Optimizing Sperm Detection and Tracking in Fluids with Equalize Class Representation Augmentation

Trong-Hieu Nguyen-Mau[1,2], Quoc-Huy Trinh[1,2], Ngoc-Linh Nguyen-Ha[1,2],
Tuong-Vy Truong-Thuy[1,2], Tuan-Anh Yang[1,2], Hai-Dang Nguyen[1,2,*],
Ngoc-Thao Nguyen[1,2,*] and Minh-Triet Tran[1,2,*]

[1]*University of Science, VNU-HCM*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

Abstract

The task of Transparent Tracking of Spermatozoa aims to detect and track sperm in a fluid environment. In addressing this challenge, we propose a framework that utilizes YOLOv8 and BoTSORT to address the issues related to the failure to detect small objects. Additionally, we suggest incorporating the equalization augmentation method to tackle problems related to imbalanced data. Our analysis results indicate that our methods can effectively resolve the imbalance issues in each data class and accurately detect small objects. This improvement significantly enhances the overall detection results.

1. Introduction

Traditional manual sperm quality assessment through microscopy faces challenges like time consumption, the need for expert skills, and variability in results. Computer-Aided-Sperm-Analysis (CASA) systems, introduced to automate sperm identification, tracking, and counting, offer an efficient alternative for male fertility evaluation. Despite their growing popularity, CASA systems often struggle with inaccuracies. Previous deep learning approaches, including those using YOLO-based models [1, 2], have shown promise in enhancing detection and tracking. Yet, these methods still grapple with detecting small objects and addressing data imbalance, leading to reduced precision in tracking spermatozoa.

To address these shortcomings, we propose a novel approach in this challenge. Our work employs YOLOv8, a supervised model with the capability to effectively detect small objects, and apply equalization augmentation to solve the problem of an imbalanced dataset. Additionally, we assess the performance of this model using a simple tracking pipeline to underscore the crucial role of the detection model in this task.

In the 2023 MediaEval challenge [3], our focus is on the Medical Multimedia Task - Transparent Tracking of Spermatozoa. The Medico 2023 task [3] is centered on the effective tracking of sperm cells in video recordings [4]. Our participation is geared towards resolving the primary challenges in the accurate detection and tracking of sperm cells, which involves tackling both Subtask 1 and Subtask 2 of the Medico 2023 challenge.

---

2. Method

2.1. Detection model

YOLOv8 [5] represents the most recent advancement in the YOLO series of object detection models by incorporating the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN). The FPN in YOLOv8 [5] operates by progressively reducing the spatial resolution of the input image while simultaneously increasing the number of feature channels. This process generates feature maps adept at detecting objects across various scales and resolutions. Conversely, the PAN architecture enhances the model's ability to capture multi-scale and multi-resolution features essential for accurately identifying objects of diverse sizes and shapes, by integrating features from different network levels using skip connections [6]. We employed YOLOv8 [5] and its various scaled versions, including YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x in the detection stage.

2.2. Equalize Class Representation Augmentation in Sperm Detection

This comparative display in Figure 1 showcases the augmentation process designed to balance class representation in a sperm detection dataset. The first column presents the original microscopic images. The second column features these images annotated with blue bounding boxes identifying sperm, green for clusters, and red for small or pinhead spermatozoa. The third column demonstrates the images post-augmentation, indicating the enhancement of dataset diversity. The final column displays the augmented images with retained and updated annotations, ensuring accurate identification across the dataset's newly diversified spectrum.
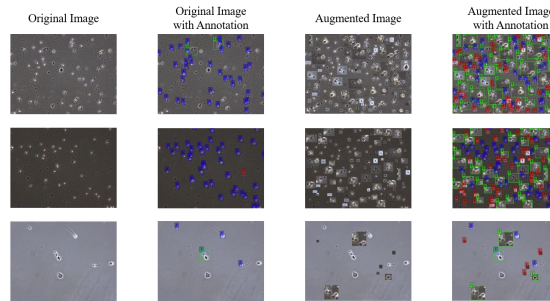

Figure 1: Visualization of the Equalize Class Representation Augmentation in Sperm Detection.

Table 1
Frequency of Different Classes

Class	Samples	Percentages
sperm	612377	93.30%
cluster	22112	3.37%
small or pinhead	21846	3.33%

Table 1 shows that in our dataset, the "sperm" class predominates at 93.30%, while "cluster" and "small or pinhead" classes are underrepresented at 3.37% and 3.33%. This imbalance highlights the need for our Equalize Class Representation Augmentation method, aimed at balancing the dataset for better model training and enhancing detection accuracy for less frequent classes. Our "Equalize Class Representation Augmentation in Sperm Detection" method

combats class imbalance by augmenting underrepresented classes to equal the dominant class's frequency. It involves cropping regions from original images and randomly pasting them at non-overlapping locations, thereby increasing the presence of rarer classes. Updated annotations ensure dataset integrity, leading to a more balanced class representation and improved accuracy and generalization of the detection model.

It is worth noting that our Equalize Class Representation Augmentation method references [7], which contextualizes our work within the broader field of data augmentation in instance segmentation. [7] also showed that pasting objects randomly is sufficient and can provide solid gains on top of strong baselines.

2.3. Tracking method

To track the detected sperm, we utilize BoT-SORT [8], an extended version of the BYTETracker class for YOLOv8, specifically designed for object tracking incorporating ReID and the GMC algorithm. One advantage of employing this tracking method over the previous one is its capability to capture motion and seamlessly integrate it to enhance the Kalman filter state vector more effectively.

The tracking system is configured with specific parameters, including an initial association threshold of 0.5, a secondary association threshold of 0.1, an initialization threshold for new tracks set at 0.6, a track buffer duration of 30, and a track matching threshold of 0.8. For BoT-SORT, the settings comprise a global motion compensation method named sparseOptFlow, a proximity threshold of 0.5, an appearance threshold of 0.25, and ReID model usage not enabled.

3. Experiment

3.1. Implementation Detail

During both the training and inference stages, our model is resized to 640, following the instructions from YOLO [5]. As for the hyperparameters, a batch size of 64 is employed, and the SGD optimizer with a learning rate of 0.001 is used. Additionally, online augmentation techniques such as flip, rotation, mixup, translation, and mosaic are applied. Our model is obtained after 300 epochs.

3.2. Experimental result

Results of our experiments on different sizes of pre-trained YOLOv8 model for the detection task with the validation set are presented in Table 2. The validation set has 5850 images, with the number of times class "sperm", "cluster", "small or pinhead" appear are 159305, 9606 and 5149, respectively.

Through experimenting with different sizes of pre-trained YOLOv8 models for detection, we found that larger models generalize better on the task, with a noticeable exception for the "cluster" class. YOLOv8x outperformed the "sperm" class, compared to smaller models - with a mAP50 of 0.719 and an mAP50-95 of 0.271 for "sperm" classes. YOLOv8x also performed best for detecting "small or pinhead" sperms with a mAP50 of 0.0919 and a mAP50-95 of 0.0361. However, the smaller the model, the better it can detect "cluster" sperms. Most notably, YOLOv8n had the highest precision and recall rate for cluster sperms - 0.253 and 0.112 respectively. By applying data augmentation, YOLOv8n detected cluster sperms best with a mAP50 of 0.14 and an mAP50-95 of 0.0384. Given the difference in model performance, ensembling is a possible choice.

Table 2
Qualitative results of our methods

Model	Data Augmentation	Class	P	R	mAP50	mAP50-95
YOLOv8n	No	all	0.288	0.254	0.219	0.0717
		sperm	0.558	0.63	0.49	0.166
		cluster	**0.253**	**0.112**	0.0797	0.0185
		small or pinhead	0.0542	0.0202	0.0281	0.0101
YOLOv8n	Yes	all	0.274	0.262	0.236	0.0861
		sperm	0.628	0.637	0.6	0.227
		cluster	0.139	0.0999	**0.14**	**0.0384**
		small or pinhead	0.0546	0.0478	0.0288	0.0129
YOLOv8s	Yes	all	0.23	0.236	0.191	0.0705
		sperm	0.59	0.628	0.52	0.199
		cluster	0.0734	0.038	0.0387	0.00739
		small or pinhead	0.0262	0.0406	0.0139	0.00503
YOLOv8m	Yes	all	0.208	0.182	0.155	0.0533
		sperm	0.522	0.505	0.411	0.141
		cluster	0.00257	0.000104	0.0013	0.00013
		small or pinhead	0.1	0.0404	0.0523	0.019
YOLOv8l	Yes	all	0.234	0.23	0.206	0.0768
		sperm	0.582	0.635	0.558	0.216
		cluster	0.0805	0.0237	0.0412	0.00595
		small or pinhead	0.0398	0.033	0.0206	0.00853
YOLOv8x	Yes	all	**0.317**	**0.272**	**0.28**	**0.108**
		sperm	**0.727**	**0.742**	**0.719**	**0.271**
		cluster	0.0586	0.00271	0.0294	0.0159
		small or pinhead	**0.166**	**0.0701**	**0.0919**	**0.0361**

4. Discussion and Outlook

In conclusion, in this challenge, we introduce a novel framework employing YOLOv8 and advancements in equalization augmentation to tackle issues related to sperm shape and class imbalance as observed in the aforementioned work. The experimental results highlight that our model effectively mitigates weaknesses in detecting small objects, ultimately yielding improved results in the tracking stage. Furthermore, incorporating our offline augmentation methods into the dataset can assist the model in partially addressing issues related to class imbalance. The results from the experiments demonstrate the promise of our method to facilitate further research in sperm detection, contributing to enhanced performance of the tracking pipeline in general.

Acknowledgment

References

[1] 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 (2022).

[2] M. Kosela, J. Aszyk, M. Jarek, J. Klimek, T. Prokop, Tracking of spermatozoa by yolov5 detection and strongsort with osnet tracker (2022).

[3] 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.

[4] T. B. Haugen, S. A. Hicks, J. M. Andersen, O. Witczak, H. L. Hammer, R. Borgli, P. Halvorsen, M. Riegler, Visem: A multimodal video dataset of human spermatozoa, in: MMSys, 2019, pp. 261–266.

[5] G. Jocher, A. Chaurasia, J. Qiu, YOLO by Ultralytics, 2023.

[6] J. Terven, D. Cordova-Esparza, A comprehensive review of yolo: From yolov1 to yolov8 and beyond, arXiv preprint arXiv:2304.00501 (2023).

[7] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, B. Zoph, Simple copy-paste is a strong data augmentation method for instance segmentation, in: CVPR, IEEE, 2021.

[8] N. Aharon, R. Orfaig, B.-Z. Bobrovsky, Bot-sort: Robust associations multi-pedestrian tracking, arXiv preprint arXiv:2206.14651 (2022).