Comparison of U-Net Based Models for Human Embryo Segmentation

Araştırma Makalesi/Research Article

Nefise UYSAL [1]*, Tahir Koray YOZGATLI [2], Ecem Nur YILDIZCAN [3], Emre KAR [2], Murat GEZER [4], Ercan BAŞTU [5]

[1]Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey
[2]School Of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey
[3]Department of Mathematics Engineering, Istanbul Technical University, Istanbul, Turkey
[4]Department of Informatics, Istanbul University, Istanbul, Turkey
[5]Department of Obstetrics and Gynecology, Acibadem Mehmet Ali Aydinlar University School of Medicine, Istanbul, Turkey
uysal18@itu.edu.tr, tahirkoray@gmail.com, ecemnuryildizcan@gmail.com, emre.kar@live.acibadem.edu.tr, murat.gezer@gmail.com, ercanbastu@gmail.com

Abstract— The quality of human embryos produced during in vitro fertilization is conventionally graded by clinical embryologists and this process is time-consuming and prone to human error. Artificial intelligence methods may be used to grade images captured by time-lapse microscopy (TLM). Segmentation of embryos from the background of TLM images is an essential step for embryo quality assessment as the background of the embryo has various artifacts which may mislead the grading algorithms. In this study, we performed a comparative analysis of automated day-5 human embryo (blastocyst) image segmentation methods based on deep learning. Four fully convolutional deep models, including U-Net and its three variants, were created using the combination of two gradient descent-based optimizers and two-loss functions and compared to our proposed model. The experimental results on the test set confirmed that our customized Dilated Inception U-Net model with Adam optimizer and Dice loss outperformed other U-Net variants with Dice coefficient, Jaccard index, accuracy, and precision of 98.68%, 97.52%, 99.20%, and 98.52%, respectively.

Keywords— U-Net, deep learning, convolutional neural network, in vitro fertilization (IVF), human embryo, segmentation

İnsan Embriyo Segmentasyonu için U-Net Tabanlı Modellerin Karşılaştırılması

Özet— Tüp bebek tedavisi sırasında üretilen insan embriyolarının kalitesi, geleneksel olarak klinik embriyologlar tarafından derecelendirilir ve bu süreç zaman alıcı olup insan hatasına açıktır. Hızlandırılmış mikroskopi (TLM) yöntemi ile alınan görüntüleri derecelendirmek için yapay zeka yöntemleri kullanılabilir. TLM görüntülerinde embriyonun arka plandan segmentasyonu, arka planın derecelendirme algoritmalarını yanlış yönlendirebilecek çeşitli artefaktlara sahip olması nedeniyle embriyo kalite değerlendirmesi için önemli bir adımdır. Bu çalışmada, derin öğrenmeye dayalı otomatikleştirilmiş 5. gün insan embriyosu (blastosist) görüntü segmentasyon yöntemlerinin karşılaştırmalı bir analizi yapılmıştır. U-Net ve üç varyantından oluşan dört tam evrişimli derin model, iki gradyan iniş tabanlı optimizasyon algoritmasının ve iki kayıp fonksiyonunun kombinasyonu kullanılarak oluşturulmuş ve önerilen modelimiz ile karşılaştırılmıştır. Test setindeki deneysel sonuçlar, optimizasyon fonksiyonu olarak Adam ve kayıp fonksiyonu olarak ise Dice kullanan özelleştirilmiş Dilated Inception U-Net modelinin, sırasıyla %98.68, %97.52, %99.20 ve %98.52'lik Dice katsayısı, Jaccard benzerlik katsayısı, doğruluk ve kesinlik ile diğer U-Net tabanlı modellerden daha iyi performans gösterdiğini doğrulamıştır.

Anahtar Kelimeler— U-Net, derin öğrenme, evrişimli sinir ağları, in vitro fertilizasyon (IVF), insan embriyosu, segmentasyon

1. INTRODUCTION

Image segmentation separates images into coherent regions based on deterministic features such as color, intensity value, and texture of pixels in the image and is usually the first step of image analysis. The main purpose of image segmentation in medical image processing applications is to optimize disease diagnosis by detecting the required region of interest (ROI) using an automated tool or algorithm [1]. Current studies show that although image segmentation is not a simple process, it is an essential step for diagnosing disease and isolating the ROI in different medical imaging modalities [2].

Infertility is estimated to affect 8-12% of couples of reproductive age worldwide [3]. In vitro fertilization (IVF) is one of the most common types of assisted reproductive technology (ART). More than 8 million babies have been born with IVF treatment since 1978 when IVF was used in the clinic successfully and the first baby was born [4]. IVF involves culturing 7 to 12 candidate embryos of the patient in vitro for 3 to 5 days and then transferring one embryo or two embryos that are selected based on their developmental and morphological qualities to the patient [5]. Since the selection of a high quality embryo affects the success of IVF treatment greatly, there is a need for a robust, reliable embryo selection method to choose the best potential candidate. Furthermore, since single embryo transfer (SET) reduces the risk of multiple pregnancies and offers health advantages for the fetus and the mother, transferring the single best embryo is the safest approach possible and is recommended whenever possible [6–8]. Traditional embryo analysis based on morphological evaluation under a microscope is time-consuming, subjective, and leads to deterioration of optimal culture conditions when the embryo is removed from the incubator to be put under the microscope [9]. Although time-lapse imaging (TLI) systems for IVF are now available as an effective tool that allows the noninvasive analysis of embryos through continuous monitoring, the assessment made with TLI systems is still prone to human error [10–12]. Human error can be minimized in embryo evaluation by developing a decision support system using artificial intelligence algorithms. Segmentation is an important preprocessing stage for these algorithms to become more effective [13], [14].

Time-lapse microscopy (TLM) embryo images can involve various artifacts such as irregular embryo shape, hatched day−5 human embryo (blastocyst) shape, fragments attached to the external surface of the embryo, other data in the image zone such as the well number and information regarding the time since the start of the incubation, and artifacts created by the image optics. Manual cropping for artifact removal can be used before classification studies that aim to determine embryo quality from large numbers of microscopic images with automatic analysis [15]. However, this process can't remove artifacts completely and is time-consuming. Once sufficiently trained, segmentation

methods automatically remove artifacts outside of the embryo region. Therefore, segmentation of the embryo image is crucial in the preprocessing step to develop a classification algorithm. In this paper, we focus on how segmentation can help improve outcomes in infertility treatment by preparing images to facilitate the grading of human embryos by automated algorithms. We tackle the problem of automatic segmentation of human embryos with U-Net based models in the blastocyst stage of embryo development.

In this study, Section 2 reviews relevant literature and summarizes the previous work on the segmentation of embryos or different embryo regions in images. Section 3 presents the implemented segmentation architectures. Section 4 describes the dataset used in the training stage and the evaluation parameters used on the testing stage. In Section 5, the experimental results are presented in detail, along with current approaches in the literature. Finally, Section 6 is the concluding section.

2. RELATED WORKS

Several attempts have been made to automate embryo segmentation, including works on automating the segmentation of the inner cell mass (ICM), trophectoderm epithelium (TE), and zona pellucida (ZP) in time-lapse microscope images.

Karlsson et al. [16], [17] proposed a variational method based on an image model for automatic segmentation of the outer and inner circumference of the zona pellucida. Morales et al. [18] presented the use of a parametric active contour model adapted for quantification and segmentation on the thickness variation of zona pellucida that could only be applied after a preprocessing step aimed at improving certain aspects of visual information. Furthermore, several semi-automated segmentation techniques have been described [19]. Filho et al. [13] proposed a semi-automatic segmentation method for segmentation of both TE and ICM. They used an ellipse fitting method for segmentation of TE's outer boundaries and variational levels set algorithm involved with initialization of a circle manually placed for both the TE's inner boundaries and ICM. Recently, Saeedi et al. [20] have published their work for automatic detection of the TE and ICM regions using a set of texture measurements in HMC human embryo images. They reported 91.3% and 86.6% accuracy for identification of ICM and TE respectively.

Singh et al. [21] utilized the retinex algorithm to deemphasize debris in the cavity area of the embryo and then applied a level-set algorithm to identify the boundaries of TE region. They achieved a mean shape accuracy of 87.8%. Recently, Rad et al. [22] presented an automatic segmentation approach of TE region in the blastocyst images by using four fully convolutional deep models. Their proposed models have reached an average accuracy of 96.9% and a Dice coefficient of 86.61% for the identification of TE.

Deep convolutional networks have made great advances possible in numerous segmentation tasks over the past few years [23], [24] especially in biomedical image processing [25]. Kheradmand et al. [26] proposed a two-stage ICM segmentation pipeline that first uses the preprocessing method and then a pre-trained 16-layer VGG network on ImageNet. Rad et al. [27] proposed a multi-resolutional ensemble of stacked dilated U-Net to segment the ICM region. The proposed method outperformed the previous work [26] by 5.1% in accuracy and 2.8% in Dice coefficient. Recently, Harun et al. [28] presented a U-Net-based deep neural network (DNN) to detect the ICM and TE regions. This method achieved a 92.5% Dice coefficient, 85.3% Jaccard index for TE segmentation, 94.3% Dice coefficient, and an 89.3% Jaccard index for ICM segmentation.

When these studies in the literature were examined, it was seen that there were relatively few studies on human embryo segmentation and they were generally aimed at the segmentation of embryo components. A study, which includes the comparison of U-Net-based embryo segmentation models with a new dataset, can be useful in the literature where studies are completed using the same dataset.

3. METHODS

In this section, details of implemented U-Net based architectures were provided.

3.1. Base U-Net

U-net developed by Ronneberger et al. [25] is commonly used in biomedical image segmentation and doesn't require very large datasets. It is an end-to-end network made of encoder blocks that extract features of images and decoder blocks that restore the features along with bridges for connecting these two types of blocks. This architecture applies convolution blocks followed by a rectified linear unit (ReLU) in both the encoder and decoder blocks.

The U-Net model was used with an input size of 256×256 pixels. The block diagram of the U-Net architecture is depicted in Figure 1. Turquoise boxes symbolize multi-channel feature maps. Input sizes and the number of channels are denoted on the left of the boxes and above each feature map, respectively. The purple boxes represent copies of the feature maps. The color arrows show the different functions.
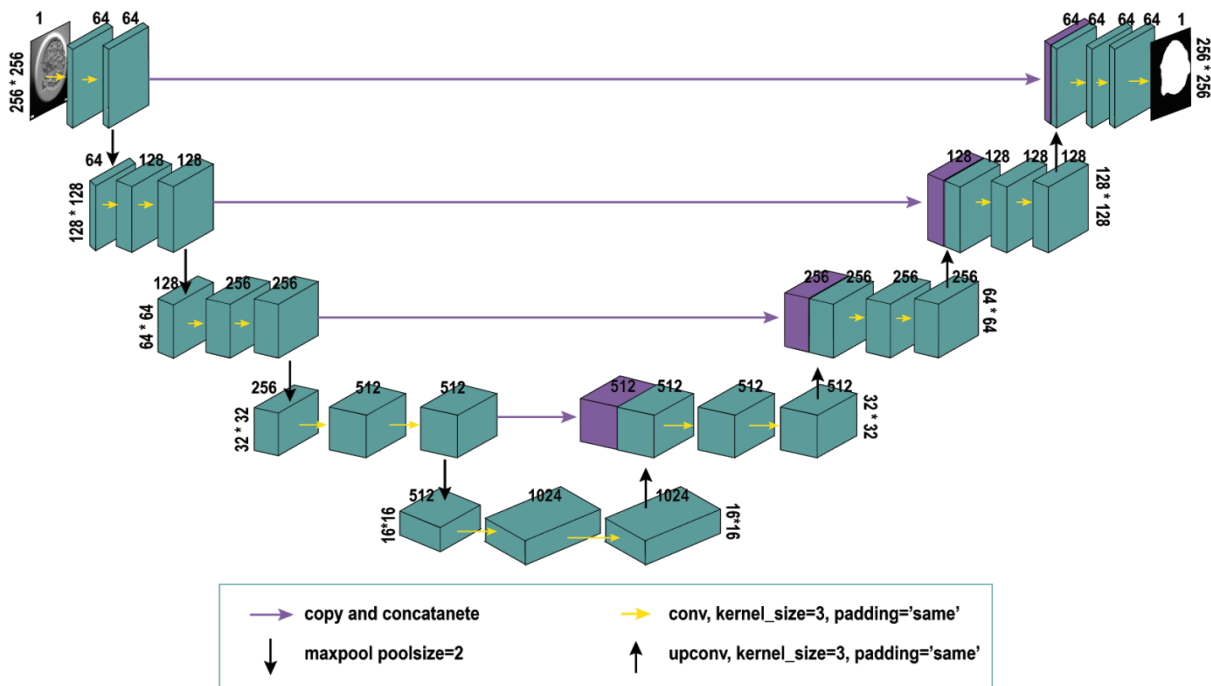

Figure 1. The U-Net architecture.

The base U-Net architecture consists of an encoder unit and a decoder unit, both with a depth of 4. The encoder unit makes up the first half of the architectural diagram. Here, four convolution blocks are applied with 64, 128, 256, and 512 kernels respectively. Each block consists of two 3x3 convolutions (unpadded convolutions) followed by a ReLU activation unit and max pooling operation performed to encode the input image into multiple different level feature representations in each block. The decoder unit makes up the second half of the architecture. The decoder unit performs up-sampling and concatenation to restore the features supplied by the encoder block onto the pixel space (higher resolution). The first decoder block produces 512 kernels, and then the number of

kernels is reduced by half after each dropout and batch normalization operation until it reaches 64 kernels, while the size of the image gradually increases using the up-convolution layer.

3.2. Inception U-Net

The inception U-Net architecture combining inception module and the U-Net is used to make the network deeper so that it can read high-level details via various kernel sizes instead of fixed-size filters for convolutions. The convolution blocks in the U-Net architecture were replaced with the inception blocks in the GoogleNet [29] achieving the state-of-the-art in the classification computer vision tasks of the ImageNet (ILVRC14). The outputs from the different filters were concatenated and transferred onto the next layer. In this study, the Inception U-Net had 4 pairs of encoder and decoder blocks. Each inception module included 3 × 3 max pooling operations and two types of convolution operations, consisting of 1 × 1 convolutions before every 3 × 3 convolution for dimensionality reduction as shown in Figure 2.
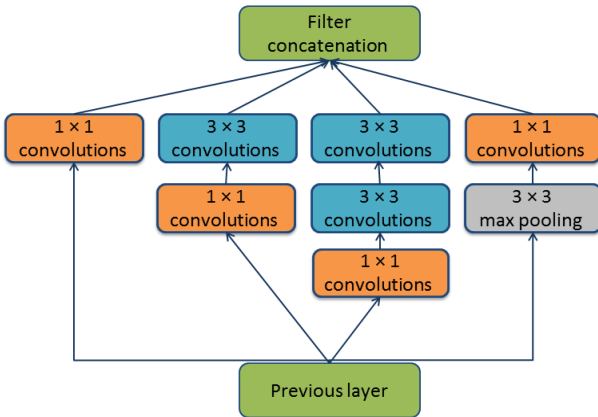

Figure 2. The inception module utilized in the Inception U-Net models.

3.3. Deep Residual U-Net (ResUNet)

Deep neural networks perform better when the depth of the network is increased however this may cause a degradation problem [30]. ResNet architecture, which has the projection shortcut connection, can decrease this problem. Zhang et al. [31] presented a deep residual U-Net that uses residual units instead of plain convolutional units as the fundamental block. In this study, residual units consisted of a projection shortcut and two convolution blocks including a batch normalization (BN) layer, a ReLU activation layer, and 3x3 convolutions as shown in Figure 3.

3.4. Dilated Inception U-Net

Shi et al. [32] incorporated dilated convolution based inception module and inception module into the encoding path of the original U-Net to create a multi-scale information learning structure. Each dilated inception

module used three dilated convolutions with a kernel size of 3×3 and different dilation factors: 1, 2, and 3. This method was used instead of the inception module with different size kernels in the Inception U-Net. Dilated convolutions increase the field of view by keeping the same size with its input so that it can concatenate these convolution outputs having different scale feature information.
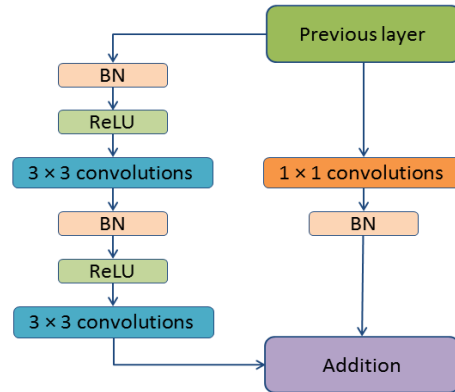

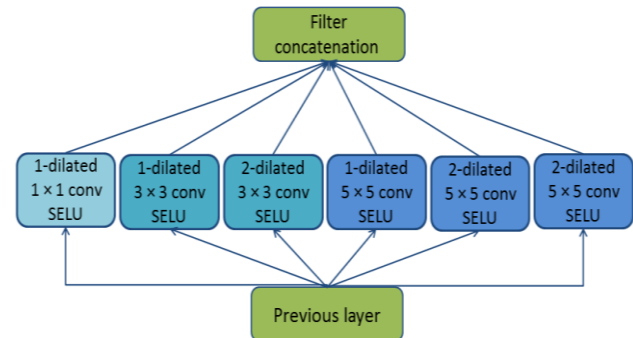Figure 3. The residual module utilized the Residual U-Net models.

Figure 4. The dilated inception module utilized in the proposed models.

Our proposed network was created with dilated inception modules consisting of convolutions with multiple kernel sizes with two different scale dilation to extract features from a wider receptive field simultaneously. Unlike the original dilated inception modules with the kernel size of a convolutional layer is 3 × 3, we used convolutions with multiple kernel sizes, group normalization, and scaled exponential linear unit (SELU). We adopted 1 × 1, 3 × 3, and 5 × 5 convolutions with the 1 and 2 kernel dilation factors as shown in Figure 4.

Network with SELUs has self-normalizing features as it enables stabilization of the mean and variance in the previous layers at each layer [33]. Networks trained with small batch sizes suffer from the incorrect prediction of the batch statistics and increase the training error dramatically when using batch normalization (BN). Therefore, we used group normalization (GN), which is independent of batch size [34].

4. EXPERIMENTS AND EVALUATIONS

In our study, the U-Net variant models were trained on 5-fold cross-validation sets divided from the private dataset consisting of 258 training set images. We chose the number of k fold as 5 to save computational complexity and time. The images were obtained as anonymized images from the Acıbadem Fulya Hospital IVF Center. The embryo images were captured continuously over five days with the integrated time-lapse imaging system Embryoscope© developed by Vitrolife (USA).

4.1. Experiment Setup

4.1.1 Dataset

In our private dataset used in this study, the embryo images were acquired using an Embryoscope system an optic system that uses Hoffman modulation contrast to perform microscopic imaging similar to an inverted microscope. In total 303 embryo images of 500× 500 size were included in the study.

We also used a publicly available human blastocyst dataset introduced by Saeedi et al. [20] to better evaluate how effective our proposed segmentation method is. This dataset consists of 249 embryo images acquired using an Olympus IX71 inverted microscope with Nomarski differential interference contrast (DIC) optics from the Pacific Centre for Reproductive Medicine (PCRM), Canada.

4.1.2 Ground Truth Acquisition

The regions of the embryo in the images were annotated as belonging to the embryo or the background by experts from Acıbadem Fulya Hospital IVF Center using The VGG Image Annotator (VIA). Ground truth images were generated by filling out all the closed boundary coordinates of the embryo in an annotation csv file using the OpenCV library.

4.1.3 Data Augmentation

Data augmentation in a dataset can prevent overfitting in training data and provide a better generalization to test data. Since the number of labeled images were limited, we utilized the augmentation methods to increase the number of samples. We enriched the data required for training by augmenting the number of samples 20 times with rotation, horizontal-vertical flip, horizontal-vertical shift, and zoom. The images were split into a training set and a test set with a ratio of 85:15. The training set in the private dataset which contained a total of 258 images of (256,256) size was increased to 5418 images using the augmentation methods details of which are listed below:

- Shifting horizontally or vertically (0.1)
- Flipping horizontally or vertically
- Zooming in or out [0.9, 1.1]
- Rotating by an angle [0, 360]

4.1.4 Hyperparameter Optimization

The Adaptive Momentum (Adam) [35] is a technique for efficient stochastic optimization to optimize all network parameters. Root Mean Square Propagation (RMSProp) [36] is an adaptive learning rate method that updates parameters using a momentum on the rescaled gradient. Dice loss [37] is a loss function based on Dice coefficient, an evaluation metric for segmentation results, Tversky loss function based on the Tversky index deals with imbalanced class problems. Tversky loss [38] is different from Dice loss because of the equal weights for false positives (FP) and false negatives (FN).

Combinations of two different loss functions and two optimizers to reduce the loss between ground truth and prediction of the network were used. We adopted Dice and Tversky as loss functions and Adam and RmsProp as optimizers with an initial learning rate of 0.0001 and 0.045, respectively. Furthermore, we used the early stopping callback to avoid overfitting when the loss value has not reduced for 15 epochs. Similarly, when the validation loss failed to improve for 5 consecutive steps, the learning rate was reduced by the factor of 0.2.

4.1.5 Implementation Framework

The model was trained and tested using an NVIDIA GeForce® RTX 2080 Ti GPU with 11 GB of memory and 32 GB of RAM. The model was implemented using Keras with a TensorFlow backend. We used a minibatch size of 4, early stopping, and maximum epochs of 50.

4.2. Evaluation Metrics

To evaluate the U-Net model's performance, predictions of the embryo segmentations were compared with their corresponding ground truth annotations. We used different common evaluation metrics such as accuracy, Jaccard index, Dice coefficient, and precision. These metrics depend on four parameters: TP represents the number of true positives (pixels correctly included as embryo pixels), FP represents the false positives (incorrectly included as embryo pixels), FN represents the false negatives (pixels incorrectly excluded although they were part of the embryo pixels) and TN represents the true negatives (pixels correctly included as background pixels).

Pixel accuracy (ACC) is the percent of pixels in the image that were classified correctly as the background and embryo.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

The Jaccard index, also called intersection over union (IOU), is a similarity measure that evaluates the number of pixels shared between the ground truth and prediction masks divided by their union.

$$\text{Jaccard index} = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP+FP+FN} \qquad (2)$$

The Dice similarity coefficient (DSC), also termed as overlap index or F1 Score [39], denotes the similarity between predicted and ground truth embryo regions. The Dice coefficient is very similar to the IOU.

$$\text{Dice coefficient} = \frac{2 \times |A \cap B|}{|A|+|B|} = \frac{2 \times TP}{2 \times TP+FP+FN} \qquad (3)$$

Precision, also known as positive predictive value (PPV), represents the percent of correctly predicted pixels among all the predicted pixels.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (4)$$

5. RESULTS AND DISCUSSION

5.1. Quantitative Results

To evaluate of segmentation performance of the proposed method, the Dice coefficient, Jaccard index, Accuracy, and precision were adopted. The results of the proposed model and U-Net variant models for the embryo segmentation in test data were tabulated, taking into account the different optimizers and loss functions utilized, and were shown in Tables 1, 2, 3, and 4.

As shown in Tables 1, 2, 3, and 4, all models trained with Adam optimizer and Dice loss showed higher performance results compared to their equivalents in the other table, while our customized Dilated Inception U-Net achieved the best overall blastocyst segmentation performance with a Jaccard index of 97.52% and a Dice score of 98.68%.

Table 1. Performance of different U-Net models with Adam optimizer and Dice loss function (in %).

Architecture	DSC	IOU	ACC	PPV
Base U-Net	98.56	97.16	99.14	98.29
Inception U-Net	98.58	97.19	99.15	98.39
Dilated Inception U-Net	98.58	97.18	99.15	98.38
ResUNet	98.63	97.29	99.18	98.40
Custom Dilated Inception U-Net	**98.68**	**97.52**	**99.20**	**98.52**

Table 2. Performance of different U-Net models with Adam optimizer and Tversky loss function (in %).

Architecture	DSC	IOU	ACC	PPV
Base U-Net	98.3	96.66	98.97	97.32
Inception U-Net	98.54	97.12	99.12	98.1
Dilated Inception U-Net	98.49	97.03	99.09	97.8
ResUNet	98.45	96.95	99.06	97.57
Custom Dilated Inception U-Net	**98.61**	**97.27**	**99.15**	**98.12**

Table 3. Performance of different U-Net models with RmsProp optimizer and Dice loss function (in %).

Architecture	DSC	IOU	ACC	PPV
Base U-Net	98.08	96.23	98.85	97.78
Inception U-Net	**98.41**	**96.87**	**99.05**	98.31
Dilated Inception U-Net	98.35	96.76	99.02	**98.73**
ResUNet	98.33	96.71	99.0	98.34
Custom Dilated Inception U-Net	98.30	96.84	99.04	97.93

Table 4. Performance of different U-Net models with RmsProp optimizer and Tversky loss function (in %).

Architecture	DSC	IOU	ACC	PPV
Base U-Net	96.32	92.91	97.72	93.09
Inception U-Net	97.98	96.04	98.77	96.59
Dilated Inception U-Net	98.29	96.63	98.97	**97.47**
ResUNet	98.15	96.37	98.88	97.20
Custom Dilated Inception U-Net	**98.25**	**96.79**	**99.00**	97.30

The data in Table 4, which includes models using RmsProp optimizer with Tversky loss function, suggests that our proposed model outperformed the commonly used Base U-Net by 1.93% in Dice coefficient and 3.88% in Jaccard index, 4.21% in precision, and 1.28% in accuracy. In our study, we showed that a model with

dilated convolution-based inception module including using group normalization and SELU can be successful.

5.2. Comparison with State of the Art

In order to demonstrate the robustness and reliability of our proposed model, the results with the public dataset and our private dataset were shown in Table 5 together with the results of the relevant studies in the literature. The second column shows the region of interest that is intended to be segmented in the study. Our customized model was compared to five present state-of-the-art approaches [20], [22], [26–28] aiming different blastocyst

components segmentation of embryo on the same public dataset. In TE segmentation, our proposed model outperformed state-of-the-art, [22], by 1.52% in Dice Coefficient and 2.18% in Jaccard index. The Dice coefficient, Jaccard index, accuracy, and precision values in the ICM segmentation of the proposed model were 91.97%, 85.20%, 98.77%, and 90.69%, respectively. These values were higher than the studies performed for ICM segmentation in the literature [20], [26], [27]. The blastocyst segmentation outcomes showed that our customized Dilated Inception U-Net is a robust model which produces higher metric values in Dice coefficient, Jaccard index, and precision than the state-of-the-art study [40].
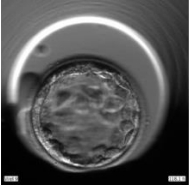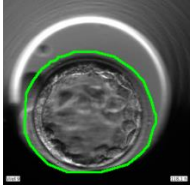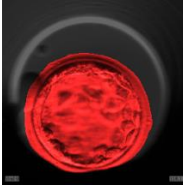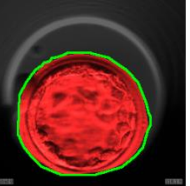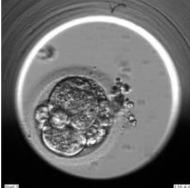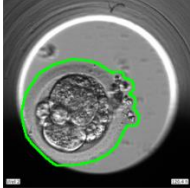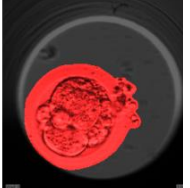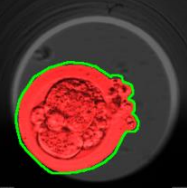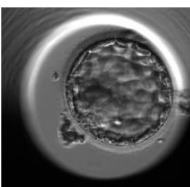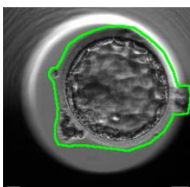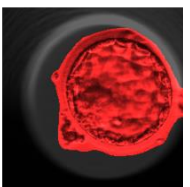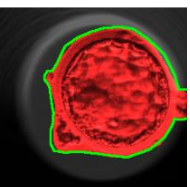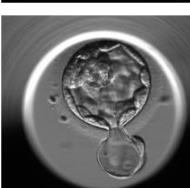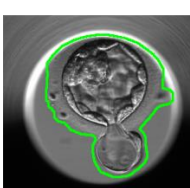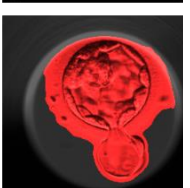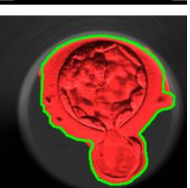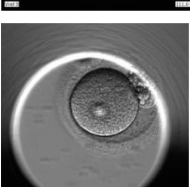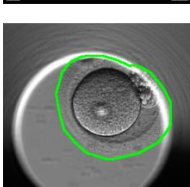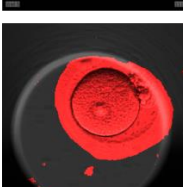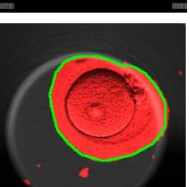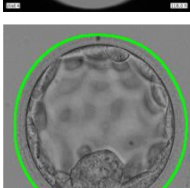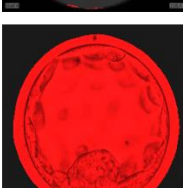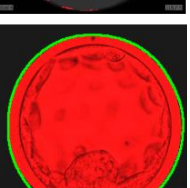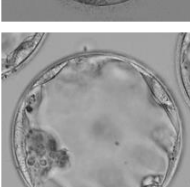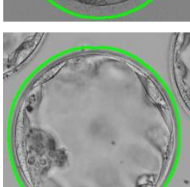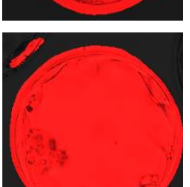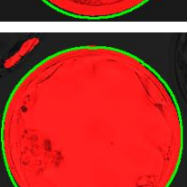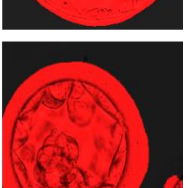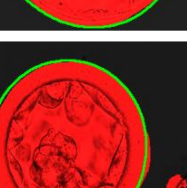
Table 5. Performance comparison between the proposed model and the state-of-the-art approaches (in %).

	Region	**DSC**	**IOU**	**ACC**	**PPV**	**Dataset**
Saeedi et al. [20]	TE	77.3	63.0	86.6	69.0	Dataset [20]
Saeedi et al. [20]	ICM	83.1	71.1	93.3	84.5	Dataset [20]
Kheradmand et al. [26]	ICM	86.7	76.5	95.6	-	Dataset [20]
Rad et al. [27]	ICM	89.5	81.6	98.3	88.6	Dataset [20]
Harun et al. [28]	ICM	94.3	89.3	99.1	94.9	Dataset [20]
Harun et al. [28]	TE	92.5	85.3	98.3	91.8	Dataset [20]
Harun et al. [40]	Zona-Ablated Blastocysts	98.4	96.9	**99.4**	98.1	Private dataset
Rad et al.[22]	TE	86.50	76.56	96.89	-	Dataset [20]
Custom Dilated Inception U-Net	Blastocysts	**98.68**	**97.52**	99.20	**98.52**	Private dataset
Custom Dilated Inception U-Net	Blastocysts	98.11	96.31	97.71	97.83	Dataset [20]
Custom Dilated Inception U-Net	ICM	91.97	85.20	98.77	90.69	Dataset [20]
Custom Dilated Inception U-Net	TE	88.02	78.74	97.31	87.08	Dataset [20]

5.3. Qualitative Evaluation

To verify the segmentation model results, the predicted embryo area was compared with the manually labeled ground truth embryo by experts. The embryo ground truth area was overlaid on the predicted embryo to evaluate the differences. Figure 5 shows the results of several human embryo images predicted by proposed models overlaid on

the ground truth images in private and public datasets. All results of the customized method were organized from the best to worst prediction samples in public and private datasets, according to the Jaccard index values. In our private data set, the best and worst Jaccard index values were 98.50% and 92.85%, respectively, while in the public dataset these values were 98.61% and 91.51%, respectively.

| Dataset	Input Image	Ground truth (GT)	Segmentation results	Comparison with GT	Performance Metrics (in %)
Private dataset					Dice: 99.25 IOU: 98.51 Accuracy:99.50 Precision:98.88
Private dataset					Dice: 98.82 IOU: 97.66 Accuracy:99.42 Precision:99.22
Private dataset					Dice: 98.26 IOU: 96.59 Accuracy:98.80 Precision:97.75
Private dataset					Dice: 97.88 IOU: 95.86 Accuracy: 98.51 Precision: 96.19
Private dataset					Dice: 96.29 IOU: 92.85 Accuracy: 98.00 Precision: 95.61
Public dataset[20]					Dice: 99.30 IOU: 98.61 Accuracy: 99.13 Precision:99.30
Public dataset[20]					Dice: 98.18 IOU: 96.43 Accuracy: 97.69 Precision:97.07
Public dataset[20]					Dice: 97.33 IOU: 94.80 Accuracy: 97.19 Precision:95.26
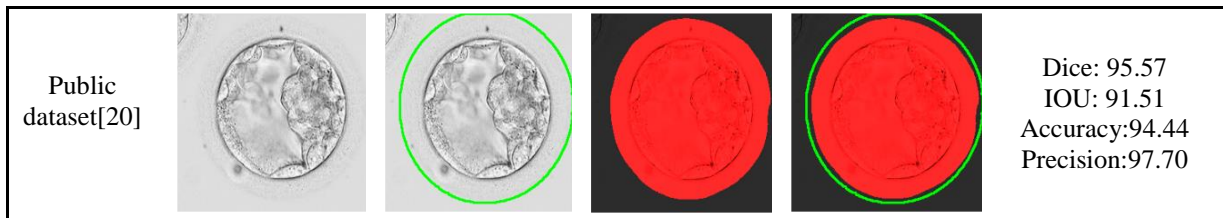
Figure 5. Sample results of the custom segmentation method: Green lines imply contour of the ground truth, red highlights the segmented blastocyst by proposed model.

6. CONCLUSION

In this study, we compared different optimization algorithms such as Adam and RmsProp and different loss functions such as Dice and Tversky in U-Net variant architectures and proposed Dilated Inception U-Net to measure the performance for embryo segmentation. Our custom models adopted group normalization and SELU activation function in place of batch normalization and ReLU, respectively. The best performance was achieved with 98.68% Dice coefficient and 97.52% Jaccard index by using our proposed Dilated Inception U-Net utilizing Adam optimizer and Dice loss. Besides, we observed that models using Adam optimizer with Dice loss showed better embryo segmentation performance than the other models using combinations of loss function and optimizer in our experiments. As a part of the evaluation, we had also implemented unweighted and weighted ensemble methods of these U-Net variant models in this study; however, as these methods performed inadequately, they were not included in the manuscript. Further work utilizing various architectures and larger data sets is needed to determine the best approach in embryo segmentation.

ACKNOWLEDGMENT

REFERENCES

[1]   K. S. Tamilselvan and G. Murugesan, "Image Segmentation", **Medical and Biological Image Analysis**, Editor: R.Koprowski, IntechOpen, London, UK, 2018.

[2]   P. Aggarwal, V. Renu, S. Bhadoria, and C. Dethe, "Role of Segmentation in Medical Imaging: A Comparative Study", *International Journal of Computer Applications*, 29(1), 54–61, 2011.

[3]   M. Vander Borght and C. Wyns, "Fertility and infertility: Definition and epidemiology", *Clinical Biochemistry*, 62, 2–10, 2018.

[4]   G. E. Crawford and W. L. Ledger, "In vitro fertilisation/intracytoplasmic sperm injection beyond 2020", *BJOG: An International Journal of Obstetrics & Gynaecology*, 126(2), 237-243, 2019.

[5]   D. Castelló, Y. Motato, N. Basile, J. Remohí, M. Espejo-Catena, and M. Meseguer, "How much have we learned from time-lapse in clinical IVF?", *Molecular Human Reproduction*, 22(10), 719–727, 2016.

[6]   A. M. Lee, M. T. Connell, J. M. Csokmay, and A. K. Styer, "Elective single embryo transfer- the power of one", *Contraception and Reproductive Medicine,* 1, 11, 2016.

[7]   D. M. Kissin, A. D. Kulkarni, V. A. Kushnir, and D. J. Jamieson, "Number of Embryos Transferred After In Vitro Fertilization and Good Perinatal Outcome", *Obstetrics and Gynecology,* 123(2 Pt 1), 239–247, 2014.

[8]   R. J. Heitmann, M. J. Hill, K. S. Richter, A. H. DeCherney, and E. A. Widra, "The simplified SART embryo scoring system is highly correlated to implantation and live birth in single blastocyst transfers", *Journal of Assisted Reproduction and Genetics*, 30(4), 563–567, 2013.

[9]   J. L. Collins, B. van Knippenberg, K. Ding, and A. V.Kofman, "Time-Lapse Microscopy", **Cell Culture**, Editor: Radwa Ali Mehanna, IntechOpen, London, UK, 2018.

[10]   K. Kirkegaard, A. Ahlström, H. J. Ingerslev, and T. Hardarson, "Choosing the best embryo by time lapse versus standard morphology", *Fertility and Sterility,* 103(2), 323–332, 2015.

[11]   L. Sundvall, H. J. Ingerslev, U. Breth Knudsen, and K. Kirkegaard, "Inter- and intra-observer variability of time-lapse annotations", *Human Reproduction,* 28(12), 3215–3221, 2013.

[12]   D. J. Kaser and C. Racowsky, "Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review", *Human Reproduction Update*, 20(5), 617–631, 2014.

[13]   E. S. Filho, J. A. Noble, M. Poli, T. Griffiths, G. Emerson, and D. Wells, "A method for semi-automatic grading of human blastocyst microscope images", *Human Reproduction,* 27(9), 2641–2648, 2012.

[14] C. Manna, L. Nanni, A. Lumini, and S. Pappalardo, "Artificial intelligence techniques for embryo and oocyte classification", *Reproductive Biomedicine Online,* 26(1), 42–49, 2013.

[15] A. A. Septiandri, A. Jamal, P. A. Iffanolida, O. Riayati, and B. Wiweko, "Human Blastocyst Classification after In Vitro Fertilization Using Deep Learning", **7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)**, Online Virtual Conference, 1–4, 2020.

[16] A. Karlsson, N. C. Overgaard, and A. Heyden, "Automatic segmentation of zona pellucida in HMC images of human embryos", **Proceedings of the 17th International Conference on Pattern Recognition (ICPR)**, Cambridge, UK, 8380971, 518-521, 2004.

[17] A. Karlsson, N. Chr. Overgaard, and A. Heyden, "A Two-Step Area Based Method for Automatic Tight Segmentation of Zona Pellucida in HMC Images of Human Embryos", **Scale Space and PDE Methods in Computer Vision**, Berlin, Heidelberg, 503–514, 2005.

[18] D. A. Morales, E. Bengoetxea, and P. Larrañaga, "Automatic Segmentation of Zona Pellucida in Human Embryo Images Applying an Active Contour Model", **Proceedings of the 12th Medical Image Understanding and Analysis (MIUA)**, Dundee, UK, 209-213, 2008.

[19] E. S. Filho, J. A. Noble, and D. Wells, "A Review on Automatic Analysis of Human Embryo Microscope Images", *The Open Biomedical Engineering Journal*, 4(1), 170–177, 2010.

[20] P. Saeedi, D. Yee, J. Au, and J. Havelock, "Automatic Identification of Human Blastocyst Components via Texture", *IEEE Transactions on Biomedical Engineering,* 64(12), 2968–2978, 2017.

[21] A. Singh, J. Au, P. Saeedi, and J. Havelock, "Automatic Segmentation of Trophectoderm in Microscopic Images of Human Blastocysts", *IEEE Transactions on Biomedical Engineering,* 62(1), 382–393, 2015.

[22] R. M. Rad, P. Saeedi, J. Au, and J. Havelock, "Trophectoderm segmentation in human embryo images via inceptioned U-Net", *Medical Image Analysis,* 62(4), 101612, 2020.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", **Proceedings of the 25th International Conference on Neural Information Processing Systems**, Red Hook, NY, USA, 1097–1105, 2012.

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, Boston, MA, USA, 3431–3440, 2015.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**, Munih, Germany, 234–241, 2015.

[26] S. Kheradmand, A. Singh, P. Saeedi, J. Au, and J. Havelock, "Inner cell mass segmentation in human HMC embryo images using fully convolutional network", **2017 IEEE International Conference on Image Processing (ICIP)**, Beijing, China, 1752–1756, 2017.

[27] R. M. Rad, P. Saeedi, J. Au, and J. Havelock, "Multi-Resolutional Ensemble of Stacked Dilated U-Net for Inner Cell Mass Segmentation in Human Embryonic Images", **25th IEEE International Conference on Image Processing (ICIP)**, Athens, Greece, 3518–3522, 2018.

[28] M. Y. Harun, T. Huang, and A. T. Ohta, "Inner Cell Mass and Trophectoderm Segmentation in Human Blastocyst Images using Deep Neural Network", **IEEE 13th International Conference on Nano/Molecular Medicine & Engineering (NANOMED)**, Gwangju, Korea (South), 214–219, 2019.

[29] C. Szegedy *et al.*, "Going deeper with convolutions", **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, Boston, MA, USA, 1–9, 2015.

[30] M. A. Kızrak and B. Bolat, Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma, *Bilişim Teknolojileri Dergisi*, 11(3), 263–286, 2018.

[31] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net", *IEEE Geoscience Remote Sensing Letters,* 15(5), 749–753, 2018.

[32] W. Shi, F. Jiang, and D. Zhao, "Single image super-resolution with dilated convolution based multi-scale information learning inception module", **2017 IEEE International Conference on Image Processing (ICIP)**, 977–981, 2017.

[33] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks", **Proceedings of the 25th International Conference on Neural Information Processing Systems**, Red Hook, NY, USA, 972–981, 2017.

[34] Y. Wu and K. He, "Group Normalization", *International Journal of Computer Vision,* 128(3), 742–755, 2020.

[35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", **Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)**, San Diego, 2015.

[36] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude", *COURSERA: Neural Networks for Machine Learning*, 4(2), 26–31, 2012.

[37] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation", **2016 Fourth International Conference on 3D Vision (3DV)**, California, USA, 565–571, 2016.

[38] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks", **Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017**, QC, Canada, 379–387, 2017.

[39] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool", *BMC Medical Imaging*, 15(29), 2015.

[40] M. Y. Harun *et al.*, "Image Segmentation of Zona-Ablated Human Blastocysts", **2019 IEEE 13th International Conference on Nano/Molecular Medicine Engineering (NANOMED)**, Gwangju, South Korea, 208–213, 2019.