifying embryo time-lapse sequence frames into cleavage, morula, and blastocyst stages. Embryo stage classification predictions were improved by incorporating synergic learning to directly compare features of pairs of input images. Adding a temporal LSTM layer after the classifier output provided short-range temporal dependency to improve morula stage classification and downstream stage onset performance. Extra image and sequence augmentations during training further improved each method by reducing generalization gap on the test set. Sequence post-processing was applied to restructure embryo stage classifications to be monotonic non-decreasing then stage onset was extracted from the restructured sequence predictions.

This algorithm can be used to automatically predict the frame at which embryos reach morula and blastocyst stage onset from time-lapse sequences. Using the time-stamp on each frame, the stage onset could be expressed in hours post-fertilization and compared to normal ranges for embryo development in vitro.

Chapter 5

Conclusions

For couples suffering from infertility issues, IVF treatment is a way to conceive their own child. The success rate remains low, and the financial and emotional cost is high. During this procedure, embryos are inspected visually to inspect quality of embryo structures and their timings to reach important cell and development stages. These morphological and morphokinetic parameters are used to assess their quality and predict their likelihood of leading to positive implantation outcome. Developing robust and objective automated embryo quality assessment tools could help embryologists select the highest quality embryo for transfer.

In this thesis, three methods were proposed to automate embryo quality assessment:

1. A multi-label multi-class CNN was developed to automatically assign ICM, TE and BE grades to blastocyst images simultaneously.
2. A structured regression CNN was used to localize cell centroids for counting cell stage in early embryo time-lapse sequences.
3. An image classification CNN with synergic loss and temporal learning was trained to classify embryo stage and detect morula and blastocyst stage onset in embryo time-lapse sequences.

Previous work for blastocyst grading simplified quality scores into two classes or required multiple networks for different grades. The proposed approach combined three grading tasks into a single network, leveraging a shared convolutional encoder to extract image features relevant to all grades. Since the experiments were performed using a small dataset from a single fertility clinic, the algorithm suffered from lack of available samples from lower quality ICM and TE grades. It could be made more robust by training on a larger dataset with more balanced distribution of samples from different grades.

Previous approaches for embryonic cell centroid and border localization were developed for analyzing single images. The proposed approach exploited temporal relationship of frames in embryo time-lapse sequences, using cell centroid information from the previous frame and cell movement between frames to localize cell centroids in the current frame. Related classification-based approaches for cell counting struggled to achieve good cell counting performance on the short 3-cell stage. Counting localized cell centroids minimized the performance drop at 3-cell stage since the system detected individual cells. The cell centroid localization network still faced challenges with embryo sequences containing significant artifacts (e.g. large, cell-like fragments and severe cell overlap). The algorithm can predict centroid locations for embryonic cell tracking and cell counts for extracting cell stage onset and duration.

Previous methods for classifying morula and blastocyst stage onset consider only a single image or rely on texture and embryo structure properties that may not transfer well to different datasets. The proposed approach extracts more robust convolutional features with pairwise learning and utilizes temporal context from embryo time-lapse sequences to predict embryo stages with subtle differences. It can be expanded to classify sequence frames into more embryo development classes (e.g. cell stages, cavitation, expanded blastocyst) for fine-grained development progress evaluation.

The methods developed in this thesis can be used to extract certain morphological parameters from blastocyst images and morphokinetic parameters from embryo time-lapse sequences. These two classes of parameters can give insight into embryo development progress and implantation potential. There are many opportunities to improve the performance and breadth of automatic morphological and morphokinetic parameter extraction. Furthermore, the use of information extracted from embryo images and time-lapse sequences for predicting implantation outcome and live birth rate remains a significant challenge.

Bibliography

[1] Mortimer Abramowitz and Michael W. Davidson. Hoffman modulation contrast basics. https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/hoffman/.

[2] Mortimer Abramowitz and Michael W. Davidson. Phase contrast microscopy - Introduction. https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/phasecontrast/phase/.

[3] Shubhra Aich and Ian Stavness. Improving object counting with heatmap regulation. *arXiv preprint arXiv:1803.05494*, 2018.

[4] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. An unsupervised learning model for deformable medical image registration. In *Conference on Computer Vision and Pattern Recognition*, June 2018.

[5] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

[6] Natalia Basile, Pablo Vime, Mireia Florensa, Belen Aparicio Ruiz, Juan A. Garcia Velasco, José Remohí, and Marcos Meseguer. 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):276–283, 2015.

[7] Charles L. Bormann, Manoj K. Kanakasabapathy, Prudhvi Thirumalaraju, Raghav Gupta, Rohan Pooniwala, Hemanth Kandula, Eduardo Hariton, Irene Souter, Irene Dimitriadis, Leslie B. Ramirez, et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. *Elife*, 9:e55301, 2020.

[8] Juan Cao, Peng Qi, Qiang Sheng, Tianyun Yang, Junbo Guo, and Jintao Li. Exploring the role of visual content in fake news detection. *Disinformation, Misinformation, and Fake News in Social Media*, pages 141–161, 2020.

[9] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[10] Tsung-Jui Chen, Wei-Lin Zheng, Chun-Hsin Liu, Ian Huang, Hsing-Hua Lai, and Mark Liu. Using deep learning with large dataset of microscope images to develop an automated embryo grading system. *Fertility & Reproduction*, 1(01):51–56, 2019.

[11] Jacques Cohen and Kay Elder. *Human preimplantation embryo selection*, volume 6. CRC Press, 2007.

[12] Joe Conaghan, Alice A. Chen, Susan P. Willman, Kristen Ivani, Philip E. Chenette, Robert Boostanfar, Valerie L. Baker, G. David Adamson, Mary E. Abusief, Marina Gvakharia, Kevin E. Loewke, and Shehua Shen. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertility and Sterility*, 100(2):412 – 419.e5, 2013.

[13] María Cruz, Nicolás Garrido, Javier Herrero, Inmaculada Pérez-Cano, Manuel Muñoz, and Marcos Meseguer. Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality. *Reproductive Biomedicine Online*, 25(4):371–381, 2012.

[14] Xiaodong Cun. Dual TVL1 optical flow . https://github.com/vinthony/Dual_TVL1_Optical_Flow, Feb 2017.

[15] Yann N. Dauphin, Harm De Vries, and Yoshua Bengio. Equilibrated adaptive learning rates for non-convex optimization. *arXiv preprint arXiv:1502.04390*, 2015.

[16] Nina Desai, Stephanie Ploskonka, Linnea R. Goodman, Cynthia Austin, Jeffrey Goldberg, and Tommaso Falcone. Analysis of embryo morphokinetics, multinucleation and cleavage anomalies using continuous time-lapse monitoring in blastocyst transfer cycles. *Reproductive Biology and Endocrinology*, 12(1):1–10, 2014.

[17] Darius Dirvanauskas, Rytis Maskeliunas, Vidas Raudonis, and Robertas Damasevicius. Embryo development stage prediction algorithm for automated time lapse incubators. *Computer Methods and Programs in Biomedicine*, 177:161–174, 2019.

[18] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.

[19] Maxime Feyeux, Arnaud Reignier, M. Mocaer, Jenna Lammers, Dimitri Meistermann, Paul Barrière, Perrine Paul-Gilloteaux, Laurent David, and Thomas Fréour. Development of automated annotation software for human embryo morphokinetics. *Human Reproduction*, 35(3):557–564, 2020.

[20] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019.

[21] David K. Gardner and Basak Balaban. Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still important? *MHR: Basic Science of Reproductive Medicine*, 22(10):704–718, 10 2016.

[22] David K. Gardner and Michelle Lane. Culture and selection of viable blastocysts: a feasible proposition for human IVF? *Human Reproduction Update*, 3(4):367–382, 07 1997.

[23] David K. Gardner, Michelle Lane, John Stevens, Terry Schlenker, and William B. Schoolcraft. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertility and Sterility*, 73(6):1155–1158, 2000.

[24] David K. Gardner and William B. Schoolcraft. In vitro culture of human blastocyst / edited by R. Jansen, D. Mortimer. In *Towards Reproductive Certainty: Fertility and Genetics Beyond 1999*, pages 378–388. Parthenon Press, 1999.

[25] Florin C. Ghesu, Bogdan Georgescu, Tommaso Mansi, Dominik Neumann, Joachim Hornegger, and Dorin Comaniciu. An artificial agent for anatomical landmark detection in medical images. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, pages 229–237, Cham, 2016. Springer International Publishing.

[26] Alessandro Giusti, Giorgio Corani, Luca Gambardella, Cristina Magli, and Luca Gianaroli. Blastomere segmentation and 3d morphology measurements of early embryos from hoffman modulation contrast image stacks. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1261–1264. IEEE, 2010.

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[28] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[29] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[30] Catherine Jacobs, Mariana Nicolielo, Renata Erberelli, Fabiana Mendez, Marina Fanelli, Livia Cremonesi, Beatriz Aiello, and Aline R. Lorenzon. 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*, 24(3):273, 2020.

[31] Aisha Khan, Stephen Gould, and Mathieu Salzmann. Deep convolutional neural networks for human embryonic cell counting. In *European Conference on Computer Vision*, pages 339–348. Springer, 2016.

[32] Salman H. Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A. Sohel, and Roberto Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3573–3587, Aug 2018.

[33] Shakiba Kheradmand, Parvaneh Saeedi, Jason Au, and John Havelock. Preimplantation blastomere boundary identification in hmc microscopic images of early stage human embryos. *arXiv preprint arXiv:1910.05972*, 2019.

[34] Shakiba Kheradmand, Amarjot Singh, Parvaneh Saeedi, Jason Au, and Jon Havelock. Inner cell mass segmentation in human HMC embryo images using fully convolutional network. In *IEEE International Conference on Image Processing*, pages 1752–1756, Sep 2017.

[35] Pegah Khosravi, Ehsan Kazemi, Qiansheng Zhan, Jonas E. Malmsten, Marco Toschi, Pantelis Zisimopoulos, Alexandros Sigaras, Stuart Lavery, Lee A. D. Cooper, Cristina Hickman, Marcos Meseguer, Zev Rosenwaks, Olivier Elemento, Nikica Zaninovic, and Iman Hajirasouliha. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *npj Digital Medicine*, 2(1):1–9, Apr 2019.

[36] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[37] Mikkel F. Kragh, Jens Rimestad, Jørgen Berntsen, and Henrik Karstoft. Automatic grading of human blastocysts from time-lapse imaging. *Computers in Biology and Medicine*, 115:103494, 2019.

[38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[39] Cristina Lagalla, Marzia Barberi, Giovanna Orlando, Raffaella Sciajno, Maria A. Bonu, and Andrea Borini. A quantitative approach to blastocyst quality evaluation: morphometric analysis and related IVF outcomes. *Journal of Assisted Reproduction and Genetics*, 32(5):705–712, Apr 2015.

[40] Tingfung Lau, Nathan Ng, Julian Gingold, Nina Desai, Julian McAuley, and Zachary C. Lipton. Embryo staging with weakly-supervised region selection and dynamically-decoded predictions. In *Machine Learning for Healthcare Conference*, pages 663–679. PMLR, 2019.

[41] Brian D. Leahy, Won-Dong Jang, Helen Y. Yang, Robbert Struyven, Donglai Wei, Zhe Sun, Kylie R. Lee, Charlotte Royston, Liz Cam, Yael Kalma, et al. Automated measurements of key morphological features of human embryos for IVF. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 25–35. Springer, 2020.

[42] TzuTa Lin. LabelImg. https://github.com/tzutalin/labelImg, Oct 2019.

[43] Zihan Liu, Bo Huang, Yuqi Cui, Yifan Xu, Bo Zhang, Lixia Zhu, Yang Wang, Lei Jin, and Dongrui Wu. Multi-task deep learning with dynamic programming for embryo early development stage classification from time-lapse videos. *IEEE Access*, 7:122153–122163, 2019.

[44] Jonas Malmsten, Nikica Zaninovic, Qiansheng Zhan, Zev Rosenwaks, and Juan Shan. Automated cell division classification in early mouse and human embryos using convolutional neural networks. *Neural Computing and Applications*, pages 1–12, 2020.

[45] Marcos Meseguer, Javier Herrero, Alberto Tejera, Karen M. Hilligsøe, Niels B. Ramsing, and José Remohí. The use of morphokinetics as a predictor of embryo implantation. *Human Reproduction*, 26(10):2658–2671, 2011.

[46] Marcos Meseguer, Irene Rubio, Maria Cruz, Natalia Basile, Julian Marcos, and Antonio Requena. Embryo incubation and selection in a time-lapse monitoring system improves pregnancy outcome compared with a standard incubator: a retrospective cohort study. *Fertility and Sterility*, 98(6):1481–1489, 2012.

[47] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 4th International Conference on 3D Vision*, pages 565–571. IEEE, 2016.

[48] Yamileth Motato, María José de los Santos, María José Escriba, Belén Aparicio Ruiz, José Remohí, and Marcos Meseguer. Morphokinetic analysis and embryonic prediction for blastocyst formation through an integrated time-lapse system. *Fertility and Sterility*, 105(2):376–384, 2016.

[49] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2018.

[50] Nathan H. Ng, Julian J. McAuley, Julian Gingold, Nina Desai, and Zachary C. Lipton. Predicting embryo morphokinetics in videos with late fusion nets & dynamic decoders. In *International Conference on Learning Representations (Workshop)*, 2018.

[51] Joseph Paul Cohen, Genevieve Boucher, Craig A. Glastonbury, Henry Z. Lo, and Yoshua Bengio. Count-ception: Counting by fully convolutional redundant counting. In *IEEE International Conference on Computer Vision (Workshop)*, pages 18–26, 2017.

[52] Reza M. Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Coarse-to-fine texture analysis for inner cell mass identification in human blastocyst microscopic images. In *Seventh International Conference on Image Processing Theory, Tools and Applications*, pages 1–5, Nov 2017.

[53] Reza M. Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Blastomere cell counting and centroid localization in microscopic images of human embryo. In *IEEE 20th International Workshop on Multimedia Signal Processing*, 08 2018.

[54] Reza M. Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. A hybrid approach for multiple blastomeres identification in early human embryo images. *Computers in biology and medicine*, 101:100–111, 2018.

[55] Reza M. Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Multi-resolutional ensemble of stacked dilated u-net for inner cell mass segmentation in human embryonic images. In *25th IEEE International Conference on Image Processing*, pages 3518–3522, Oct 2018.

[56] Reza M. Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Cell-net: Embryonic cell counting and centroid localization via residual incremental atrous pyramid and progressive upsampling convolution. *IEEE Access*, 7:81945–81955, 2019.

[57] Reza M. Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Trophectoderm segmentation in human embryo images via inceptioned u-net. *Medical Image Analysis*, 62:101612, 2020.

[58] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[59] Vidas Raudonis, Agne Paulauskaite-Taraseviciene, Kristina Sutiene, and Domas Jonaitis. Towards the automation of early-stage human embryo development detection. *Biomedical Engineering Online*, 18(1):1–20, 2019.

[60] Laura Rienzi, Danilo Cimadomo, Arantxa Delgado, Maria Giulia Minasi, Gemma Fabozzi, Raquel del Gallego, Marta Stoppa, Jose Bellver, Adriano Giancani, Marga Esbert, et al. Time of morulation and trophectoderm quality are predictors of a live birth after euploid blastocyst transfer: a multicenter study. *Fertility and Sterility*, 112(6):1080–1093, 2019.

[61] Nelida Rodriguez-Osorio, Sule Dogan, Erdogan Memili, et al. Epigenetics of mammalian gamete and embryo development. *Livestock epigenetics*, 1, 2012.

[62] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[63] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[64] Parvaneh Saeedi, Dianna Yee, Jason Au, and Jon Havelock. Automatic identification of human blastocyst components via texture. *IEEE Transactions on Biomedical Engineering*, 64(12):2968–2978, Dec 2017.

[65] E. Santos Filho, J. Alison Noble, Maurizio Poli, Tracey Griffiths, Gerri Emerson, and Darren Wells. A method for semi-automatic grading of human blastocyst microscope images. *Human Reproduction*, 27(9):2641–2648, 06 2012.

[66] E. Santos Filho, J. Alison Noble, and Darren Wells. Toward a method for automatic grading of microscope human embryo images. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1289–1292, April 2010.

[67] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *IEEE conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016.

[68] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.

[69] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[70] Amarjot Singh, Jason Au, Parvaneh Saeedi, and Jon Havelock. Automatic segmentation of trophectoderm in microscopic images of human blastocysts. *IEEE Transactions on Biomedical Engineering*, 62(1):382–393, Jan 2015.

[71] Amarjot Singh, John Buonassisi, Parvaneh Saeedi, and Jon Havelock. Automatic blastomere detection in day 1 to day 2 human embryo images using partitioned graphs and ellipsoids. In *International Conference on Image Processing*, pages 917–921. IEEE, 2014.

[72] Canadian Fertility & Andrology Society. Canadian assisted reproductive technologies register plus (CARTR plus). https://cfas.ca/cartr-annual-reports.html, Sep 2019.

[73] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[74] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[75] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[76] Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.

[77] Jian Wang, Hengde Zhu, Shui-Hua Wang, and Yu-Dong Zhang. A review of deep learning on medical image analysis. *Mobile Networks and Applications*, pages 1–30, 2020.

[78] Shoujin Wang, Wei Liu, Jia Wu, Longbing Cao, Qinxue Meng, and Paul J. Kennedy. Training deep neural networks on imbalanced data sets. In *International Joint Conference on Neural Networks*, pages 4368–4374, July 2016.

[79] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to model the tail. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 7029–7039. Curran Associates, Inc., 2017.

[80] Weidi Xie, J. Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):283–292, 2018.

[81] Feng Xiong, Qing Sun, Guangui Li, Zhihong Yao, Peilin Chen, Caiyun Wan, Huixian Zhong, and Yong Zeng. Association between the number of top-quality blastocysts and live births after single blastocyst transfer in the first fresh or vitrified–warmed IVF/ICSI cycle. *Reproductive BioMedicine Online*, 2020.

[82] Yao Xue, Nilanjan Ray, Judith Hugh, and Gilbert Bigras. Cell counting by regression using convolutional neural network. In *European Conference on Computer Vision*, pages 274–290. Springer, 2016.

[83] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime TV-L1 optical flow. In *Joint Pattern Recognition Symposium*, pages 214–223. Springer, 2007.

[84] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[85] J. Zhang, M. Liu, and D. Shen. Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. *IEEE Transactions on Image Processing*, 26(10):4753–4764, 2017.

[86] Jianpeng Zhang, Yutong Xie, Qi Wu, and Yong Xia. Medical image classification using synergic deep learning. *Medical Image Analysis*, 54:10–19, 2019.

[87] Kaihao Zhang, Yongzhen Huang, Yong Du, and Liang Wang. Facial expression recognition based on deep evolutional spatial-temporal networks. *IEEE Transactions on Image Processing*, 26(9):4193–4203, 2017.

[88] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*, pages 286–301, 2018.