Medico Multimedia Task at MediaEval 2023: Transparent Tracking of Spermatozoa

Vajira **Thambawita**[1,*], Andrea M. **Storås**[1,2], Tuan-Luc **Huynh**[3,4], Thien-Phuc **Tran**[3,4], Hai-Dang **Nguyen**[3,4], Minh-Triet **Tran**[3,4], Trung-Nghia **Le**[3,4], Pål **Halvorsen**[1,2], Michael A. **Riegler**[1,2] and Steven **Hicks**[1]

[1]*SimulaMet, Norway*

[2]*OsloMet, Norway*

[3]*University of Science, VNU-HCM, Vietnam*

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

Abstract

The Medico Multimedia Task returns for its seventh iteration as a segment of MediaEval 2023. The challenge comprises three main tasks: sperm tracking, sperm detection, and sperm motility predictions. Additionally, this year, we have broadened our focus by incorporating new graph data derived from sperm-bounding boxes and unique identifiers taken from manually annotated data. We invite participants to employ innovative methods, diverging from traditional ones, to study sperm using machine learning. The dataset includes video recordings of spermatozoa, complemented with annotations and graph data.

1. Introduction

The 2023 Medico task builds on the previous Medico edition about transparent tracking of spermatozoa in videos [1, 2, 3]. While infertility is increasing on a global basis, optimizing techniques used for treating this medical condition is becoming increasingly more important. A central part of selecting the appropriate treatment is analysis of semen samples through a microscope, which is performed by a medical expert. However, manual examinations are time consuming, and the results are subjective and highly dependent on the experience of the medical expert. Computer-aided systems for automatic analysis have been developed, but they do not work well in clinical settings. Consequently, there is a need for improved methods for identifying, tracking and counting spermatozoa in fresh semen samples.

The goal of the 2023 Medico task is to encourage the participants to track individual spermatozoa in real-time and combine different data sources to predict common measurements used for sperm quality assessment, specifically the motility of the spermatozoa. Solving this task successfully might pave the way for developing improved computer-aided systems to assist medical experts in the fertility clinic.

Annotated videos from the VISEM-Tracking dataset [4] are used in the task. The provided development dataset contains 20 videos which have frame-by-frame bounding box annotations, each being 30 seconds long. In addition, we provide a set of sperm characteristics (hormone levels, fatty acid data, etc.), anonymized study participant data, and motility and morphology data aligned with World Health Organization (WHO) recommendations. Finally, we provide graph data, i.e., data containing nodes and edges that represent the spatial and temporal relationships

between the sperm, extracted from the original VISEM-Tracking dataset. Based on this data, the participants will be asked to solve the following four subtasks, where Subtask 4 is optional:

Subtask 1: The goal of this subtask is localization and tracking of sperm cells in a given semen video. Specifically, the subtask focuses on examining microscopic videos of sperm, where experts have manually annotated spermatozoa. Participants are tasked with detecting individual sperm cells by providing bounding box coordinates and tracking them by assigning unique IDs. The required format for the bounding box coordinates should align with the structure used in the development datasets.

Subtask 2: This subtask requires the participants to further refine the methodologies employed in Subtask 1, emphasizing not only high prediction accuracy but also computational efficiency and inference time. Participants must provide reports on the average Frames Per Second (FPS) and Floating-Point Operations Per Second (FLOPS) while conducting inference with a batch size of 1.

Subtask 3: The goal of this task is to anticipate sperm motility[1] in terms of the percentage of progressive and non-progressive spermatozoa. The prediction needs to be performed patient-wise, resulting in a singular value for each patient pertaining to the predicted attribute. To address this subtask, sperm tracking or bounding boxes obtained from Subtasks 1 and/or 2 are indispensable. Participants are strongly encouraged to consider the temporal dimension of the videos, since temporal information propagated from previous frames are crucial for extrapolating properties in subsequent frames (i.e., sperm motility). This is important due to the fact that an analysis based solely on individual frames will be insufficient to capture the movement or motility of sperm, which contains vital information necessary for accurate predictions.

Subtask 4: This task is experimental in nature and asks the participants to generate graphs representing the predicted tracks for the spermatozoa in order to assess the sperm motility. The participants are asked to employ graph data structures as input to a model that predicts the level of motility in sperm samples. The construction of graph structures can be facilitated using the predicted bounding boxes. Graphs for training models to predict sperm motility are provided, while the graphs required for testing the final models must be generated from the prediction models in Subtasks 1 and/or 2.

2. Dataset Details

As in the 2022 edition [1], the 2023 Medico task uses the VISEM-Tracking dataset [4]. VISEM-Tracking is based on the VISEM dataset [5], where males aged 18 years and older were examined with respect to fertility. The participants provided written informed consent to participate in the study. The project was approved by the Norwegian data authority and the Regional Medical Ethics Committee of South-East Norway (REK). For this task, we include a development set consisting of 20 videos from VISEM-Tracking. Each video is 30 seconds long and contains detailed frame-by-frame annotations of individual spermatozoa using bounding boxes. Five additional videos without annotations are provided for testing.

For each patient, we include a video of live sperm (video and extracted frames), manually annotated bounding box details for each spermatozoon (sample frames are presented in Figure 1), a set of measurements from a standard semen analysis for the whole sample, a sperm fatty acid profile, the fatty acid composition of serum phospholipids, study participants-related data, and WHO analysis data. The bounding box coordinates are provided in two separate folders: one

---

[1]Motility is the ability to move independently, where a progressive spermatozoon is able to "move forward" and a non-progressive would move for example in circles without any forward progression.

folder has bounding box coordinates in YOLO format [6] and the other folder contains feature identifiers in addition to the bounding box coordinates. These feature identifiers can be used to identify the same bounding box in different frames in a video. Each video has a resolution of 640×480 pixels and runs at 50 frames per second (FPS). The dataset contains six CSV files in total. One row in each CSV file represents a participant. The six CSV files are:

- semen_analysis_data: the results of standard semen analysis.
- fatty_acids_spermatozoa: the levels of several fatty acids in the sperm of the participants.
- fatty_acids_serum: the serum levels of the fatty acids of the phospholipids (measured from the blood of the participant).
- sex_hormones: the serum levels of sex hormones measured in the blood of the participants.
- study_participant_related_data: general information about the participants such as age, abstinence time, and Body Mass Index (BMI).
- videos: overview of which video file belongs to what participant.

Regarding Subtask 4, which is optional, graph data[2] is provided. The structure of the graphs is depicted in Figure 2. The graphs represent spatial and temporal relationships between sperm in a video. Spatial edges (e_s) connect sperms within the same frame, while temporal edges (e_t) connect sperms across different frames. The graphs have been generated with varying spatial threshold (Tr) values, where each threshold value determines the maximum distance between two nodes for them to be connected in the graph. The following threshold values are used: 0.1, 0.2, 0.3, 0.4, and 0.5. The graph data contains separate folders with graphs generated using each spatial threshold, i.e., there are five folders in total. In each threshold folder, there is a subfolder including separate graphs for individual frames in the video and a GraphML file containing the graph for the complete video.

3. Evaluation

In order to evaluate the proposed solutions thoroughly, several evaluation metrics are computed for each subtask.

Subtask 1: We use the widely recognized COCO evaluation metrics to evaluate the performance of sperm detection methods. This comprehensive assessment involves precision, recall, mAP@50 (mean average precision @50), and mAP@50-95, providing a multidimensional perspective on the accuracy of the methods employed in this critical domain. Furthermore, we leverage Jonathan Luiten's TrackEval library [7], encompassing crucial metrics such as the Higher Order Tracking Accuracy (HOTA) [8] and a spectrum of other multi-object tracking (MOT) evaluation criteria, to offer a profound analysis of tracking performance. This addition broadens the scope of evaluation.

Subtask 2: We utilize the aforementioned evaluation metrics as in Subtask 1. However, the performance evaluation considers the inference speed of the methods as a weighted factor. This integration adds a new dimension to the evaluation, weighing the performance against the efficiency and agility of inference, thus illuminating the intricate interplay between accuracy and speed in sperm detection and tracking.

Subtask 3 and **Subtask 4:** Evaluation of regression sperm motility prediction performance involves the use of Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). These metrics provide insights into the accuracy of the predictions.
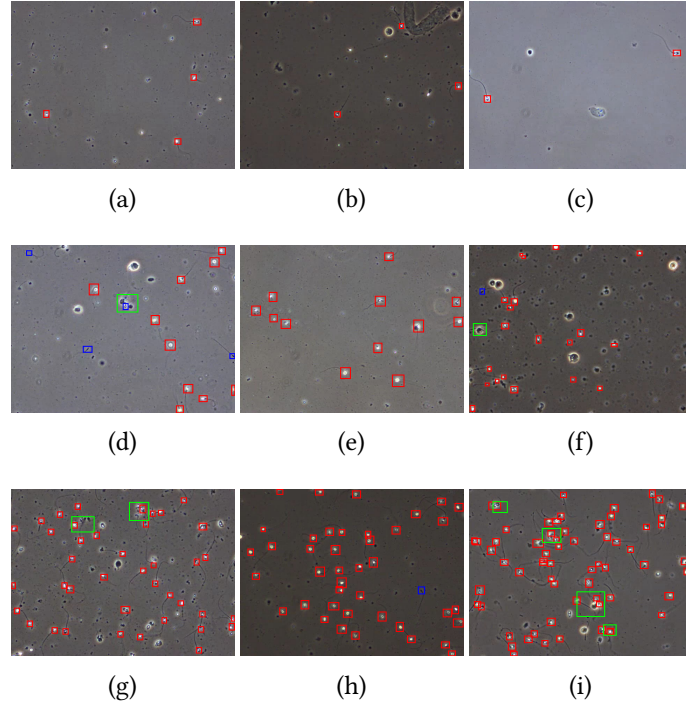
---

[2]https://huggingface.co/datasets/SimulaMet-HOST/visem-tracking-graphs

Figure 1: Sample frames from data with corresponding manually annotated bounding boxes. Colors represent different classes - red: sperm, green: sperm cluster, and blue: small or pinhead sperm. (a, b, c), (d, e, f), and (g, h, i) are sample frames from low, moderate, and high concentrate sperm samples, respectively.
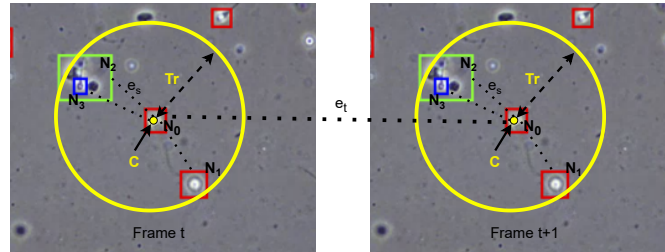

Figure 2: Graph structure. Tr: Threshold value used to find neighbors. N_n: Graph nodes with different class labels (sperm, cluster, or pinhead). C: center of objects to calculate Tr values which is euclidean distance. e_s: spatial edges which connect objects in the same frame. e_t: temporal edges that connect the same objects in the next frame.

4. Discussion and Outlook

This year, we introduced supplementary graph data derived from the details of bounding boxes and feature IDs. These IDs facilitate the identification of the same sperm cell across multiple frames of a video. We believe that participants can explore innovative approaches that diverge from conventional methods in the realm of sperm analysis through machine learning. Alongside experimental Subtask 4, we retained the foundational tasks associated with sperm tracking and the prediction of sperm motility, denoted as Subtasks 1, 2, and 3.

References

[1] V. Thambawita, S. A. Hicks, A. M. Storås, J. M. Andersen, O. Witczak, T. B. Haugen, H. Hammer, T. Nguyen, P. Halvorsen, M. A. Riegler, Medico Multimedia Task at MediaEval 2022: Transparent Tracking of Spermatozoa, in: Proceedings of MediaEval 2022 CEUR Workshop, 2022. URL: https://2022.multimediaeval.com/paper5501.pdf.

[2] 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: Proceedings of MediaEval 2022 CEUR Workshop, 2023. URL: https://2022.multimediaeval.com/paper6101.pdf.

[3] M. Kosela, J. Aszyk, M. Jarek, J. Klimek, T. Prokop, Tracking of Spermatozoa by YOLOv5 Detection and StrongSORT with OSNet Tracker, in: Proceedings of MediaEval 2022 CEUR Workshop, 2023. URL: https://2022.multimediaeval.com/paper7367.pdf.

[4] 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.

[5] 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: Proceedings of the 10th ACM Multimedia Systems Conference, MMSys '19, Association for Computing Machinery, 2019, p. 261–266. doi:10.1145/3304109.3325814.

[6] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. doi:10.1109/CVPR.2016.91.

[7] A. H. Jonathon Luiten, TrackEval, https://github.com/JonathonLuiten/TrackEval, 2020.

[8] 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 (2021) 548–578. doi:10.1007/s11263-020-01375-2.