

Human Blastocyst's Zona Pellucida segmentation via boosting ensemble of complementary learning

Reza Moradi Rad^{a,*}, Parvaneh Saeedi^a, Jason Au^b, Jon Havelock^c

^a School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

^b Pacific Centre for Reproductive Medicine, Burnaby, BC, Canada

^c Reproductive Endocrinology and Infertility Special Interest Group, Canadian Fertility Andrology Society, Montreal, QC, Canada

ARTICLE INFO

Keywords:

Zona pellucida
Human embryo
IVF
Medical image analysis
Deep neural network

ABSTRACT

Characteristics of Zona Pellucida (ZP), particularly its thickness, is a key indicator of human blastocyst (day-5 embryo) quality. Therefore, ZP segmentation is an important step towards achieving automatic embryo quality assessment. In this paper, a novel approach based on boosting ensemble of hybrid complementary learning is proposed to segment Zona Pellucida in human blastocyst images. First, an inner-ZP localization method is proposed to separate the ZP from the heavily textured area inside a blastocyst. Then, a deep Hierarchical Neural Network (HiNN) is proposed to segment ZP area. The hierarchical nature of the proposed network enables learning features with respect to their spatial location in the embryo. Finally, a Self-supervised Image-Specific Refinement (SISR) strategy is proposed as a complementary step to boost the performance. The proposed system is a hybrid approach in the sense that the HiNN learns the intra-correlation among images, while the SISR takes into account the inter-correlation within the query image. Experimental results confirm that the proposed method is capable of identifying ZP area with average *Precision*, *Recall*, *Accuracy* and *Jaccard Index* of 85.2%, 92.0%, 95.6% and 78.1%, respectively. The proposed HiNN system outperforms state of the art by 4.9% in *Precision*, 11.2% in *Recall*, 3.6% in *Accuracy* and 10.7% in *Jaccard Index*.

1. Introduction

The infertility rate in Canada currently is about 16%. This translates into one in every six couples experiencing difficulties in conceiving, three times more than that in 1980 [1]. This situation is more pronounced in developing countries as the World Health Organization (WHO) announced that on average one in every four couples is suffering from infertility [2]. In Vitro Fertilization (IVF) is one of the most effective and common forms of assisted reproductive technology to address the infertility problem. IVF is the process of manual fertilization of female eggs outside uterus. The fertilized eggs (refers to as embryos) are cultured for 3 – 5 days inside an incubator. Then, embryos with the highest quality according to their morphological attributes are chosen and transferred into the patient's uterus.

IVF is widely practiced around the world with more than one million treatments per year [3]. Unfortunately, after four decades, since the first successful IVF, it still has rather low success rate specially in older patients. IVF's relatively low success rate has made it the subject of active research [4–16]. Since embryo quality is an important factor affecting IVF success (pregnancy outcome), choosing the most

promising embryos results in a greater success rate [8,9]. Embryo quality assessment is a challenging task since embryos are variable in their morphological features. Furthermore, manual human monitoring is time consuming and heavily dependent on the embryologist's experience and therefore subjective. Automating this process offers several advantages, including preventing human error, eliminating subjectivity, saving manpower and reducing training costs.

Identification of embryo components is a crucial step in performing automatic embryo quality assessment. A protective glycoprotein layer called Zona Pellucida (ZP) encapsulates the oocyte and maintains close cell proximity between blastomeres as the zygote develops. The space between the plasma membrane and the ZP is called the perivitelline space. When the embryo has reached the blastocyst stage, the blastocoel will be filled with fluid causing the trophectoderm to expand and press against the inner surface of the ZP. ZP decreases in thickness as embryo develops [17]. Characteristics of ZP, specially its thickness, are the key indicators in assessing embryo's quality [4–7]. The relationship of ZP thickness with morphological assessment and implantation outcome were studied in Refs. [9] and [10]. These studies suggest that embryo with thinner ZP are more advanced and their implantation potential

* Corresponding author.

E-mail address: rmoradir@sfu.ca (R.M. Rad).

<https://doi.org/10.1016/j.imu.2018.10.009>

Received 13 August 2018; Received in revised form 18 October 2018; Accepted 18 October 2018

Available online 25 October 2018

2352-9148/ © 2018 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

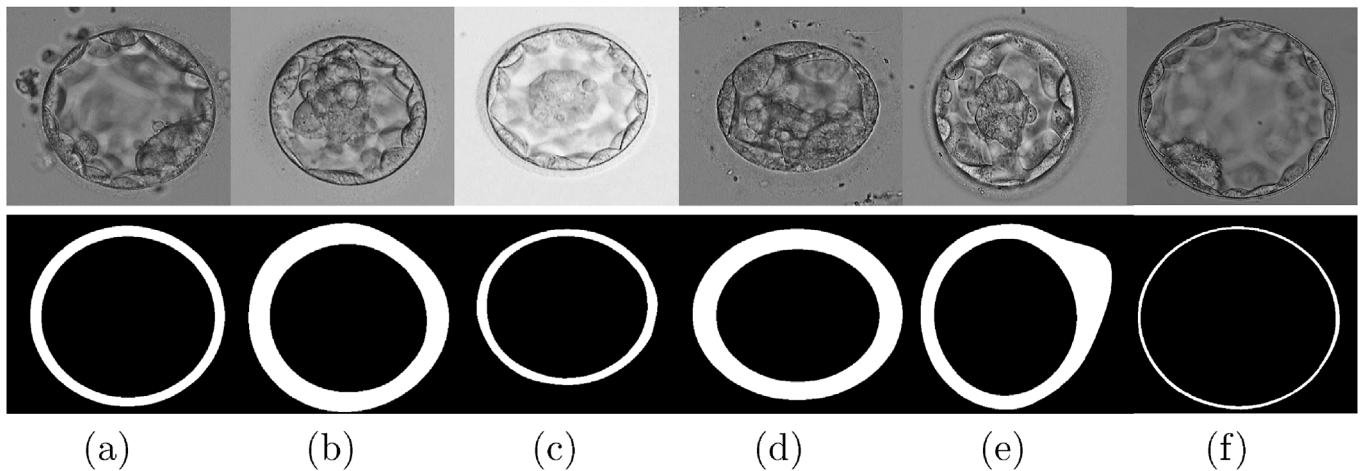


Fig. 1. First row: six human blastocyst images; Second row: corresponding ground truth masks for ZP regions.

seem to be significantly higher. A more recent study [11] also confirms that the thickness and the density of ZP region are correlated with human embryo's quality and implantation success. Therefore, not only accurate segmentation of ZP is a crucial step towards achieving automatic embryo quality assessment, but also a useful tool to validate theoretical assumptions about relationship between ZP features and pregnancy outcome.

Analyzing ZP in human embryos through 2D microscopic images is a challenging task. Fig. 1 shows five sample blastocyst images and their corresponding ground truth ZP regions. Blastocyst in Fig. 1-a contains multiple artifacts. ZP boundary in the blastocyst of Fig. 1-b is slowly fading with no explicit borders. Blastocyst in Fig. 1-c has brightness inhomogeneity over the ZP structure. Texture of the ZP and its background in the image of Fig. 1-d are very similar. Embryo in Fig. 1-e developed an asymmetric ZP where the thickness varies along its boundaries. Blastocyst in Fig. 1-f is a fully expanded one with a very thin ZP band. As shown through these examples, precise identification of ZP region is a complex task. These complexities make most of the standard segmentation techniques unsuitable for ZP identification.

The main contributions of the proposed method include:

1. A novel Hierarchical Neural Network (HiNN) that receives concentric input patches and provides scalable context view on local regions enabling the network to learn context-aware intensity features. This helps in coping with artifacts, which may share common texture features with the ZP area.
2. A Self-supervised Image-Specific Refinement (SISR) technique that boosts the learning process. This can be utilized as a complementary step for any other applications, specially in medical research, where labelled data is limited and/or hard to collect.
3. A data set of 235 images of human blastocysts from the actual patients at the PCRM with the ZP's ground truth manually identified by expert embryologists is made available to the public and research community.

2. Related work

A few methods have been reported in the literature for segmentation of ZP region at various stages of embryonic development during incubation. These methods can be classified into two general categories: unsupervised and supervised. A brief description of them is presented next.

Unsupervised methods generally use either model-based or optimization-based techniques to segment human embryo images. Karlsson et al. [12] reported one of the earliest works for segmentation of ZP in, where a semi-automatic variational method on a model image was

proposed to segment the ZP region by finding the inner and outer boundaries. This work was further developed in Ref. [13], to handle artifacts and occlusions. Morales et al. [4] proposed an active contour based method to identify the ZP area in human embryo of day 2 to day 3. Yee et al. [14] introduced a method to measure the thickness of ZP on day 5 based on ellipsoidal models and edge maps. Wang et al. [5] proposed a method to select ZP dissection position for performing embryo biopsy. They first estimate inner and outer ZP boundaries by transferring the image into polar coordinates. Then, the internal and external ZP boundaries were estimated via fitting an ellipsoidal model to image edges. This work was further extended in Refs. [6] and [7]. Recently, Wang et al. [18] proposed a method for auto-focusing and polar body detection. As part of this method, ZP mask is created by 15 pixel dilation of the contour. Wong et al. [19] presented a method for automation and optimization of multipulse laser ZP drilling. In the first stage of the method, the inner and the outer boundaries of ZP are estimated using polar transformation and graph search. All the above mentioned methods, except [14], attempted to segment ZP at early-stage embryos not at blastocyst stage. This task is expected to be less challenging as generally thickness of ZP decreases by time. In general, unsupervised methods have rather limited applications as they provide very specific solutions and may not perform well on larger datasets with variant characteristics.

Supervised methods utilize the labelled ground truth data to learn the relationship between the input raw images and the output binary segmented masks. Although, supervised segmentation is a well-studied discipline for medical image analysis, it is a new trend for human embryo image analysis. Over the past few years, a few supervised techniques are reported in the literature. Conventional supervised methods work with a set of hand-crafted features along with a classifier. Kheradmand et al. [15] developed a supervised segmentation method for three main blastocyst components (ZP, TE and ICM) by training an ANN using hand-crafted DCT features. Generally, hand-crafted feature design is heavily dependent on experience and requires tuning, therefore delivers limited performance for complicated tasks. There is no supervised method that offers automatic feature design in the literature for ZP segmentation. However, there are a few supervised methods for cell counting and Inner Cell Mass (ICM) segmentation in human embryo images. Kheradmand et al. [20], proposed an ICM segmentation technique by employing a 16-layer Visual Geometry Group (VGG) network [21]. Moradi et al. [22] proposed multi-resolutional stacked dilated convolutional network to perform ICM segmentation. Moradi et al. [23] developed a cell counting technique for human embryo images using a deep convolutional neural network. Supervised methods generally deliver better performance compared to the unsupervised ones. To the best of our knowledge, the presented work in this paper is the first work

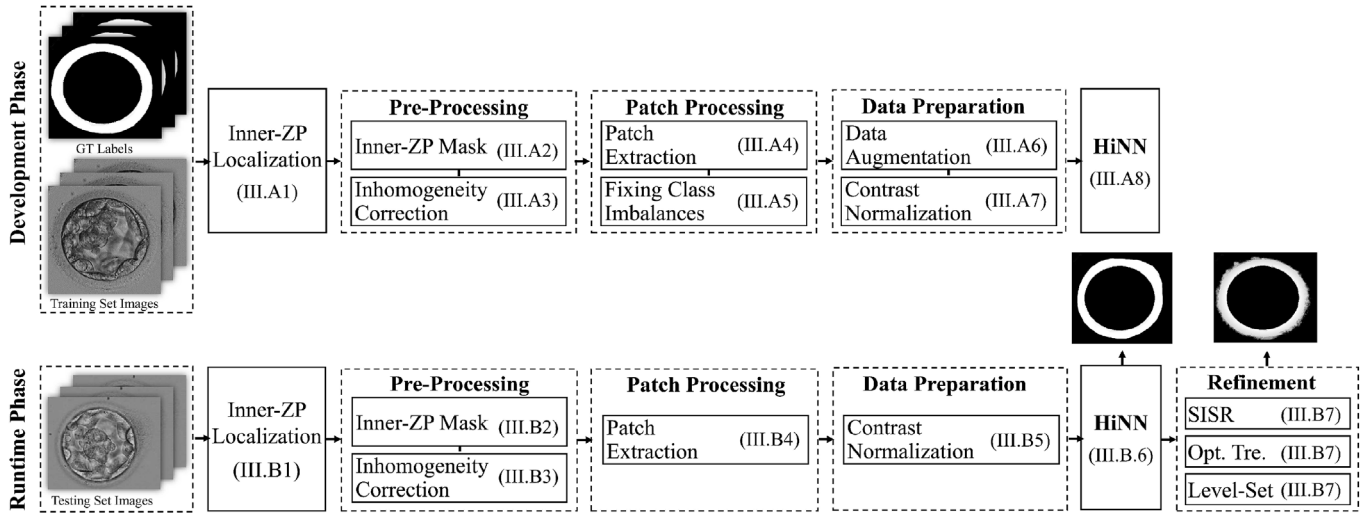


Fig. 2. Block diagram of the proposed method.

using a deep hierarchical neural network for segmentation of ZP in human blastocyst images.

3. Proposed method

Block diagram of the entire proposed system is depicted in Fig. 2. The proposed ensemble system consists of two main phases: development and runtime. During the development phase, the inner-ZP localization method is used to mask out the internal structures of blastocyst before ZP segmentation. Then, image patches are extracted and augmented to be fed into the HiNN network. In the runtime phase, a test image runs through the same processes as development phase except for Sections 3.1.5 and 3.1.6. The last step in the runtime phase is a network output refinement strategy that acts as a complementary learning process on the output of the conventional machine learning approach. All building blocks of the proposed system are described in the following sub-sections.

3.1. HiNN development phase

3.1.1. Inner-ZP localization

We propose a supervised classification for inner-ZP boundary identification using a mixture of pixel-level, local and global features. ZP often contains a fading boundary, similar to ridge-like edge profiles. Frangi vesselness filter [24] is employed to generate pixel-level features that describe the inner-ZP boundary features. It was originally introduced to enhance blood vessels in retina images. This filter uses second order information (Hessian matrix of the pixels) to compute vessel likelihood and utilizes a range of depth values to measure the vessel likelihood of ridges. Next, three features, including standard deviation, entropy and range of intensities are utilized as statistical measures of texture level by measuring the degree of variability of each pixel in its local region. Furthermore, Euclidean distance from the center of blastocyst is considered as a global feature to capture spatial characteristics of the inner-ZP's boundary. This spatial indicator is specially useful since the inner-ZP's boundary is always a circular region, connected to the embryo and located not too close to the center of blastocyst. Performance of the inner-ZP Localization is reported in Table 1. A three-layer feed-forward network with sigmoid hidden and softmax output neurons is used as the classifier. This network is trained with scaled conjugate gradient back-propagation method and cross-entropy loss function.

Table 1

Inner-ZP localization results.

	Precision	Recall	Accuracy	Jaccard Index	Error(%)
Yee et al. [14]	–	–	–	–	7.6
Inner-ZP Localization	.991	.983	.995	.974	2.6

3.1.2. Inner-ZP mask

The purpose of pre-processing is to enable the network to learn more efficiently from the raw image data. Table 1 shows that the proposed inner-ZP localization method can reliably localize inner-ZP boundaries. Using the identified boundaries of the inner-ZP, a mask is generated by filling out all the pixels within the closed boundaries of the inner-ZP. Separating this region from the rest is beneficial as the region inside a blastocyst can potentially confuse the network due to its dense texture.

3.1.3. Illumination in-homogeneity correction

Although an optical enhancement technique, known as Differential Interference Contrast (DIC), is employed during the image acquisition, the appearance of the ZP may vary in intensity from light to dark (as shown in Fig. 3). This is due to the enhanced areas of light and shadow, which add distortion to the image. An energy minimization based on multiplicative intrinsic component optimization method [25] is implemented to correct the intensity in-homogeneity in blastocyst images. Some selected regions are depicted in Fig. 3 (before and after the correction) to highlight the effect of illumination inhomogeneity correction.

3.1.4. Patch extraction

Every blastocyst image is decomposed into patches by extracting square patches of five different sizes. These image patches are used for training the proposed HiNN network.

3.1.5. Fixing class imbalances

Number of training patches per class might vary between the ZP and the background classes. Such imbalanced distribution may cause a bias in favour of the larger class. In order to balance the classes, a fixed number of patches were extracted in Ref. [26]; that may lead to under-utilization of the samples. Over-sampling on the smaller class or under-sampling on the larger class are other standard approaches to deal with the imbalanced class distributions. In case of human blastocyst images, the background class is usually larger (up to three times) compared to the ZP class. Since oversampling the minority class may increase the

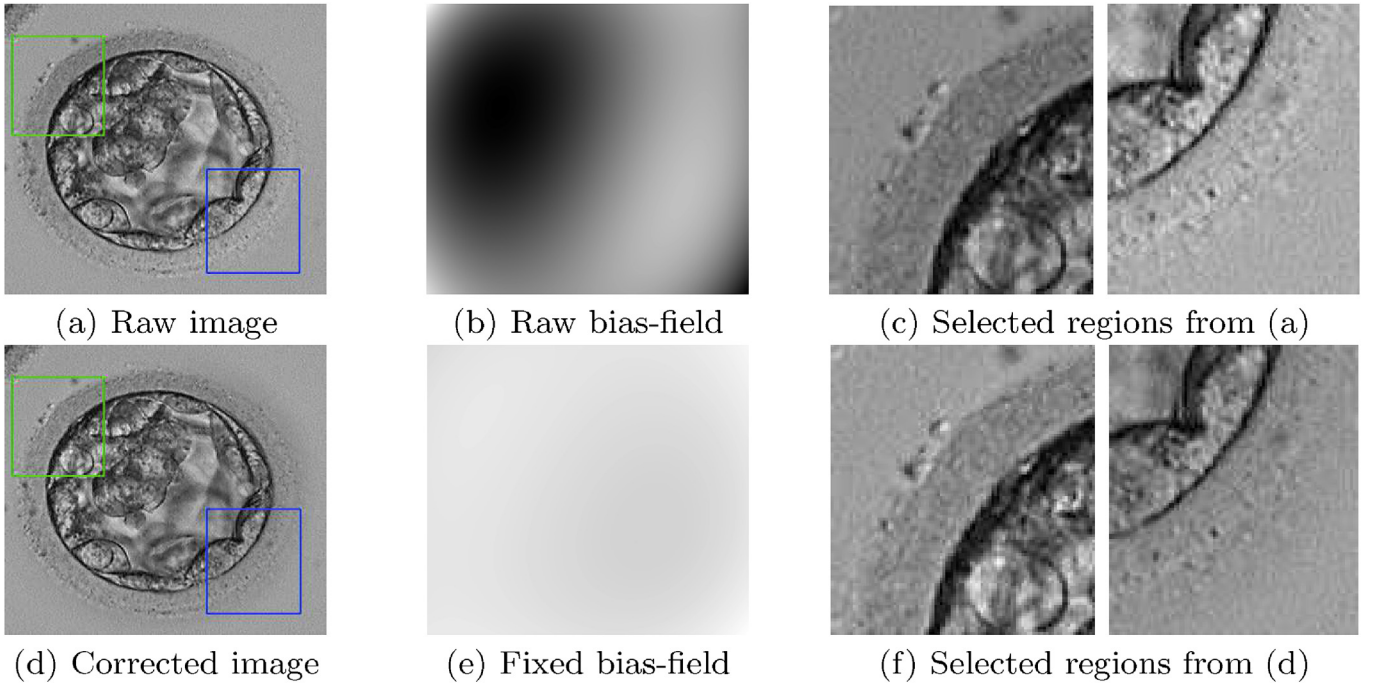


Fig. 3. Illumination in-homogeneity correction by removing the bias field.

likelihood of over-fitting, under-sampling of the background is preferred in here. However, unlike [26], where a defined number of patches were extracted for each class, a dynamic number (size of minor class) is used to under-sample the major class to minimize the information loss for an effective training.

3.1.6. Data augmentation

Authors in Ref. [27] showed that one of the main reasons for the above issue is a relatively small number of training samples comparing to the network parameters. Furthermore, quantity is not always equal to quality, specifically in the context of classification. Data augmentation generally refers to generating additional samples by synthetically transforming the existing ones to increase the number of training samples and/or enhance their quality. Augmenting training images by the right amount of noise not only makes the model robust to noise but modifies image histogram distribution. Here, the data augmentation was performed prior to the training. Each patch underwent three independent transformations sequentially including four randomized process from the following list:

- Flipping horizontally or vertically
- Rotating by an angle [0: 90: 360]
- Zooming in or out [0.8 1.2]
- Adding light Gaussian [$m = 0$, $v = 0.0003$] or Speckle noises [$m = 0$, $v = 0.0006$]

This increased size of the training set by a factor of three. Table 2 summarizes detailed statistics of patches in the training set and Fig. 4 depicts augmented examples of positive and negative patches. Here, the term *Positive* refers to the patches that are labelled as ZP and the term *Negative* refers to those labelled as background or non-ZP. Each augmented patch has received the same class label as the original GT patch.

3.1.7. Contrast normalization

Since the brightness can simply vary across the blastocyst image, all the training patches are normalized to have zero mean and unit variance using Eq. (1).

Table 2

Detailed statistics of patches in the training set.

	Original Patches		Augmented Patches	
	Positive/Negative	Total	Positive/Negative	Total
1 st Branch	862,135	1,724,270	2,586,405	5,172,810
2 nd Branch	860,408	1,720,816	2,581,224	5,162,448
3 rd Branch	857,848	1,715,696	2,573,544	5,147,088
4 th Branch	854,152	1,708,304	2,562,456	5,124,912
5 th Branch	848,697	1,697,394	2,546,091	5,092,182
Total	4,283,240	8,566,480	12,849,720	25,699,440

$$X_{new} = \frac{X - \mu}{\sigma} \quad (1)$$

Here, μ refers to the mean intensity and σ indicates the variance.

3.1.8. HiNN architecture

In this paper, inspired by Ref. [26] which achieved impressive results for retinal vessel segmentation, a deep neural network is proposed. The proposed Hierarchical Neural Network (HiNN) contains a total of five branches at five different depths. The input and output layers have 49 (7×7), 121 (11×11), 225 (15×15), 361 (19×19) and 529 (23×23) units at depths of 1–5, respectively. The hidden layers have 100 units with Leaky Rectifier Linear Unit LReLU [28] activation function (as defined in Eq. (2)) and one softmax in the last layer. Each leaky ReLUs introduces a small non-zero gradient when the unit is not active ($x < 0$) instead of zero (as done by ReLU in Eq. (3)). Here, α is a leakiness parameter.

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{otherwise} \end{cases} \quad (2)$$

$$f(x) = \max(0, x) \quad (3)$$

The hierarchical design of the network (as depicted in Fig. 5) enables the multi-scale feature learning to capture intensity features with respect to spatial characteristics. For each patch size, a separate network branch is developed. Therefore, the weights and biases in the

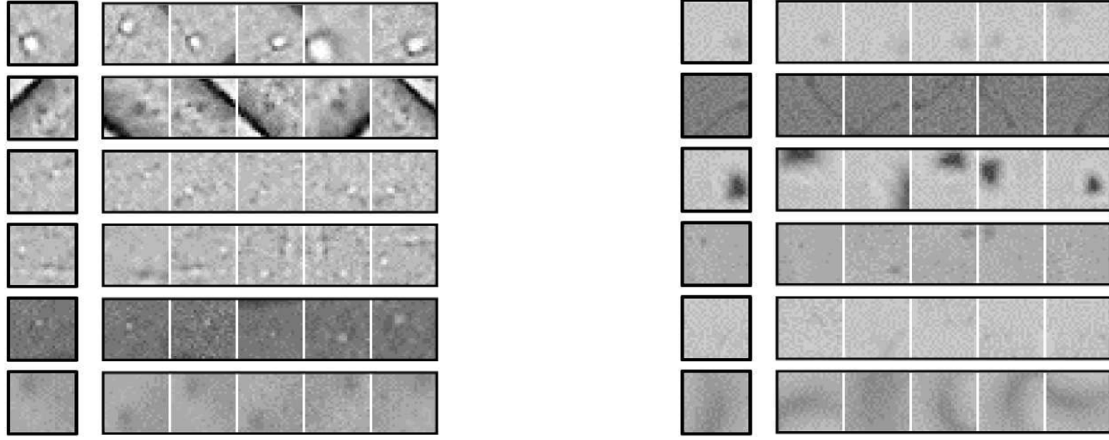


Fig. 4. Augmented patch examples of positive (left) and negative patches (right); Each row depicts five random augmentation of the leftmost patch.

proposed HiNN are optimized exclusively for each patch size. To learn useful features for transformed versions of the input instead of recovering the input and to avoid vanishing or exploding the gradients, a two-step training procedure (pre-training of the first layer and overall training) is adopted from Ref. [26]. First, the idea of denoising auto-encoder is generalized and used to initiate the first hidden layer of the proposed HiNN. Then, other hidden layers are initialized using the Nguyen-Widrow algorithm [29] and the overall network training is done using scaled conjugate gradient back-propagation. An auto-encoder is traditionally used to reduce the dimensionality or feature learning by mapping the input x to the hidden representations and then generating \hat{x} , the reconstructed version of the x . On the other hand, in the denoising auto-encoder, a distorted version of input vector, indicated by x' , is used to generate an undistorted version of the input x by minimizing the following error:

$$\theta = \operatorname{argmin}_{\theta} \frac{1}{n} \sum_{i=1}^n \|x_i - x'_i\|^2 \quad (4)$$

Denoising auto-encoders have proven ability to learn how to reconstruct the original input from the distorted input [30,31]. The distortion is often introduced to the original input by setting some components of the input to zero. This idea is generalized by using a zero mask as the distorted version of the segmented mask, suggested by Ref. [26].

The last step in segmentation via HiNN is construction of a probability map for ZP regions. Unlike conventional supervised segmentation, where probability is resulted from a single input image patch, hence there are multiple input patches trained through different branches of the proposed HiNN. Note that the probability map generation here is different from the strategy employed in Ref. [26] in two ways. First, the supporting patches are coming from different branches instead of neighboring patches. Second, the output of each branch is dynamically weighted by the error introduced in the development phase instead of averaging the probability of all patches containing a query pixel. The final decision to label a pixel is made by:

$$P(X_m \in C_1) = \sum_{i=1}^N W_i \times P(X_m \in C_1 | i) \quad (5)$$

Here, N is the number of branches at different depths, $P(X_m \in C_1)$ is the probability of pixel X_m belonging to class C_1 and W_i is the weight of branch i . Furthermore, a normalized 2D Gaussian filter with same dimension as the probability map and $\sigma = \sim 100$ is applied on the overall estimated map to comply with the biological fact that probability of a point being on the ZP band constantly decreases by its radial distance from the corresponding inner-ZP boundary point.

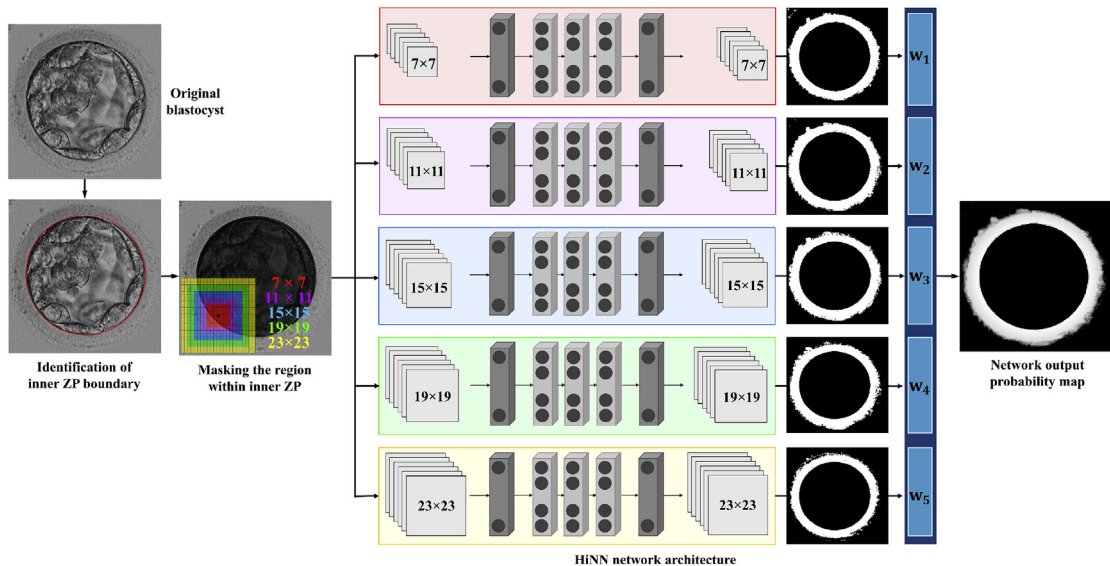


Fig. 5. Illustration of the proposed HiNN network.

3.2. HiNN runtime phase

3.2.1. Inner-ZP localization

Inner-ZP boundary is identified using the proposed inner-ZP localization approach explained in Section 3.1.1.

3.2.2. Inner-ZP mask

The identified inner-ZP boundary is utilized to generate inner-ZP mask as described in Section 3.1.2.

3.2.3. In-homogeneity correction

The masked blastocyst image is analyzed for intensity in-homogeneity and get fixed accordingly using the technique summarized in Section 3.1.3.

3.2.4. Patch extraction

Patches in multiple sizes are extracted from the blastocyst image, similar to the process in Section 3.1.4.

3.2.5. Contrast normalization

Blastocyst images are normalized for integrity and training efficiency as described in Section 3.1.7.

3.2.6. HiNN architecture

The pre-processed blastocyst images are fed into the trained HiNN network that was introduced in Section 3.1.8 to generate the prediction map.

3.2.7. Self-supervised Image-Specific Refinement (SISR)

Song et al. in Ref. [32] proved that applying an effective post-processing can improve performance of their network. Unlike some of the previous work, where a simple and application-specific post-processing is applied (like size-based filtering using a fixed threshold in Ref. [33]), we apply a more general automatic strategy. The proposed network learns the inter-correlation (i.e., among training images) in the data set. In contrast to conventional methods, which terminate the learning process at this stage, additional steps are introduced to further improve the performance. As observed in Fig. 1, every human embryo image has distinctive characteristics and features that make it difficult to apply the learned knowledge from one image to another. Therefore, in addition to the standard inter-correlation learning procedure, intra-correlation within each embryo image can be learned. As a complementary step, an automatic network output refinement is performed to further improve the performance.

Network output is then used to perform an exclusive supervised classification that is automatically designed for each image in the test set. Training samples are obtained from the query test image. For class 1 (ZP) the interior of ZP area is extracted from the network output (network probability ≥ 0.9) and for class 2 (background) samples are taken from four corners of the image, as shown in Fig. 6. Then, an ANN is utilized as the classifier with a hand-crafted feature set. This feature

set contains eighteen texture descriptors, including Local Binary Pattern (LBP), Histograms of Oriented Gradients (HOG), Local Phase Quantization (LPQ), information measure of correlation, auto-correlation, entropy, energy, correlation, homogeneity, contrast, variance, dissimilarity, inverse difference moment, angular second moment, inverse difference, inverse difference normalized, and inverse difference moment normalized. Finally, an ellipsoidal model is designed to effectively examine the neighborhood avoiding expensive pixel level processing. We start from the ZP's interior and let the ellipse grow outward. To reduce computation time, angles are sampled uniformly with steps of 2° . The search area is shown in Fig. 7-a.

In order to evaluate the proposed supervised refinement, two alternative approaches are implemented. First, we utilize an optimized threshold technique, which is the most common method for network refinement and employed by Refs. [26] and [34]. Second, a region based level-set approach [35] is implemented to find the boundary points according to each individual ZP.

- **Region Based Level-Set:** A region-based active contour model based on a variational level set formulation is adopted [35]. In noisy images with weak boundaries, region-based models such as [35] expected to outperform edge-based models such as [36] since they rely on global information to direct the contour. Here, image regions are described by Gaussian distributions with different means and variances, which enables the technique to handle in-homogeneities and multiplicative noises [32], [37].

A set of N disjoint image regions are denoted by $\{\Omega_i\}_{i=1}^N$ and a circular neighborhood is denoted by $\Theta_x \equiv \{y; \|x - y\| \leq \rho\}$. Segmentation of Φ_x is achieved based on maximum a posteriori probability (MAP):

$$\prod_{i=1}^N \prod_{y \in \Omega_i \cap \Theta_x} P_{i,x}(I(y)) \quad (6)$$

Taking the logarithm, maximizing Eq. (6) is equal to minimizing the following equation:

$$E_x^{LGF} = \sum_{i=1}^N \int_{\Omega_i \cap \Theta_x} -\log P_{i,x}(I(y)) dy \quad (7)$$

For a precise identification of ZP boundaries, this level-set technique is applied to further evolve the initial area corresponding to ZP interior (network probability ≥ 0.9) from network output.

- **Optimized Threshold:** For each embryo image in the test set, a probability map is obtained from the trained HiNN network. Then, a binary ZP segmentation is generated based on an optimized threshold. The optimized threshold is set as the one at which the maximum *Jaccard Index* is achieved on the training set. This the most common method to generate the binary segmentation map as employed by Refs. [26], [34].

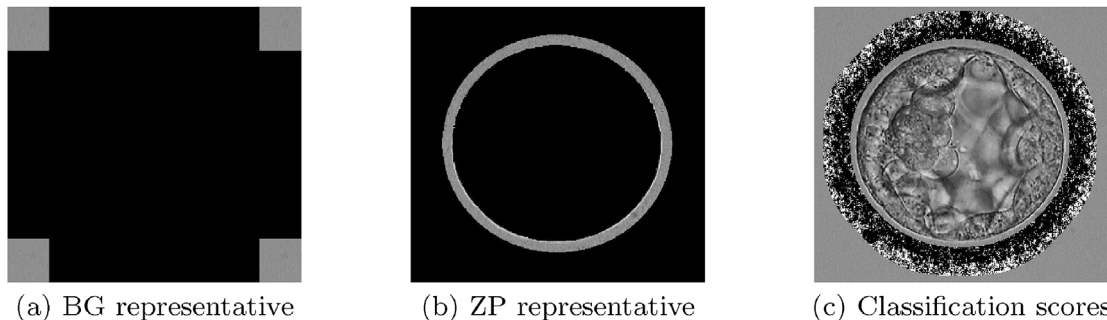


Fig. 6. Illustration of the self-Supervised image-specific texture classification.

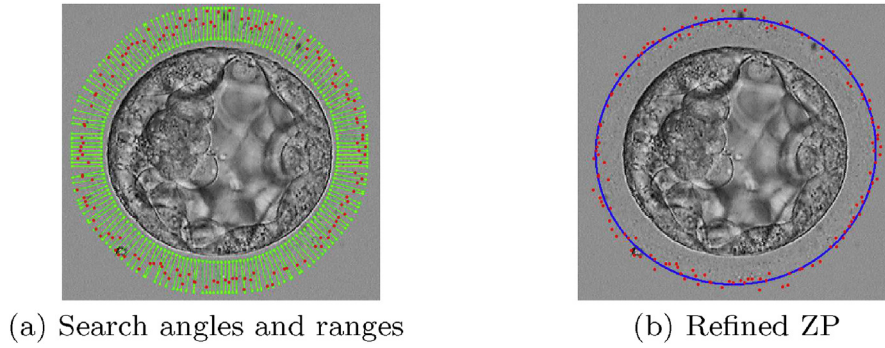


Fig. 7. ZP outer boundary localization in SISR.

4. Experimental setup

4.1. Dataset

The blastocyst dataset introduced in Ref. [17] is utilized for experimental purposes, as the only publicly available dataset for human embryo components to date. This benchmark blastocyst dataset consists of 235 blastocyst images from recent patients at the Pacific Center for Reproductive Medicine (PCRM). These images have been acquired using an Olympus IX71 inverted microscope that employs Nomarski optical enhancement technique (DIC). The training set consists of patches from 77 images ($\sim 1/3$ of the total images in our dataset) and the test set includes patches from the remaining 158 images.

4.2. Ground truth

Performance of the proposed method is compared against GT that generated manually by expert embryologists [17]. The ZP boundaries are precisely outlined by a one pixel wide contour to enable accurate measuring of the segmentation performance.

4.3. Evaluation criterion

To measure the quality of the identified ZP at the pixel level, we utilize *Precision*, *Recall*, and *Accuracy* as well as the *Jaccard Index* (as defined in Eq. (8)). In this context, *Precision* indicates how much of the identified ZP are relevant to the GT (or how precise the identification results are).

Recall indicates how much of the ZP's GT are covered by the identified ZP (or how much of the ZP ground truth is identified). *Accuracy* measures overall accuracy of the segmentation (binary classification). *Jaccard Index* takes into account the effect of both falsely called and missed regions to measures the overall quality of the identification.

$$\text{Jaccard Index} = \frac{TP}{TP + FP + FN} \quad (8)$$

here TP indicates the number of pixels correctly identified as ZP, FP implies the number of pixels incorrectly identified as ZP, FN signifies the number of pixels that are missed and TN is the number of pixels that are correctly skipped.

5. Experimental results and discussion

The proposed HiNN network is evaluated both qualitatively and quantitatively and is compared against [4] [14], and [15].

5.1. Qualitative evaluation

In this section, performance of the proposed method is presented for visual inspection. Table 3 contains sample results of the proposed method with various applied refinements. All results are clustered into

three classes, worst, average and best, according to the *Jaccard Index* values. Then, two representative samples are extracted and displayed for each class. From Table 3, the proposed HiNN + SISR performs well for blastocyst images with fading boundaries (i.e., first row) and corrects the network output where ZP has an asymmetrical shape (i.e., second row). Third, fourth and sixth rows of Table 3 show that the proposed HiNN + SISR can cope with heavy artifacts. The worse performance achieved where test images contain multiple artifacts, brightness in-homogeneity, fading boundaries and non-discriminating textures (fifth and sixth rows).

5.2. Quantitative evaluation

The evaluation criterion introduced in Section 4.3 are considered to measure the quality of the segmented ZP in a quantitative way. The experiments are conducted five times with different training/test sets and the average performance are reported in Tables 4 and 6. Kheradmand et al. [15] is the only reported supervised ZP segmentation method in the literature. Table 4 compares the performance of six variants of the proposed method with that of [15]. The proposed method when SISR refinement is applied is at its best with an average of 85.2%, 92.0%, 95.6% and 78.1% for *Precision*, *Recall*, *Accuracy* and *Jaccard Index*, respectively. The HiNN + SISR variant outperforms [15] by 4.9% in *Precision*, 11.2% in *Recall*, 3.6% in *Accuracy* and 10.7% in *Jaccard Index* on the same dataset. It is worth mentioning that [15] was trained on 130 and tested on 63 images. The proposed method is trained only on 77 (33%) and tested on 158 images, which implies the learning power of the proposed network. Furthermore, Table 1 suggests that the inner-ZP localization error here is less than half of that in Ref. [14]. Moreover, to compare performance of the proposed method with that of [4] for further evaluation, 80 early-stage human embryo (day 2 – 3 embryo) images are considered. According to Table 5, the proposed method delivers mean *Jaccard Index* of 91.0% that is comparable with *Jaccard Index* of 91.6% achieved by Ref. [4] where 60 and 20 images were used for training and testing, respectively. This result indicates that segmentation of ZP in early-stage embryo images is less of a challenge as they are less developed (i.e., thicker) when compared to blastocyst (late-stage embryo) images.

5.3. Validation of key components

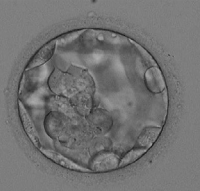
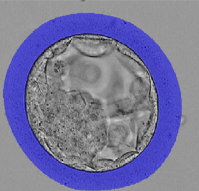
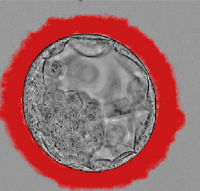
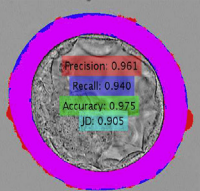
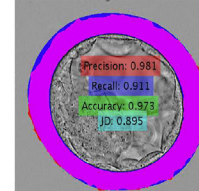
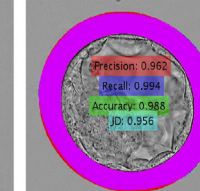
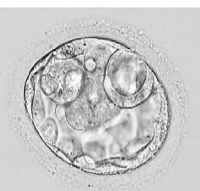
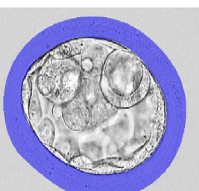

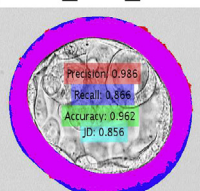
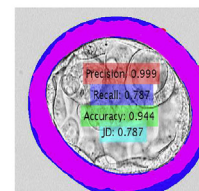
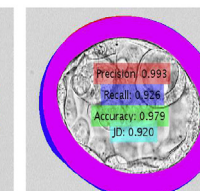

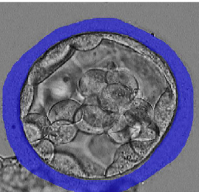

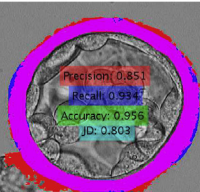
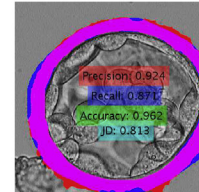
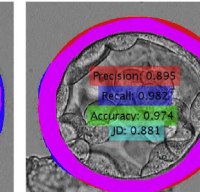

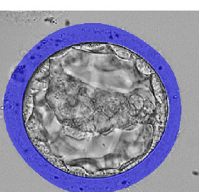
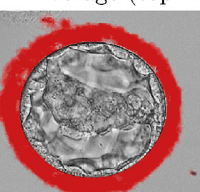
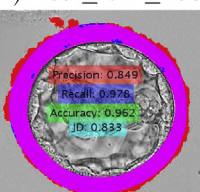
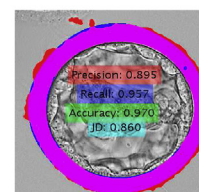
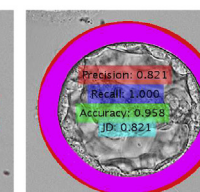
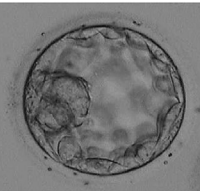
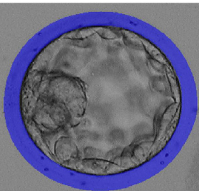
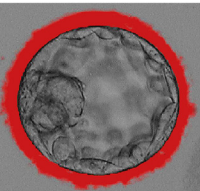
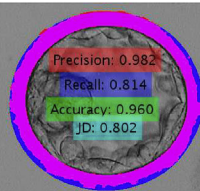
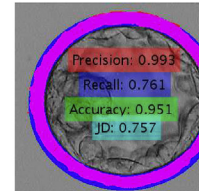
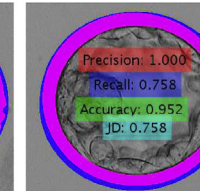

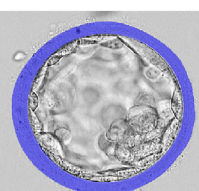
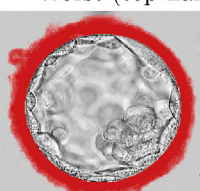
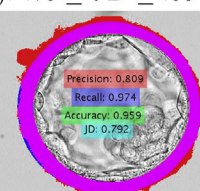
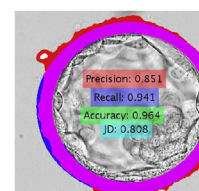
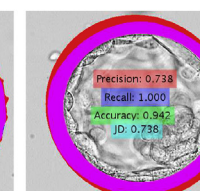
Here, the effect of each component of the proposed approach is studied using their contributions to the overall performance. First, all the metrics are computed using the proposed method on the test set. Then, the component under study is either removed or substituted with alternatives.

5.3.1. Effect of pre-processing and augmentation

As shown Table 4 both pre-processing and augmentation techniques employed in this work contributed to the better performance. All the

Table 3

Sample results for the proposed method (best viewed in color). Here, blue implies the ground truth, red highlights the identified ZP (TP + FP), and magenta depicts the TP..

	Original image	Ground truth	HiNN output	HiNN + Opt.Tre.	HiNN + Level-Set	HiNN + SISR
Best ($.90 \leq JI \leq 1.0$)						
	Best (top half): $.95 \leq JI \leq 1.0$					
						
	Best (bottom half): $.90 \leq JI \leq .95$					
Average ($.80 \leq JI \leq .90$)						
	Average (top half): $.85 \leq JI \leq .90$					
						
	Average (bottom half): $.80 \leq JI \leq .85$					
Worst ($.70 \leq JI \leq .80$)						
	Worst (top half): $.75 \leq JD \leq .80$					
						
	Worst (bottom half): $.70 \leq JD \leq .75$					

metrics dropped when either pre-processing or augmentation were eliminated from the pipeline. The lowest performance is achieved when no augmentation is applied where the number of training patches reduced three times.

5.3.2. Effect of hierarchical design

The effect of the proposed HiNN is evaluated by comparing the

performance of the HiNN as a whole with each of the branches. From Table 6, for higher branch depths, *Precision* increases while *Recall* decreases. This is because higher depth branches have better visual context over the image leading to better *Precision* values. On the other hand, lower depth branches have more detailed local information contributing to higher *Recall* values.

Table 4
Performance comparison of proposed method variants with state of the art.

	Precision	Recall	Accuracy	Jaccard Index
Kheradmand et al. [15]	.803	.808	.920	.674
HiNN	.794	.915	.951	.741
HiNN (No Augmentation)	.767	.892	.935	.702
HiNN (No PreProcessing)	.783	.908	.948	.726
HiNN + Opt.Tre.	.821	.911	.953	.755
HiNN + Level-Set	.847	.903	.953	.763
HiNN + SISR	.852	.920	.956	.781

Table 5
Performance on early-stage human embryo.

	Precision	Recall	Accuracy	Jaccard Index
Morales et al. [4]	–	–	–	.916
HiNN + SISR	.920	.989	.976	.910

Table 6
Performance breakdown for HiNN.

	Precision	Recall	Accuracy	Jaccard Index
1 st Branch	.746	.936	.949	.737
2 nd Branch	.753	.931	.945	.719
3 rd Branch	.770	.917	.949	.737
4 th Branch	.785	.900	.951	.738
5 th Branch	.807	.887	.951	.731
Mean	.776	.913	.949	.732
HiNN	.794	.915	.951	.741

5.3.3. Effect of SISR refinement strategy

The effect of the proposed SISR refinement on the ZP segmentation is evaluated by comparing with two alternative refinement strategies. Experimental results confirmed that the proposed SISR refinement strategy has improved the performance by 5.8% in *Precision*, 0.5% *Recall*,

0.5% in *Accuracy* and 4.0% in *Jaccard Index* compared to the widely used optimized threshold. Box plots for all evaluation criterion are presented in Fig. 8 to compare the proposed SISR with the other two alternative methods. These box plots facilitate a detailed comparison by describing the distribution using a six number summary: minimum, maximum, median, mean, second quartile box, and third quartile box.

6. Conclusion

In this paper, a system based on boosting ensemble of hybrid complementary learning was proposed to segment of Zona Pellucida in human blastocyst images. First, a deep Hierarchical Neural Network was proposed to segment Zona Pellucida in microscopic images of human blastocyst with a pre-processing method specifically designed for this application. The proposed method outperformed previous work by 4.9% in *Precision*, 11.2% in *Recall*, 3.6% in *Accuracy* and 10.7% in *Jaccard Index*. Second, a Self-supervised Image-Specific Refinement was proposed that can be utilized for other applications to maximize the performance, specially in medical image analysis, were labelled data for training is either limited or hard to collect. It was shown that ensemble of these two models boosts the performance and copes with artifacts and weak boundaries associated with the human blastocyst images. Furthermore, the proposed method can be employed to validate theoretical assumptions about relationship between ZP characteristics and pregnancy outcomes.

Ethical statement

None Declared.

Conflicts of interest

None Declared.

Acknowledgement

Not Applicable.

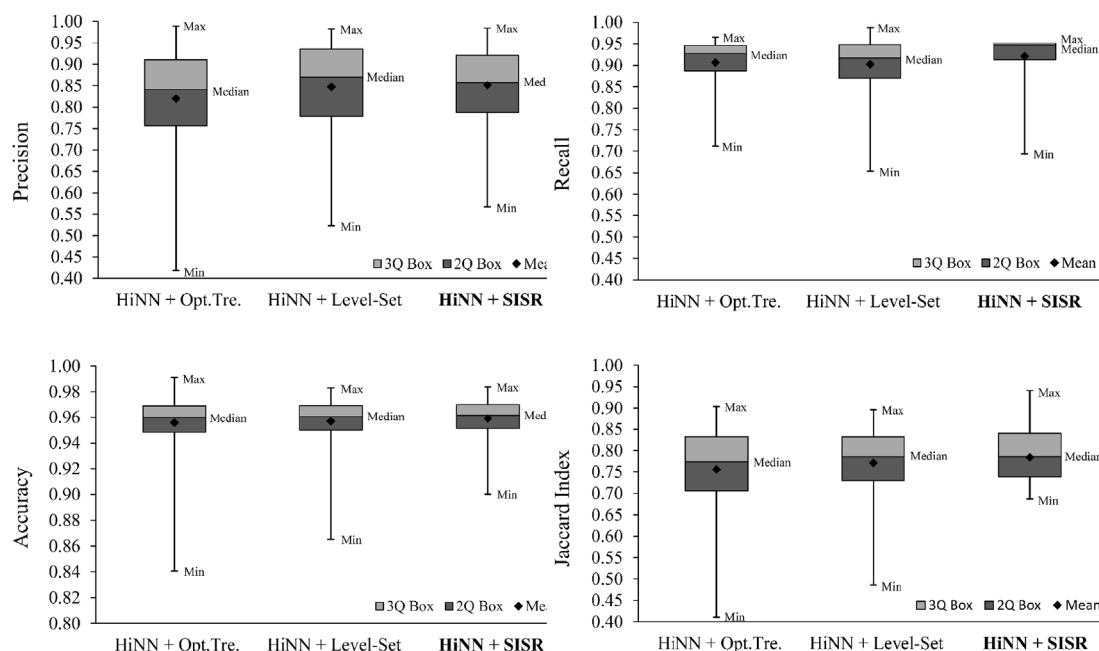


Fig. 8. Boxplots for a detailed comparison of various refinement strategies on *Precision*, *Recall*, *Accuracy* and *Jaccard Index*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2018.10.009>.

References

- [1] Bushnik T, Cook JL, Yuzpe AA, Tough S, Collins J. Estimating the prevalence of infertility in Canada. *Hum Reprod* 2013;27(3):738–46.
- [2] Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med* 2012;9(12):1–12. <https://doi.org/10.1371/j.pmed.1001356>.
- [3] Filho ES, Noble JA, Wells D. Estimating the prevalence of infertility in Canada. *Open Biomed. Eng J* 2010;4:170–7.
- [4] Morales DA, Bengoetxea E, Larranaga P. Automatic segmentation of zona pellucida in human embryo images applying an active contour model. *Proc. Of the 12th annu. Conf. On med. Image understanding and anal. MIUA*; 2008. p. 209–13.
- [5] Wang Z, Ang WT. Automatic zona pellucida dissection position selection for embryo biopsy in pre-implantation genetic diagnosis. *Proc. Of the IEEE int. Conf. On robotics and biomimetics (ROBIO)*. 2014.
- [6] Wang Z, Ang WT, Tan SYM, Latt WT. Automatic segmentation of zona pellucida and its application in cleavage-stage embryo biopsy position selection. *Proc. Int. Conf. Of the IEEE eng. In medicine and biology soc. EMBC*; 2015. p. 3859–64.
- [7] Wang Z, Ang WT. Automatic dissection position selection for cleavage-stage embryo biopsy. *IEEE Trans Biomed Eng* 2016;63(3):563–70.
- [8] Tomsu M, Sharma V, Miller D. Embryo quality and IVF treatment outcomes may correlate with different sperm comet assay parameters. *Hum Reprod* 2002;17(7):1856–62. [arXiv:oup/backfile/content_public/J./humrep/17/7/10.1093/humrep.17.7.1856/1/171856.pdf](https://arxiv.org/abs/1707.18561).
- [9] Paternot G, Wetzels AM, Thonon F, Vansteenbrugge A, Willems D, Devroe J, Debrock S, D'Hooghe TM, Spiessens C. Intra- and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: a multicentre study. *Reprod Biol Endocrinol* 2011;9(1):127.
- [10] Thompson SM, Onwubalili N, Brown K, Jindal SK, McGovern PG. Blastocyst expansion score and trophoctoderm morphology strongly predict successful clinical pregnancy and live birth following elective single embryo blastocyst transfer (ESET): a national study. *J. of assisted reproduction and genetics* 2013;30(12):1577–81.
- [11] Zhao J, Zhang NY, Xu ZP, Chen LJ, Zhao X, Zeng HM, Jiang YQ, Sun HX. Effects of abnormal zona pellucida on fertilization and pregnancy in IVF/ICSI-ET. *Reprod Contracept* 2015;26(2):73–80.
- [12] Karlsson A, Overgaard NC, Heyden A. Automatic segmentation of zona pellucida in hmc images of human embryos. *Proc. Int. Conf. On pattern recognition (ICPR)*, vol. 3. 2004. p. 518–21.
- [13] Karlsson A, Overgaard NC, Heyden A. A two-step area based method for automatic tight segmentation of Zona Pellucida in HMC images of human embryos. *Berlin, Heidelberg: Springer Berlin Heidelberg*; 2005. p. 503–14.
- [14] Yee D, Saeedi P, Havelock J. An automatic model-based approach for measuring the zona pellucida thickness in day five human blastocysts. *Proc. Int. Conf. On image process., comput. Vision and pattern recognition (IPCV)*. 2013. p. 877–80.
- [15] Kheradmand S, Saeedi P, Bajic I. Human blastocyst segmentation using neural network. *Proc. IEEE canadian conf. On elect. And comput. Eng. CCECE*; 2016. p. 1–4.
- [16] Rad RM, Saeedi P, Au J, Havelock J. A hybrid approach for multiple blastomeres identification in early human embryo images. *Comput Biol Med* 2018;101:100–11. <https://doi.org/10.1016/j.compbiomed.2018.08.001>.
- [17] P. Saeedi, D. Yee, J. Au, J. Havelock, Automatic identification of human blastocyst components via texture, *IEEE Trans Biomed Eng*.
- [18] Wang Z, Feng C, Ang WT, Tan SYM, Latt WT. Autofocusing and polar body detection in automated cell manipulation. *IEEE Trans on Biomed Eng* 2017;64(5):1099–105.
- [19] Wong CY, Mills JK. Automation and optimization of multipulse laser zona drilling of mouse embryos during embryo biopsy. *IEEE Trans on Biomed Eng* 2017;64(3):629–36.
- [20] Kheradmand S, Singh A, Saeedi P, Au J, Havelock J. Inner cell mass segmentation in human embryo images using fully convolutional network. *Proc. Of the IEEE int. Conf. On image process. ICIP*; 2017.
- [21] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *CoRR abs/1409.1556*.
- [22] Rad RM, Saeedi P, Au J, Havelock J. Multi-resolutional ensemble of stacked dilated u-net for inner cell mass segmentation in human embryonic images. 2018 25th IEEE international conference on image processing (ICIP) 2018. p. 3518–22. <https://doi.org/10.1109/ICIP.2018.8451750>.
- [23] Rad RM, Saeedi P, Au J, Havelock J. Blastomere cell counting and centroid localization in microscopic images of human embryo. 2018 IEEE 20th international workshop on multimedia signal processing (MMSp). 2018. p. 1–5.
- [24] Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. *Berlin, Heidelberg: Springer Berlin Heidelberg*; 1998. p. 130–7.
- [25] Li C, Gore JC, Davatzikos C. Multiplicative Intrinsic Component Optimization (MICO) for MRI bias field estimation and tissue segmentation. *Magnetic resonance Imag* 2014;32(7):913–23.
- [26] Li Q, Feng B, Xie L, Liang P, Zhang H, Wang T. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans on Med Imag* 2016;35(1):109–18.
- [27] A. Dosovitskiy, J. T. Springenberg, M. A. Riedmiller, T. Brox, Discriminative unsupervised feature learning with convolutional neural networks, *CoRR abs/1406.6909*.
- [28] Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. *Proc. Int. Conf. On mach. Learn. (ICML)*, vol. 30. 2013.
- [29] Nguyen D, Widrow B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proc. Int. Joint conf. On neural networks (IJCNN)*. 1990. p. 21–6.
- [30] Bengio Y, Yao L, Alain G, Vincent P. Generalized denoising auto-encoders as generative models. *Proc. Int. Conf. On neural inform. Process. Syst. (NIPS)*. 2013. p. 899–907.
- [31] Alain G, Bengio Y. What regularized auto-encoders learn from the data-generating distribution. *J Mach Learn Res* 2014;15(1):3563–93.
- [32] Song Y, Zhang L, Chen S, Ni D, Lei B, Wang T. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. *IEEE Trans Biomed Eng* 2015;62(10):2421–33.
- [33] Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Trans on Med Imag* 2016;35(5):1240–51.
- [34] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans on Med Imag* 2016;35(11):2369–80.
- [35] Wang L, He L, Mishra A, Li C. Active contours driven by local Gaussian distribution fitting energy. *Signal Process* 2009;89(12):2435–47. [special Section: Visual Inform. Anal. for Security].
- [36] Li C, Xu C, Gui C, Fox MD. Level set evolution without re-initialization: a new variational formulation. *Proc. IEEE computer soc. Conf. On computer vision and pattern recognition (CVPR)*. 2005. p. 430–6.
- [37] Yu L, Guo Y, Wang Y, Yu J, Chen P. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. *IEEE Trans Biomed Eng* 2017;64(8):1886–95.