I2CNet: An Intra- and Inter-Class Context Information Fusion Network for Blastocyst Segmentation

Hua Wang[1,2] , **Linwei Qiu**[1,2] , **Jingfei Hu**[1,2†] and **Jicong Zhang**[1,2*†]
[1]School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China
[2]Hefei Innovation Research Institute, Beihang University, Hefei 230012, China
hwang0609@163.com, qiulinwei@buaa.edu.cn, hujingfei24@163.com, jicongzhang@buaa.edu.cn

Abstract

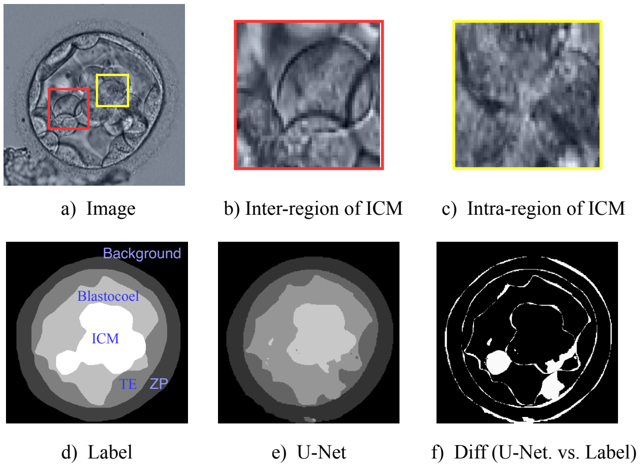
The quality of a blastocyst directly determines the embryo's implantation potential, thus making it essential to objectively and accurately identify the blastocyst morphology. In this work, we propose an automatic framework named I2CNet to perform the blastocyst segmentation task in human embryo images. The I2CNet contains two components: IntrA-Class Context Module (IACCM) and InteR-Class Context Module (IRCCM). The IACCM aggregates the representations of specific areas sharing the same category for each pixel, where the categorized regions are learned under the supervision of the groundtruth. This aggregation decomposes a K-category recognition task into K recognition tasks of two labels while maintaining the ability of garnering intra-class features. In addition, the IRCCM is designed based on the blastocyst morphology to compensate for inter-class information which is gradually gathered from inside out. Meanwhile, a weighted mapping function is applied to facilitate edges of the inter classes and stimulate some hard samples. Eventually, the learned intra- and inter-class cues are integrated from coarse to fine, rendering sufficient information interaction and fusion between multi-scale features. Quantitative and qualitative experiments demonstrate that the superiority of our model compared with other representative methods. The I2CNet achieves accuracy of 94.14% and Jaccard of 85.25% on blastocyst public dataset.

1 Introduction

Infertility is a disorder that characterized by failure to establish a clinical pregnancy following normal and unprotected intercourse within twelve months [Zegers Hochschild *et al.*, 2009]. About 10%-15% of couples suffer from infertility around the world [Tamrakar and Bastakoti, 2019].

In-Vitro Fertilization (IVF) is one of the most common infertility treatments. During IVF process, the fertilized eggs (embryos) are first need to be cultured under constant conditions in vitro until they develop into blastocysts (about 5 days) and then high-quality embryos are selected for implantation based on morphological characteristics. However, human participation in such quality assessments may result in higher inter- and intra-observer variability. Furthermore, it is a laborious and error-prone process that requires domain knowledge and expertise. Therefore, it is necessary to explore automatic segmentation methods for different tissue regions in blastocysts.

In recent years, the development of deep learning has continuously improved the performance of medical image segmentation tasks. The most commonly used network is U-Net [Ronneberger *et al.*, 2015] and its derivative variants. There are some works that have attempted to segment blastocyst tissue, but the segmentation performance still needs to be improved. In blastocyst images, on the one hand, there are large differences in intra-class characteristics and adhesion of various tissues, as shown in the region of interest (ROI) of Fig. 1 a), b), and c); on the other hand, the inter-class structural distinction is not obvious, especially the edge area, as shown in Fig. 1 b). Since the blastocyst is a transparent object, and the image is collected under low-light conditions,



a) Image b) Inter-region of ICM c) Intra-region of ICM

d) Label e) U-Net f) Diff (U-Net. vs. Label)

Figure 1: The blastocyst images and labels. a) Original image; b) inter-region of ICM; c) intra-region of ICM; d) labels of different tissues for blastocyst; e) predicted map of U-Net; f) the difference map between U-Net and label.

[*] Corresponding Author.

[†] Co-corresponding Authors.

this exacerbates the intra-class blur and the indistinct features at the edges. The results of traditional U-Net for blastocysts segmentation show that intra-class is no clustering, such as some areas of inner cell mass (ICM) cannot be recognized, some areas of zona pellucida (ZP) are incorrectly recognized as blastocoel, and the inter-class structures are unclear (the errors are mainly concentrated on the boundary and some hard samples' pixels), as shown in Fig. 1 a), d), e), and f).

To address the aforementioned problems, we propose an I2C module, which contains an IntrA-Class Context Module (IACCM) and an InteR-Class Context Module (IRCCM). IACCM adaptively enhances the same category and suppresses the feature weights of other categories, thereby the degree of aggregation within the class is enhanced and the segmentation results are more unified. The advantage of the IACCM is to simplify and decompose a K-category recognition task into K recognition tasks of two labels. But this also brings two problems: one is that the IACCM will be more inclined to learn easy samples, a large number of easy samples dominate the gradient. How to make up for hard samples to learn weights is particularly important, such as the pixels in the edge area account for a small proportion of the entire image and the characteristics are not sharp; the second is that although the decomposition task reduces the complexity of model learning, it lacks the structural relevance of inter-class samples. Consequently, we introduce the IRCCM, through a weighted mapping function to continuously focus on the unsolved difficulties, and these are the boundary of the category and hard samples. Meanwhile, the interaction of different classes also makes up for the shortcomings of intra-class modules.

The I2C module is equipped at each level of the network, and top-to-down continuously introduces supervision information in a coarse-to-fine manner, rendering sufficient information interaction and fusion between multi-scale features. At the high level of network, the topological structures of different categories are integrated based on semantic information. At the bottom layer of network, the image edges are processed in a more refined manner according to low-level statistical features such as color, edge, and texture. Intra-class and inter-class interactions enable the model to better learn latent features of embryos.

In summary, the main contributions of this work can be summarized as follows:

- A general architecture framework named I2CNet is proposed in this work, which reveals how to leverage information of intra-class and inter-class to consistently boost the segmentation performance of blastocyst image.
- We design a simple and effective IACCM to enhance the aggregation of the same class and a novel IRCCM to capture the relation information of different class, respectively. Experimental results demonstrate the effectiveness of our method. Especially, the IRCCM contains a weighted mapping function, which can promote the represents of boundary and hard pixels.
- The top-to-down feature enhancement pathway which couples the backbone encoder increases the representation power of feature maps with intra-class enhancement and inter-class enhancement in a coarse-to-fine manner.

2 Related Work

In recent years, there are many works that have attempted to segment blastocyst image [Filho *et al.*, 2010; Khan *et al.*, 2016], but mainly focus on the identification of ICM and trophectoderm (TE).

The semi-automatic segmentation method includes the level set idea [Filho *et al.*, 2012]. By defining the initial contour, the gradient information is applied to segment the ICM and TE. However, in clinical practice, the zona pellucida of the blastocyst is ruptured and the ICM will be outside the zona pellucida. In this method, it is difficult to define contour information and the recognition effect is not significant. Moreover, it requires human intervention. On this basis, Singh *et al.* [Singh *et al.*, 2014] applied level set combined with the correction of Retinex to automatically identify blastocysts, mainly for TE segmentation. It achieved an accuracy of 87.8% and a recall rate of 78.7%. Rad *et al.* [Rad *et al.*, 2017] utilized image texture features (Gabor and DCT features) and combined level sets to determine ICM boundaries, obtaining a Jaccard index of 70.3%. Saeedi *et al.* [Saeedi *et al.*, 2017] also applied texture features, but combined k-means clustering and watershed to identify ICM and TE, where ICM and TE achieved Jaccard index of 71.1% and 63%, respectively. Kheradmand *et al.* [Kheradmand *et al.*, 2016] combined Discrete Cosine Transform (DCT) and used a two-layers neural network for feedback propagation to identify ZP, TE and ICM, but only obtained Jaccard indices of 47.7%, 58.9% and 67.4%. Kheradmand *et al.* [Kheradmand *et al.*, 2017] used a Fully Convolutional Neural (FCN) network to segment the ICM and achieved a Jaccard Index of 76.5%. The ICM recognition methods that FCN-based [Leahy *et al.*, 2020] and U-Net based [Rad *et al.*, 2018] outperform [Kheradmand *et al.*, 2016; Rad *et al.*, 2017; Saeedi *et al.*, 2017; Kheradmand *et al.*, 2017]. Yu and Koltun [Yu and Koltun, 2015], Rad *et al.* [Rad *et al.*, 2019] proposed a U-Net variant by adding dilated convolution, which only increases the perception field of a single pixel, but not pay attention to the attribute relationship between pixels, and Jaccard index is below 82%.

The above methods are either dedicated to the segmentation of a single tissue or identification of several tissues. It is necessary to focus on the overall identification of a single tissue and the structural associations of all tissues. There is still a research gap of how to improve the segmentation performance of blastocyst.

3 Method

An overview of the proposed framework is illustrated in Fig. 2. The I2CNet incorporates the intra-class contextual information and inter-class contextual information for blastocyst image segmentation.

3.1 Formulation of I2CNet

Given a set of images S, Let $S = \{(X_m, Y_m), m = 1, \cdots, M\}$, where X_m denotes the original input embryo image and Y_m denotes the corresponding ground truth. Since

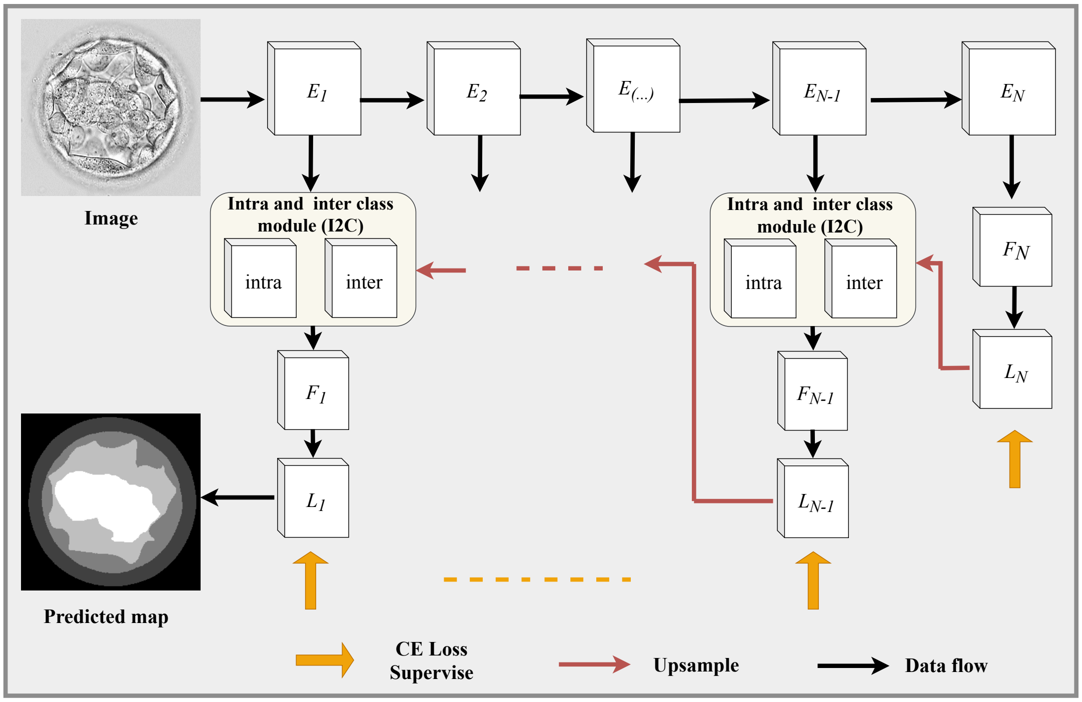


Figure 2: Architecture of the proposed I2CNet. It is an encoder-decoder structure, which contains intra and inter class modules. Deep supervision is applied for each level during the training.

each image X_m is processed separately, the subscript m is omitted for simplicity in the following sections.

We define the encoding block as $E = \{E_n(\cdot), n = 1, 2, \cdots, N\}$, where N is the numbers of encoding blocks. $E_n(\cdot)$ contains two convolutional layers, which contains two Batch Normalization (BN) layers, two ReLU activation layers and a max pooling layer. We first use E to project the pixels in X into a non-linear embedding space so that we can obtain the pixel representations R_E^n

$$R_E^n = \begin{cases} E_n(X), n = 1 \\ E_n(R_E^{n-1}), n = 2, \cdots, N, \end{cases} \tag{1}$$

where R_E^n is a matrix of size $C \times \frac{H}{2^n} \times \frac{W}{2^n}$, C is the channel of R_E^n.

We use $F(\cdot)$ squeeze R_E^n to representations R_F^n that with K channel (K is the number of class), where $F(\cdot)$ contains two 1×1 convolutional layers. The process can be formulated as follows

$$R_F^n = \begin{cases} F_n(R_E^n), n = N \\ F_n(R_{I2C}^n), n = 1, 2, \cdots, N-1, \end{cases} \tag{2}$$

where the R_{I2C}^n is the features representation that aggregate the intra- and inter-class information from the whole image in different scale by I2C module. I2C module $F_{n-1}^{I2C}(\cdot)$ contains intra-class context model $F_{n-1}^{intra}(\cdot)$ and inter-class context model $F_{n-1}^{inter}(\cdot)$, as shown in Fig. 3.

After squeezing the channel of representations by $F(\cdot)$, we use $L(\cdot)$ to get the supervised logit map R_L^n of $K \times \frac{H}{2^n} \times \frac{W}{2^n}$ scale images

$$R_L^n = L_n(R_F^n), n = 1, 2, \cdots, N, \tag{3}$$

where $L(\cdot)$ contains a convolutional layer, BN and ReLU. Following above expectation, as indicated in Fig. 2, before gettinng the aggregate information of the intra- and inter-module, we get the probability map from the lower level logit map R_L^n.

To match the scale of high level representations, we use upsampled function to get the probability map P^n with size of $(K \times H' \times W', H' = \frac{H}{2^{n-1}}, W' = \frac{W}{2^{n-1}})$

$$P^n = \text{Upsampled}(\sigma(R_L^n)), n = 1, 2, \cdots, N, \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid function. Then, I2C module $F_{n-1}^{I2C}(\cdot)$ can use the P^n to enhance the R_E^{n-1} features $(K \times H' \times W', H' = \frac{H}{2^{n-1}}, W' = \frac{W}{2^{n-1}})$:

$$R_{intra}^{n-1} = F_{n-1}^{intra}(R_E^{n-1}, P^n), n = 1, 2, \cdots, N-1, \tag{5}$$

$$R_{inter}^{n-1} = F_{n-1}^{inter}(R_E^{n-1}, P^n), n = 1, 2, \cdots, N-1, \tag{6}$$

$$R_{I2C}^{n-1} = F_{n-1}^{I2C}(R_{intra}^{n-1}, R_{inter}^{n-1}), n = 1, 2, \cdots, N-1. \tag{7}$$

Finally, the output predicted map in different scale is define as O_n:

$$O_n = R_L^n, n = 1, 2, \cdots, N. \tag{8}$$

3.2 Intra-Class Context Module

In I2C module, intra-class context module F_{intra} is designed to capture the context information from the same class of the

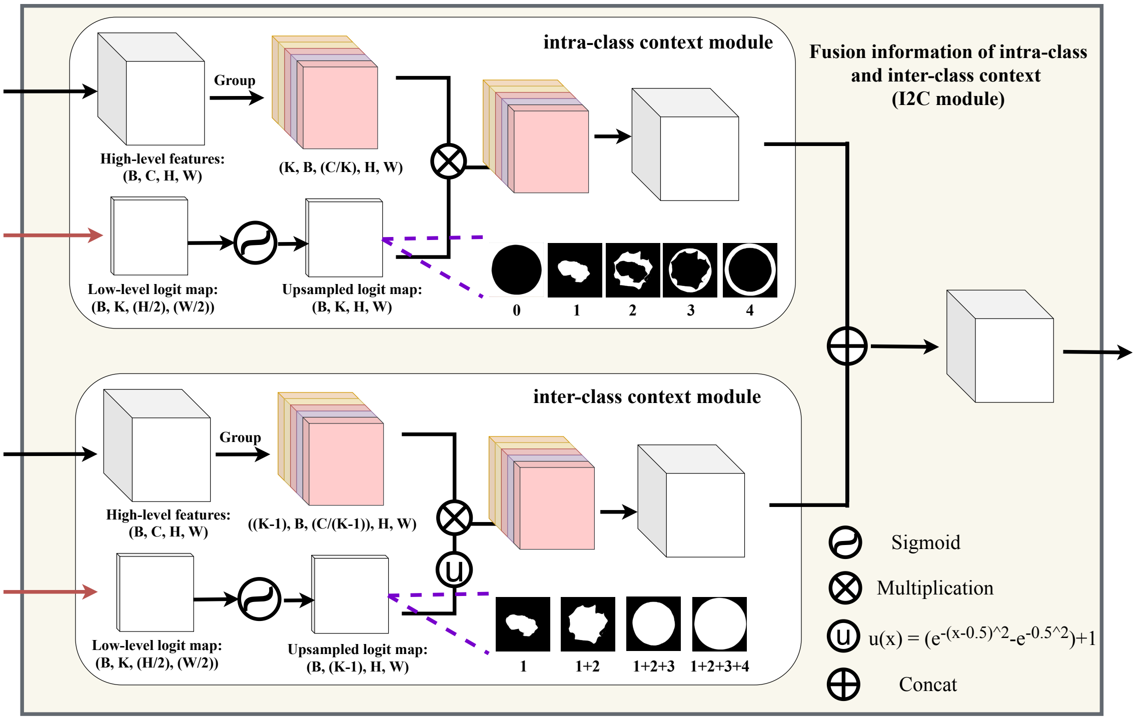

intra-class context module
Group
High-level features: (B, C, H, W)
(K, B, (C/K), H, W)
Low-level logit map: (B, K, (H/2), (W/2))
Upsampled logit map: (B, K, H, W)
0
1
2
3
4
Fusion information of intra-class and inter-class context (I2C module)
inter-class context module
Group
High-level features: (B, C, H, W)
((K-1), B, (C/(K-1)), H, W)
Low-level logit map: (B, K, (H/2), (W/2))
Upsampled logit map: (B, (K-1), H, W)
1
1+2
1+2+3
1+2+3+4
Sigmoid
Multiplication
u(x) = (e-(x-0.5)^2-e-0.5^2)+1
Concat

Figure 3: Details flow of the I2C module. The top diagram is the intra-class context module and the bottom denotes the inter-class context module. These two modules share the same input, *i.e.* the results of corresponding encoding block and the upsampled probability map. The final output of our I2C module is the information fusion of intra-class and inter-class context module.

whole image, as shown in the top of Fig. 3. There are feature enhancements both in the same scale and across different scales. Before intra-scale enhancement, every level logit map passes a supervised module before fusion. Then, these logit maps are activated with neighboring high level features to focus on the context information from the same class.

To make the adjacent high-level features match the probability map of the K-class tissues, we orderly grouped R_E^{n-1} into K groups, $(R_E^{n-1})^k$ is a matrix of size $C_k \times \frac{H}{2^{n-1}} \times \frac{W}{2^{n-1}}$, where C_k denotes the number of channels belonging to the category k. In this way, the number of parameters can be effectively reduced

$$C_k = \begin{cases} \lfloor \frac{C}{K} \rfloor, k = 1, 2, \cdots, K-1, \\ C - (K-1) \times \lfloor \frac{C}{K} \rfloor, k = K. \end{cases} \tag{9}$$

For the convenience of presentation, we define the $(P^n)^k$ is the logit map belonging to category k. To aggregate contextual information from the same class, we compute the region representation for each class k as follows:

$$(R_{intra}^{n-1})_i^k = (R_E^{n-1})_i^k \cdot (P^n)^k, i = 1, 2, \cdots, C_k. \tag{10}$$

Eventually, we orderly concatenate the all $(R_{intra}^{n-1})_i^k$ and use a 1×1 convolution to get the final R_{intra}^{n-1} features. The R_{intra}^{n-1} features contain the context information from the same class of the whole image. By grouping, we decompose a multi-category recognition task into multiple two-category recognition tasks, thereby reducing the learning difficulty of the model, allowing the model to focus on learning features of the same category without interference from other categories. Moreover, the amount of parameters is reduced to a certain extent.

3.3 Inter-Class Context Module

Although IACCM effectively solves the problem of feature representation and learning within the same class, it ignores the inter-class interconnection. As shown in Fig. 1 f), it can be seen that the predictive error between classes accounts for a relatively large proportion, including some pixels with unobvious structural features. To address this problem, inter-class context module F_{inter} is designed to deal with the relation information of different class and some hard pixels , as shown in the bottom of Fig. 3.

The inter-class and intra-class modules share the same high-level features and follow the same grouping principle. The intra-class grouping is equal to the number of tissues categories, but the inter-class grouping considers the correlation between tissues, which is $K-1$ group. And the tissues order of the blastocyst from inside to outside is ICM, Blastocoel, TE, ZP, and Background. So we only considered four organizational relationships: 1, $1+2$, $1+2+3$, and $1+2+3+4$, where 1, 2, 3, 4 represent ICM, Blastocoel, TE, ZP, respectively.

To aggregate the context information from the inter class, we calculate the region representation as follows:

$$\begin{aligned}(R_{inter}^{n-1})_i^k = (R_E^{n-1})_i^k \cdot U(\textstyle\sum_{j=1}^{i} (P^n)^j), \\ i = 1, 2, \cdots, C_k - 1,\end{aligned} \tag{11}$$

IACCM	IRCCM w/o $U(\cdot)$	IRCCM w $U(\cdot)$	Accuracy(%)	Precision(%)	Recall(%)	Dice(%)	Jaccard(%)
			91.67	88.52	89.09	88.53	80.28
✓			93.88	90.23	91.91	91.26	84.68
	✓		93.04	90.47	90.79	90.19	83.28
		✓	93.58	91.17	91.97	91.30	84.77
✓		✓	**94.14**	**91.18**	**92.32**	**91.56**	**85.25**

Table 1: The ablation study results of blastocyst segmentation. Both IACCM and IRCCM have greatly improved the performance. Moreover, the weighted mapping function $U(\cdot)$ benefits the segmentation.

where $U(\cdot)$ is a weighted mapping function, which can enhance the represents of the border and hard pixels. According to Eq. 11, we can get the structural feature map of the correlation between the inside-out (ICM to ZP) tissues.

In IRCCM, according to the characteristics of cross entropy loss, as the probability approaches 0.5, the greater the loss, which means that the sample is harder to recognize. For embryo samples, the samples of the boundary between different tissues is small. Moreover, the blastocyst is a transparent object and the image is collected under low-light conditions, resulting in unclear boundary features. Therefore, we designed a weighted mapping function $U(\cdot)$ to redistribute the probabilities weights of each category, that is, the weight value of the sample pixels that are not easy to identify is close to $1+\alpha(1-e^{-(\frac{1}{2})^{\beta}})$, and the weight value of the sample pixels that are easy to identify is close to 1, which makes the model more effective for learning and reasoning at edges and hard pixels. The $U(\cdot)$ is defined as:

$$U(\cdot) = \alpha(e^{-(x-\frac{1}{2})^{\beta}} - e^{-(\frac{1}{2})^{\beta}}) + 1, \tag{12}$$

where α and β are the activation factors, and set as 1 and 2 in our experiment.

Finally we orderly concatenate the $(R^{n-1}_{inter})^{k}_{i}$ and use convolution with kernel size 1×1 to get the final R^{n-1}_{inter} features. The R^{n-1}_{inter} features contain the context information from the different class of the whole image.

3.4 Loss Function

In this work, we use cross entropy loss to supervise logit map, it can be defined as Eq. 13:

$$\mathcal{L}_n = \frac{2^n}{H\times W}\sum\nolimits_{i,j}\mathcal{L}_{ce}((RS^n_L)^{K\times\frac{H}{2^n}\times\frac{W}{2^n}}_{|*,i,j|}, Y^n_S), \quad n = 1,2,\cdots,N, \tag{13}$$

where Y^n_S represents the one-hot label $(K\times H\times W)$ scale to the size of $K\times\frac{H}{2^n}\times\frac{W}{2^n}$, and $(RS^n_L)^{K\times\frac{H}{2^n}\times\frac{W}{2^n}}$ is calculated as Eq. 14:

$$RS^n_L = \text{Softmax}(R^n_L), \tag{14}$$

$\mathcal{L}_{ce}$ denotes the cross entropy loss and $\sum_{i,j}(\cdot)$ denotes that summation is calculated over all locations on the input image X. The total cross entropy loss define as:

$$\mathcal{L}_{total} = \frac{1}{N}\sum\nolimits_{n=1}^{N}\mathcal{L}_n, n = 1,2,\cdots,N. \tag{15}$$

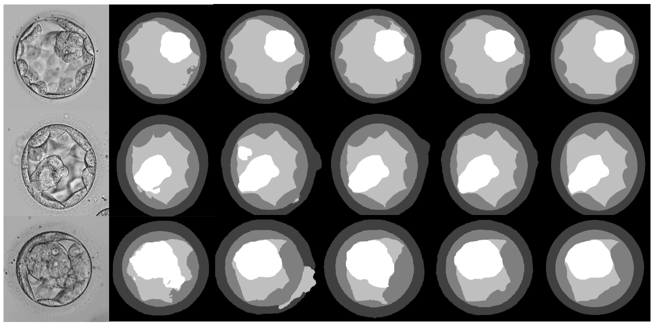


Figure 4: Qualitative results of ablation study. With the help of IACCM and IRCCM, our method can learn representations of same classes and model the relationships of different classes, leading to more accurate segmentation results.

4 Experiments

4.1 Experimental Setup

Dataset and Ground Truth. The dataset is introduced in [Saeedi *et al.*, 2017], which is the only public dataset on human blastocyst. This benchmark consists of 249 blastocyst images and their Ground Truth (GT) masks provided by the Pacific Centre for Reproductive Medicine (PCRM). The training set comprises 199 images (80%) and the testing set contains 50 images (20%).

Training Details. In this work, the number of encoding blocks is 5. The backbone and two integrated context modules are initialized randomly by kaiming. The model was trained with 2000 epochs and the Adam was used as the optimizer. Initial learning rate and weight decay were set to 1×10^{-4} and 1×10^{-5}. Poly learning rate policy with factor $(1-\frac{epoch}{total\_epoch-epoch})^{0.9}$ is performed for training. Synchronized batch normalization implemented by PyTorch is enabled during training. For data augmentation, we process each sample with random scaling in the range of $[0.5, 2.0]$, random rotating in the range of $[-15, 15]$ degrees, random cropping and left-right flipping in training stage. The input image size is 256×256. By default, no tricks (*e.g.* multi-scale and multi-rotation with flipping testing) are adopted during the testing stage.

Reproducibility. The proposed framework is implemented on PyTorch (version $\geq$ 1.3) and trained on four NVIDIA Tesla V100 GPUs with a 32 GB memory per-card. And all the testing procedures are performed on a single NVIDIA Tesla V100 GPU. To provide full details of our framework, our code will be made publicly available.

Type	Methods	Accuracy(%)	Precision(%)	Recall(%)	Dice(%)	Jaccard(%)
CNN-based	U-Net [Ronneberger *et al.*, 2015]	93.15	90.62	90.99	90.49	83.65
	Blast-Net [Rad *et al.*, 2019]	93.11	90.65	91.23	90.66	83.74
	AttUnet [Oktay *et al.*, 2018]	93.22	90.87	91.22	90.79	84.05
	Unet++ [Zhou *et al.*, 2018]	93.22	90.67	90.76	90.44	83.72
	Deeplabv3+ (resnet101) [Chen *et al.*, 2018]	93.35	91.09	91.15	90.88	84.17
	Unet3+ [Huang *et al.*, 2020]	93.06	90.36	91.50	90.62	83.70
	Pranet (res3net50) [Fan *et al.*, 2020]	93.28	90.73	91.45	90.83	84.04
	nnUnet [Isensee *et al.*, 2018]	93.41	90.94	91.63	91.00	84.28
Transformer-based	SETR (ViT) [Zheng *et al.*, 2021]	92.99	90.11	90.93	90.28	83.12
	MedT [Valanarasu *et al.*, 2021]	92.86	89.98	90.99	90.21	83.05
	TransUnet [Chen *et al.*, 2021]	93.21	90.78	91.20	90.73	83.86
	Swin-Unet [Cao *et al.*, 2021]	92.66	89.73	90.33	89.80	82.40
Ours	**I2CNet**	**94.14**	**91.18**	**92.32**	**91.56**	**85.25**

Table 2: Comparison of performance with state-of-the-art methods. Compared with these CNN-based and transformer-based approaches, we have achieved the best performance among five metrics.

Evaluation Metrics. In order to evaluate the performance of the network, there are five commonly used indicators: Accuracy, Precision, Recall, Dice Coefficient and Jaccard Index [Csurka *et al.*, 2004; Zhu *et al.*, 2016; Thoma, 2016; Taha and Hanbury, 2015].

4.2 Ablation Study

Effectiveness of IACCM. In embryo images, there is a phenomenon of intra-class continuous aggregation, in which the inner cell mass is the most core continuous region, while other tissues are band-like regions but also continuous. It is very important to ensure that the model can learn to understand and recognize the features of the same category more accurately during the training process. As shown in the third column of Fig. 4, we can see that the proposed IACCM can effectively focus on learning the features of the same category and model the continuous dependencies of pixels in the same category, making the intra-class tissues more clustered. In Table 1, we can see that the addition of IACCM improves the overall performance of the network, in which the Dice and Jaccard indicators increase by 2.73% and 4.4% respectively. The experimental results show that the IACCM is effectiveness for increasing intra-class segmentation performance.

Effectiveness of IRCCM. In medical images, especially for embryo images, focusing on the logical relationship between different tissues is also very helpful for accurate segmentation of embryo tissues. Since the tissue types of embryo cells from inside to outside are arranged according to developmental law, such as the inner circle of the trophoblast must be connected to the cyst and the outer circle must be connected to the zona pellucida. Therefore, we designed an IRCCM to make the model learn and express the features of embryo images more efficiently by modeling inter-class relationships progressively. In particular, we add a weighted mapping function $U(\cdot)$ to increase the weights of boundaries and hard samples. From Table 1, we observe that $U(\cdot)$ improves the performance of IRCCM, increasing the Accuracy from 93.04% to 93.58%, the Dice from 90.19% to 91.30% and the Jaccard from 83.28% to 84.77%. The introduction of IRCCM with $U(\cdot)$ improves the backbone performance, which improves Dice by 2.77% and Jaccard by 4.49%. The visualization results are shown in the fourth column of Fig. 4, clearly showing that the IRCCM is effectiveness for improving the inter-class recognition error. These improvements suggest that introducing IRCCM component can model the relationship between classes well and increase the learning ability of the model.

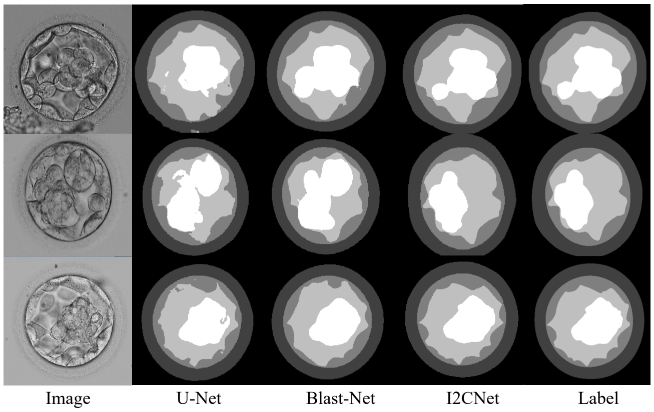


Figure 5: Comparison of model performance of the proposed method. Our model can recognize the difficult borders of embryo tissues.

Effectiveness of IACCM & IRCCM. In this work, we integrate IACCM and IRCCM to get the I2C module. As illustrated in Table 1, we can find that combining IACCM and IRCCM outperforms the backbone by 4.97% in terms of Jaccard. The improvement is much higher than applying single IACCM (4.97% *v.s.*4.40%) or single IRCCM (4.97% *v.s.* 4.49%). Moreover, from the fifth column of Fig. 4, we can see that after combining the IACCM and IRCCM modules, not only the intra-class is more aggregated, but the inter-class relationship is also more reasonable and the boundary is clearer. These results indicate that IACCM and IRCCM can complement and promote each other, which well demonstrates the reliability of the basic motivation, and the effectiveness of the designed framework in this paper.

4.3 Comparison with State-of-the-Art Methods

Table 2 lists the blastocyst segmentation performance comparison between the proposed framework and related methods in the literature. Among the compared methods, there

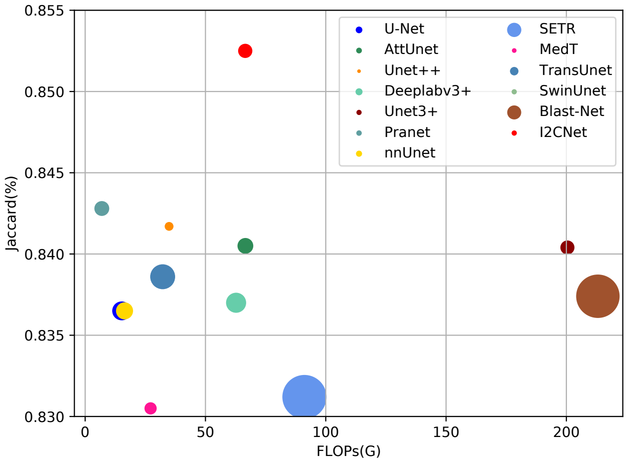


Figure 6: Comparison of Parameters, FLOPs and Jaccard for different methods. Note that parameters and FLOPs are calculated on a 256×256 input image.

are models including the classic CNNs and the more popular transformers. Deeplabv3+, SETR, and Pranet apply pretrained models, which are noted in the back brackets. Others are trained from scratch and we have optimized their hyperparameters. It can be seen from the Table 2 that the proposed I2CNet achieves the state-of-the-art performance compared with other methods, among which Accuracy, Precision, Recall, Dice and Jaccard achieve 94.14%, 91.18%, 92.32%, 91.56%, 85.25%, respectively. Fig. 5 shows some qualitative results for U-Net, Blast-Net and we proposed I2CNet. As seen, the results show that our proposed I2CNet achieves the most similar results to the ground truth. These experimental results demonstrate that the effectiveness of the proposed method.

In addition, Fig. 6 shows the parameters, FLOPs and Jaccard for the comparison methods and our proposed method. The size of the circle represents the parameter amount of the model (the larger circle means a larger number of parameters in the model). It can be intuitively seen that I2CNet has achieved the best Jaccard index, meanwhile the parameters and FLOPs are fewer, which further illustrates the superiority of our proposed method.

5 Conclusions

In this work, we studied the intra-class and inter-class relationship of embryo image, and proposed an IntrA-Class Context Module and an InteR-Class Context Module. The IACCM transforms a multi-class recognition task into multiple binary classification tasks, and employe the supervised probability map to guide the learning of high-level features, which not only reduces the complexity of the model but also improves the continuity of intra-class tissues. The IRCCM transmits the relationship probability map of different tissues to the high-level to guide the learning of the model. Meanwhile, the probability map is remapped and assigned new weights to enhance the weights of boundary and hard samples, so that the model can better focus on valuable sample pixels instead of blindly learning simple samples. The proposed I2CNet integrates both IACCM and IRCCM to achieve state-of-the-art performance. Finally, we qualitatively and quantitatively verify the effectiveness of our proposed model through ablation experiments and comparative experiments with other state-of-the-art methods.

Acknowledgments

This work was supported in part by the Beijing Natural Science Foundation under Grant Z200024; in part by Hefei Innovation Research Institute, Beihang University; and in part by the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-044.

References

[Cao *et al.*, 2021] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021.

[Chen *et al.*, 2018] Liang Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[Chen *et al.*, 2021] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[Csurka *et al.*, 2004] Gabriela Csurka, Diane Larlus, Florent Perronnin, and France Meylan. What is a good evaluation measure for semantic segmentation? *IEEE PAMI*, 26(1), 2004.

[Fan *et al.*, 2020] Deng Ping Fan, Ge Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 263–273. Springer, 2020.

[Filho *et al.*, 2010] Efraim Santos Filho, Julia Alison Noble, and Darren Wells. A review on automatic analysis of human embryo microscope images. *The open biomedical engineering journal*, 4:170, 2010.

[Filho *et al.*, 2012] Efraim Santos Filho, Julia 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, 2012.

[Huang *et al.*, 2020] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020.

[Isensee *et al.*, 2018] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[Khan *et al.*, 2016] Aisha Khan, Stephen Gould, and Mathieu Salzmann. Segmentation of developing human embryo in time-lapse microscopy. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pages 930–934. IEEE, 2016.

[Kheradmand *et al.*, 2016] Shakiba Kheradmand, Parvaneh Saeedi, and Ivan Bajic. Human blastocyst segmentation using neural network. In *2016 IEEE Canadian conference on electrical and computer engineering (CCECE)*, pages 1–4. IEEE, 2016.

[Kheradmand *et al.*, 2017] 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 *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1752–1756. IEEE, 2017.

[Leahy *et al.*, 2020] 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.

[Oktay *et al.*, 2018] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

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

[Rad *et al.*, 2018] Reza Moradi 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 *2018 25th IEEE international conference on image processing (ICIP)*, pages 3518–3522. IEEE, 2018.

[Rad *et al.*, 2019] Reza Moradi Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Blast-net: Semantic segmentation of human blastocyst components via cascaded atrous pyramid and dense progressive upsampling. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1865–1869. IEEE, 2019.

[Ronneberger *et al.*, 2015] 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.

[Saeedi *et al.*, 2017] 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, 2017.

[Singh *et al.*, 2014] 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, 2014.

[Taha and Hanbury, 2015] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.

[Tamrakar and Bastakoti, 2019] Suman Raj Tamrakar and Rashmi Bastakoti. Determinants of infertility in couples. *Journal of Nepal Health Research Council*, 17(1):85–89, 2019.

[Thoma, 2016] Martin Thoma. A survey of semantic segmentation. *arXiv preprint arXiv:1602.06541*, 2016.

[Valanarasu *et al.*, 2021] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 36–46. Springer, 2021.

[Yu and Koltun, 2015] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[Zegers Hochschild *et al.*, 2009] Fernando Zegers Hochschild, Geoffrey David Adamson, Jacques de Mouzon, Osamu Ishihara, Ragaa Mansour, Karl Nygren, Elisabeth Sullivan, and Sher van der Poel. The international committee for monitoring assisted reproductive technology (icmart) and the world health organization (who) revised glossary on art terminology, 2009. *Human reproduction*, 24(11):2683–2687, 2009.

[Zheng *et al.*, 2021] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890, 2021.

[Zhou *et al.*, 2018] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[Zhu *et al.*, 2016] Hongyuan Zhu, Fanman Meng, Jianfei Cai, and Shijian Lu. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *Journal of Visual Communication and Image Representation*, 34:12–27, 2016.