Transparent Tracking of Spermatozoa with YOLOv8

Bao-Tin Nguyen[1,2,†], Van-Loc Nguyen[1,2,†] and Minh-Triet Tran[1,2,3,*]

[1]University of Science, VNUHCM, Vietnam
[2]Viet Nam National University Ho Chi Minh City, Vietnam
[3]John von Neumann Institute, VNU-HCM, Vietnam

Abstract

Accurate spermatozoa tracking is crucial for understanding fertilization and reproductive health and developing novel sperm-based diagnostics and therapies. This paper explores the application of YOLOv8, a state-of-the-art object detection model, for automated and robust spermatozoa tracking in microscopic videos. Our approach employs transfer learning, fine-tuning the pre-trained YOLOv8 model on the VISEM-Tracking dataset of labeled spermatozoa images. We evaluate the performance of the proposed method on a specific part of the dataset, which does not participate in the training step. By providing a reliable and efficient tool for automated spermatozoa tracking, this work paves the way for further research, advancements, and applications in the field of reproductive medicine.

1. Introduction

Evaluating sperm under a microscope manually takes a long time and requires expensive experts with a lot of training. In addition, manual analysis is not very dependable because it is tough to consistently track, identify, and count sperm in fresh samples. While computer-aided sperm analyzer (CASA) systems were designed to address these issues, recent studies have pointed out that they struggle with the inconsistency of semen samples in real clinical settings.

In recent years, deep learning-based approaches have emerged as powerful tools for automated and objective analysis of spermatozoa. Among various deep learning object detection models, YOLOv8 [1] has gained significant attention due to its high accuracy and real-time performance.

We participate in the Medical Multimedia Task - Transparent Tracking of Spermatozoa [2], one task of the MediaEval 2023 challenge. This paper explores the potential of YOLOv8 for accurate and efficient spermatozoa tracking in microscopy images. We train and evaluate the performance of YOLOv8 on the VISEM-Tracking dataset [3] of sperm images and demonstrate its effectiveness in detecting and tracking individual spermatozoa. Our results on both sperm detection and tracking, i.e., Subtask 1, and efficient detection and tracking, i.e., Subtask 2 suggest that YOLOv8 can be a valuable tool for automated and objective analysis of spermatozoa.

In conclusion, the use of deep learning-based object detection models like YOLOv8 has the potential to revolutionize sperm analysis by providing faster, more accurate, and cost-effective results. This study underscores YOLOv8's significance in automating spermatozoa tracking for potential applications in clinical settings.

2. Related Work

Spermatozoa tracking in microscopic videos demands both laser-sharp detection and seamless connection across frames. This challenge divides computer vision algorithms into two main

MediaEval'23: Multimedia Evaluation Workshop, February 1–2, 2024, Amsterdam, The Netherlands and Online
*Corresponding author.
†These authors contributed equally.
20120596@student.hcmus.edu.vn (B. Nguyen); 20120131@student.hcmus.edu.vn (V. Nguyen); tmtriet@fit.hcmus.edu.vn (M. Tran)
0009-0009-0347-5099 (B. Nguyen); 0000-0001-9351-3750 (V. Nguyen); 0000-0003-3046-3041 (M. Tran)

camps: the methodical two-stage approach and the lightning-fast one-stage approach.

2.1. The Two-Stage Approach

The two-stage approach in object detection involves initially proposing candidate regions in an image and subsequently classifying and refining those regions. This approach operates with meticulous precision. Algorithms under this paradigm meticulously examine each frame, carefully identifying individual sperm. Examples include advanced methods like DeepSORT [4], an extension of SORT [5] integrating deep appearance features for improved accuracy. Furthermore, there are methods with Deep Convolutional Neural Networks (DCN) [6], and variants of the Region-based Convolutional Neural Network (R-CNN) family, such as Faster R-CNN and Mask R-CNN, are noteworthy representatives [7] [8] [9]. These methods excel in accuracy, especially in challenging environments with complex backgrounds or overlapping sperm. However, its meticulous nature comes at the cost of processing power, limiting its real-time capabilities.

2.2. The One-Stage Approach with YOLOv8 in Spermatozoa Tracking

In contrast, the one-stage approach in spermatozoa tracking employs a single, high-speed model to both detect and trace sperm in microscopic videos. YOLO (You Only Look Once), a renowned object detection system, exemplifies this approach with its real-time efficiency [10]. In the context of the Transparent Tracking of Spermatozoa task at MediaEval 2022, various approaches utilizing YOLO-based models have been successful. For instance, Huynh et al. developed a simple efficient framework for tracking sperm cells using a YOLOv7 model [11], particularly focusing on tail-aware sperm detection [10]. In the same task, Kosela et al. solved the problem using YOLOv5 for object detection and StrongSORT with the OSNet tracking algorithm [9]. In our work, we opt for YOLOv8 due to its balanced performance, offering a favorable trade-off between speed and accuracy. YOLOv8's ease of implementation, coupled with the potential for optimization through fine-tuning and data augmentation, makes it an ideal choice for real-time applications such as sperm motility analysis in microscopic videos. Notably, its improved anchor-free detection enhances robustness in challenging scenarios with complex backgrounds or overlapping sperm.

3. Approach

3.1. Dataset splitting

By examining the VISEM-tracking dataset, and from the official implementation of the dataset [3], we noticed that this dataset is well-formed for being processed with YOLO models. In the training dataset provided by task organizers, there are 20 videos with spermatozoa, each with a 30-second length. We divide this into 2 parts: 80% of the dataset is used for training, and the rest 20% is used for validation.

3.2. Subtask 1: Sperm detection and tracking

As discussed, regarding the compatibility of the dataset format with the YOLO model, we employed the YOLOv8 small (s) configuration, we conducted the training duration to 100 epochs to further perform the detection and tracking task.

Our decision to use YOLOv8 small and extend the training duration was driven by the need to strike a balance between sensitivity and adaptability. By setting the Confidence Threshold at 0.25 and the NMS IoU Threshold at 0.7, we aimed to achieve several objectives:

- Balanced Sensitivity and Adaptability to Challenging Conditions: We intend our model to be sensitive enough to detect spermatozoa accurately, even in challenging conditions. With a lower Confidence Threshold, we want to ensure a higher recall rate and minimize the false negatives, as well as make the model more robust and reliable, which is crucial in medical applications.
- Accuracy and Precision: At the same time, we aim to maintain a balance between accuracy and precision. This ensures that the detected spermatozoa are indeed spermatozoa and not false alarms.
- Redundancy Removal: The NMS IoU Threshold at 0.7 allowed us to eliminate redundant or overlapping bounding boxes. This step is crucial for enhancing tracking accuracy, as it ensures that each spermatozoon is represented by a single bounding box, preventing multiple detections of the same sperm.

In summary, our approach with YOLOv8 small and the specified threshold values was designed to strike a careful balance between sensitivity, precision, adaptability, and tracking accuracy for the challenging task of spermatozoa detection and tracking in microscopy images, ultimately contributing to more accurate and efficient sperm analysis in clinical practice.

3.3. Subtask 2: Efficient detection and tracking

While YOLOv8s works well in accurate sperm identification and tracking, its inherent processing demands may not be readily compatible with the efficiency of system requirements of Task 2. Recognizing this potential limitation, we opted to prioritize Task 1's accuracy and robustness, ensuring a dependable baseline for further improvement.

4. Results and Analysis

4.1. Experiments on Validation set

The table below displays our YOLOv8-s model submission using the IoU threshold set at 0.7.

Table 1
Detection result on validation set

Submission	Precision	Recall	mAP50	mAP50-95
YOLOv8s+Conf@.25	0.5	0.628	0.506	0.191

The validation set shows a precision of 0.5 and a recall of 0.628, indicating our approach detects 62.8% of spermatozoa in microscopy images. However, the 50% false positive rate suggests room for model optimization. For insights into accuracy and robustness across overlap thresholds, we considered *mAP50* and *mAP50-95*. *mAP50* at 0.506 signifies a mean average precision of 50%, showcasing reasonable detection accuracy. *mAP50-95* at 0.191 emphasizes the need for improved robustness, especially at stricter overlap thresholds.

4.2. Submission results on the Test set of the Medico 2023 Challenge

In the evaluation of the sperm tracking task using YOLOv8, the performance metrics reveal noteworthy insights. The HOTA (Higher Order Tracking Accuracy) [12] and the OWTA (Open-World Tracking Accuracy) [13] scores demonstrate the overall tracking performance, which

Table 2
Detection results of our YOLOv8-s submission on the test set of the Medico 2023 Challenge

Metrics	Video 66	Video 68	Video 76	Video 80	Average
Precision+Thresh@.50	0.308	0.536	0.581	0.415	0.460
Precision+Thresh@.75	0.011	0.045	0.043	0.016	0.028
Recall+Thresh@.50	0.034	0.073	0.121	0.022	0.063
Recall+Thresh@.75	0.244	0.299	0.277	0.180	0.250

Table 3
Tracking results of our YOLOv8-s submission on the test set of the Medico 2023 Challenge

Submission	DetRe	DetPr	AssRe	AssPr	LocA	HOTA [12]	OWTA [13]	FPS
YOLOv8s	0.410	0.506	0.597	0.691	0.707	0.415	0.460	54.38

takes into account several of metrics of evaluating the tracker, with an overall HOTA score of 0.415 and OWTA score of 0.46. One of the strengths of the model is its precision, as demonstrated by the high DetPr (0.506) and AssPr (0.691) scores. These scores suggest that the model is very good at identifying true sperm tracks and avoiding false positives. However, the recall rates, as measured by DetRe (0.41) and AssRe (0.597), are somewhat lower. This means that the model may be missing some sperm tracks, particularly during the association phase.

Besides, the localization accuracy (LocA) of the model is also high at 70.7, indicating that it can accurately estimate the positions of the sperm cells. Moreover, our model performs inference with an FPS (frame per second) metric of 54.38 and an FLOPS (floating-point operations per second) score of 8,570,207,776, which indicates that our model is very competitive in terms of inference time.

5. Discussion and Outlook

YOLO provided some techniques for model augmentation like HSV augmentation, Image Flip, Image FixUp, Segment CopyPaste. These augmentation methods could potentially enhance the model's performance. Moreover, utilizing tools like Ray Tune, used for parameter optimization could also refine the model's performance. In the extension of our method, we can consider using these augmentation and hyperparameter tuning methods to improve the result returned by YOLOv8.

Even though the dataset is organized to be compatible with YOLO models, different methods can still be efficiently implemented to perform the sperm object detection and tracking task, resulting in the understanding of fertilization. In future work, we can find some experiments to perform transparent tracking and detection of spermatozoa with different methods, such as Deep CNN or Faster R-CNN to evaluate and compare the results between models.

Moreover, the result from this task can also be inferred to be used for the prediction of motility, the ability of an organism to move independently. And from that, we can observe whether a spermatozoon can "move forward", or can only move around in circles.

Acknowledgment
This research is funded by Viet Nam National University Ho Chi Minh City (VNU-HCM) under grant number DS2020-42-01.

References

[1] Ultralytics, YOLOv8 — docs.ultralytics.com, https://docs.ultralytics.com/models/yolov8/, 2023. [Accessed 15-12-2023].

[2] V. Thambawita, A. M. Storås, T.-L. Huynh, H.-D. Nguyen, M.-T. Tran, T.-N. Le, P. Halvorsen, M. A. Riegler, S. Hicks, Medico Multimedia Task at MediaEval 2023: Transparent Tracking of Spermatozoa, in: Proceedings of MediaEval 2023 CEUR Workshop, 2023.

[3] V. Thambawita, S. A. Hicks, A. M. Storås, T. Nguyen, J. M. Andersen, O. Witczak, T. B. Haugen, H. L. Hammer, P. Halvorsen, M. A. Riegler, Visem-tracking, a human spermatozoa tracking dataset, Scientific Data 10 (2023) 1–8. doi:10.1038/s41597-023-02173-4.

[4] N. Wojke, A. Bewley, D. Paulus, Simple online and realtime tracking with a deep association metric, in: 2017 IEEE international conference on image processing (ICIP), IEEE, 2017, pp. 3645–3649.

[5] A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, Simple online and realtime tracking, in: 2016 IEEE international conference on image processing (ICIP), IEEE, 2016, pp. 3464–3468.

[6] T. B. Haugen, O. Witczak, S. A. Hicks, L. Björndahl, J. M. Andersen, M. A. Riegler, Sperm motility assessed by deep convolutional neural networks into who categories, Scientific Reports 13 (2023). doi:10.1038/s41598-023-41871-2.

[7] S. A. Hicks, J. M. Andersen, O. Witczak, V. Thambawita, P. Halvorsen, H. L. Hammer, T. B. Haugen, M. A. Riegler, Machine learning-based analysis of sperm videos and participant data for male fertility prediction, Scientific Reports 9 (2019). doi:10.1038/s41598-019-53217-y.

[8] D. Somasundaram, M. Nirmala, Faster region convolutional neural network and semen tracking algorithm for sperm analysis, Computer Methods and Programs in Biomedicine 200 (2021) 105918. doi:10.1016/j.cmpb.2020.105918.

[9] M. Kosela, J. Aszyk, M. Jarek, J. Klimek, T. Prokop, Tracking of spermatozoa by yolov5 detection and strongsort with osnet tracker, in: MediaEval 2022 Workshop, 2022.

[10] T.-L. Huynh, H.-H. Nguyen, X.-N. Hoang, T. T. P. Dao, T.-P. Nguyen, V.-T. Huynh, H.-D. Nguyen, T.-N. Le, M.-T. Tran, Tail-aware sperm analysis for transparent tracking of spermatozoa, in: MediaEval 2022 Workshop, 2022.

[11] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464–7475.

[12] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixé, B. Leibe, Hota: A higher order metric for evaluating multi-object tracking, International Journal of Computer Vision 129 (2020) 548–578. URL: http://dx.doi.org/10.1007/s11263-020-01375-2. doi:10.1007/s11263-020-01375-2.

[13] Y. Liu, I. E. Zulfikar, J. Luiten, A. Dave, D. Ramanan, B. Leibe, A. Ošep, L. Leal-Taixé, Opening up open-world tracking, 2021. arXiv:arXiv:2104.11221.

[14] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). doi:10.1109/cvpr.2016.91.

[15] M. R. Mohammadi, M. Rahimzadeh, A. Attar, Sperm detection and tracking in phase-contrast microscopy image sequences using deep learning and modified csr-dcf (2020).

[16] J. B. You, C. McCallum, Y. Wang, J. Riordon, R. Nosrati, D. Sinton, Machine learning for sperm selection, Nature Reviews Urology 18 (2021) 387–403. doi:10.1038/s41585-021-00465-1.

[17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.