Zona pellucida thickness measurement via instance segmentation using compound loss functions

Ming-Kai Cheng
Department of Electrical Engineering
National Taiwan Ocean University
Keelung 20224, Taiwan
kyle1kyle4@gmail.com

Ming-Jer Chen
Division of Infertility
LeeWomen's Hospital
Taichung 40705, Taiwan
mingjerchen@gmail.com

Yu-Chiao Yi
Division of Reproductive Endocrinology and Infertility
Veterans General Hospital (VGHTC)
Taichung 40705, Taiwan
yuchiaoyi@gmail.com

Shih-Kai Lee
Department of Electrical Engineering
National Taiwan Ocean University
Keelung 20224, Taiwan
wax78216@gmail.com

Ren-Jie Huang
Department of Electrical Engineering
National Taiwan Ocean University
Keelung 20224, Taiwan
jajs8869@gmail.com

Jung-Hua Wang*
AI Research Center
National Taiwan Ocean University
Keelung 20224, Taiwan
jhwang@email.ntou.edu.tw

***Abstract*—This paper focuses on the measurement of zona pellucida thickness (ZPT) in In Vitro Fertilization(IVF) treatment. A novel compound loss functions-based instance segmentation approach, which is useful for measuring ZPT for time-lapse incubator (TLI) images of embryos, is proposed. Many previous studies conjecture that ZPT, as one salient feature, can be used in conjunction with TLI images for training machine learning (ML) models to assess the outcome of IVF treatment. The performance of instance segmentation can be greatly improved by using the compound loss functions presented in this paper. Ablation experiments conducted on loss functions of Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), and IoU Boundary (B-IoU) are used to identify the optimal combination for segmenting TLI images. Our method allows the deep learning model to adapt to task objective differences, resulting in better segmentation outcomes that are justified through the visualization of segmentation results. Using YOLOv7 as a demonstrative ML model, parametric features of ZPT and its variance (ZPTV) are derived. Although there indeed exists a correlation relation between pregnancy outcome and ZPT (ZPTV as well), significance testing on 57 embryos (20 pregnant, 37 not pregnant) shows that p-values are all greater than 0.05 for all ZPT-related features.**

Keywords—Compound loss functions, Deep learning, Zona pellucida thickness, Vitro Fertilization, Instance Segmentation

I. INTRODUCTION

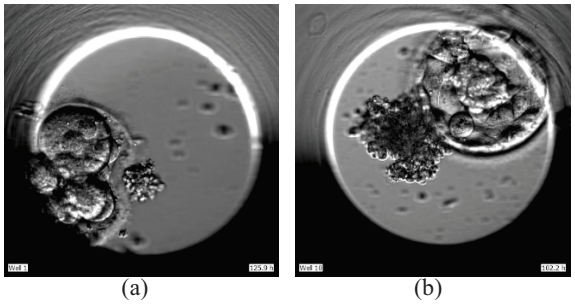
Zona pellucida thickness (ZPT) has been studied in infertility treatment research for nearly two decades, with numerous relevant studies intermittently published during this period. As early as 2001[1], research confirmed the correlation between ZPT and pregnancy rates, demonstrating that combining information on ZPT variations increased the probability of clinical pregnancy by 2.5 times for embryos with great scores compared to embryos with less-than-ideal scores based on classic embryo grading standards. By 2012[2], studies began focusing on the impact of ZPT on fertilization failure and its effect on implantation and live births in rabbits (Oryctolagus cuniculus). While ZPT significantly influences pregnancy processes, it does not affect live births. Previous research suggests that changes in zona pellucida thickness before embryo implantation may impact pregnancy outcomes.

Recently, more and more infertility treatment research has adopted embryo imaging in studies, mainly using static microscopic images lacking temporal information. We can barely gather insight into the embryo-growing procedure through these images. This information gap complicates embryo development observation, leading many medical facilities to adopt time-lapse incubators for cultivating embryos during treatment. Time-lapse incubators(TLI) provide a stable environment for embryo cultivation, enabling continuous observation and recording of the entire growth process without disrupting development. Distinct from traditional microscopic images, TLI generates time-lapse videos with temporal information that can enhance our understanding of embryo development.

Hence, analyzing crucial growth process information in time-lapse videos might be a trend in future research in infertility treatment. As emphasized in prior work, it's important to identify more temporally relevant features in TLI images to aid in related studies[3]. For instance, we can use instance segmentation through TLI images, gathering the area of zona pellucida (ZP) and computing the ZPT using the segmentation result, thereby obtaining the zona pellucida thickness variance (ZPTV) at any given time point.

Due to the vast number of time-lapse incubator images recorded, it's difficult for us to measure ZPTV or specific ZPT at different embryo growth stages one by one. To handle this, we employed deep learning instance segmentation methods to segment complete embryos from TLI images and analyze them further.

TLI images offer advantages in temporal information compared to static microscopic images but pose challenges from a fixed viewing angle, resulting in embryo container masking by the edge. Embryos obscured by the edge of the cell-culture dish or sperm clusters (Fig. 1(a), (b)), can complicate observation, called “partially occluded data



(a) (b)

Fig. 1 (a) Embryo image corrupted by the edge of the culture dish
(b) Images corrupted by the edge of the culture dish and sperm clusters.

2835-5776/24/$31.00 ©2024 IEEE
DOI 10.1109/IRI62200.2024.00057

(POD)". Segmenting such images presents difficulties and may lead to inaccuracies in embryo range or ZPT calculations. It's difficult to solve the issue of POD by using the original YOLOv7[4][5] instance segmentation model. Through review of various loss functions for medical image instance segmentation[6] clearly indicates that no single loss function can perform consistently well for all task objectives. However, compound loss functions are robust losses. Therefore, our study focuses on enhancing instance segmentation performance, specifically in the case of POD appearance, by refining the loss function module.

While recent research focuses on modifying network architectures[7][8][9][10], our approach prioritizes optimizing the loss function block to enhance instance segmentation performance. This addresses the practical challenge of the POD issue in TLI imaging. Our method offers a robust solution to this issue, ensuring segmentation accuracy even with POD. This paper proposes a novel approach to enhance instance segmentation performance in TLI images, focusing on optimizing the loss function block. This addresses a crucial practical issue and provides a robust, accurate embryo component segmentation solution.

II. MATERIALS AND METHODS

This study analyzes the association between the sequence of zona pellucida thickness (ZPT) value across all stages of embryo growth and pregnancy outcome. We design a series of processes (Fig. 2) to obtain the sequence of ZPT value by performing instance segmentation on TLI images and then calculating the ZPT.

First, we optimize the YOLOv7 instance segmentation model performance with compound loss functions during model training. Then we put TLI images into the optimal model, for embryo segmentation. The segmentation results are divided into two masks: $MaskA$ representing the entire embryo and $MaskB$ representing all areas within the zona pellucida. Subtracting $MaskB$ from $MaskA$ yields $MaskZP$, which is input into the calculation of ZPT. After obtaining the ZPT, statistical analysis is conducted to remove outliers that may cause errors, and the above steps are repeated to obtain sequences of ZPT values for all embryos throughout their growth process.

A. Data collection

Our dataset used in this study for deep learning algorithms consists of de-identified TLI embryo image provided between 2021 to 2023 by the Department of Obstetrics and Gynecology at the Veterans General Hospital Taichung Branch (VGHTC) in Taiwan. According to Article 12, paragraph 2 of the Taiwan Human Research Act, "*Research protocol shall obtain the consent of participating research subjects as approved by the IRB(Institutional Review Board). But the research protocol within the scope of exemption categories for consent requirements, as announced by the competent authority, shall not apply*." The Taiwan Centers for Disease Control and Prevention issued an official document (code:1010265083). The method of obtaining the research data complies with the second and third items of the above announcement, so Waiver of informed consent. The annotation of this proprietary dataset was assisted by the expertise of three embryo specialists within the hospital.

B. Instance segmentation

Currently, mainstream AI image segmentation models are divided into semantic and instance segmentation. Semantic segmentation focuses on labeling image pixels by category but lacks differentiation between different instances. Instance segmentation not only considers pixel-level class labels but distinguishes between different instances of the same object. In this study, we plan to utilize the state-of-the-art instance segmentation technique YOLOv7 to obtain masks for segmenting various parts of embryos in TLI images.

C. Loss function

In deep learning, the loss function is crucial in guiding model training. It measures the difference between model predictions and actual targets and transforms this difference into an optimizable objective. A suitable loss function effectively guides model parameter optimization, enabling the model to converge quickly and accurately to the optimal solution. Conversely, an unsuitable or poorly designed loss function can lead to training difficulties or converging failure.

Therefore, choosing an appropriate loss function is essential for model performance and training success. Loss functions can also be customized according to different tasks and requirements. For various application scenarios, different forms of loss functions may be needed to address specific problems, such as classification, regression, bounding box detection, or segmentation. The following discussions will elaborate on the composition of the compound loss functions used in this study and their rationality.

$$Compound\ Loss = BCE\ Loss + DoU\ Loss + Dice\ Loss$$

1) BCE(Binary cross entropy) Loss

The BCE Loss is the default primary loss function in YOLOv7, designed for binary classification tasks, which aligns well with the instance segmentation task. While doing instance segmentation, each pixel is classified as foreground or background. BCE loss provides a direct and effective way to measure the difference between predicted and accurate pixel labels at the pixel level. Additionally, BCE loss

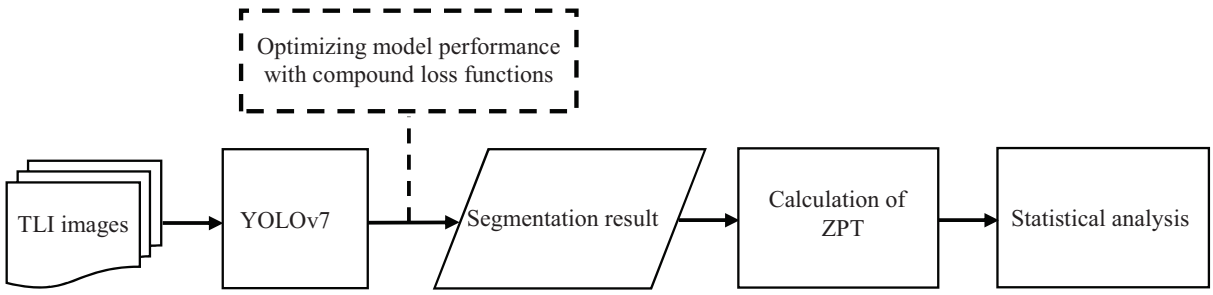


Fig. 2 Flowchart of the proposed mathod of measuring zona pellucida thickness.

independently calculates the loss for each pixel, allowing the model to focus on correctly classifying individual pixels rather than the entire image. This is crucial in instance segmentation, where accurately delineating objects at the pixel level is the primary objective.

2) DoU(Boundary difference over union) Lreoss

The main task of this study is to segment the entire embryo region in TLI images. Although BCE loss has performed well in general image segmentation tasks, it fails to achieve the expected performance in scenarios where sperm clusters and the edge of the cell-culture dish occur(Fig. 1). This error severely affects subsequent ZPT calculations.

This study identifies this problem as boundary segmentation and aims to improve the accuracy of boundary segmentation by introducing the DoU loss[11]. DoU loss is designed to provide an effective and stable calculation method for boundary segmentation tasks (1). It can even adaptively adjust the attention to the boundary based on the size of the detected target (2). Since the proportion of each boundary area to the entire target varies for different-sized boundary areas, when the segmentation target is larger, the proportion of the boundary is relatively smaller, making the target's interior easier to segment. In such cases, using a larger α value to focus more on boundary segmentation is preferable. Conversely, the interior and boundary areas are difficult to segment when the segmented target is smaller. In such cases, a smaller α value is preferable. Here, C represents the length of the target boundary, and S represents the size of the boundary.

$$L_{DoU} = \frac{G \cup P - G \cap P}{G \cup P - \alpha * G \cap P} \tag{1}$$

$$\alpha = 1 - 2 \times \frac{C}{S},\ \alpha \in [0, 1), \tag{2}$$

3) Dice Loss

In time-lapse incubator (TLI) images, embryos undergo contractions and expansions over time. In later stages, embryos may expand to sizes close to the edge of the cell-culture dish's edge(Fig. 1), with texture features significantly different from before expansion(Fig.3). Considering the changing foreground-background ratio over time, we chose Dice loss as a component of the compound loss function.

Dice loss calculates the similarity between segmentation results without relying on pixel class(3), effectively handling situations where foreground-background ranges vary over time. This metric measures the similarity between predicted and actual labels, providing an intuitive understanding of model training and parameter optimization. This enables the model to adjust effectively, thereby enhancing instance segmentation accuracy.

$$Dice\ Loss = 1 - \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3}$$

D. Zona pellucida thickness calculation

In this study, two segmentation models were trained on time-lapse incubator (TLI) images. Through the stage of

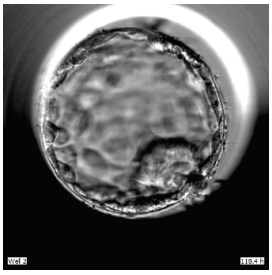


Fig. 3 Expanded-stage embryos.

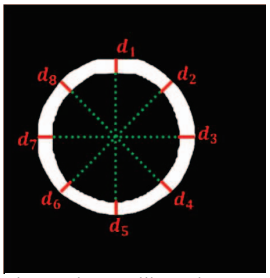


Fig. 5 Diagram illustrating zona pellucida thickness calculation.

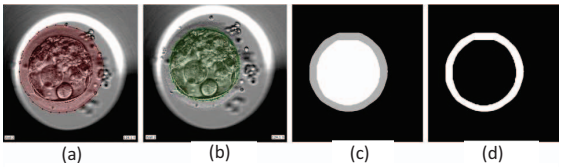


Fig. 4 Illustration demonstrating the process of obtaining the zona pellucida mask.

calculating ZPT(see **Pseudo code** below), where the parameter "i" represents the number of embryos in the test data, and parameter "j" represents the total number of TLI images for a particular embryo, $MaskA$ represents the mask for the complete embryo(Fig. 4(a)). $MaskB$ represents all areas within the zona pellucida(Fig. 4(b)). Subtracting $MaskB_{ij}$ from $MaskA_{ij}$ (Fig. 4(c)) yields the zona pellucida mask, $MaskZP_{ij}$ (Fig. 4(d)), $MaskZP_{ij}$ is then inputted into $Function1$ to determine the minimum enclosing circle and the centroid, C_{ij}, located within the zona pellucida. $MaskZP_{ij}$ and C_{ij} are subsequently fed into $Function2$, which computes the number of overlapping pixels between $MaskZP_{ij}$ and radial rays drawn every θ degrees around C_{ij}(Fig. 5). In this study, θ is set to 10°, but here, the θ is set to 45° for explanatory purposes.A scale conversion converts the resulting pixel count to actual zona pellucida thickness. The ZPT values obtained from eight different radial angles are averaged to obtain d_{ij}. The maximum value of d_{ij} is taken as $D_{i_{max}}$, and the minimum

Pseudo code for computing ZPT

$Input$: $\{MaskA_{ij}, MaskB_{ij}\}$, i=1, 2,..k, j=1, 2,..l
$Output$: D_{max}、D_{min}、$ZPTV$

```
for i = 1:k
    for j = 1:l
        MaskZP_ij = MaskA_ij − MaskB_ij
        C_ij = Function1(ZP_ij)
        d_ij = Function2(ZP_ij, C_ij)
    D_imax = max(d_i1, d_i2, ⋯, d_il)
    D_imin = min(d_i1, d_i2, ⋯, d_il)
    ZPTV_i = D_imax − D_imin
D_max = {D_1max, D_2max, ⋯, D_kmax}
D_min = {D_1min, D_2min, ⋯, D_kmin}
ZPTV = {ZPTV_1 + ZPTV_2 + ⋯ + ZPTV_k}
```

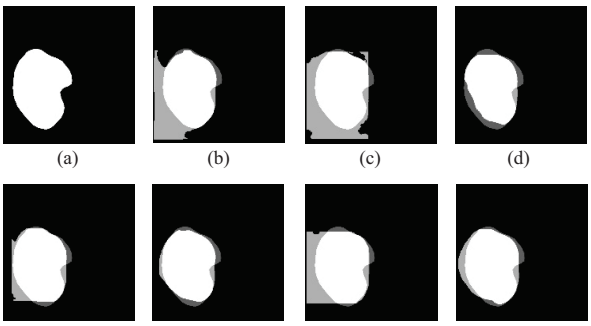


Fig. 6 Visual representation of the instance segmentation results for the loss function ablation experiment in Fig. 1(a); (a)Ground truth, (b)BCE, (c)Dice, (d)DoU, (e)BCE+Dice, (f)BCE+DoU, (g)Dice+DoU, (h)BCE+Dice+Dou

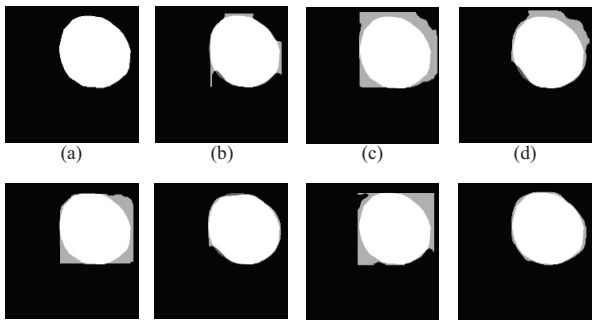


Fig. 7 Visual representation of the instance segmentation results for the loss function ablation experiment in Fig. 1(b); (a)Ground truth, (b)BCE, (c)Dice, (d)DoU, (e)BCE+Dice, (f)BCE+DoU, (g)Dice+DoU, (h)BCE+Dice+Dou

value is $D_{i_{min}}$. The difference between $D_{i_{max}}$ and $D_{i_{min}}$ is calculated as $ZPTV_i$.

E. Significance test of data pre-processing

Our study employed quantitative statistical analysis methods to process the data. A 95% confidence interval was utilized to mitigate the influence of outliers during the data preprocessing of ZPT measurements. Extreme values beyond this interval were excluded from the analysis to ensure the robustness and reliability of the results. After measuring ZPT, three preliminary sets of data were compiled: the 95% average D_{max}, 95% average D_{min}, and 95% average $ZPTV$, made of 57 sets of ZPT data, along with corresponding pregnancy outcomes, were subjected to chi-square tests. The chi-square statistic and degrees of freedom were calculated to determine the p-value[12], evaluating their association. This study evaluates the segmentation results of Testsets A, B, and C by three metrics, including the Dice Similarity Coefficient (DSC) [13]、Hausdorff Distance(HD)[14], and IoU Boundary(B-IoU)[15].

III. RESULT

A. Experiment environment

Hardware: CPU i7-11700 with NVIDIA GPU/RTX-3090 and RAM/32G. Software: Windows11, Python 3.9.13, Cuda 11.3, cuDNN 8.2.1, Pytorch 1.11.0.

B. Qualitative results

Our study used 229 test images with a resolution of 500x500 pixels. The testing dataset included 83 TLI images of embryos that were occluded by the edge of the cell-culture dish or sperm clusters (Testset A) and 146 unobstructed embryo TLI images (Testset B). The combined Testset A+B consists of a total of 229 TLI images (Testset C). The dataset used in chi-square tests comprised 57 embryos with clear pregnancy outcomes: 20 were pregnant (Group 1), and 37 were not pregnant (Group 2).

Ablation experiments were conducted on the employed loss functions ($BCE\ Loss,\ DoU\ Loss,\ Dice\ Loss$). We selected $MaskA$ as the result of qualitative visualizations of segmentation, which can be used in all IVF treatment-related studies. Compared to using single-loss functions, compound loss functions optimized the instance segmentation results, particularly for handling POD. Fig. 1(a) and 1(b) show the original images corresponding to Fig. 6 and Fig. 7, respectively. Fig. 1(a) illustrates an embryo occluded by the edge of the cell-culture dish, while Fig. 1(b) shows embryos occluded by sperm clusters and the edge of the cell-culture dish simultaneously.

We use the instance segmentation results(Fig. 6 and Fig. 7) of these two TLI images, Fig. 1(a) and 1(b)), to serve as examples to illustrate the outcomes of this study. Fig. 6 and Fig. 7, respectively demonstrate different instance segmentation results in the ablation experiment. Fig.6, 7 (a) represent a ground truth label, while Fig.6, 7 (b) to (h) represent the segmentation results based on different loss function combinations. Fig. 6(d), (f), and (h) are closely match the actual label ranges. The superior result in Fig. 7(d) was obtained using a single loss function model (DoU Loss). Upon revisiting Fig. 1(a), it is evident that sperm clusters did not heavily occlude this image, and the boundary obscured by the edge of the cell culture dish is relatively smooth. Since DoU Loss is specifically designed for boundary segmentation, the comparable segmentation performance of this model, as good as other compound loss function models, is foreseeable.

In Fig. 1(b), the edge of the cell-culture dish and a sperm cluster obstruct the embryo, causing significant interference. Despite these challenges, the segmentation results in Fig. 7(f) and (h) closely match the ground truth label, indicating robust performance. However, Fig. 7(d) falls short compared to Fig. 7(f) and (h), highlighting the limitations of using a single loss function when compound interfering factors are present.

To demonstrate the effectiveness of our algorithm optimization, we compare $MaskZP$ result before and after using our method. Fig. 8(a) shows the interference caused by both the edge of the cell-culture dish and sperm clusters, while Fig. 8(b) shows a deformed ZP mask due to interference. After applying our method, Fig. 8(c) reveals a more accurate ZP instance segmentation result, significantly reducing thickness calculation errors.

C. Quantitative results

By examining TABLE Ⅰ, TABLE Ⅱ, TABLE Ⅲ we found that the compound loss functions of the BCE+DOU+Dice model and the BCE+DOU model perform best for $MaskA$ segmentation, achieving higher DSC and B-IoU scores. Specifically, the BCE+DOU+Dice model achieves over 90% DSC similarity in Testset A(TABLE I), indicating its superiority over single loss function models in instance segmenting of POD. However, for $MaskB$

TABLE I The performance of quantitative metrics on Test Dataset A

	MaskA			*MaskB*			*MaskZP*		
	***DSC*↑**	***HD*↓**	***B-IoU*↑**	***DSC*↑**	***HD*↓**	***B-IoU*↑**	***DSC*↑**	***HD*↓**	***B-IoU*↑**
BCE	88.21	8.78	79.7	93.16	5.65	87.59	65.12	9.15	49.98
DOU	89.94	**6.30**	82.33	**94.27**	**5.29**	**89.41**	**68.51**	7.80	**53.23**
Dice	84.97	7.01	74.48	86.50	6.32	77.43	49.69	9.96	34.38
BCE+Dice	88.37	6.90	79.93	91.49	5.62	84.81	63.43	8.46	47.68
BCE+DOU	89.84	6.76	82.22	92.94	5.35	87.26	67.40	7.88	52.19
Dice+DOU	87.02	6.75	77.65	89.71	5.98	81.31	57.27	8.64	41.35
BCE+DOU+Dice	**90.44**	6.33	**83.1**	91.27	5.82	84.46	65.78	**7.80**	49.98

TABLE II The performance of quantitative metrics on Test Dataset B

	MaskA			*MaskB*			*MaskZP*		
	***DSC*↑**	***HD*↓**	***B-IoU*↑**	***DSC*↑**	***HD*↓**	***B-IoU*↑**	***DSC*↑**	***HD*↓**	***B-IoU*↑**
BCE	96.14	4.77	92.67	98.24	3.00	96.53	83.98	5.86	73.17
DOU	96.04	4.61	92.51	98.14	3.46	96.34	83.58	**5.71**	72.60
Dice	91.48	6.21	84.49	94.14	6.93	88.92	63.23	8.84	47.62
BCE+Dice	95.43	4.63	91.43	98.05	3.00	96.17	82.42	5.82	71.15
BCE+DOU	**96.24**	4.67	92.86	**98.47**	3.46	**96.99**	**84.52**	5.83	**73.88**
Dice+DOU	94.31	4.54	89.47	96.56	3.16	93.36	77.23	6.29	64.17
BCE+DOU+Dice	96.09	**4.61**	**95.59**	98.17	**3.00**	96.40	83.51	5.78	72.43

TABLE III The performance of quantitative metrics on Test Dataset C

	MaskA			*MaskB*			*MaskZP*		
	***DSC*↑**	***HD*↓**	***B-IoU*↑**	***DSC*↑**	***HD*↓**	***B-IoU*↑**	***DSC*↑**	***HD*↓**	***B-IoU*↑**
BCE	93.45	6.13	88.27	96.44	4.11	93.36	77.81	6.90	65.60
DOU	93.97	**5.18**	89.06	**96.73**	**4.08**	**93.83**	78.31	6.42	65.86
Dice	89.25	6.48	81.07	90.92	6.13	83.96	58.24	9.26	42.81
BCE+Dice	93.04	5.40	87.53	95.73	4.10	92.17	75.61	6.78	62.79
BCE+DOU	94.07	5.37	89.25	96.28	4.13	93.10	**78.47**	6.54	**66.21**
Dice+DOU	91.77	5.31	85.35	93.87	4.49	88.99	69.69	7.18	55.64
BCE+DOU+Dice	**94.02**	5.25	**89.12**	95.35	4.30	91.76	78.36	**6.36**	65.92

TABLE IV Average D_{max} 、Average D_{min} 、Average $ZPTV$ 、Percentage of Average $ZPTV$ values

	Average D_{max}	**Average D_{min}**	**Average $ZPTV$**	**Percentage of Average $ZPTV$**
Group 1	15.84	4.77	11.07	69.59%
Group 2	16.03	5.63	10.40	64.94%

TABLE V P-value result between Average D_{max} 、Average D_{min} 、Average $ZPTV$, and pregnancy outcomes

	Average D_{max}	**Average D_{min}**	**Average $ZPTV$**
p-value	0.332	0.327	0.363

segmentation, the compound loss function only performs optimally in Testset B(TABLE II). Notably, experimental combinations using the Dice loss function for segmenting $MaskB$ in Testset A (TABLE I) demonstrate poor performance, attributed to the lack of labels for images that occluded by sperm clusters and the edge of the cell-culture dish in the training data for the $MaskB$ model. In the case of using Dice loss, which calculates similarity based on segmentation results rather than pixel classification, the advantage mentioned in (II.C.3)) actually becomes a disadvantage here. Consequently, the instance segmentation models, which include Dice loss, result in lower generalization ability when segmenting $MaskB$ in Testset A and C(TABLE I, TABLE Ⅲ).

Overall, the compound loss function model HD, DSC, and B-IoU performance surpasses that of single-loss function models. These findings highlight the effectiveness of applying a compound loss function model to segment POD. Note that$MaskZP$ is indirectly obtained through the Algorithm for Measuring ZPT, representing the combined performance of $MaskA$ and $MaskB$. Therefore, we focus on providing objective performance evaluation metrics as a reference for performance assessment, refraining from commenting on the direct impact of the instance segmentation training process on $MaskZP$.

D. Statistical analysis

We also conducted statistical analysis on D_{max}, D_{min}, $ZPTV$, values. These analyses produced average D_{max}, average D_{min}, average $ZPTV$, then analyzed via Chi-square testing against pregnancy outcomes. Results in TABLE IV reveal that pregnant embryos had larger average $ZPTV$ values but smaller average D_{max} and average D_{min} values than non-pregnant ones.

Despite TABLE V showing no significant difference in p-values, Group 1 demonstrated a smaller average D_{min} and larger average $ZPTV$ than Group 2, which is consistent with the literature[16]. We acknowledge that this prospective study's sample size is currently limited. As TLI image data with pregnancy outcomes increases, more considerable datasets applied to our methods might offer a different perspective or results.

IV. CONCLUSION

Significant improvements were observed in segmenting $MaskA$ on Testset C, with the DSC increasing to 94.07, HD decreasing to values below 5.4, and B-IoU increasing to 89.25.

(a) (b) (c)

Fig. 8 The comparison of $MaskZP$ before and after applying our method (a) Ground truth, (b) Segmentation result of original model, (c) Segmentation result of our model

In Testset A, the DSC improved to 90.44, HD decreased to values not exceeding 6.9, and B-IoU increased to 83.1. Visualized instance segmentation results also demonstrated notable enhancements (Fig. 8(b) 、8(c)).

A key aspect of our study is the use of compound loss function in improving the performance of deep learning instance segmentation models. Specific weight coefficients were not assigned to the compound loss function components, as determining optimal coefficients requires extensive experimentation. Given hardware limitations and the significant time cost of training deep learning models, further optimization methods can be explored in the future. We also examined the correlation between ZPT and pregnancy rates and calculated p-values for ZPT parameters that did not show significant differences. Pregnant embryos exhibited greater average changes in ZPT during growth, as well as both smaller maximum and minimum thicknesses over the non-pregnant embryos, which is in line with previous studies[2][17]. Although a correlation exists between pregnancy outcome and ZPT (ZPTV), significance testing on 57 embryos (20 pregnant, 37 not pregnant) shows that p-values are all greater than 0.05 for all ZPT-related features. There still seems to be room for discussion on the impact of zona pellucida thickness on pregnancy outcomes. Nowadays, many features in TLI images remain unexplored. The greatest strength of this study is that, as a practical analysis tool, the method proposed in this article can analyze other key information in TLI images that may affect infertility treatment intuitively and interpretably.

References

[1] A. Gabrielsen, et al, “The impact of the zona pellucida thickness variation of human embryos on pregnancy outcome in relation to suboptimal embryo development. A prospective randomized controlled study,” *Human Reproduction*, Vol.16, No.10, pp.2166-2170, 2001.

[2] F. Marco-Jiménez, et al, “Influence of zona pellucida thickness on fertilization, embryo implantation and birth,” *Animal reproduction science*, Vol.132, No.1-2, pp.96-100, 2012.

[3] Y. Alhelou, et al., “Novel image analyzer-assisted morphometric methodology offer unique opportunity for selection of embryos with potential for implantation,” *Bmc pregnancy and childbirth*, Vol. 23, No. 1, 2023.

[4] C. Y. Wang, et al, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[5] C. Y. Wang, et al., “Designing Network Design Strategies Through Gradient Path Analysis,” *Journal of Information Science and Engineering*, 2023.

[6] J. Ma, et al., “Loss odyssey in medical image segmentation,” *Medical image analysis*, Vol. 71, 2021.

[7] M. Ishaq, et al., “Assisting the Human Embryo Viability Assessment by Deep Learning for In Vitro Fertilization,” *Mathematics.* Vol. 11, No.2023, 2023.

[8] A. Mushtaq, et al., “Artificial Intelligence-Based Detection of Human Embryo Components for Assisted Reproduction by In Vitro Fertilization,” *Sensors*, Vol. 22, No. 7418, 2022.

[9] M. Arsalan, et al., “Human Blastocyst Components Detection Using Multiplescale Aggregation Semantic Segmentation Network for Embryonic Analysis,” *Biomedicines*, Vol. 10, No. 1717, 2022.

[10] M. Arsalan, et al., “Detecting Blastocyst Components by Artificial Intelligence for Human Embryological Analysis to Improve Success Rate of In Vitro Fertilization,” *Journal of personalized medicine,* Vol. 12, No. 124, 2022.

[11] F. San, et al., “Boundary Difference over Union Loss for Medical Image Segmentation,” *Medical Image Computing and Computer-Assisted Intervention*, pp.292-301.

[12] Y. Benjamini, et al., “ASA President’s Task Force statement on statistical significance and replicability,” Chance, Vol. 34, No. 4, pp. 10-11,2021.

[13] Lee R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, Vol. 26, No. 3, pp. 297-302, 1945

[14] F. Hausdorff, "Grundzüge der Mengenlehre," *von Veit,* Vol.7, 1914.

[15] B. Cheng, et al., “Boundary IoU: Improving object-centric image segmentation evaluation,” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15334-15342, 2021.

[16] Cohen, J., et al., ”Videocinematography of fresh and cryopreserved embryos: a retrospective analysis of embryonic morphology and implantation,” *Fertility and Sterility*, Vol. 51, No. 5, pp 820-827, 1989

[17] Erin I, Lewis, et al., “Use of imaging software for assessment of the associations among zona pellucida thickness variation, assisted hatching and implantation of day 3 embryos,” *Journal of assisted reproduction and genetics,* Vol. 34, pp.1261-1269, 2017