Occlusion-Aware Multi-Object Tracking for Point-of-Care Semen Analysis Using Smartphone Imaging

Aojun Jiang, Miao Hao, Yiqian Li, Jiaqi Wang, Chunfeng Yue, Zongjie Huang, Ying Yang, Meng Ma, Qifeng Lyu, Yu Sun, *Fellow, IEEE*, Zhuoran Zhang, *Member, IEEE*

***Abstract*— Semen analysis is the gold standard method for assessing male reproductive capacity. Although clinical semen analysis routinely uses Computer-Assisted Semen Analysis (CASA) systems to measure sperm concentration and motility, clinical visits for semen analysis are not always feasible due to the unavailability of the high-cost CASA systems and the mental stress for clinical visits. Here, we present a point-of-care semen analysis method using smartphone imaging. It follows the same visual tracking scheme as laboratory CASA systems, aiming to provide semen analysis results that are accurate and consistent with clinical standards. To address the challenge of robustly tracking multiple sperm posed by smartphone imaging and the point-of-care nature, we propose an Occlusion-Aware Joint Probabilistic Data Association Filter (OA-JPDAF). It combines sperm head contour obtained from a boundary-sensitive segmentation network and a kinematics-based probabilistic model to determine the occlusion status of both targets and measurements, excluding unreasonable association events to enable robust data association. Experimental results showed that the OA-JPDAF outperformed the state-of-the-art multi-sperm tracking algorithms, and achieved a success rate of 95.14% for tracking occluded sperm, a mean error of 2.03 million/ml for sperm concentration measurement, and a mean error of 1.58% for sperm motility measurement. Clinical tests involving 50 participants revealed that the point-of-care method showed good agreement in sperm concentration and motility measurement (Spearman rank correlation coefficients of 0.94 and 0.89, respectively) with that of the laboratory CASA, even when used by inexperienced users.**

***Index Terms*— Joint Probabilistic Data Association, Point-of-care, Semen Analysis, Smartphone imaging**

I. Introduction

THE World Health Organization (WHO) estimates over 48 million couples worldwide suffer from infertility [1]. Male factor infertility accounts for 40%-50% of all infertility cases in clinical practice [2], mostly attributing to the failure to produce a sufficient quantity of motile sperm. For clinical diagnosis of male infertility, semen analysis is routinely performed to evaluate sperm concentration and motility. Sperm concentration measures the number of sperm per unit volume of the ejaculate, and a low sperm concentration (oligospermia) may indicate undescended testicles [3], genetic disorders [4], dysfunction in spermatogenesis [5], or even azoospermia [6]. Sperm motility measures the percentage of sperm with progressive movement, and a low sperm motility (asthenozoospermia) is an indicator for varicocele [7], sperm mitochondrial dysfunction [8], or testicular damage [9]. Quantitative measurement of sperm concentration and motility and accurate semen analysis results are the basis for diagnosis and treatment of male infertility.

Computer-Assisted Semen Analysis (CASA) has been gradually recognized as a gold standard for laboratory semen analysis [10]. A CASA system visually recognizes and locates sperm in microscopic images of a semen sample, and tracks multiple sperm over a period of time to obtain sperm counts as well as kinematics and thus to calculate concentration and motility. Since CASA systems use a high-grade microscope with an integrated high-performance computer for data processing, visualization and storage, CASA instruments are expensive (tens of thousands of dollars) and require professional technicians to operate.

Availability, affordability, and quality of the diagnosis of male infertility remain a challenge globally [1]. The demand for low-cost, professional-free, accurate semen analysis, or in other words, point-of-care semen analysis is high worldwide. In the past decade, several point-of-care semen analysis methods have been proposed, for instance, paper-based colorimetric assay [11], lateral flow immuno-chromatography [12], centrifugal separation analysis [13], and smartphone-based visual tracking [14],[15]. Among these methods, smartphone-based visual tracking is considered a promising approach for point-of-care semen analysis. It shares the same technical nature as a laboratory CASA, i.e., automated multi-sperm tracking based on sperm detection and location. Differently, it uses a smartphone camera and magnification accessories to image semen samples instead of a high-grade microscope and certified medical camera. The advantages over laboratory CASA systems include significantly lower cost, ease of use, and high

A. Jiang and Y. Sun are with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S3G8, Canada; M. Hao, C. Yue, and Z. Huang are with Suzhou Boundless Medical Technology Company Ltd., Suzhou 215000, China; Y. Li, J. Wang, and Z. Zhang are with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China; Y. Yang, M. Ma, and Q. Lyu are with the Department of Assisted Reproduction, Shanghai 9th People's Hospital, Shanghai 200011, China. Corresponding authors: Zhuoran Zhang (zhangzhuoran@cuhk.edu.cn), Yu Sun (sun@mie.utoronto.ca), and Qifeng Lyu (lyuqifeng@126.com). A. Jiang and M. Hao contributed equally to this manuscript.

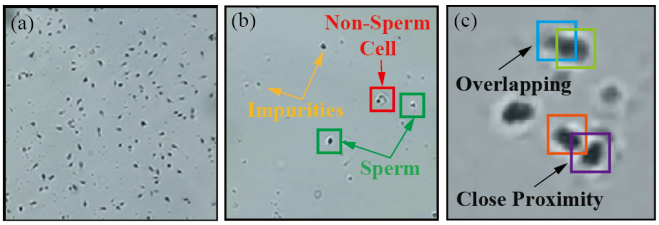

Fig. 1. Technical challenges of point-of-care semen analysis with smartphone imaging: (a) sperm are small in size, making the sperm tail (<1 m) even invisible under smartphone imaging; (b) non-sperm impurities are prevalent and need to be distinguished by the algorithm; (c) semen analysis uses undiluted raw semen where sperm are in close proximity and overlapping/occlusion of sperm is frequent (e.g., >100 sperm moving simultaneously within a field of view of 1024 pixels × 1024 pixels), thus requiring robust multi-sperm tracking algorithms to track individual sperm.

accessibility that mentally helps mitigate embarrassment and stress caused by clinical visits.

However, constrained by cost and portability requirements, smartphone-based microscopic imaging typically has lower magnification and inferior imaging quality compared to laboratory microscopy imaging. Thus, the following technical challenges must be overcome for smartphone-based semen analysis to achieve as accurate sperm concentration and motility measurement as a laboratory CASA.

1) The size of a sperm is small (**Fig. 1a**). Typical tail diameter is smaller than 1 m, making it hardly visible under smartphone imaging (200×). The small size of sperm requires the algorithm to rely mainly on features of the sperm heads to identify and distinguish each sperm.

2) Non-sperm impurities are prevalent in raw semen samples (**Fig. 1b**), whose quantity may exceed the number of sperm by more than two-fold. These impurities often exhibit similar sizes, shapes, and colors to sperm, rendering them challenging to differentiate from sperm.

3) Sperm frequently overlap and occlude each other (**Fig. 1c**). Under a typical concentration of 50-80 million sperm/ml [18], hundreds of sperm move simultaneously within a field of view. As shown in **Section V.D**, failure to track individual sperm leads to significant errors in semen motility evaluation, calling for a robust multi-sperm tracking algorithm.

This paper reports a novel point-of-care semen analysis method using smartphone imaging to address the technical challenges above and improve robustness in tracking occluded targets to achieve accurate sperm concentration and motility measurements. Contributions include:

1) A novel smartphone-based imaging modality for semen analysis was developed. Through this modality, continuous bright-field imaging at an equivalent microscope 200Œ magnification of a controlled volume of raw semen samples can be obtained at 4K 30 fps.

2) A boundary sensitive segmentation network was proposed for specifically identifying, segmenting the sperm head, and separating different sperm in close proximity under smartphone imaging.

3) An occlusion-aware multi-object tracking algorithm was developed to reduce occlusion-caused tracking errors. By incorporating information from both the track (sperm kinematics) and measurements (segmented contours), the occurrence, persistence, and termination of an occlusion event between sperm targets can be predicted and determined. With occlusion awareness, the constraints in enumerating feasible joint events are dynamically relaxed during occlusion to allow the same measurement to match multiple targets. Experiments confirmed the occlusion-aware strategy significantly increased accuracy for tracking occluded sperm.

4) Theoretical analysis was performed to reveal how failures in multi-sperm tracking affect sperm concentration and motility evaluation. Tracking failures were classified into six types, and the effect of each failure type on sperm concentration and motility measurement was quantified. To the best of our knowledge, this is the first work to quantify the effect of tracking failures, especially failures in distinguishing occluded sperm, on semen analysis, which is applicable for both point-of-care and clinical semen analysis.

II. Related Works

A. Methods for Sperm Identification

Accurate identification of sperm cells is the basis for sperm counting and subsequent tracking, which can be achieved by either object detection methods or semantic segmentation methods. Compared to detection methods, target segmentation algorithms enable higher localization precision and provide additional target features (e.g., target contour) that can facilitate robust sperm tracking in complex environments. In this subsection, existing or potential segmentation methods for sperm identification are reviewed. The CNN-based semantic segmentation algorithms, including U-Net [16], DeepLabv3 [17], and Mask-RCNN [18] are the state-of-the-art approaches for sperm segmentation. Compared with conventional thresholding methods such as binary pixel intensity thresholding [19], enhanced-Otsu thresholding [20], and adaptive local threshold [21], CNN-based methods are capable of extracting high-level features to distinguish the background and impurities, resulting in significant improvement in segmentation accuracy. Chen et. al. [22] evaluated popular methods for sperm segmentation on the public Sperm Videos and Images Analysis dataset, among which U-Net demonstrated the top dice score. Although improvements to U-Net and other CNN-based methods for sperm segmentation have been proposed [23]-[25], these methods have not been specifically optimized for smartphone imaging conditions. Moreover, U-Net shares a prominent mis-segmentation issue with other CNN-based methods that multiple sperm in close proximity tend to be segmented as a cluster [22] although they are distinguishable by human eyes. This problem is worsened under smartphone imaging, due to the poorer imaging quality of non-diluted raw semen samples. In this paper, a boundary sensitive segmentation network is proposed to address this problem, as described in **Section IV**.

B. Vision Methods for Sperm Motility Measurement

As stipulated by WHO [26], sperm motility measurement is to calculate the percentage of motile sperm (velocity > 5 m/s) in a semen sample. The technical requirement for sperm motility measurement is to simultaneously track multiple sperm to obtain their trajectories, which is essentially a multiple-object

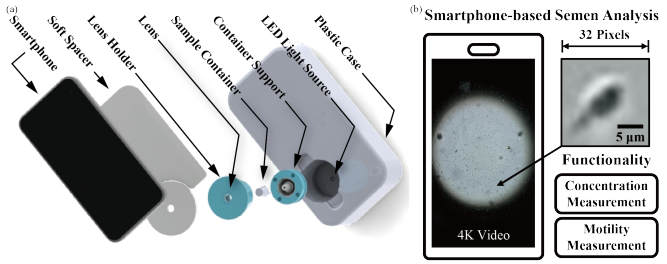

Fig. 2. (a)Diagram of parts breakdown of the developed smartphone semen imaging modality; (b) Raw images captured by the modality.

tracking (MOT) problem. Tracking-by-detection is the most extensively utilized paradigm for such purpose. Its nature is a data association problem, i.e., to associate sperm positions (measurements obtained by the front-end detection or segmentation method) to an existing target in previous frames.

To date, most commercial CASA and point-of-care semen analysis systems use Nearest Neighbor for data association in sperm tracking [27]. It associates a target to a measurement according to minimum inter-frame Euclidean distances. Advances have been achieved by Global Nearest Neighbor (GNN) tracking, which adopts distance-based global optimal association. However, limited by the greedy data association principle, both NN and GNN-based tracking are occlusion sensitive, i.e., the only measurement produced by two or more targets in occlusion cannot be assigned to multiple targets, leading to increased association errors and risk of target loss. Although point-of-care semen analysis with smartphone imaging has been attempted, the NN and GNN-based tracking algorithms applied in such attempts resulted in large errors in motility measurement, up to 100% [14], [28].

Multi-Hypothesis Tracking (MHT) [29], Markov Chain Monte Carlo Data Association (MCMCDA) [30] and Joint Probabilistic Data Association (JPDAF) [31] employ advanced data association principles to accommodate higher tolerance for temporary loss of measurements due to occlusion. MHT uses all past measurements to perform data association, whose heavy computational complexity makes it impractical to track sperm in high concentration samples, despite its high tracking accuracy. MCMCDA uses Markov Chain Monte Carlo to merge track partitions originating from the same target, caused by occlusion. However, this method does not track well for high sperm concentrations (e.g., >60 million/ml) [20] due to the difficulty of correctly grouping too many similar tracks.

The JPDAF tracker represents the state-of-the-art technique for multiple sperm tracking [32]. Its high accuracy is attributed to its enumeration mechanism, i.e., feasible target-measurement association events are enumerated, then the probability of each event is calculated for weighted update of target states. The enumeration strictly follows the one-to-one correspondence assumption, i.e., a target generates at most one measurement and a measurement originates from at most one target. However, this assumption is not always valid, e.g., during sperm overlapping a single measurement (contour) originates from multiple targets. This assumption prevents the correct association event to be enumerated, causing significant association bias and even target loss during occlusion.

Another type of data association strategy is to employ deep learning networks to extract features for feature similarity-based data association [33]–[35]. However, low magnification, lack of color features under smartphone microscopic imaging, and high similarity in sperm appearance (especially during occlusion) may hinder the strength of feature similarity-based data association. Highly similar features also make it challenging for joint-detection-and-tracking methods [36],[37]. Although existing methods [38]-[41] incorporating Transformer may provide enhanced feature extraction ability, under smartphone imaging mainly the sperm heads are visible, and it remains unknown whether these methods can keep distinguishing a specific sperm using the limited features of sperm heads during occlusion. To the best of our knowledge, no studies have validated the performance of such methods for the challenging task of tracking multiple sperm with highly similar appearance.

In addition to multi-sperm tracking, another type of method uses deep learning to take sperm videos as input to directly predict/classify sperm motility, e.g., CNN-based prediction [42] and motilitAI [43]. Such methods bypass multi-sperm tracking, and thus fail to provide sperm trajectories. The black-box classification approach lacks interpretability and makes it challenging to identify potential errors in motility predictions, thus increasing the risk of misdiagnosis.

In this work, we advance the state-of-the-art multi-sperm tracking algorithm, JPDAF. Through occlusion-awareness, the original constraint is dynamically relaxed to ensure robust tracking under occlusion. Note that occlusion-aware JPDAF is not a black box. By analyzing the obtained trajectories, how tracking failure affects motility measurement is revealed.

III. SMARTPHONE MICROSCOPIC IMAGING MODALITY FOR SEMEN ANALYSIS

A novel imaging modality was developed to enable smartphone-based microscopic imaging of a fixed volume of semen samples. As shown in **Fig. 2**, the proposed imaging modality comprises a smartphone, a soft spacer, an objective lens and its holder, a semen sample container and its support, an LED light source, and a plastic case. In contrast to other smartphone-based semen analysis systems that rely on microfluidic devices to maintain a consistent sample volume for imaging, our system utilizes two small, identical, and transparent containers that mimic the glass slide-cover slip structure. The semen sample containers were engineered to accommodate a fixed volume (0.02μL) of semen and allow light to pass through for imaging. To prepare the sample, an excess of raw semen is first added to one of the containers, and the other container is then inserted, causing excess semen to be squeezed out, thus leaving a consistent volume of semen confined to the space between the two containers. The objective lens is mounted in a holder with a small magnet that holds it in place with the container support and keeps the optical path centered. During imaging, the rear camera of the smartphone is aligned with the optical axis of the objective lens by the operator. To accommodate various smartphone models, a soft spacer is custom-designed to provide adequate clearance for the protruding smartphone lenses during imaging. The built-in video recording software of the smartphone enables the

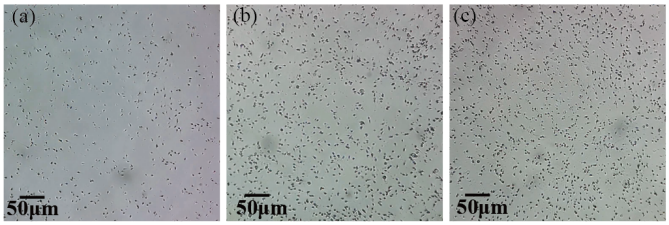

Fig. 3. The FoVs of samples with different sperm concentrations using the smartphone imaging modality. The size for each FoV is 1024 pixels Œ 1024 pixels. Each FoV contains (a) 326 sperm (concentration: 84 millioin/ml); (b) 597 sperm (concentration: 154 million/ml); (c) 912 sperm (concentration: 235 million/ml).

recording of 4K (3648 by 2056 pixels) videos at 30 frames per second. As calibrated using a commercial calibration slide (MR095, AmScope, USA), the smartphone microscopic imaging modality achieved a magnification that is equivalent to 200× of a standard laboratory microscope. The subsequent concentration and motility analysis in this study was based on these settings for a video duration of three seconds.

IV. OCCLUSION-AWARE JPDAF FOR MULTI-SPERM TRACKING

A. Problem Formulation

Accurate visual measurement of sperm concentration and motility is essentially a multi-object (multiple sperm) tracking problem. Multi-sperm tracking using smartphone imaging modality is challenging due to 1) the use of raw, undiluted semen samples in which sperm frequently cross over and occlude each other, as shown in **Fig. 3**; 2) the inherent limitations in imaging quality of smartphone technology, which further exacerbates the challenge of accurately tracking each sperm.

As discussed in **II. Related Work**, among existing data association algorithms, the joint probability data association filter (JPDAF) achieves the highest accuracy for tracking multiple sperm under occlusion. It achieves high accuracy mainly because it enumerates all feasible joint events (association of targets and measurements). However, JPDAF was originally designed for radar-based tracking of multiple flying objects such as aircrafts. In such scenarios, measurement (i.e., physical position) of each target is obtained from radars and such measurement is rarely affected by target occlusion. The JPDAFs basic assumption, i.e., the one-to-one association between measurement and target, is a global optimization hypothesis to enumerate a feasible joint event corresponding to the real situation.

To address these limitations, a novel occlusion-aware JPDAF (OA-JPDAF) is proposed, aiming to realize robust multi-sperm tracking under occlusion for smartphone-based semen analysis. As shown in the workflow in **Fig. 4a**, a dynamic constraint relaxation mechanism is proposed to relax the constraint in the original assumption of one-to-one association between measurement and target. An occlusion evaluator, which incorporates information from both measurements (sperm contour) and targets (kinematic model), is utilized as a front module to enable efficient occlusion awareness. This allows for the initial filtering of feasible joint events that do not match reality. Feasible joint event exclusion are then applied to further refine the enumerated feasible joint events and improve the tracker's robustness against clutter.

B. Relaxation of Feasible Joint Events Enumeration Constraint

JPDAF enumerates feasible joint events based on a validation matrix Ω, which is defined as:

$$\Omega = [\omega_{ij}], \qquad i = 1, \ldots, m, \qquad j = 1, \ldots, n \tag{1}$$

where m and n are the number of measurements and targets in the validation matrix, ω_{ij} is given by:

$$\omega_{ij} = \begin{cases} 1, & \text{if } M_i \in G_j \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where G_j is the gating region of T_j (target j), which means ω_{ij} equals to 1 only when M_i (measurement i) falls into the gating region of T_j.

Feasible joint events of a validation matrix are enumerated based on the basic assumption of original JPDAF that measurement and target are one-to-one corresponding, which can be converted to the feasible enumeration constraints as

$$\sum_{i=1}^{m} \omega_{ij} \leq 1 \text{ and } \sum_{j=1}^{n} \omega_{ij} \leq 1 \tag{3}$$

where $\sum_{i=1}^{m} \omega_{ij} \leq 1$ constrains that one target can generate at most one new measurement, and $\sum_{j=1}^{n} \omega_{ij} \leq 1$, constrains that one measurement can originate from at most one target.

When occlusion occurs, however, the contours of two or more sperm merge into one larger contour, thus providing only one measurement. Using the original constraint **(3)**, no enumerated feasible joint events correspond to the real case. As a result, target states would be incorrectly updated with biased association. As one can expect, if sperm concentration become higher and occlusion between targets becomes more frequent, biased associations would become more common, leading to increased tracking errors.

Assuming a single measurement might originate from multiple targets, the constraints in **(3)** is relaxed to

$$\sum_{i=1}^{m} \omega_{ij} \leq 1 \tag{4}$$

However, directly relaxing the constraints by using **(4)** causes the total number of enumerated feasible joint events to significantly increase, causing more biased associations. With increased biased associations, the target position prediction becomes less accurate with a larger gating region, leading to increased size of validation matrix (increased computational efforts in enumeration which is already exponential time complexity [44]) and wrong state updates. Further restrictions should be imposed on **(4)** for practicable enumeration of feasible joint events.

This side effect stems from the fact that it is unknown whether a target is occluded. Even for targets that are not occluded, it may share the same measurement with other targets in certain feasible joint events. Assume the occlusion status of the targets can be obtained (e.g., by using methods in **Section IV.C**), the relaxation of the constraint can be strictly limited to occluded targets, and the number of feasible joint events can be considerably reduced, also reducing the risk of biased associations. Hence, the first restriction is proposed.

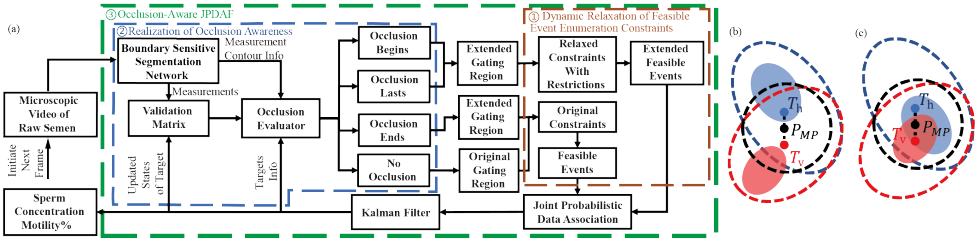

Fig. 4. (a)Workflow of our proposed occlusion-aware JPDAF for robust sperm tracking in the scenario of smartphone-based point-of-care semen analysis. (b) demonstrates an example that no occlusion/occlusion ends is confirmed from measurements by the occlusion evaluator, whereas occlusion begins/ lasts is confirmed in (c). The blue and red dots are the predicted position of $\hat{\boldsymbol{T}}_{\boldsymbol{h}}$ and $\hat{\boldsymbol{T}}_{\boldsymbol{v}}$, surrounded by an individual ellipsoidal gating region. Red and blue shadowed areas are the contour measurement of target $\boldsymbol{T}_{\boldsymbol{h}}$ and $\boldsymbol{T}_{\boldsymbol{v}}$ from the segmentation network. The circular region $\boldsymbol{\Lambda}$ is denoting as the area inside the black dashed line, whose center is at $\boldsymbol{P}_{\boldsymbol{MP}}$ (midpoint of $\hat{\boldsymbol{T}}_{\boldsymbol{h}}$ and $\hat{\boldsymbol{T}}_{\boldsymbol{v}}$).

Restriction 1: The relaxed constraint **(4)** is only applied to occluded targets.

The side effect of significantly increased feasible joint events also originates from the lack of information to determine if a measurement is generated by a single target or multiple occluded targets. Even for a measurement that is not caused by target occlusion, it could be shared by two or multiple targets in certain feasible joint events. If the occlusion status of a measurement can be obtained (e.g., by comparing contour areas across frames which detailed in **Section IV.C**), the feasible joint events enumerated by the **(4)** with **Restriction 1** can be further refined. Hence the second restriction is proposed.

Restriction 2: The relaxed constraint (4) is only applied to measurements that originate from occluded targets.

With the **Restriction 2**, for a measurement from a non-occluded target, feasible joint events that associate this measurement with two or more targets would be excluded. Also, targets within the same occlusion cluster should avoid being associated with different measurements in feasible joint events, which is granted by the third restriction.

Restriction 3: Targets in the same cluster should be simultaneously associated with either a same measurement or null.

The determination of whether targets are within the same enclosed cluster can be obtained pairwise using the method in **Section IV.C**. When different pairs of targets have one or more shared target(s), these targets are obviously within the same cluster. This restriction further reduces the number of enumerated feasible joint events.

For the tracking of occluded targets, the position of an immotile sperm is apparently easier to determine than that of a motile sperm, even if it may present subtle movement due to collisions. For this reason, immotile targets that have been confirmed as in occlusion should be excluded for feasible joint event enumeration until occlusion ends. Here the fourth restriction is proposed.

Restriction 4: Immotile targets in occlusion are excluded for feasible joint event enumeration.

During occlusion, the immotile target is updated using the state before the occlusion. This design further reduces the number of targets in validation matrix and thus contributes to fewer feasible joint events to be enumerated. Moreover, it helps avoid incorrect position updates for immotile targets that are in occlusion, and avoid tracking errors such as target ID switching with moving targets. Following the end of occlusion, such design also facilitates correct re-matching of occluded targets. The above four restrictions can be summarized as the new constraint (C') in **(5)**

$$C' = \begin{cases} \sum_{i=1}^{m} \omega_{ij} \leq 1 \text{ and } \omega_{ih} = \omega_{iv}, \text{if} \\ M_i \in S_{OM}, T_j, T_h, T_v \in S_{OT} \cap S_{MT} \\ \sum_{i=1}^{m} \omega_{ij} \leq 1 \text{ and } \sum_{j=1}^{n} \omega_{ij} \leq 1, \text{otherwise} \end{cases} \tag{5}$$

where T_h and T_v are pair of occlude targets in the same cluster at the current frame. S_{OT} is the set of targets in occlusion, S_{MT} is the set of targets that are motile, and S_{OM} is the set of measurements that originate from the occlusion of two or more targets.

With the above restrictions, the number of feasible joint events can be significantly reduced and the risk of biased association is reduced. Give that the **Restriction 1,2** and **3** rely on the awareness of occlusion of target and measurement, it is necessary to realize occlusion-awareness, which will be detailed in the following section.

C. Realization of Occlusion-awareness

Here we propose an occlusion evaluator driven by multi-frame information fusion to realize predicting, monitoring, and determining the occurrence, persistence, and termination of an occlusion event of multiple sperm targets. To detect when an occlusion event occurs, from the perspective of measurement, the contour of a single sperm differs from that of a cluster of sperm. Another approach is to predict the occlusion of a moving target based on its kinematic model. Hence, the proposed occlusion evaluator assesses an occlusion event from two aspects, i.e., occlusion prediction from existing targets and occlusion validation from measurements.

(i) Occlusion-awareness from Targets: Using the kinematic model and the existing target states, the region where the target appears in the next frame is predictable and the probability of the targets location follows a two-dimensional normal distribution. Therefore, the probability of occlusion is obtained by probabilistically integrating the elliptical overlap part of the predicted position probabilities between targets. For example, for occlusion prediction for target h (T_h) and target v (T_v), the possibility of occlusion can be derived as

$$P_{Occul,T_hT_v} = P\left(T_{h,pred}\right) P\left(T_{v,pred}\right) \iint f_{T_h} f_{T_v} dxdy \tag{6}$$

where $P\left(T_{h,pred}\right)$ and $P\left(T_{v,pred}\right)$ are the probability of successful gating a correct measurement for targets T_h and T_v, which are preset empirical parameters for the formation of the validation matrix. f_{T_h} and f_{T_v} are the probability density function of the target threshold ellipse, which follows

$$f_{T_h}(x,y) = \frac{e^{-\frac{1}{2}[(x,y)-(\mu_x,\mu_y)]^T \Sigma^{-1}[(x,y)-(\mu_x,\mu_y)]}}{2\pi\sqrt{\det(\Sigma)}} \quad (7)$$

where Σ is covariance-dependent eigenvalue matrix. When P_{Occul,T_hT_v} exceeds the set threshold, T_h and T_v will be proposed as occluded target pair in the next frame.

ii) Occlusion-awareness from Measurements: Measurements are first obtained using image segmentation for subsequent occlusion evaluation. This image segmentation step serves two purposes. The first is to detect/locate sperm from a FoV mixed with non-sperm cells and other impurities. The second is to extract contour of a spermatozoa or a sperm cluster to provide measurement for occlusion evaluation.

For the first purpose of detecting/locating sperm, an enhanced U-Net network using ResNet-34 as the backbone is proposed. Segmentation is preferred over object detection because in smartphone imaging, sperm are densely distributed small objects, and accurate target bounding box for object detection method relies on a good selection of non-maximal suppression parameters, while image segmentation methods do not require this delicate parameter adjustment. To improve the capacity of segmentation network for identifying sperm, the segmentation network is desired to extract the more generalizable features to exclude sperm-like impurities but to include sperm groups containing two or more overlapping sperm (see **Fig. 1c**).

Hence, our network adopts ResNet-34 as the backbone. The introduction of ResNet-34 is expected to solve the problem of vanishing gradients and mitigate the degradation (accuracy saturation) problem of deeper neural network models, leading to better learning of generalizable sperm features. For the second purpose, accurately segmenting and distinguishing single sperm and sperm clusters is critical, because it determines the accuracy of subsequent occlusion evaluation. Hence, the enhanced U-Net should be further improved in terms of boundary sensitivity to separate single sperm that is close to other sperm but do not form a cluster due to occlusion. As shown by our results, the original U-Net sometimes fails. To address this issue, a loss function based on boundary-weighted maps is introduced. While the original U-Net does not elevate weights for boundaries of isolated objects (objects that are far away from other objects), the weight map generation method consistently enables the highest elevated weights on the sperm contour boundary line, followed by pixels close to the contour line, and no elevation for other pixels. This enables the segmentation network to learn sperm boundaries with higher weights even if there are no other sperm nearby, facilitating the network to learn more boundary information. The boundary-sensitive loss function is

$$\mathcal{L} = \frac{1}{N}\sum_{i=1} -w_i[y_i \log(p_i) + (1-y_i)\log(1-p_i)] \quad (8)$$

where y_i is the labeled value of pixel (1 for sperm, 0 for others). p_i is the predicted probability of a pixel belonging to sperm, w_i is the weight value of the pixel i in the pre-generated weight map, and N is the total number of pixels in an image.

Based on precise segmentation results, an occlusion-awareness method utilizing information from both targets and measurements powered by multi-frame information fusion is proposed. Initially for an occlusion event, in the $(t-1)^{th}$ frame, P_{Occul,T_hT_v} is first calculated followed by **(6)** and **(7)**. Once T_h and T_v are proposed as a potential occluded target pair in the t^{th} frame, an examination of contour area changes in a circular searching region Λ is subsequently performed to confirm. The searching region Λ (illustrated in **Fig. 4b&c**) centers at the midpoint (P_{MP}) of the predicted positions of T_h and T_v in t^{th} frame, with a radius that is proportional to the velocity of T_h or T_v, which is defined as

$$\Lambda = \left\{p \middle| (p-P_{MP})^2 < R^2,\ R \propto \max(v_h, v_v)\right\} \quad (9)$$

where p is a point on and within the region Λ, and v_h and v_v are the absolute velocity of T_h and T_v, respectively.

There are two potential circumstances for T_h and T_v in the t^{th} frame: 1) occlusion begins and 2) no occlusion. The probability of occlusion p_t in the t^{th} frame is calculated by comparing the size of the largest contour in the circular region Λ with the size of the contour of the target sperm in the preceding frame, which is given by

$$p_t = c\sum_{i=t-2}^{t-1}[(\log(\frac{A_t}{A_{h,i}})\log(\frac{A_t}{A_{v,i}})] \quad (10)$$

where A_t is the measurement of the largest contour area of sperm or sperm cluster within the circular searching region Λ, and $A_{h,i}$ and $A_{v,i}$ are the measurement of sperm contour area in i^{th} frame of target h and v, and c is a scaling coefficient.

The change of occlusion conditions alters the value of p_t. Thus, by comparing p_{t-1} and p_t, different occlusion conditions (Occlusion begins: when $p_t \geq T_H$; No occlusion: when $p_t \geq T_H$; Occlusion ends: when $p_{t-1} \geq T_L$ and $p_t < T_L$ after occlusion begins; Occlusion lasts: when $p_{t-1} \geq T_L$ and $p_t \geq T_L$ after occlusion begins. p_{t-1} is the possibility of occlusion occur in $(t-1)^{th}$ frame, T_L is the low threshold and T_H is the high threshold).

After confirmation by the above method, the pair of targets proposed to be in occlusion by **(6)** can be finally labeled as occluded pair of targets (for **Restriction 1** and **3**). The measurement with largest contour in the circular region Λ can be identified as a measurement generated by occluded targets (for **Restriction 2**). By the proposed method, accurate awareness of occlusion state of both targets and measurements is obtained.

D. Calculation of Sperm Concentration and Motility

The sperm concentration is calculated by

$$\text{Concentration} = \frac{N_{sperm}}{L \times W \times \mu^2 \times H} \quad (11)$$

where N_{sperm} is the final number of targets in the tracker, L (unit: pixel) is the length of the FoV, W (unit: pixel) is the width of the FoV, μ is a calibrated coefficient converting the length of each pixel from image coordinate to real world length, H is the depth of semen container. In the current setup, L = W = 1024 pixels, = 0.43pixel/μm, H = 20μm. This method is used to calculate concentration instead of averaging the number of measurements provided by the segmentation network in each frame because it avoids the reduction of sperm

count due to occlusion, and also avoids the effect of clutter on the count that only presents intermittently for a few frames.

The motility is calculated as the percentage of motile sperm (velocity > 5 $\mu m/s$ as defined by WHO) in a semen sample:

$$\text{Motility} = \frac{N_{sperm|v>5\mu m/s}}{N_{sperm}} \tag{12}$$

where $N_{sperm|v>5\mu m/s}$ is the number of sperm with velocity greater than $5\mu m/s$.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Human sperm samples used in experiments were obtained from Shanghai Ninth Peoples Hospital (Shanghai, China). Informed consent was obtained from all subjects (50 patients enrolled). The study was approved by the Institutional Review Board of the Shanghai Ninth Peoples Hospital (IRB: SH9H-2021-T14-1). In experiments, all semen videos were recorded using an Android phone (Huawei P20, China) and the developed imaging modality. Following the WHO guideline, the duration of each video was set to 3 seconds.

A. Evaluation of the Segmentation Network for Sperm Recognition

TABLE I
PERFORMANCE OF THE SEGMENTATION NETWORK FOR DISTINGUISHING SPERM AND NON-SPERM CELLS

Sample	1	2	3	4	5	6	7	8	9	10	Overall
Sperm Count	114	137	83	165	158	295	260	150	105	141	1608
Non-sperm Count	201	69	119	126	123	129	120	192	154	247	1480
True Positive	106	128	77	159	153	279	251	132	97	131	1513
False Positive	5	3	8	14	16	14	16	8	12	13	109
False Negative	8	9	6	6	5	16	9	18	8	10	95
Precision	0.95	0.98	0.91	0.92	0.91	0.95	0.94	0.94	0.89	0.91	0.93
Recall	0.93	0.93	0.93	0.96	0.97	0.95	0.97	0.88	0.92	0.93	0.94

To evaluate the performance of the proposed segmentation network for distinguishing sperm and non-sperm impurities, ten raw semen samples with high impurities were used for analysis, in which non-sperm impurities accounted for at least 33% and up to 64% of the total identifiable objects. The number of sperm and non-sperm impurities of each tested sample were manually counted by experienced lab technicians.

In these challenging scenarios, the proposed segmentation network achieved a mean precision of 93% and a mean recall rate of 94% among the 10 samples (**Table I**). As for error cases, totally 109 false recognitions (false positive) and 95 missing recognitions (false negative) were found in the evaluation, accounting for 3.1% and 3.5% of total identifiable objects in all samples, with the worst case 5.7% and 5.3%, respectively. In addition, no trend of increase in precision and recall rate was found as non-sperm rate or sperm count increases. The results demonstrated that our proposed algorithm remained robust even in the case of high impurities.

The capability of our boundary-sensitive segmentation network was also qualitatively evaluated and compared to the original U-Net. For comparison purposes, the boundary sensitive network and the original U-Net were both trained on the same dataset containing 50 images (12,229 sperm head contours labeled by experienced technicians). As shown in **Fig. 5**, our boundary-sensitive segmentation network trained with boundary weighted map was able to obtain separated contours of sperm with close distances. While for the original U-Net, these sperm in close proximity were segmented as a cluster with a large contour. Note that the goal in semen analysis is to separate, as much as possible, all sperm that are at a close distance without occlusion. The improvement in boundary sensitivity of the proposed segmentation network contributes to a lower computational complexity and a more accurate occlusion evaluation for the subsequent Occlusion-Aware JPDAF tracker.

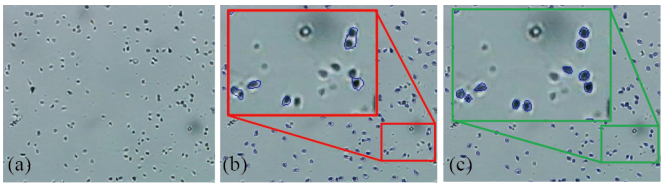

Fig. 5. An example segmentation result of the boundary sensitive segmentation network. (a) Original image of a semen sample obtained by the smartphone imaging modality; (b) Segmentation result of the original U-Net, sperm in close proximity were segmented as a whole contour; (c) Segmentation result of our proposed boundary sensitive segmentation network, sperm in close proximity were segmented as individual and separated contours.

B. Tracking Occluded (Overcrossing) Sperm

Most existing multi-sperm trackers can robustly track sperm in non-occluded scenarios (i.e., individual sperm swimming freely). Hence, we focused on evaluating the tracking performance of the proposed Occlusion-Aware JPDAF tracker (OA-JFDAF) for occluded sperm (i.e., multiple sperm crossing over and overlapping each other). The performance of OA-JPDAF was compared to that of the original JPDAF (a state-of-the-art multi-sperm tracking benchmark) and Nearest Neighbor (NN, the most commonly used tracking method for commercial CASA systems). Our boundary sensitive segmentation network was used to provide accurate sperm head contours and head centroid coordinates for all the three trackers.

To quantitatively evaluate the success rate of the three tracking algorithms, a total 329 cases of overcrossing of sperm from 8 video clips were included, with results summarized in **Table II**. A success case is defined as the sperm continuing to be tracked by the tracker with the same ID as before and after occlusion. The proposed OA-JPDAF showed the highest success rate and was capable of tracking 95.14% of sperm through occlusions. The original JPDAF was ranked next, with a success rate of 82.98%. While the NN algorithm performed the worst among the three, with a tracking success rate of 75.08%. It is worth noting that the achieved tracking success rate of 95.14% is for occlusion cases only. In more common samples that include both occlusion and non-occlusion cases, the overall success rate would be even higher. To further analyze tracking failure cases of each algorithm, the failure cases were classified into six categories according to the type of errors and occlusions, as detailed in the next section.

C. Analysis of Tracking Failure Cases

Depending on how the target sperm ID is affected, failure cases were divided into the six cases, as shown in **Fig. 6**.

(1) Extra sperm ID generation: an extra sperm ID is generated after the sperm has been occluded, and the existing sperm ID exists simultaneously with the new sperm ID. Obviously, this type of error results in a higher than actual sperm concentration. If this type of error occurs between

TABLE II
PERFORMANCE OF TRACKING OCCLUDED TARGET

Method	Overcrossing Case Count	Error Case Count						Success Case Count	Success Rate
		a	b	c	d	e	f		
Occlusion-aware JPDAF	329	6	3	4	1	0	2	313	95.14%
Original JPDAF	329	24	8	7	1	12	2	275	83.59%
Nearest Neighbor	329	16	7	36	14	5	4	247	75.08%

motile and immotile sperm (tracking error **a**), it results in a statistically higher motility rate. If it occurs between motile sperm (tracking error **b**), the motility values may either increase or decrease depending on the original speed of the two sperm. The main reason for the occurrence of type **a** error is the biased association caused by sperm occlusion. One of the targets in occlusion may associate with clutter or measurements originated from other sperm, altering the kinematic model. This causes the measurement that should have been matched to exceed the gating area of the target when the occlusion ends, and the measurement is established as a new target. Similarly, type **a** error happens when motile sperm collides with non-motile sperm, causing the non-motile sperm to move and altered the velocity vector of the motile sperm. Although it is the most frequent error in OA-JPDAF, its frequency was only 37.5% of the NN algorithm and 25% of the original JPDAF algorithm. This significant improvement was enabled by the improved data association mechanism for occlusion targets supported by occlusion awareness. When occlusion is predicted for a target and confirmed based on the size change of the contour, OA-JPDAF stops updating the state of the immotile target with new measurements until the end of occlusion (**Restriction 4** of the relaxed constraint) to ensure that the states of the motile and immotile targets do not converge due to moving together followed by collisions. Moreover, OA-JPDAF essentially alleviates biased association during occlusion by using the relaxed constraint with extra restrictions **(5)**, ensuring significantly reduced tracking error when handling occluded targets.

(2) Sperm ID switch: the occluding sperm IDs are switched or exchanged after occlusion. The total number of sperm remains unchanged, and this would not cause a statistical error in sperm concentration. However, when this error occurs between motile and immotile sperm (tracking error **c**), the motility values are incorrectly increased. When the error occurs between motile sperm (tracking error **d**), although the change in motility value cannot be determined qualitatively, it changes the true trajectory of the sperm.

Type **c** and **d** errors are mainly prevalent in NN due to the fact that the data association of the NN algorithm is based on the nearest distance principle only. Therefore, sperm ID exchange often occurs when the targets are occluded. The JPDAF algorithm introduces the Mahalanobis distance and kinematic state updated by Kalman Filter to facilitate correct target-measurements matching. As a result, it can significantly reduce target ID exchange between motile and immotile sperm, thus leading to fewer type c and d errors than NN (**Table II**). However, when the motile sperm push the immotile sperm to move, their kinematic states may be still

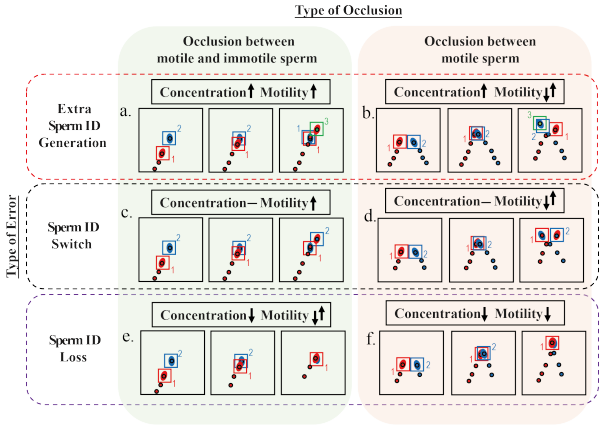

Fig. 6. Schematics of six common tracking errors when tracker handling occluded targets as in Table II. The effect of different tracking errors on concentration and motility measurement are given by ↑ , ↓ , ↓↑ , and –, which represents increase, decrease, indeterminable, and no effect.

prone to converge, and then sperm ID exchange still occurs in the original JPDAF. OA-JPDAF inherits the advantages of the original JPDAF while it stops updating the state vectors of stationary targets in occlusion state based on occlusion awareness (**Restriction 4** of the relaxed constraint), and alleviates biased data association by implementing an occlusion-aware based relaxed constraint **(5)**, ensuring that the ID switching is further reduced. Accordingly, OA-JPDAF further reduced the occurrence of type **c** error by over 40% compared to the original JPDAF.

(3) Sperm ID loss: after occlusion, one of the sperm IDs is deleted and the two sperm share a single ID. Evidently, this type of error leads to a decrease of the sperm concentration. For this type of error occurring between motile and immotile sperm (tracking error **e**), depending on whether the disappearing ID originally belonged to a motile or immobile sperm, the motility value can be either incorrectly decreased or increased. For such errors occurring between motile sperm (tracking error **f**), the concentration will be incorrectly reduced because the percentage of motile sperm is decreased.

Type **e** and **f** errors are mainly due to the inability of the trackers data association mechanism to correctly assign a single measurement caused by occlusion to two or more sperm. For the NN algorithm, this single measurement would be assigned to any target that was closer to the measurement in the previous frame.

If the occlusion duration is long, a target has not associate with any measurements for a period of time will be deleted, leading to type **e** and **f** errors. For the original JPDAF algorithm, ideally, when only a single measurement falls into the gating regions of two occluded targets at the same time, both targets can use the same measurement for state update based on probability weighted data association. However, when targets are dense or clutters exist in the surrounding area, the gating region of the occluded target may include clutter or measurement of other targets, causing the incorrect updates of target position. If no other measurement falls into its gating region after incorrect state updates for certain period of time, it would be deleted, resulting in errors of type **e** and **f**. OA-JPDAF employs an occlusion-aware based relaxed constraint **(5)**, which senses the occlusion status of

both targets and measurements, and only retains feasible joint events that are most likely corresponding to the true target-measurement matching relationship. The application of this novel data association mechanism of OA-JPDAF resulted in a 50% reduction in the number of type **f** errors relative to the original JPDAF and NN. Combining with the mechanism of stopping state updates of stationary occluded targets with measurements, this resulted in no occurrence of type **e** errors in all of the counted overcrossing cases.

D. Effects of Tracking Failure on Sperm Concentration and Motility Evaluation

To further quantify the effect of these tracking errors on sperm concentration and motility measurement, another eight videos were captured from eight participants and the results of smartphone measurement were benchmarked to manually counting and measurement results. To provide accurate manual benchmarking, experienced laboratory technicians counted the number of sperm in each video with best care and then converted to concentration and motility according to **(11)** and **(12)**. Similarly, the trajectory of each sperm presented in the last frame was retrospectively reviewed by the technician to calculate sperm velocity and motility results.

For concentration (**Fig. 7a**), the OA-JPDAF showed the least mean error of 2.03 million/ml, followed by NN (2.90 million/ml) and original JPDAF (3.70 million/ml). All the three tracking algorithms gave a higher concentration than manual measurement. This is in agreement with the analysis of tracking failure cases in **Section V.C** because error type **a** and **b** increase the concentration, while type **e** and **f** decrease the concentration. All three algorithms have more type **a** and **b** errors than type **e** and **f** errors. Despite this, no significant differences were observed among the three groups. This could be because the same segmentation network was used to identify sperm and to provide head contours for all the three trackers.

As for motility (**Fig. 7b**), the OA-JPDAF showed a significantly lower mean motility error than original JPDAF (1.58% vs. 7.73%, p<0.001) and NN (1.58% vs. 21.40%, p<0.001). These data also align with the tracking error case analysis in **Section V.C**. Both Original JPDAF and NN had significantly more type **a** and **c** errors than OA-JPDAF. Given that type **a** and **c** errors lead to an increase in motility, the average motility obtained by Original JPDAF and NN was significantly higher than that of OA-JPDAF. These statistical results illustrate that OA-JPDAF outperformed the other two algorithms in motility measurement accuracy. Furthermore, concentration and motility results of the algorithms were highly correlated with their tracking failure cases, suggesting that tracking errors in occluded targets had a significant impact on the measurement of concentration and motility.

E. Comparison of Smartphone-Based Semen Analysis by Inexperienced End-Users with Laboratory CASA on 50 Clinical Samples

In order to fully test the performance under point-of-care scenarios, all samples for smartphone-based semen analysis were prepared by users without relevant experience, and the

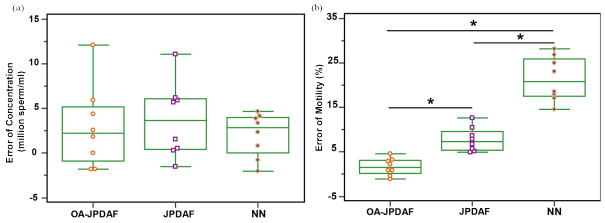

Fig. 7. Sperm concentration and motility errors of our proposed smartphone-based semen analysis system benchmarked to manual measurement, where * denotes significant difference (p<0.001) is found between two groups.

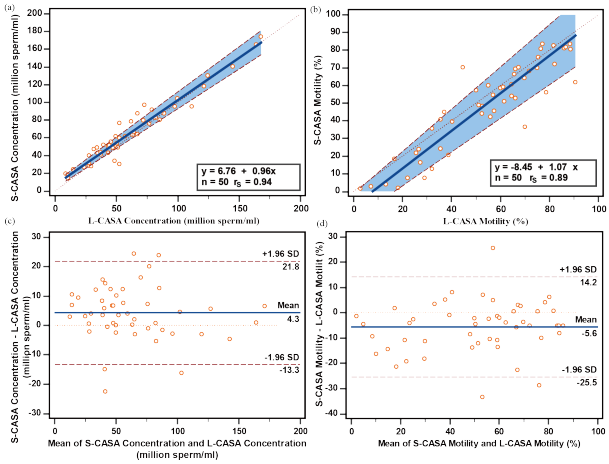

Fig. 8. Evaluation of clinical samples (n=50). First row: Passing-Bablok regression for the concentration (a) and motility results (b) between the proposed smartphone-based CASA and laboratory CASA; Second row: Bland-Altman analysis comparing the sperm concentration (c) and motility (d) obtained by smartphone-based CASA and laboratory CASA

obtained results were compared to laboratory CASA. From each sample, one aliquot was tested using our smartphone-based system by an inexperienced user, while another aliquot was tested using laboratory CASA (BEION S5, BEION, China) by an experienced laboratory technician, for side by side comparison. Fifty patients were enrolled in this study.

Passing-Bablok regression (n = 50) was conducted to evaluate the agreement of sperm concentration (**Fig. 8a**) and motility (**Fig. 8b**) calculation between the proposed smartphone-based semen analysis system and laboratory CASA. For the regression of concentration, a slope of 0.96 (95% Confidence Interval CI: 0.90 to 1.02) and an intercept of 6.76 (95% CI: 2.64 to 9.64) were achieved. Model linearity was validated by Cusum test with P = 0.99. A Spearman rank correlation coefficient of 0.94 (95% CI: 0.9 to 0.97) with P <0.0001 was obtained. For the regression of motility, a slope of 1.07 (95% CI: 0.95 to 1.20) and an intercept of -8.45 (95% CI: -15.95 to -1.44) were achieved. Model linearity was validated by Cusum test with P = 0.89. A Spearman rank correlation coefficient of 0.89 (95% CI: 0.81 to 0.94) with P <0.0001 was obtained. The above results confirmed that our point-of-care semen analysis system for sperm concentration and motility calculation based on smartphone raw semen images agreed well with conventional laboratory CASA.

The Bland-Altman analysis (n = 50), which compared the sperm concentration obtained by smartphone-based CASA and laboratory CASA (**Fig. 8c**), showed a mean bias of 4.25 million sperm/ml with an SD of 8.95 million sperm/ml. The limits

of agreement ranged from 13.34 to 21.85 million sperm/ml. The Bland-Altman analysis on sperm motility between two methods (**Fig. 8d**) showed a mean bias of -5.62% with an SD of 10.13%. The limits of agreement ranged from 25.45 to 14.20. As shown in **Fig. 8c&d**, as sperm concentration or motility increased, no fixed difference or trend of error was found, suggesting the proposed method can accommodate samples with a wide range of sperm concentrations to provide accurate sperm concentration and motility measurement.

VI. Conclusion

This paper presented a novel semen analysis method with smartphone imaging to address the challenges of robust point-of-care sperm concentration and motility measurement. A smartphone imaging modality was developed to enable clear imaging of a fixed volume of raw semen samples. A boundary sensitive segmentation network was proposed to segment sperm contours and to distinguish sperm with non-sperm cells or impurities in raw semen. To tackle the challenging of frequent overlapping and occlusion between sperm, an Occlusion-Aware JPDAF was developed for robust tracking of multiple sperm. Comparative studies confirmed that the proposed Occlusion-Aware JPDAF outperformed state-of-the-art tracking algorithm, making significantly less tracking errors when handling occluded targets. The effect of tracking failure on sperm concentration and motility measurement was quantified, which is applicable to both point-of-care and clinical semen analysis for male infertility diagnosis.

References

[1] WHO, *Infertility Prevalence Estimates, 1990–2021*. Apr. 2023.

[2] N. Kumar and A. K. Singh, "Trends of male factor infertility, an important cause of infertility: A review of literature," *J.Hum. Reprod. Sci.*, vol. 8, no. 4, pp. 191–196, 2015.

[3] M. J. Mathers *et al*, "The Undescended Testis: Diagnosis, Treatment and Long-Term Consequences," *Dtsch. Ärztebl. Int.*, vol. 106, no. 33, pp. 527–532, Aug. 2009.

[4] R. Flannigan and P. N. Schlegel, "Genetic diagnostics of male infertility in clinical practice," *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 44, pp. 26–37, Oct. 2017.

[5] B. D. Anawalt, "Approach to Male Infertility and Induction of Spermatogenesis," *J. Clin. Endocrinol. Metab.*, vol. 98, no. 9, pp. 3532–3542, 2013.

[6] A. Agarwal *et al.*, "Male infertility," *The Lancet*, vol. 397, no. 10271, pp. 319–333, Jan. 2021.

[7] A. Agarwal *et al*, "Effect of varicocele on semen characteristics according to the new 2010 WHO criteria: a systematic review and meta-analysis," *Asian J. Androl.*, vol. 18, no. 2, pp. 163–170, 2016.

[8] T. Folgerø *et al*, "Mitochondrial disease and reduced sperm motility," *Hum. Reprod. Oxf. Engl.*, vol. 8, no. 11, pp. 1863–1868, Nov. 1993.

[9] R. Jacobsen *et al.*, "Risk of testicular cancer in men with abnormal semen characteristics: cohort study," *BMJ*, vol. 321, no. 7264, pp. 789–792, Sep. 2000.

[10] J. Lammers *et al.*, "Automated semen analysis…the new gold standard? a comprehensive study comparing manual and automated semen analysis," *Fertil. Steril.*, vol. 98, no. 3, Supplement, p. S150, Sep. 2012.

[11] Y. E. Yoon *et al.*, "Validation of SwimCount™, a Novel Home-Based Device That Detects Progressively Motile Spermatozoa: Correlation with World Health Organization 5th Semen Analysis," *World J. Mens Health*, vol. 38, no. 2, pp. 191–197, Apr. 2020.

[12] M. A. Coppola *et al.*, "SpermCheckő Fertility, an immunodiagnostic home test that detects normozoospermia and severe oligozoospermia," *Hum. Reprod. Oxf. Engl.*, vol. 25, no. 4, pp. 853–861, Apr. 2010.

[13] C.-Y. Chen *et al.*, "Sperm quality assessment via separation and sedimentation in a microfluidic device," *Analyst*, vol. 138, no. 17, pp. 4967–4974, Jul. 2013.

[14] M. K. Kanakasabapathy *et al.*, "An automated smartphone-based diagnostic assay for point-of-care semen analysis," *Sci. Transl. Med.*, vol. 9, no. 382, p. eaai7863, Mar. 2017.

[15] A. Agarwal *et al.*, "Home sperm testing device versus laboratory sperm quality analyzer: comparison of motile sperm concentration," *Fertil. Steril.*, vol. 110, no. 7, pp. 1277–1284, Dec. 2018.

[16] O. Ronneberger *et al*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, 2015, pp. 234–241.

[17] L.-C. Chen *et al*, "Rethinking Atrous Convolution for Semantic Image Segmentation." arXiv, Dec. 05, 2017.

[18] K. He *et al*, "Mask R-CNN." arXiv, Jan. 24, 2018.

[19] M. J. Tomlinson *et al.*, "Validation of a novel computer-assisted sperm analysis (CASA) system using multitarget-tracking algorithms," *Fertil. Steril.*, vol. 93, no. 6, pp. 1911–1920, Apr. 2010.

[20] G. Jati *et al*, "Multi-sperm tracking using Hungarian Kalman Filter on low frame rate video," in *2016 International Conference on Advanced Computer Science and Information Systems*, Oct. 2016, pp. 530–535.

[21] P. Hidayatullah *et al*, "Automatic sperms counting using adaptive local threshold and ellipse detection," in *2014 International Conference on Information Technology Systems and Innovation*, Nov. 2014, pp. 56–61.

[22] A. Chen *et al.*, "SVIA dataset: A new dataset of microscopic videos and images for computer-aided sperm analysis," *Biocybern. Biomed. Eng.*, vol. 42, no. 1, pp. 204–214, Jan. 2022.

[23] Q. Lv, *et al.*, "An Improved U-Net for Human Sperm Head Segmentation," *Neural Process. Lett.*, vol. 54, no. 1, pp. 537–557, Feb. 2022.

[24] V. Valiukait *et al*, "Deep Learning Based Evaluation of Spermatozoid Motility for Artificial Insemination," *Sensors*, vol. 21, no. 1, p. 72, 2020.

[25] R. Melendez *et al.*, "Sperm Cell Segmentation in Digital Micrographs based on Convolutional Neural Networks using U-Net Architecture," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, Jun. 2021, pp. 91–96.

[26] *WHO laboratory manual for the examination and processing of human semen*. World Health Organization, 2021.

[27] H. O. Ilhan *et al*, "Sperm Motility Analysis by using Recursive Kalman Filters with the smartphone based data acquisition and reporting approach," *Expert Syst. Appl.*, vol. 186, p. 115774, Dec. 2021.

[28] W. H. Cheon *et al.*, "Validation of a smartphone-based, computer-assisted sperm analysis system compared with laboratory-based manual microscopic semen analysis and computer-assisted semen analysis," *Investig. Clin. Urol.*, vol. 60, no. 5, pp. 380–387, Sep. 2019.

[29] I. J. Cox and S. L. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 138–150, Feb. 1996.

[30] S. Oh, S. Russell, and S. Sastry, "Markov Chain Monte Carlo Data Association for Multi-Target Tracking," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 481–497, Mar. 2009.

[31] Y. Bar-Shalom *et al.*, "The probabilistic data association filter," *IEEE Control Syst. Mag.*, vol. 29, no. 6, pp. 82–100, Dec. 2009.

[32] L. F. Urbano *et al*, "Automatic Tracking and Motility Analysis of Human Sperm in Time-Lapse Images," *IEEE Trans. Med. Imaging*, vol. 36, no. 3, pp. 792–801, Mar. 2017.

[33] N. Wojke *et al*, "Simple Online and Realtime Tracking with a Deep Association Metric." arXiv, Mar. 21, 2017.

[34] F. Yu *et al*, "POI: Multiple Object Tracking with High Performance Detection and Appearance Feature." arXiv, Oct. 19, 2016.

[35] J. Xu *et al*, and H. Hu, "Spatial-Temporal Relation Networks for Multi-Object Tracking." arXiv, Apr. 25, 2019.

[36] P. Bergmann *et al*, "Tracking without bells and whistles," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 941–951.

[37] Z. Wang *et al.*, "Towards Real-Time Multi-Object Tracking." arXiv, Jul. 14, 2020.

[38] P. Sun *et al.*, "TransTrack: Multiple Object Tracking with Transformer." arXiv, May 04, 2021.

[39] T. Meinhardt *et al*, "TrackFormer: Multi-Object Tracking with Transformers." arXiv, Apr. 29, 2022.

[40] F. Zeng *et al*, "MOTR: End-to-End Multiple-Object Tracking with Transformer." arXiv, Jul. 19, 2022.

[41] A. Vaswani *et al.*, "Attention Is All You Need." arXiv, Dec. 05, 2017.

[42] Priyansi *et al*, "Predicting Semen Motility using three-dimensional Convolutional Neural Networks." arXiv, Jan. 14, 2021.

[43] S. Ottl *et al*, "motilitAI: A machine learning framework for automatic prediction of human sperm motility," *iScience*, vol. 25, no. 8, p. 104644, Aug. 2022.

[44] D. F. Crouse, "Algorithms for Tracking in Clutter and for Sensor Registration," Ph.D., University of Connecticut, United States -- Connecticut.