

**Abstract citation ID:** deac107.104

**P-108 Real-time deep learning based multi object tracking of spermatozoa in fresh samples**

**V. Thambawita<sup>1</sup>, S.A. Hicks<sup>1</sup>, A. Storås<sup>1</sup>, O. Witczak<sup>2</sup>, J.M. Andersen<sup>2</sup>, H.L. Hammer<sup>3</sup>, P. Halvorsen<sup>1</sup>, M.A. Riegler<sup>1</sup>, T.B. Haugen<sup>2</sup>**

<sup>1</sup>Simula Metropolitan Center for Digital Engineering, Department of Holistic Systems, Oslo, Norway

<sup>2</sup>Faculty of Health Sciences- OsloMet – Oslo Metropolitan University, Department of Life Sciences and Health, Oslo, Norway

<sup>3</sup>Faculty of Technology- Art and Design- OsloMet – Oslo Metropolitan University, Department of Computer Science, Oslo, Norway

**Study question:** Can real-time deep learning model track hundreds of spermatozoa simultaneously?

**Summary answer:** The state-of-the-art deep learning detection model YOLOv5 shows possibilities of multi-sperm tracking with high sensitivity and precision.

**What is known already:** Computer-aided sperm analysis (CASA) systems can be used for the evaluation of sperm motility by applying tracking algorithms. However, the presence of cellular debris and/or cell aggregations in human semen samples makes the tracking of spermatozoa difficult, resulting in unreliable motility assessment. Thus, there is a need for an improved methodology. There are several studies on detecting spermatozoa in real-time, such as DeepSperm, which is trained and tested on bull semen samples. However, differences in characteristics between human and animal spermatozoa imply the requirement of a multi-sperm tracking system adapted to human spermatozoa.

**Study design, size, duration:** We used the open-access VISEM dataset consisting of video recordings of human semen samples from 85 participants. We selected three videos with low sperm counts to perform manual annotations (bounding boxes around sperms) and create a training dataset. Only the first 30 seconds of each video were extracted for annotations. Then, all the spermatozoa of the three videos were annotated with bounding boxes using LabelBox. The annotations will be made public in a future study.

**Participants/materials, setting, methods:** We used two different object detection models, YOLOv5 Nano and YOLOv5 XLarge to detect spermatozoa and performed transfer learning without layer freezing. Each model was trained for a maximum of 300 epochs, stopping if the validation loss did not improve for the previous 100 epochs. Leave-one-out cross-validation was performed to obtain the presented results. Precision, sensitivity, and mean average precision (mAP) are the quantitative metrics measured to evaluate our detection models.

**Main results and the role of chance:** Our models are able to detect a large number of human spermatozoa simultaneously in a given video in real-time. The model YOLOv5 Nano took around 45 min and the YOLOv5

XLarge model took around 270 min to train on one fold. In terms of training and detection, the Nano model is faster because it has a lower number of trainable parameters compared to the XLarge model, but at the expense of precision. YOLOv5 Nano has capability of predicting 200 frames per second (fps), and YOLOv5 XLarge can predict 56 fps. Both models show high prediction rates which are greater than the real-time prediction 25 fps threshold. Performance-wise, the Nano model shows an average recall value of 0.8545 which is better than the average recall value of 0.7632 of XLarge model. In contrast, the XLarge model shows a higher precision value of 0.8020 compared to the low precision of 0.7149 of the Nano model. Furthermore, the XLarge model has an average mAP (mAP<sub>0.5</sub>) of 0.7632 which is larger than the mAP value of the Nano model of 0.7027. The current prediction errors in this stage might be a result of the small training dataset.

**Limitations, reasons for caution:** The low amount of data makes validating the models difficult, especially evaluating how well they generalize to unseen samples. Furthermore, the training data consisted of samples containing low sperm counts, making the performance on samples with high sperm concentration uncertain.

**Wider implications of the findings:** Sperm tracking is integral to achieving accurate and less subjective motility assessment. The detections can be used to analyse the individual spermatozoa, leading to better performance. Deep feature extraction, trajectory prediction, and trajectory extraction can be used for future studies like generating synthetic spermatozoa for training generalizable machine learning models.

**Trial registration number:** not applicable