2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-2923-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISBI52829.2022.9761654

EARLY PREDICTION OF BLASTOCYST DEVELOPMENT VIA TIME-LAPSE VIDEO ANALYSIS

Xiang Xie[1*], *Pengxiang Yan*[1*], *Fang-Ying Cheng*[2*], *Feng Gao*[3†], *Qingyun Mai*[2†], *Guanbin Li*[1†]

[1] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China
[2] The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[3] The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

ABSTRACT

Advances in assisted reproductive technology allow more and more infertile patients to benefit from in vitro fertilization (IVF) treatment. For an IVF treatment, selecting embryos that have developed into the blastocyst stage is a crucial step for the subsequent embryo transfer. In the clinic, it requires embryologists to keep observing the developmental status of embryos, which not only highly depends on the professional level of reproductive experts but greatly increases patients' economic cost. To address this problem, we propose a novel task termed *early blastocyst development prediction*, which aims at predicting the potential that an embryo can develop into the blastocyst stage by partially observing its early development status so as to assist embryologists with early embryo selection. To achieve this goal, we collect a new IVF dataset with 2,898 time-lapse videos of embryo development. Based on this, we also propose a benchmark solution named Attentive Multi-focus Selection Network (AMSNet). Specifically, AMSNet is a deep learning-based time-lapse video analysis method that includes a separate attention mechanism to exploit the features of embryoscope images captured at multiple focal planes and a temporal feature channel shift operation to obtain memory capability over time-lapse videos. Experimental results demonstrate the effectiveness and clinical significance of our proposed AMSNet on our IVF dataset.

Index Terms— early blastocyst development prediction, time-lapse video analysis.

1. INTRODUCTION

With the development of assisted reproductive technology, in vitro fertilization (IVF) has become an important technique in the treatment of infertility. Due to the limited survival rate of embryos in vitro culture, it generally requires to simultaneously culture multiple embryos in each IVF cycle. All IVF embryos are cultured in the time-lapse [1] incubator, as it can provide a stable culture environment, and continuously record the morphological information of embryos. Fig. 1 illustrates the developmental stages of an embryo. After being placed in incubator, the zygote will divide continuously and enter the morula and blastocyst stages successively. Most embryos take about 5 days to develop into blastocysts, and a few take 7 days. During this period, some embryos may stop developing and fail to enter the blastocyst stage, and only blastocysts will be considered for transfer [2]. Therefore, identifying whether an embryo has become a blastocyst is a critical step in IVF and lies an important foundation for the subsequent embryo transfer.

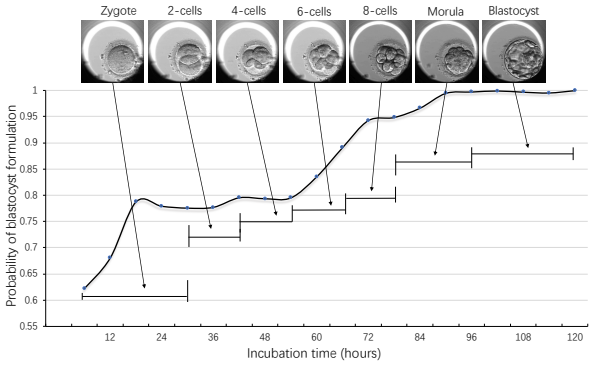

Fig. 1. An example of early blastocyst development prediction.

In an IVF cycle, to determine whether an embryo develops into the blastocyst state, embryologists need to carefully observe the development of the embryo under the microscope at certain time points (e.g., Day 5 and Day 6). This highly depends on the embryologists' experience. Moreover, failure to identify the embryos with low blastocyst development potential early will severely limit the size of the group that can benefit from high quality IVF and greatly increase the economic expenditure. Therefore, an automatic and accurate blastocyst development prediction technique is of great clinical significance for IVF. To achieve this goal, we draw inspiration from a sub-direction in the field of computer vision called early action prediction [3, 4, 5, 6] and propose a novel task named *early blastocyst development prediction*, which aims at providing an early prediction of the probability of an embryo

*Each author contributes equally to this work.
†Corresponding authors.

developing into a blastocyst in an IVF cycle. Instead of recognizing the blastocyst stage by observing the whole culture process, this task tries to predict the potential of blastocyst formation by observing the early incubation. Therefore, it can assist the embryologists to stop the culture of embryos with low potential for blastocyst formation in an earlier stage, so as to release more IVF space and alleviate the financial burden.

Time-lapse incubator can take embryoscope images of embryos at a regular interval from 7 focal planes and combine those images into a time-lapse video. Benefiting from time-lapse videos, algorithms could be easily applied to assist embryo selection. Recently, several algorithms [7, 8, 9] have been proposed to apply the morphokinetic analysis to detect the occurrence of key developmental events including the blastocyst formation. Although developmental event detection can benefit embryo selection, these methods still rely on embryologists to annotate a large number of developmental events by observing the morphological changes of embryos. Moreover, it's unfeasible for these methods to capture the full temporal and spatial richness of time-lapse videos with only a few handcrafted morphological parameters. To solve this problem, Tran *et al.* [10] proposed a deep learning-based model to analyze the entire time-lapse videos and predict the fetal heart (FH) pregnancy probability of embryos to assist the embryo selection. Nevertheless, a successful FH pregnancy also depends on the maternal pregnancy environment. Unfortunately, none of existing methods are able to explore the developmental potential in the early stage of embryos and predict the probability of blastocyst formation for embryos, which means they are powerless for stopping the culture for embryos with low developmental potential in an early stage.

In this paper, to improve the above-mentioned issues and realize the goal of early blastocyst development prediction, we propose a new time-lapse video dataset, as well as a benchmark solution. Specifically, we propose Attentive Multi-focus Selection Network (AMSNet), which is built upon a ResNet-50 [11] with two major components, i.e., a multi-focus feature selection (MFS) module and a temporal shift module (TSM) [12]. Firstly, MFS takes the embryoscope images shot at multiple focal planes as input in every moment. It includes a channel-wise attention module to selectively fuse the multi-focus feature channels and a Gaussian non-local [13] mechanism to model pair-wise spatial correlation. Secondly, TSM is responsible to partially shift the feature channels along the temporal dimension, which endows AMSNet with memory capability to achieving a temporal understanding of time-lapse videos. As shown in Fig. 1, the AMSNet can continuously read time-lapse data and provide a high-accurate blastocyst development prediction to assist embryologists with early embryo selection. In summary, the main contributions of this paper include: (1) We propose a new research task named early blastocyst development prediction, which has clinical significance for early embryo selection in clinical IVF. (2) We collect a new time-lapse incubation dataset from clinical IVF and provide a benchmark solution for early blastocyst development prediction. (3) Experimental results demonstrate the effectiveness of our proposed AMSNet on our constructed time-lapse dataset.

2. DATASET

We collaborate with the First Affiliated Hospital of Sun Yat-sen University to collect 2,898 embryo data samples and build a new IVF dataset. Specifically, we used two EmbryoScope time-lapse machines as the IVF incubators to fertilize and culture embryos. During the incubation process, these two incubators acquired the microscope images of all embryos every 10 or 15 minutes at seven focal planes. And the resolution of an embryoscope image is 500 × 500. All the embryos would be cultured until they reached the blastocyst stage or stopped developing, which generally occurs during Day 5 to Day 7 of the incubation. Experienced embryologists would determine the timing to stop the incubation of an embryo and provide an annotation of whether the embryo has developed into a blastocyst or not. Statistically, the collected IVF dataset is composed of 2,898 time-lapse videos of embryo development, of which 1,634 embryos had developed into the blastocyst stage and 1,264 had not (blastocyst: 56.38%, nonblastocyst: 43.62%). We randomly divide the dataset into three parts, including 1,746 for training, 576 for validation, and 576 for testing.

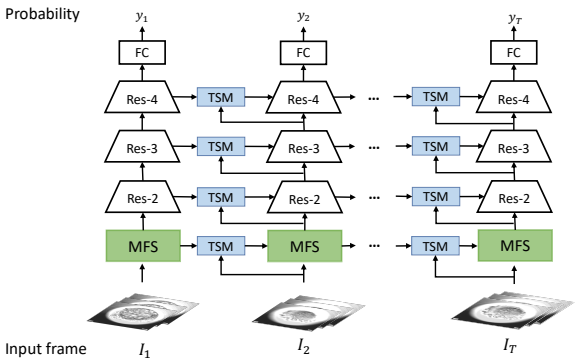

Fig. 2. Overview of AMSNet. For each moment t, it takes the embryo images I_t from multiple focal planes as input and generate the probability of blastocyst formulation $y_t \in [0, 1]$.

3. METHOD

Given a time-lapse video of an IVF treatment V, we sample T frames from the video I_1, I_2, ..., I_T. For each frame I_t, we utilize the embryo images taken at m focal planes, i.e., $I_t = \{F_t^1, F_t^2, \ldots, F_t^m\}$. To achieve the goal of early blastocyst development prediction, we propose the Attentive Multi-focus Selection Network (AMSNet). As shown in Fig. 2, AMSNet adopts an ResNet-50 [11] as the backbone, which consists of four blocks, i.e., Res-1, Res-2, Res-3, Res-4 and a fully connected (FC) layer. For each moment $t \in \{1, 2, \ldots, T\}$, AMSNet takes the multi-focus images of I_t as input. The features from multiple focal planes will

be selectively integrated through our proposed Multi-focus Feature Selection Module (MFS) by exploiting both channel-wise and spatial attention. Then the multi-focus features produced by MFS will be fed into Res-2 and the rest parts of ResNet-50 to predict the probability $y_t \in [0, 1]$ that the embryo can develop into a blastocyst at the current moment t. Furthermore, we inject the temporal shift module (TSM) [12] into AMSNet to endow ResNet-50 with memory capability so as to utilize temporal information more effectively.

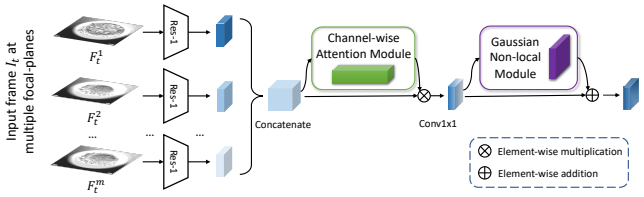

Fig. 3. Multi-focus Feature Selection Module (MFS)

3.1. Multi-focus Feature Selection Module(MFS)

For each moment, Our IVF dataset provides the embryoscope images of multiple focal planes. To effectively consider the image features of different focal planes, we propose the MFS module to selectively exploit the multi-focus features via an attention mechanism. Inspired by [14], we propose to decompose the attention module into channel-wise attention and spatial attention as a separate attention scheme is much more efficient to process the 3D multi-focus feature maps [14]. As shown in Fig. 3, without adding any parameters, the MFS module first adopts m individual Res-1 branches with shared weights to process the input images from m focal planes. Then the features from m branches are concatenated and enhanced by the channel-wise attention module. Next, the channel dimension of the channel-wise enhanced feature is reduced to as same as that of the output feature map of Res-1. Last, it is further enhanced by the spatial attention module via a residual connection pattern.

As for channel-wise attention, we exploit the inter-channel relationship of multi-focus features. Specifically, we first concatenate the feature maps of m weight-shared Res-1 branches into $U = [u_1, u_2, ..., u_{mc}]$, where c denotes the number of channels in each feature map. Then we apply both average pooling and maximum pooling for each channel u_i to obtain the channel features $V \in \mathbf{R}^{mc}$ and $M \in \mathbf{R}^{mc}$, and forward them to a shared MLP. Then we use element-wise addition to merge the output features of shared MLP and get the channel attention feature $W \in R^{mc}$ by a sigmoid function. After element-wise multiplying W with U and reducing the channel dimension, we obtain the channel-wise enhanced feature $F \in \mathbf{R}^{c \times H \times W}$.

As for spatial attention, we propose to utilize a 2D Gaussian non-local module[13] to capture the spatial correlation from channel-wise enhanced feature F. By computing interactions between any two positions on feature map F, regardless of their position distance, Gaussian non-local can achieve 2D pair-wise spatial correlation learning, and thus enhence F in spatial dimension. Experimental results demonstrate that, by adding few parameters, 2D Gaussian non-local can improve model performance.

3.2. Temporal Shift Module (TSM)

To produce a reliable prediction of blastocyst formulation, model not only need to learn the appearance features from embryoscope images but also need to model the developmental status of embryos along temporal dimension of time-lapse videos. Therefore, we refer to a widely used residual temporal shift module (TSM)[12] and embed each residual block in AMSNet with a temporal channel shift operation. Without adding any parameters, by transferring part of features forward, temporal channel shift operation equips the AMSNet with temporal modeling and memory abilities. Specifically, as shown in Fig. 4, for each residual block, TSM shifts part of the channels of its input feature at moment $t-1$, i.e., X_{t-1}, into that at moment t, i.e., X_t , to obtain a temporally combined feature X_{t-1}^t. Then, TSM enhances X_t with X_{t-1}^t via a residual addition to obtain a temporally enhanced output feature Y_t. Thus, the appearance features of the embryo at each moment contain the features of previous developmental state.

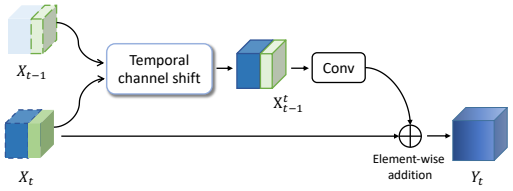

Fig. 4. Residual Temporal Shift Module (TSM)

4. EXPERIMENT

4.1. Experiment Setup

Implementation Details. The AMSNet is implemented on PyTorch[15], a flexible framework for deep learning. During the training process, we split each time-lapse video clip into small segments at a pace of 6h and randomly sample one frame from each segment. In total, we sample the time-lapse data of 7 days (168h), i.e, T = 28 frames from each video. For those samples with fewer than 168h of time-lapse data, we pad the number of sample frames to T = 28 and ignore the loss back-propagation of the padded frames. For each frame, the AMSNet uses the embryoscope images shot at seven focal-planes at most. The resolution of each video clip is resized and randomly cropped to 224×224, and the format of each video clip is converted to grayscale as embryo is transparent. We initialize AMSNet with the ResNet50 pretrained on the Kinetics [16] dataset, and choose SGD as the optimizer during the training process.

Evaluation Metrics. Our formulated early blastocyst development prediction problem can be regarded as a video binary

No.	Methods	TSM	Channel	Spatial	#F	Day0.5	Day1	Day1.5	Day2	Day2.5	Day3	Day3.5	Day4	Day4.5	Day5	Day5.5	Day6	Day6.5	Day7	AUC
1	Baseline (ResNet-50 [11])				1	62.15	64.24	68.40	70.83	71.88	71.60	69.69	74.74	77.18	81.04	80.99	82.33	84.21	**100.00**	0.6987
2	Baseline+TSM	√			1	64.24	68.58	72.57	71.53	74.31	73.69	73.87	76.31	81.01	86.23	89.92	89.27	90.18	**100.00**	0.7354
3	Baseline+TSM	√			3	66.49	67.88	71.53	73.09	72.92	72.65	74.74	77.35	84.15	86.23	88.97	89.27	88.42	**100.00**	0.7360
4	Baseline+TSM	√			5	67.54	69.27	74.13	73.09	74.48	74.91	74.74	76.66	81.53	84.79	87.07	88.01	89.47	**100.00**	0.7371
5	Baseline+TSM	√			7	67.54	68.92	71.88	73.26	73.61	74.74	73.35	79.62	81.36	85.69	88.40	88.33	89.12	**100.00**	0.7372
6	Baseline+TSM+Channel	√	√		3	67.54	70.66	74.13	75.69	74.48	74.56	78.75	81.71	83.80	86.94	88.78	89.27	88.77	**100.00**	0.7509
7	Baseline+TSM+Channel	√	√		5	67.01	71.01	75.00	75.17	**76.91**	78.05	78.05	78.92	83.10	88.01	90.11	88.33	89.12	**100.00**	0.7538
8	Baseline+TSM+Channel	√	√		7	67.88	71.70	**75.69**	75.52	**76.91**	77.53	78.57	78.75	82.75	87.12	89.73	88.96	90.53	**100.00**	0.7555
9	Baseline+TSM+MFS (AMSNet)	√	√	√	3	**68.92**	71.70	75.17	75.87	73.44	76.48	**79.09**	81.53	**85.19**	88.19	91.45	**90.22**	**90.88**	**100.00**	0.7598
10	Baseline+TSM+MFS (AMSNet)	√	√	√	5	68.58	**72.57**	75.52	**76.91**	76.74	77.88	78.75	80.31	84.15	87.84	91.26	88.96	90.18	**100.00**	0.7609
11	Baseline+TSM+MFS (AMSNet)	√	√	√	7	67.88	**72.57**	75.00	**76.91**	76.39	**78.40**	78.57	**82.23**	84.15	**88.55**	**91.83**	89.27	**90.88**	**100.00**	**0.7633**

Table 1. Ablation study using accuracy (%) and AUC on our IVF dataset. The values in each column indicate the accuracy of different models at a particular incubation time point. "Channel" and "Spatial" denote the channel-wise attention module and Gaussian non-local module of MFS, respectively. "#F" denotes the number of input focal planes.

classification task by observing different portions of time-lapse frames. To quantitatively evaluate the performance of the proposed method, we report the accuracy scores of the early prediction with respect to different observation times as well as the area under curve (AUC). In addition, we also report the receiver operating characteristic (ROC) curve and its AUC at several time points, including Day 3, Day 4, Day 5.

4.2. Results on the IVF Dataset

We report the results of the AMSNet on the test set of our IVF dataset. The input of AMSNet is time-lapse data with 7 focal planes. As shown in Fig. 5 and Fig. 6, the AMSNet improves the performance of its baseline model (ResNet-50 [11]) on both Accuracy and ROC curves by a large margin. Specifically, the resulting AUC of its Accuracy curve reaches 0.7633. The resulting AUC scores of its ROC curves to predict the blastocyst formulation on Day 3, Day 4, and Day 5 are 0.819, 0.874 and 0.958, respectively. Experimental results demonstrate that the proposed AMSNet can provide clinically significant prediction earlier than the blastocyst formulation, such as on Day 4 (AUC of ROC: 0.874).

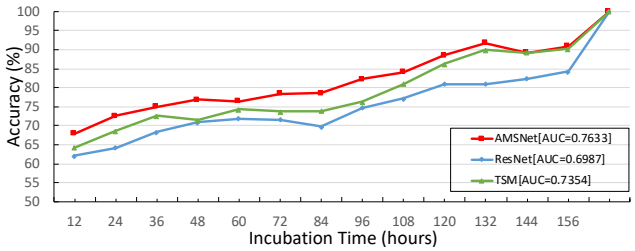

Fig. 5. Results of Accuracy curves on our IVF dataset. [AUC= *] denotes the area under the Accuracy curve.

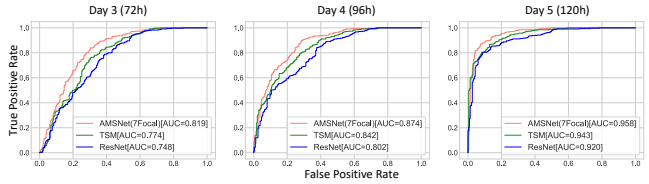

Fig. 6. Results of ROC curves of different observation time points on our IVF dataset. [AUC= *] denotes the area under the ROC curve.

4.3. Ablation Study

Effectiveness of TSM. As shown in Table 1, the baseline injected with TSM (Baseline+TSM) can outperform the baseline (ResNet-50) by 3.67% w.r.t the AUC of Accuracy curve. Moreover, as shown in Fig. 5 and Fig. 6, its Accuracy and ROC curves are generally better than those of the baseline model. These experimental results validate the effectiveness and necessity of the TSM module.

Effectiveness of Using Data With More Focal Planes. As show in No.3 to No.11 rows of Table 1, the more focal planes the input data contains, the heigher prediction accuracy each model can obtain. Specifically, these models in No.3 to No.5 rows use the directly concatenated multiple focal images as the input of TSM. These experimental results prove that the introduction of more focal plane information will inevitably bring greater model advantages and generalization.

Effectiveness of MFS. As shown in the No.2 and No.11 rows of Table 1, by considering multiple focal planes as input the MFS module can further assist AMSNet to outperform the performance w.r.t Baseline+TSM by a large margin, i.e, 2.79% w.r.t AUC of Accuracy curve. The better performance of Accuracy and ROC curves of AMSNet presented in Fig. 5 and Fig. 6 also validate the effectiveness of MFS. In addition, we provide the comparison between MFS and its two variants in the No.3 to No.5 and No.6 to No.8 rows of Table 1. We observe that the channel-wise attention and 2D Gaussian non-local [13] module involved in the MFS module both benefit the model performance and are complementary to each other.

5. CONCLUSION

In this paper, we propose a novel task named early blastocyst development prediction, which aims at assisting embryologists to select embryos at an early stage of IVF. To achieve this goal, we propose a new IVF dataset with 2,898 embryo time-lapse videos and a benchmark solution named AMSNet. Experimental results demonstrate the effectiveness of our proposed method. In future work, we will expand the scale of the proposed dataset. And we plan to further exploit the physiological and genetic information of patients to provide a more explainable algorithm for embryo selection and fetal heart pregnancy prediction.

6. ACKNOWLEDGE

This work is supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant No.2020B15 15020048, in part by the National Natural Science Foundation of China under Grant No.61976250, in part by the Guangzhou Science and Technology Project under Grant 202102020633, in part by National Natural Science Found of China (No: 81270750), in part by Natural Science Found of Guangdong China (No: 2019A1515011845).

7. REFERENCES

[1] Sarah Armstrong, Priya Bhide, Vanessa Jordan, Allan Pacey, Jane Marjoribanks, and Cindy Farquhar, "Time-lapse systems for embryo incubation and assessment in assisted reproduction," *Cochrane Database of Systematic Reviews*, , no. 5, 2019.

[2] David K Gardner, William B Schoolcraft, Lyla Wagley, Terry Schlenker, John Stevens, and John Hesla, "A prospective randomized trial of blastocyst culture and transfer in in-vitro fertilization.," *Human Reproduction*, vol. 13, no. 12, pp. 3434–3440, 1998.

[3] Michael S Ryoo, "Human activity prediction: Early recognition of ongoing activities from streaming videos," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 1036–1043.

[4] Mohammad Sadegh Aliakbarian, Fatemeh Sadat Saleh, Mathieu Salzmann, Basura Fernando, Lars Petersson, and Lars Andersson, "Encouraging lstms to anticipate actions very early," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 280–289.

[5] Yu Kong, Zhiqiang Tao, and Yun Fu, "Deep sequential context networks for action prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1473–1481.

[6] Xionghui Wang, Jian-Fang Hu, Jian-Huang Lai, Jianguo Zhang, and Wei-Shi Zheng, "Progressive teacher-student learning for early action prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3556–3565.

[7] Yamileth Motato, María José de los Santos, María José Escriba, Belén Aparicio Ruiz, José Remohí, and Marcos Meseguer, "Morphokinetic analysis and embryonic prediction for blastocyst formation through an integrated time-lapse system," *Fertility and Sterility*, vol. 105, no. 2, pp. 376–384, 2016.

[8] Simon Fishel, Alison Campbell, Sue Montgomery, Rachel Smith, Lynne Nice, Samantha Duffy, Lucy Jenner, Kathryn Berrisford, Louise Kellam, Rob Smith, et al., "Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth," *Reproductive Biomedicine Online*, vol. 37, no. 3, pp. 304–313, 2018.

[9] Miriam J Haviland, Lauren A Murphy, Anna M Modest, Matthew P Fox, Lauren A Wise, Yael I Nillni, Denny Sakkas, and Michele R Hacker, "Comparison of pregnancy outcomes following preimplantation genetic testing for aneuploidy using a matched propensity score design," *Human Reproduction*, vol. 35, no. 10, pp. 2356–2364, 2020.

[10] D Tran, S Cooke, PJ Illingworth, and DK Gardner, "Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer," *Human Reproduction*, vol. 34, no. 6, pp. 1011–1018, 2019.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[12] Ji Lin, Chuang Gan, and Song Han, "Tsm: Temporal shift module for efficient video understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7083–7093.

[13] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.

[14] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[15] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer, "Automatic differentiation in pytorch," 2017.

[16] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al., "The kinetics human action video dataset," *arXiv preprint arXiv:1705.06950*, 2017.