

MDL-DA Track: Self-supervised Sperm Motility Direction Learning and Density-Adaptive Association for Sperm Tracking

Xiaoyu Zeng¹, Yang Zhao^{1*}, Xu Wang¹, Longlong Fu², Jiahui Wu³, Jihong Pei⁴, and Xuan Yang¹

¹ College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China 2410104004@email.szu.edu.cn, {zhaoyang1990, wangxu, yangxuan}@szu.edu.cn

² NHC Key Laboratory of Frontiers and Technologies in Reproductive Health, China fullpumc@163.com

³ College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China wujiahui@szu.edu.cn

⁴ College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China jhpei@szu.edu.cn

Abstract. Existing multi-object tracking methods typically rely on motion models or appearance features for target association. However, in sperm multi-target tracking tasks, the appearance of the targets is highly similar and the motion patterns are complex. It is difficult for both appearance and motion features to exhibit stable tracking performance in targets with different distribution densities simultaneously. Additionally, it is difficult to obtain the motility direction label of sperm targets. This paper proposes Self-supervised Sperm Motility Direction Learning and Density-Adaptive Association for Sperm Tracking (MDL-DA Track). In the sperm detection module, a multi-task learning framework combining self-supervised Sperm Motility Direction Learning (SMDL) and object detection is constructed. Through self-supervised learning of temporal bounding boxes, motion direction information is captured. YOLOX is jointly employed to model both sperm target locations and motion directions. In the sperm tracking module, a motion direction-enhanced Density-adaptive Object Association (DOA) strategy is proposed to handle differences between sparse and dense target distributions. The DOA strategy adaptively selects association rules according to local target density. For targets with sparse distributions, matching is dominated by motion information. For targets with dense distributions, association is achieved by combining motion, appearance, and direction features, thereby enhancing tracking robustness in complex scenarios. Experimental results on the public VISEM-Tracking dataset and Electron Microscopy Sperm dataset show that MDL-DA Track outperforms several state-of-the-art methods in multi-object tracking. On the VISEM-Tracking dataset, higher order tracking accuracy and multiple object tracking accuracy are improved by 2.5%-3%. Ablation studies further

* Corresponding author

demonstrate that SMDL and DOA effectively reduce erroneous associations in dense scenarios and enhance robustness in sparse scenarios.

Keywords: Multi-object Tracking · Motion Direction Learning · Density-aware Association

1 Introduction

Multi-object tracking (MOT) is a core problem in computer vision with broad applications in medical image analysis [1]. In particular, computer-aided sperm analysis (CASA) systems [2,3] rely on accurate detection and tracking to extract clinically relevant motion parameters, such as velocity, linearity, and amplitude, providing objective reproductive health indicators.

Conventional MOT research has largely focused on large, visually distinctive targets, as reflected in datasets such as MOT16 [4], MOT20 [5], and DanceTrack [6]. These datasets feature prominent appearance cues and moderately complex motion. In contrast, sperm tracking presents an extreme case of *homogeneous small-object tracking under significant density variation*. Targets are small, visually similar, and exhibit complex, non-linear motion with frequent crowding, posing unique challenges for reliable association.

Traditional MOT methods, such as Simple Online and Realtime Tracking (SORT) [7] and DeepSORT [8], rely on motion models or appearance embeddings for frame-to-frame association. While effective for large distinctive targets, they struggle in sperm tracking, where appearance cues are weak and illumination-sensitive in sparse regions, and motion-based matching becomes unreliable in dense regions due to occlusions and frequent interactions. Observation-Centric SORT (OCSort) [10] incorporates motion direction derived from frame-to-frame displacement to improve association, but such cues are noise-sensitive in crowded scenarios and fail to capture intrinsic geometric orientation of sperm. Parallel methods such as TOPICTrack [11] adopt a two-stage parallel matching strategy, independently computing motion and appearance associations and resolving conflicts via trajectory-level feature selection; however, they still do not explicitly model local density or sperm-specific geometric orientation, which can provide complementary cues in both sparse and dense settings.

To address these challenges, we propose MDL-DA Track, a direction-enhanced, density-adaptive framework for sperm multi-object tracking. The approach consists of two key components: (1) the Sperm Motility Direction Learning (SMDL) module in detection, and (2) the Density-Adaptive Object Association (DOA) strategy in tracking. Specifically, in detection, the Sperm Motility Direction Learning (SMDL) module builds a self-supervised orientation prior per target, using pseudo-labels from short-term motion to supervise a lightweight direction branch within YOLOX [14], enabling the detector to learn motion direction information from image features. Unlike OCSort, this models geometric orientation independently of displacement, providing a stable cue under dense crowding or rapid movement. In tracking, we propose the Density-Adaptive Object Association (DOA) strategy. The strategy estimates local density and adjusts feature

dominance: motion-dominant matching in sparse regions, while dense regions integrate motion, appearance, and learned direction features with adaptive weighting, reducing identity switches from overlapping visually similar targets.

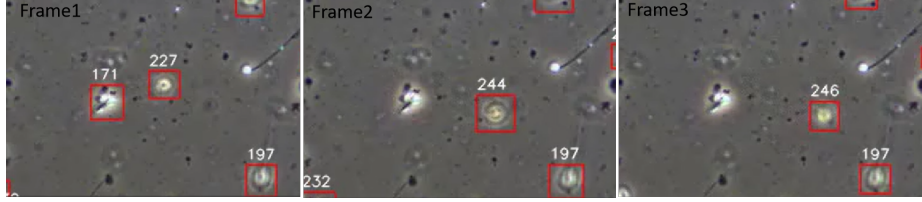


Fig. 1. Identity switch caused by appearance variation in sparse regions.

Overall, MDL-DA Track explicitly models motion direction and adapts association strategies to local density, improving tracking stability in sparse and crowded scenarios. By integrating self-supervised SMDL in detection and density-adaptive DOA in tracking, it effectively reduces identity switches and association errors.

2 Methodology

In sperm multi-object tracking, targets exhibit highly homogeneous appearance, small size, and frequent local crowding, leading to severe identity switches under traditional motion- or appearance-dominated association strategies. To address these challenges, we propose a two-stage framework (Fig. 2) that integrates displacement-supervised Sperm Motility Direction Learning (SMDL) with Density-Adaptive Object Association (DOA).

SMDL learns geometry-aware direction representations based on the biological assumption that sperm headtail orientation aligns with short-term motion direction, and embeds them into detection features. while DOA dynamically adjusts association cues according to local density and trajectory reliability, ensuring robust matching in both sparse and dense microscopic regions.

At frame t , the detector produces detections $\mathcal{D}_t = \{d_i^t\}$, where each detection contains center $\mathbf{x}_i^t$, bounding box (w_i^t, h_i^t) , appearance embedding $\mathbf{f}_i^t$, and predicted direction $\hat{\mathbf{d}}_i^t$.

2.1 Multi-Task Learning Framework for Self-Supervised Sperm Motility Direction Prediction and Target Detection

Biologically, a sperm’s short-term motion direction aligns with its head–tail orientation. We explicitly model geometric orientation as a learnable representation independent of displacement magnitude.

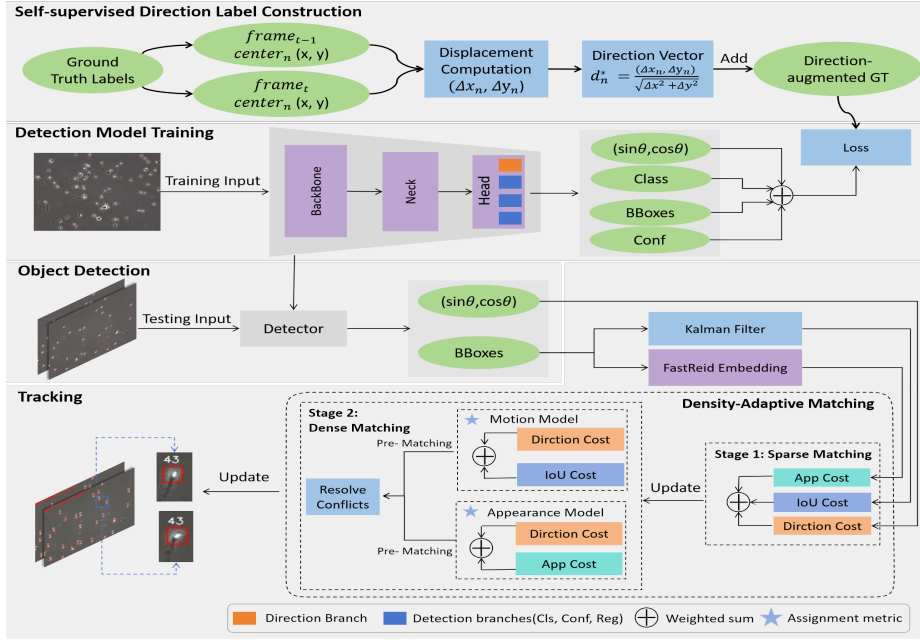


Fig. 2. Overview of MDL-DA Track. Detection uses SMDL to predict bounding boxes and motion directions, while DOA adaptively associates targets based on local density.

Self-Supervised Sperm Motility Direction Learning (SMDL) Sperm heads exhibit diverse shapes, making single-frame orientation estimation unreliable. Biologically, head orientation aligns with short-term motion. We obtain direction supervision by computing the temporal displacement of the same sperm objects between adjacent frames in the dataset. This formulation does not require explicit orientation annotations.

For detection d_i^t with center $\mathbf{x}_i^t$, the target direction is:

$$\mathbf{d}_i^{*t} = \frac{\mathbf{x}_i^t - \mathbf{x}_i^{t-1}}{\|\mathbf{x}_i^t - \mathbf{x}_i^{t-1}\|_2}. \quad (1)$$

Direction supervision is applied only for positive samples with significant displacement, following the same assignment as box regression.

The training objective is:

$$L = L_{\text{YOLOX}} + \lambda_{\text{dir}} \frac{1}{N_{\text{pos}}} \sum_i \left(1 - \langle \hat{\mathbf{d}}_i^t, \mathbf{d}_i^{*t} \rangle \right), \quad (2)$$

where N_{pos} counts valid positive samples, and λ_{dir} balances detection and direction learning.

Compared with raw displacement, SMDL provides geometry-aware orientation cues that remain stable under dense interactions and abrupt motion, improving association reliability.

Multi-Task Learning Framework The original YOLOX [14] detection head consists of classification, bounding box regression, and objectness branches. To incorporate geometric orientation cues without altering the standard detection pipeline, we introduce a lightweight direction branch as an auxiliary task. The branch shares backbone and neck features with the existing detection heads and is implemented using two 3×3 convolutions followed by a 1×1 convolution, predicting two unnormalized components $(t_{\sin,i}^t, t_{\cos,i}^t)$ for each detection.

To avoid angular discontinuity at 0° and 360° , direction is represented in vector form $(\sin \theta, \cos \theta)$ rather than direct angle regression. The predicted vector is L2-normalized as $\hat{\mathbf{d}}_i^t = (t_{\sin,i}^t, t_{\cos,i}^t) / \|(t_{\sin,i}^t, t_{\cos,i}^t)\|_2$, ensuring smooth angular representation on the unit circle.

Direction supervision is applied only to positive samples and follows the same assignment strategy as box regression, ensuring spatial consistency between detection and orientation prediction.

2.2 Motility direction-enhanced Density adaptive Object Association strategy

Estimation of Local Density of Sperm Targets To adapt association strategies to local crowding, we classify each track as sparse or dense based on its spatial neighborhood. For track k with predicted center $\hat{\mathbf{x}}_k^t$ and box size $(\hat{w}_k, \hat{h}_k)$, we define an adaptive neighborhood radius $R_k = \max(R_{\min}, \gamma \sqrt{\hat{w}_k \hat{h}_k})$, where γ is a scale factor relative to target size and $R_{\min}$ is a lower bound. The local density of track k at frame t is the number of detections within R_k , $N_k^t = \sum_j 1(\|\mathbf{x}_j^t - \hat{\mathbf{x}}_k^t\|_2 \leq R_k)$, where j indexes all detections in the current frame. If $N_k^t \leq \tau_s$, the track is treated as *sparse* and processed by sparse matching; otherwise, it is *dense* and handled by the dense association stage. The threshold τ_s is fixed for all experiments.

Two-Stage Sperm Target Association Each trajectory maintains its predicted location, direction features, and appearance features, which are used in the subsequent sparse and dense matching stages within the DOA framework.

Stage 1: Sparse-Region Sperm Matching For sparse trajectories, the association cost is

$$C_{kj}^{(1)} = w_{\text{IoU}}(1 - \text{IoU}_{kj}) + w_{\text{emb}} \frac{1 - \cos(\mathbf{f}_k^t, \mathbf{f}_j^t)}{2} + w_o \frac{1 - \langle \mathbf{d}_k^t, \hat{\mathbf{d}}_j^t \rangle}{2}, \quad (3)$$

where $\mathbf{f}_k^t$ and $\mathbf{d}_k^t$ denote the appearance feature and direction feature of track k , respectively, and $\hat{\mathbf{d}}_j^t$ is the predicted direction of detection j .

Stage 2: Dense-Region Sperm Matching For dense or unmatched trajectories, we use the parallel motion–appearance framework with α -gate conflict resolution [11]. In the *appearance branch*, the cost is $C_{kj}^{(2)} = (1 - w_o)(1 - \cos(\mathbf{f}_k^t, \mathbf{f}_j^t)) +$

$w_o(1 - \langle \mathbf{d}_k^t, \hat{\mathbf{d}}_j^t \rangle)$. In the *motion branch*, the matching score is $S_{kj}^{\text{motion}} = \text{IoU}_{kj} + w_{\text{dir}}D_{kj} + w_{\text{emb}}\cos(\mathbf{f}_k^t, \mathbf{f}_j^t)$, where D_{kj} denotes the direction-consistency score between track k and detection j . Both branches explicitly incorporate the learned direction, improving association under occlusion and crowding. Compared with the sparse stage, motion is weighted less while the direction term remains, and branch conflicts are resolved by the α -gate [11].

3 Experiments

3.1 Datasets and Splits

VISEM-Tracking dataset Experiments are conducted on the VISEM-Tracking dataset [22], containing 20 microscopic videos (640×480 , 45–50 FPS) and 29,196 frames. Only the Sperm category is used. Four sequences (IDs 12, 13, 21, 36) exhibit high sperm density and dynamic motion (rapid movement, frequent direction changes, and interactions), while the others mostly contain sparsely distributed or low-activity sperm. To ensure challenging test conditions and avoid the majority of test sequences being static or sparsely populated, we define three splits. Two mixed-density splits are constructed by randomly sampling three dense sequences and one standard sequence as the test set, while the remaining one dense sequence and the other 15 standard sequences form the training set. Additionally, we include one dense-only split where all four dense sequences are used as the test set to evaluate performance under fully dense conditions.

Electron Microscopy Sperm dataset The electron microscopy dataset [23] consists of 16 microscopic videos (1024×640 , 400 frames) containing seven object categories; only sperm are used for tracking. Three random 10:6 train:val splits are generated, attempting to cover sperm densities and motion behaviors as comprehensively as possible.

3.2 Implementation Details

YOLOX [14] pretrained on COCO [16] serves as the base detector. For VISEM-Tracking, models are trained on 640×480 frames for 40 epochs, with a direction regression branch added for sperm orientation learning (loss weight 0.01). For the Electron Microscopy Sperm dataset [23], the VISEM-Tracking model initializes training and is fine-tuned on 1024×640 frames for 60 epochs with standard augmentations. Tracking follows TOPICTrack [11], using motion prediction and appearance embeddings; FastReID [17] and other hyperparameters follow TOPICTracks default implementation. All baselines use the same YOLOX pretrained on COCO, without the direction branch. Detection confidence is 0.05, NMS threshold 0.4, and $\alpha_{\text{gate}} = 0.9$. Results are reported as mean $\pm$ standard deviation across splits. Experiments run on a single NVIDIA RTX 4090 GPU (24GB) with PyTorch 1.8 and CUDA 11.1, consuming 21GB GPU memory. The code will be released at: https://github.com/hjhxy/MICCAI_MDL-DATrack.

Table 1. Multi-object tracking results on the VISEM-Tracking (VIS) and the electron microscopy sperm tracking dataset [23] (EM-Sperm). $\uparrow/\downarrow$ indicate higher/lower is better. All methods use a YOLOX backbone pretrained on COCO. Results are reported as mean $\pm$ standard deviation across splits. Bold denotes the best mean performance.

Dataset	Method	HOTA $\uparrow$	MOTA $\uparrow$	IDF1 $\uparrow$	AssA $\uparrow$	IDs $\downarrow$
VIS	DeepSort [9]	45.87 $\pm$ 1.56	41.95 $\pm$ 1.20	57.82 $\pm$ 1.29	44.96 $\pm$ 1.55	2267 $\pm$ 187
	OCSort [10]	52.25 $\pm$ 1.09	61.93 $\pm$ 1.05	69.82 $\pm$ 0.96	52.65 $\pm$ 1.15	302 $\pm$ 58
	ByteTrack [12]	52.43 $\pm$ 1.75	62.31 $\pm$ 1.09	69.44 $\pm$ 1.00	52.23 $\pm$ 1.23	328 $\pm$ 35
	DeepOCSort [13]	53.16 $\pm$ 1.23	62.81 $\pm$ 1.38	69.98 $\pm$ 1.00	53.12 $\pm$ 1.29	354 $\pm$ 25
	BoT-SORT [15]	51.35 $\pm$ 1.40	54.56 $\pm$ 1.19	66.26 $\pm$ 1.24	51.11 $\pm$ 1.29	334 $\pm$ 28
	TOPICTrack [11]	52.95 $\pm$ 1.53	62.81 $\pm$ 1.01	70.15 $\pm$ 1.03	52.95 $\pm$ 1.07	235 $\pm$ 31
	MOTR [18]	47.13 $\pm$ 1.80	41.88 $\pm$ 1.29	56.24 $\pm$ 1.36	45.48 $\pm$ 1.57	493 $\pm$ 71
	TrackFormer [20]	47.30 $\pm$ 1.62	56.25 $\pm$ 1.20	58.93 $\pm$ 1.31	43.36 $\pm$ 1.38	656 $\pm$ 122
	TransTrack [19]	47.06 $\pm$ 1.63	48.51 $\pm$ 1.17	59.46 $\pm$ 2.28	45.13 $\pm$ 1.31	512 $\pm$ 67
	MeMOTR [21]	51.18 $\pm$ 1.75	44.52 $\pm$ 1.27	64.18 $\pm$ 1.32	54.11 $\pm$ 1.62	169 $\pm$ 16
	MDL-DA Track	55.65 $\pm$ 1.41	65.32 $\pm$ 1.15	74.36 $\pm$ 0.99	57.10 $\pm$ 1.08	202 $\pm$ 18
EM-Sperm	DeepSort	69.72 $\pm$ 0.40	79.56 $\pm$ 0.38	89.58 $\pm$ 0.42	74.22 $\pm$ 0.35	22 $\pm$ 2
	ByteTrack	71.32 $\pm$ 0.25	82.51 $\pm$ 0.33	91.39 $\pm$ 0.30	75.86 $\pm$ 0.28	3 $\pm$ 1
	OCSort	71.05 $\pm$ 0.34	80.41 $\pm$ 0.22	90.78 $\pm$ 0.31	75.96 $\pm$ 0.29	3 $\pm$ 1
	DeepOCSort	71.21 $\pm$ 0.32	80.42 $\pm$ 0.30	90.65 $\pm$ 0.28	75.89 $\pm$ 0.28	4 $\pm$ 1
	BoT-SORT	72.58 $\pm$ 0.36	83.51 $\pm$ 0.34	90.92 $\pm$ 0.30	76.55 $\pm$ 0.31	15 $\pm$ 2
	TOPICTrack	70.94 $\pm$ 0.12	81.22 $\pm$ 0.31	90.88 $\pm$ 0.29	75.92 $\pm$ 0.28	5 $\pm$ 1
	MDL-DA Track	72.93 $\pm$ 0.30	83.95 $\pm$ 0.28	91.98 $\pm$ 0.25	76.88 $\pm$ 0.27	2 $\pm$ 1

3.3 Results

Quantitative Comparison We compare MDL-DA Track with several representative tracking methods using a shared YOLOX backbone pretrained on COCO. Our method introduces an additional direction regression branch for multi-task learning, while baselines follow the standard detection pipeline. On VISEM-Tracking, MDL-DA Track achieves superior performance across HOTA, MOTA, IDF1, and AssA. Improvements in association metrics reflect enhanced identity consistency, while reductions in false negatives and fragmentation indicate better trajectory continuity. Although ID switches are not the lowest, they remain competitive. On the Electron Microscopy Sperm dataset [23], MDL-DA Track achieves the best overall performance, demonstrating robust identity preservation under dense interactions, which can facilitate more reliable quantification of sperm motility for applications such as Computer-Assisted Sperm Analysis (CASA).

Ablation Study Table 2 quantifies each module’s contribution. Adding DOA (no direction) improves AssA (+0.86) and IDF1 (+0.79) with negligible change in detection metrics, showing adaptive association alone aids identity preservation. Enabling SMDL (direction branch) boosts DetA (+1.39), DetRe (+1.39), and MOTA (+2.77), indicating improved geometric consistency and detection

Table 2. Ablation study on the VISEM-Tracking test set. Modules are incrementally added to the baseline. “+SMDL (direction only)”: direction branch in detection only; “SMDL”: direction used in both detection and tracking. $\uparrow$ indicates higher is better.

Method	HOTA $\uparrow$	DetA $\uparrow$	AssA $\uparrow$	DetRe $\uparrow$	AssRe $\uparrow$	MOTA $\uparrow$	IDF1 $\uparrow$
Baseline	52.65	53.42	52.65	60.93	58.88	62.08	69.75
+DOA (no direction)	53.05	53.35	53.51	60.85	59.38	62.04	70.54
+SMDL (direction only)	54.40	54.81	54.96	62.32	61.35	64.85	71.87
+SMDL	54.82	54.86	55.62	62.36	61.71	64.86	73.05
Full Model	55.33	54.86	56.69	62.35	62.57	64.85	73.76

Table 3. Sensitivity of the direction loss weight in SMDL on the VISEM-Tracking test set. $\uparrow$ indicates higher is better.

Weight	HOTA $\uparrow$	DetA $\uparrow$	AssA $\uparrow$	DetRe $\uparrow$	AssRe $\uparrow$	LocA $\uparrow$	MOTA $\uparrow$	IDF1 $\uparrow$
No branch	52.65	53.42	52.65	60.93	58.88	73.84	62.08	69.75
0 (with branch)	53.82	53.96	54.45	61.19	60.29	73.99	63.06	71.38
0.01	55.33	54.86	56.69	62.35	62.57	74.19	64.85	73.76
0.1	53.99	54.30	54.49	61.60	60.57	74.10	64.15	72.19

robustness. The full model (SMDL + DOA) further increases HOTA (+0.51), AssA (+1.07), and IDF1 (+0.71) compared to individual modules.

Hyperparameter Sensitivity Analysis We analyze the sensitivity of the direction loss weight λ_{dir} in SMDL on the VISEM-Tracking dataset with sparse association enabled. We compare four settings: three direction weights $\{0, 0.01, 0.1\}$ with the direction branch, and a variant without the direction branch (Table 3).

With $\text{dir} = 0$, retaining the direction branch shows a slight improvement over removing it, which may be attributed to multiple factors, including the interaction with the sparse association mechanism, the implicit regularization effect of the auxiliary supervision, and the stochasticity introduced during training. A small positive weight (0.01) further improves performance and achieves the best HOTA, AssA, and IDF1, suggesting that moderate directional cues are beneficial for identity association. In contrast, larger weights degrade performance, likely due to overemphasis on direction signals that conflicts with motion and appearance cues.

Overall, performance remains relatively stable within 0–0.1, with fluctuations within $\pm 1.5\%$, indicating that SMDL is robust to hyperparameter selection.

4 Conclusion

We present MDL-DA Track, a sperm multi-object tracking framework that integrates self-supervised motion direction learning (SMDL) with density-adaptive object association (DOA). By explicitly modeling sperm head-tail orientation

and adapting association strategies to local target density, our method improves identity consistency and trajectory continuity. Experiments on the VISEM-Tracking dataset and the Electron Microscopy Sperm dataset demonstrate consistent improvements over existing methods, with HOTA and MOTA improving by 2.5%–3%. Future work will focus on more robust pseudo-label generation under challenging imaging conditions. More reliable direction supervision is expected to enhance the learning of discriminative sperm motion patterns. We will further explore integrating directional, geometric, and appearance cues to improve robustness.

Acknowledgments. This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation (2024A1515010977, 2025A1515060013), Scientific Foundation for Youth Scholars of Shenzhen University, Guangdong Provincial Key Laboratory (2023B1212060076, 2017B030314073).

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

References

1. F. J. Baldán, D. García-Gil, and C. Fernandez-Basso: Revolutionizing Sperm Analysis with AI: A Review of Computer-Aided Sperm Analysis Systems. *Computation* **13**(6), 132 (2025)
2. R. P. Amann and D. Waberski: Computer-Assisted Sperm Analysis (CASA): Capabilities and Potential Developments. *Theriogenology* **81**(1), 5–17 (2014)
3. W. Zhao, P. Ma, C. Li, X. Bu, S. Zou, T. Jiang, and M. Grzegorzczek: A Survey of Semen Quality Evaluation in Microscopic Videos Using Computer Assisted Sperm Analysis. arXiv preprint arXiv:2202.07820 [cs.CV] (2022)
4. A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler: MOT16: A Benchmark for Multi-Object Tracking. arXiv preprint arXiv:1603.00831 [cs.CV] (2016)
5. P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé: MOT20: A Benchmark for Multi-Object Tracking in Crowded Scenes. arXiv preprint arXiv:2003.09003 [cs.CV] (2020)
6. P. Sun, J. Cao, Y. Jiang, Z. Yuan, S. Bai, K. Kitani, and P. Luo: DanceTrack: Multi-Object Tracking in Uniform Appearance and Diverse Motion. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20993–21002. IEEE, Los Alamitos (2022)
7. A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft: Simple Online and Realtime Tracking. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3464–3468. IEEE, Los Alamitos (2016). <https://doi.org/10.1109/ICIP.2016.7533003>
8. N. Wojke and A. Bewley: Deep Cosine Metric Learning for Person Re-Identification. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 748–756. IEEE, Los Alamitos (2018). <https://doi.org/10.1109/WACV.2018.00087>
9. N. Wojke, A. Bewley, and D. Paulus: Simple Online and Realtime Tracking with a Deep Association Metric. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3645–3649. IEEE, Los Alamitos (2017). <https://doi.org/10.1109/ICIP.2017.8296962>

10. J. Cao, X. Weng, R. Khirodkar, J. Pang, and K. Kitani: Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. arXiv preprint arXiv:2203.14360 [cs.CV] (2022)
11. X. Cao, Y. Zheng, Y. Yao, H. Qin, X. Cao, and S. Guo: TOPIC: A Parallel Association Paradigm for Multi-Object Tracking Under Complex Motions and Diverse Scenes. *IEEE Transactions on Image Processing* **34**, 743–758 (2025). <https://doi.org/10.1109/TIP.2025.3526066>
12. Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang: ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In: European Conference on Computer Vision (ECCV), Springer (2022)
13. G. Maggolino, A. Adnan, J. Cao, and K. Kitani: Deep OC-SORT: Multi-Pedestrian Tracking by Adaptive Re-Identification. arXiv preprint arXiv:2302.11813 [cs.CV] (2023)
14. Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun: YOLOX: Exceeding YOLO Series in 2021. arXiv preprint arXiv:2107.08430 [cs.CV] (2021)
15. N. Aharon, R. Orfaig, and B.-Z. Bobrovsky: BoT-SORT: Robust Associations for Multi-Pedestrian Tracking. arXiv preprint arXiv:2206.14651 [cs.CV] (2022)
16. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick: Microsoft COCO: Common Objects in Context. In: European Conference on Computer Vision (ECCV), LNCS, Springer (2014)
17. L. He, X. Liao, W. Liu, X. Liu, P. Cheng, and T. Mei: FastReID: A PyTorch Toolbox for General Instance Re-Identification. In: Proceedings of the 31st ACM International Conference on Multimedia, pp. 9664–9667. ACM, New York (2023)
18. F. Zeng, B. Dong, Y. Zhang, T. Wang, X. Zhang, Y. Wei: MOTR: End-to-End Multiple-Object Tracking with Transformer. In: ECCV 2022, Springer, Heidelberg (2022)
19. P. Sun, J. Cao, Y. Jiang, R. Zhang, E. Xie, Z. Yuan, C. Wang, and P. Luo: TransTrack: Multiple-Object Tracking with Transformer. arXiv preprint arXiv:2012.15460 [cs.CV] (2020)
20. T. Meinhardt, A. Kirillov, L. Leal-Taixé, and C. Feichtenhofer: TrackFormer: Multi-Object Tracking with Transformers. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8844–8854. IEEE, Los Alamitos (2022)
21. R. Gao and L. Wang: MeMOTR: Long-Term Memory-Augmented Transformer for Multi-Object Tracking. In: Proc. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9901–9910. IEEE, Los Alamitos (2023)
22. V. Thambawita, S. A. Hicks, A. M. Storås, T. Nguyen, J. M. Andersen, O. Witczak, T. B. Haugen, H. L. Hammer, P. Halvorsen, and M. A. Riegler: VISEM-Tracking, a Human Spermatozoa Tracking Dataset. *Scientific Data* **10**(1), 260 (2023)
23. Pei, J., Yang, Z., Zhao, Y., Tian, M., Zeng, N., Kang, X., Wu, J., Yang, X.: Multi-object sperm detection and tracking based on enhanced YOLOv4 and improved DeepSORT. *Scientific Reports* **15**(1), 42445 (2025)