A Novel Embryo Morphology Evaluation Based on Improved YOLOv8 Object Detection Model

Ouafa Talha[1], Wenju Zhou[1](✉), Yuan Xu[2], Qiang Liu[3], and Jethro Odeyemi[4]

[1] Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China
zhouwenju@shu.edu.cn

[2] Department of Obstetrics and Gynecology, First People's Hospital, Shanghai Jiaotong University, Shanghai, China

[3] School of Engineering and Technology, Faculty of Engineering, University of Bristol, Bristol, UK

[4] Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, Canada

Abstract. In this study, an improved YOLOv8 object detection model (YOLOv8-C2f-CA) is proposed to automate the detection and quantification of embryonic cells, supporting embryologists in accurately assessing embryo morphology. By integrating a Coordinate Attention (CA) mechanism into the YOLOv8 architecture, our model achieved 87.8% mean average precision (mAP), 83.9% precision, and 76.4% recall, outperforming the baseline YOLOv8. The lightweight CA mechanism is incorporated into both the backbone and neck networks of YOLOv8, bolstering the model's capacity to identify the key morphological features of embryos without substantially impacting its size or computational efficiency. This method facilitates a rapid and precise evaluation process, minimizing the need for extensive time and human resources while maintaining precise accuracy.

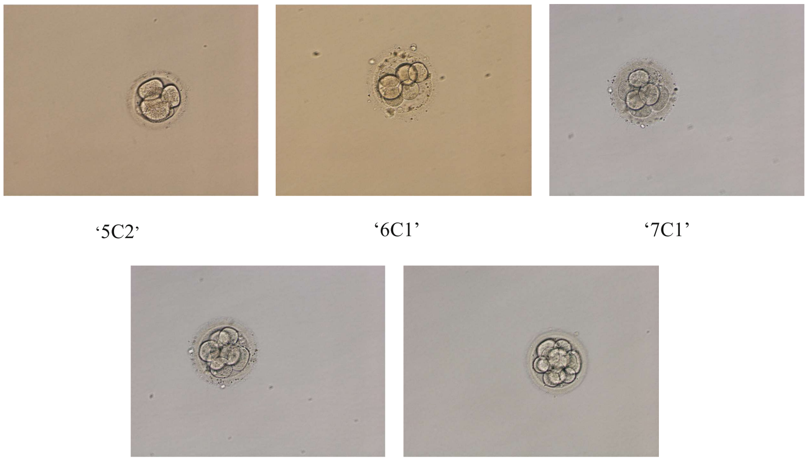
Keywords: Embryo morphology evaluation · Object detection · Assisted reproductive technology

1 Introduction

In the domain of Assisted Reproductive Technologies (ART) [1], selecting the optimal embryo for transfer is a crucial process that requires an intricate analysis of its morphological characteristics such as size, shape, and cell count during the cleavage stages. A critical aspect of this assessment involves monitoring the change in the number of embryonic cells at various developmental stages, from the initial 2-cell stage to the 14-cell stage. The number of embryonic cells at each stage is a key indicator of normal embryonic development, which helps identify potential abnormalities and determine the optimal transfer timing [2]. Figure 1 illustrates the developmental stages of the embryo, with different embryonic cell counts at each stage. Traditionally, these evaluations have been manually performed by embryologists [3]. However, this process is prone to variability and inaccuracies due to intra- and interobserver differences.

J. Gu et al. (Eds.): LSMS/ICSEE 2024, CCIS 2220, pp. 193–204, 2024.
https://doi.org/10.1007/978-981-96-0313-8_15

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) in medical imaging has shown potential in addressing the complexities of manual analysis in human embryo images. These technologies automate various tasks such as grade classification, viable embryo identification, and morphological features segmentation [4]. Key studies include Kragh et al. [5], who applied a self-supervised method for embryo viability prediction, and Liu et al. [6], who used a multi-task deep learning algorithm for classifying embryonic development stages. Further contributions include the application of AI in morphological feature analysis, including automatic blastomere cell recognition and counting, and the segmentation of trophectoderm (TE) and inner cell mass (ICM) regions [7–9]. Building on these advancements, object detection models have also emerged as powerful tools for automating complex image analysis tasks in various fields [10, 11]. These models excel in identifying detailed features in images, making them particularly suited to embryology's demands for precise observation of morphological and dynamic features.



'5C2' '6C1' '7C1'

'8C2' '10C2'

Fig. 1. Human embryos at different development stages.

This study introduces a novel embryo evaluation approach using an improved YOLOv8 object detection model, YOLOv8-C2f-CA, designed for the automated identification and counting of embryonic cells in human embryo images. This proposed approach has demonstrated the ability to accurately identify and count embryonic cells at different developmental stages. A comprehensive performance evaluation across six distinct metrics reveals that the improved model outperforms the baseline YOLOv8 model in terms of accuracy while maintaining minimal increases in model size and computational demands. The effectiveness of our approach, validated against embryologists' manual annotations, demonstrates its potential to reduce embryologists' workload by providing a reliable, automated tool for embryo evaluation.

2 Related Work

2.1 YOLO-Based Object Detection

Object detection methods are broadly categorized into two-stage algorithms and one-stage algorithms. Two-stage algorithms first generate a series of candidate bounding boxes as samples and then classify these samples using a convolutional neural network. In contrast, one-stage algorithms approach object detection as a regression task, directly predicting the bounding box and classifying objects across multiple locations within the entire image.

In 2016, Redmon et al. [12] introduced YOLO (You Only Look Once) into the field of object detection, revolutionizing the performance and efficiency of detection methods. YOLO is a single-stage detector that employs a grid-based approach for both object localization and classification, significantly outperforming previous detection methods.

The most recent iteration, YOLOv8 [13], introduces several enhancements, including an anchor-free design, a more efficient backbone network, and advanced loss functions. These enhancements make YOLOv8 a state-of-the-art solution for a wide range of object detection tasks [14].

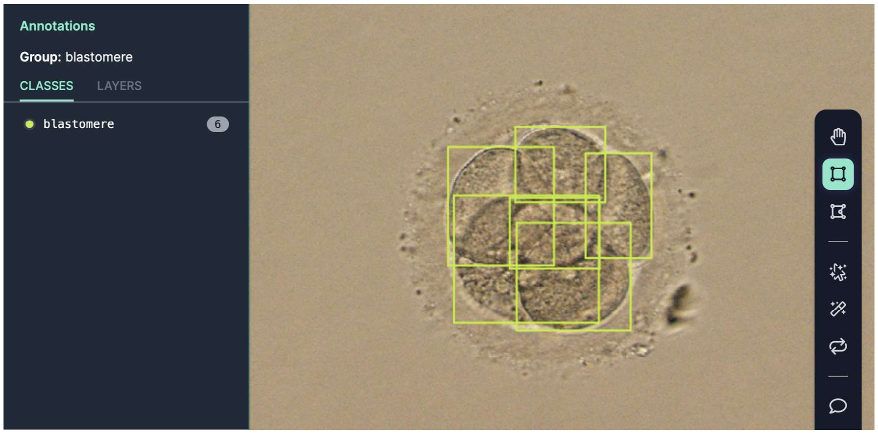
2.2 Attention Mechanism

Attention mechanisms, initially proposed by Bahdanau et al. in 2014 for neural machine translation [15], have since been widely adopted across various computer vision applications [16]. These mechanisms enable models to selectively focus on relevant regions of the input data, thereby improving feature representation and enhancing detection accuracy. Among these, the Squeeze and Excitation (SE) attention mechanism [17] stands out for its ability to adaptably recalibrate feature map channels based on their relative importance. However, the SE method primarily focuses on recalibrating channel-wise features by compressing global spatial information into channel descriptors, which makes it difficult to preserve the precise location details essential for capturing spatial structures in visual tasks. Therefore, the Coordinate Attention (CA) mechanism [18] is proposed to account for both inter-channel relationships and spatial information. This mechanism incorporates adaptive average pooling, concatenation, and the computing of attention coefficient for feature map recalibration, demonstrating its superior performance and lightweight characteristics compared to alternative attention mechanisms [19, 20].

3 Methods and Materials

3.1 Data

In this study, we analyzed a collection of 250 human embryo images obtained from Shanghai General Hospital in China using an Olympus microscope. These images depict embryos cultured from 2 to 5 days after fertilization. Embryologists graded each embryo using the ‘n-C-g’ notation based on cell quality, where ‘n’ denotes the cell count and ‘g’ indicates the grade. For instance, ‘6C3’ represents an embryo with six cells and quality level of 3. To prepare for YOLO model training, we labeled the embryonic cells in the images using the Roboflow data labeling tool [21]. Each image underwent careful labeling and verification by embryologists to ensure accuracy. Prior to augmentation, the original dataset was partitioned into training, validation, and testing sets using an 80:15:5 ratio. Subsequently, we applied data augmentation techniques, including rotations and adjustments in brightness, to expand our dataset to 503 images. The detailed labeling process is shown in Fig. 2.



Group: blastomere

● blastomere 6

Fig. 2. Labeling process of embryonic cells. A '6C3' grade human embryo image with 6 labeled blastomeres.

3.2 Improved YOLOv8: YOLOv8-C2f-CA

Building upon the YOLOv8 architecture, this study introduces an improved version, YOLOv8-C2f-CA, optimized for the detection and counting of embryonic cells in human embryo images. As illustrated in Fig. 3, the proposed network architecture consists of three main components: the backbone network (backbone), the bottleneck network (neck), and the detection layer (head). The backbone network incorporates the standard convolution module (CBS), C2f module, Coordinate Attention (CA) mechanism, and Spatial Pyramid Pooling Module (SPPF). The bottleneck network is composed of CBS and C2f modules along with the coordinate attention mechanism, followed by a series of concatenation operations. The key improvement in the YOLOv8-C2f-CA model over the original YOLOv8 architecture is the strategic incorporation of the coordinate attention mechanisms that directly follow each C2f module in both the backbone network and neck network. The C2f module is a fundamental feature extraction unit that constitutes the entire network. The subsequent inclusion of the CA module enhances the model's ability to extract complex features, enabling a more focused analysis of relevant morphological characteristics in embryo images.

C2f Module. Figure 4 illustrates the C2f module, a critical component in the feature extraction process of our network architecture. This module incorporates a CBS module,

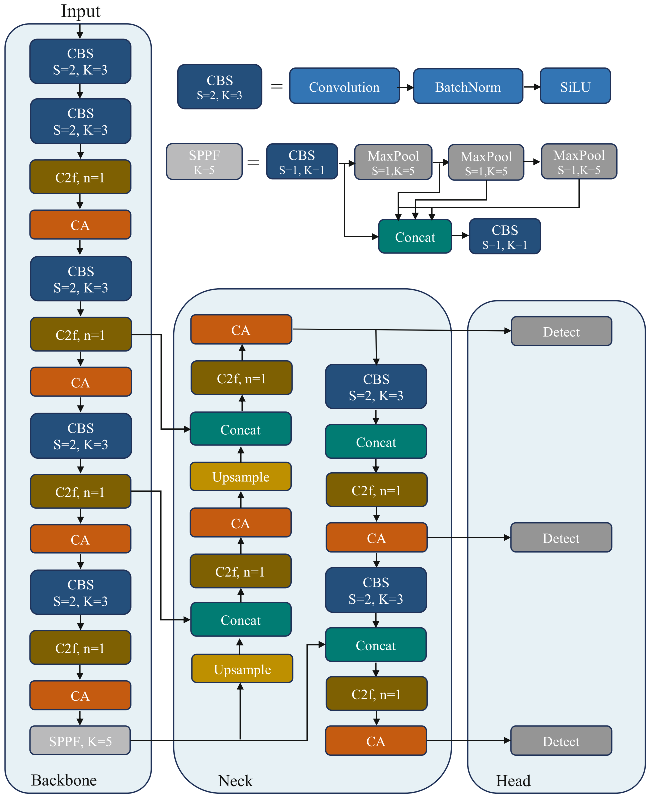


Fig. 3. The architecture of the improved YOLOv8.

which is essential for the initial phase of feature extraction. It consists of convolution, batch normalization, and the SiLU activation function. These components are responsible for performing down-sampling, dimensionality adjustment, normalization, and introducing non-linearity into the feature maps. Subsequently, the feature maps undergo a sequence of transformations through a series of bottlenecks, transitioning from basic representations to complex feature maps. The C2f module adeptly combines detailed and contextual information from various scales through strategic residual connections between different feature levels. Initially, these maps are rich in detail but lack broader context. As they advance, they gain contextual depth at the potential expense of finer details. To augment the feature representation capability of the network, particularly for the complex task of analyzing human embryo images, the coordinate attention mechanism is incorporated following the C2f module. This integration focuses on enhancing the spatial information handling within the module, enabling the network to prioritize relevant spatial details in the overlapping structures observed in embryo images.

Coordinate Attention Mechanism. When processing images, it is crucial for the object detection model to focus on the target regions rather than the entire image. The attention mechanism enables the model to concentrate on these target regions, enhancing the

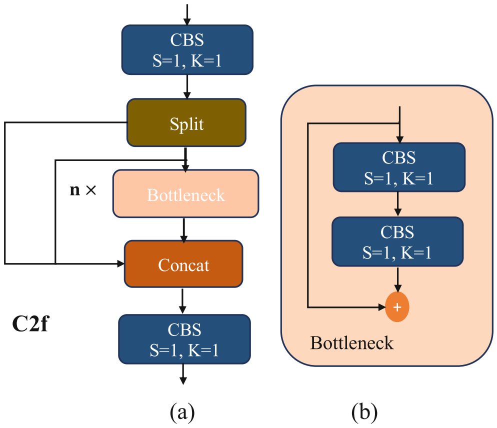


Fig. 4. The architecture of the C2f module: (a) C2f module; (b) Bottleneck in C2f.

detection performance. To ensure that the proposed model focuses on embryonic cell features and minimizes the influence of irrelevant exfoliated cells and fragmentations in the background, the coordinate attention modules are incorporated into both the backbone and neck networks of YOLOv8. By applying the CA module to features extracted from each C2f module, more relevant embryonic information is obtained. The CA mechanism extends traditional attention by considering both inter-channel dependencies and spatial information. It is designed to efficiently capture relevant features without adding significant computational overhead. The structure diagram of the CA mechanism is shown in Fig. 5, where H and W denote the height and width of the feature map, C represents the number of channels, and r is the reduction ratio.

For an input feature map X of dimensions $C \times H \times W$, we apply pooling kernels of size $H \times 1$ and $1 \times W$ along the horizontal and vertical coordinates, respectively. This process generates feature maps, f_h and f_w, with dimensions $C \times H \times 1$ and $C \times 1 \times W$. These maps are concatenated and adaptively modulated via a convolutional layer that includes batch normalization and a ReLU activation function, producing a combined feature map of dimensions $C \times 1 \times (W + H)$. Attention coefficients g^h and g^w, sized $C \times H \times 1$ and $C \times 1 \times W$ respectively, are then calculated through convolutional operations with sigmoid activations. These coefficients recalibrate the input feature map X via element-wise multiplication, emphasizing critical spatial locations while suppressing less relevant ones. The output of this CA mechanism can be defined as follows:

$$Y(i,j) = X(i,j) \otimes g^h(i) \otimes g^w(j) \tag{1}$$

where $\otimes$ denotes the element-wise product.

3.3 Performance Evaluation

To evaluate the performance of the improved YOLOv8 model, we used six essential metrics: mean average precision (mAP), precision, recall, inference time, model size,

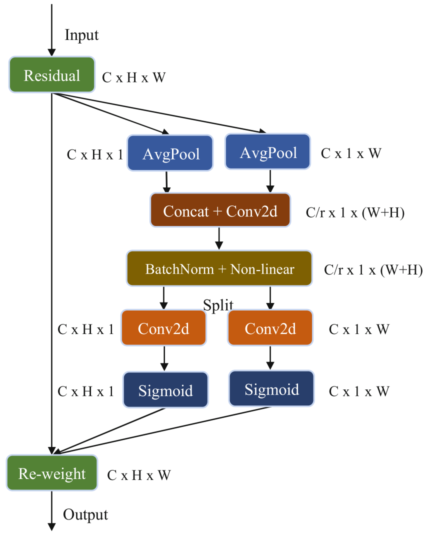


Fig. 5. The architecture of the Coordinate Attention (CA) mechanism.

and giga floating point operations per second (GFLOPs). The mAP is the average precision (AP) across different classes as shown in Eq. (2), AP itself is derived from the area under the precision-recall curves, indicated in Equation (). Precision, outlined in Eq. (4), measures the accuracy of the model's positive predictions, while recall, detailed in Eq. (5), evaluates the model's capability to identify all correct instances of embryonic cells. In these equations, 'P' and 'R' represent precision and recall, respectively, with 'TP' indicating true positives, 'FP' denoting false positives, and 'FN' standing for false negatives.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{2}$$

$$AP = \frac{1}{N}\sum_{R_i} PR_i \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

Inference time measures the network's speed to process an input image and generate predictions. The complexity of the model, known as the model size, is quantified by aggregating all trainable parameters across all layers, as expressed in Eq. (6).

$$Model\ size = \sum_{i=1}^{N} l_i \tag{6}$$

where l_i indicates the number of trainable parameters in the i^{th} layer, with N denoting the total number of layers. Additionally, GFLOPs evaluate the computational efficiency by quantifying the number of floating-point operations a model can perform in one second, measured in billions (G).

3.4 Experimental Setup and Parameters Settings

The experiments were performed on Intel HD Graphics 630. The software setup included Python 3.9.12 and PyTorch 1.13.1 versions. During the training process, specific parameters were adopted to ensure optimal performance. As depicted in Table 1, the size of the input images were uniformly adjusted to 640 × 640 pixels, and we used a consistent batch size of 16. Optimization was carried out using the Adam optimizer, which was set with a learning rate of 0.002 and a momentum value of 0.9. The training process was conducted for 25 iterations for all models.

Table 1. Parameters settings.

Parameter	Value
Image size	640 × 640
Batch size Optimizer	16 Adam
Learning rate	0.002
Momentum	0.9
Epochs	25

4 Results and Discussion

4.1 Detection Results

To illustrate the improvements made to the YOLOv8 network, this paper compares and analyzes the embryonic cell detection results of YOLOv8-C2f-CA and the original YOLOv8. As observed in Fig. 6, the baseline YOLOv8 exhibits instances of false negatives (FN) and false positives (FP). Notably, it fails to detect all embryonic cells across various embryo grades. In addition, it incorrectly identifies and classifies non blastomere-cells as blastomeres, potentially overestimating the embryonic cell count and introducing inaccuracies in embryo analysis. However, the YOLOv8-C2f-CA model effectively addresses these limitations, improving accuracy in accurately identifying the correct number of embryonic cells across various stages of embryo development.

On the other hand, Table 2 provides a quantitative evaluation of the YOLOv8-C2f-CA model compared to the baseline YOLOv8. This comparative analysis showcases a notable improvement in detection accuracy with YOLOv8-C2f-CA, as demonstrated by a 5.66% increase in mAP, along with improved precision and recall rates by 3.2%

and 7.76%, respectively. This improvement indicates that the identification of embryonic cells is more reliable, with reduced false positives and false negatives, while maintaining minimal increases in model size and computational efficiency.

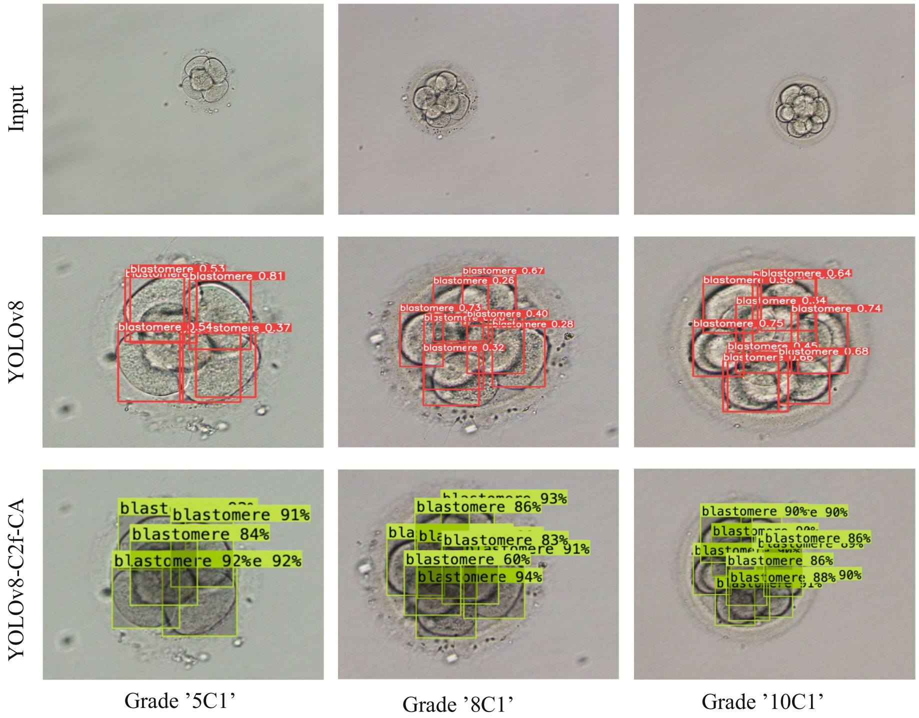


Fig. 6. Comparative analysis of detection results using the original YOLOv8 and the improved YOLOv8-C2f-CA across different embryonic development stages.

Table 2. Performance comparison of YOLOv8 and YOLOv8-C2f-CA.

Method	Precision	Recall	mAP0.5	Model Size	GFLOPs	Inference Time
YOLOv8	0.813	0.709	0.831	6.3 MB	12.2	134.8 ms
YOLOv8-C2f-CA (ours)	0.839	0.764	0.878	7.2 MB	15.7	165 ms

4.2 Effect of CA Position on Network Performance

To evaluate the impact of integrating the coordinate attention mechanism at various positions within the network, a detailed performance analysis is conducted, with findings presented in Table 3. The baseline YOLOv8 model includes four C2f modules in its backbone network and an additional four in the neck network. The designation YOLOv8-C2f-CA (N) refers to the integration of the CA mechanism after the Nth C2f module, with N ranging from 1 to 8, reflecting the total count of C2f modules within the network. Meanwhile, YOLOv8-C2f-CA (all) indicates the integration of the CA mechanism following every C2f module in both the backbone network and neck network. Our results indicate that the placement of the CA mechanism significantly affects both the accuracy and size of the model. Optimal performance is observed with YOLOv8-C2f-CA (all), achieving the highest accuracy without an increase in computational demands, as indicated by lower GFLOPs and model size. Conversely, the least effective configurations are YOLOv8-C2f-CA (4) and YOLOv8-C2f-CA (8), which increased the model's computational efficiency without enhancing its accuracy.

Table 3. Performance comparison of introducing the CA module in different positions of the network.

Method	Precision	Recall	mAP0.5	Model Size	GFLOPs	Inference Time
YOLOv8-C2f-CA (1)	0.81	0.734	0.853	8.8 MB	19.4	190.3 ms
YOLOv8-C2f-CA (2)	0.811	0.701	0.833	8.8 MB	19.6	190.8 ms
YOLOv8-C2f-CA (3)	0.821	0,716	0.845	8.8 MB	19.8	191 ms
YOLOv8-C2f-CA (4)	0.787	0.705	0.858	11.8 MB	21.2	202.2 ms
YOLOv8-C2f-CA (5)	0.814	0.729	0.851	10.6 MB	19.2	191.2 ms
YOLOv8-C2f-CA (6)	0.828	0.732	0.863	8.9 MB	17.9	178.8 ms
YOLOv8-C2f-CA (7)	0.824	0.741	0.867	10.6 MB	20.7	196.4 ms
YOLOv8-C2f-CA (8)	0.818	0.706	0.856	11.9 MB	21.2	202.7 ms
YOLOv8-C2f-CA (all)(ours)	0.839	0.764	0.878	7.2 MB	15.7	165 ms

5 Conclusion

In this paper, we propose an improved YOLOv8 object detection model to automate the detection and quantification process of embryonic cells in human embryo images. This novel approach supports embryologists in accurate embryo morphology evaluation in assisted reproductive technologies (ART). By integrating a coordinate attention mechanism into the baseline YOLOv8, the accuracy of embryonic cell detection is improved, and the precise quantification of the cells across different developmental stages is achieved, providing an efficient alternative to the traditional manual techniques that are prone to subjectivity and variability.

The comprehensive evaluation, employing six performance metrics, reveal that the improved YOLOv8-C2f-CA model not only outperforms the baseline YOLOv8 model in terms of mean average precision, precision, and recall but also maintains an optimal model size, ensuring practical applicability in clinical settings. The validation of the proposed approach against the evaluations of experienced embryologists underscores its reliability and potential to greatly reduce the workload involved in embryo morphology assessment. Due to dataset limitations, we did not conduct experiments on additional datasets. Future work will focus on enhancing the model's robustness and generalization by incorporating larger and more diverse datasets of embryo images, including those from varied sources to introduce depth information.

Acknowledgments. This work is supported by Shanghai Pujiang Program (No. 22PJ1403800) and the Natural Science Foundation of China (NSFC) under Grant 62203290, and 111 Project (No. D18003) and Natural Science Foundation of China (61877065).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Jain, M., Singh, M.: Assisted reproductive technology (ART) techniques (2022)
2. Fernando, P., Sophie, D., Josephine, G.L., Inge, A.: The cleavage stage embryo. Hum. Reprod. (2012)
3. Gardner, D.K., Balaban, B.A.: Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still important? Mol. Hum. Reprod. **22**(10), 704–718 (2016)
4. Riegler, M., et al.: Artificial intelligence in the fertility clinic: Status, pitfalls and possibilities. Hum. Reprod. **36**, 2429–2442 (2021)
5. Kragh, M.F., Rimestad, J., Lassen, J.T., Berntsen, J., Karstoft, H.: Predicting embryo viability based on self-supervised alignment of time-lapse videos. IEEE Trans. Med. Imaging **41**(2), 465–475 (2021)
6. Liu, Z., et al.: Multi-task deep learning with dynamic programming for embryo early development stage classification from time-lapse videos. IEEE Access **7**, 122153–122163 (2019)
7. Rad, R.M., Saeedi, P., Au, J., Havelock, J.: Blastomere cell counting and centroid localization in microscopic images of human embryo. In: 2018 IEEE 20th International Workshop on Multimedia Signal Processing, Vancouver, Canada, pp. 1–6. IEEE (2018)
8. Rad, R.M., Saeedi, P., Au, J., Havelock, J.: Trophectoderm segmentation in human embryo images via inceptioned U-Net. Med. Image Anal. **62**, 101612 (2020)
9. Saeedi, P., Yee, D., Au, J., Havelock, J.: Automatic identification of human blastocyst components via texture. IEEE Trans. Biomed. Eng. **64**(12), 2968–2978 (2017)
10. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)

11. Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X.: Object detection with deep learning: a review. IEEE Trans. Neural. Netw. Learn. Syst. **30**(11), 3212–3232 (2019)
12. Redmon, J., Divvala, D., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.779–788 (2016)
13. Jocher, G., Chaurasia, A., Qiu, J.: YOLO by Ultralytics. https://github.com/ultralytics/ultralytics (2023). Accessed 29 Feb 2024
14. Li, Y., Huang, H., Han, Z., Gu, Q.: A modified YOLOv8 detection network for UAV aerial image recognition. Drones **7**(5), 304 (2023)
15. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
16. Guo, M.H., et al.: Attention mechanisms in computer vision: a survey. Comp. Vis. Media **8**(3), 331–368 (2022)
17. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. IEEE (2018)
18. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13713–13722 (2021)
19. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534–11542 (2020)
20. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
21. Dwyer, B., Nelson, J., Solawetz, J.: Roboflow (version 1.0) [software] (2022)