Computational and Structural Biotechnology Journal 24 (2024) 176–184

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

Multidimensional morphological analysis of live sperm based on multiple-target tracking

Hao Yang [a,1], Mengmeng Ma [a,1], Xiangfeng Chen [c,d,e,1], Guowu Chen [f,1], Yi Shen [b], Lijun Zhao [b], Jianfeng Wang [b], Feifei Yan [b], Difeng Huang [b], Huijie Gao [b], Hao Jiang [b], Yuqian Zheng [b], Yu Wang [a], Qian Xiao [c,d,e], Ying Chen [f], Jian Zhou [g], Jie Shi [g], Yi Guo [a,*], Bo Liang [h,*], Xiaoming Teng [a,*]

[a] Department of Assisted Reproduction, Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, Shanghai, China
[b] Basecare Medical Device Co., Ltd., Suzhou, Jiangsu, China
[c] Center for Reproductive Medicine, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[d] Shanghai Key Laboratory for Assisted Reproduction and Reproductive Genetics, Shanghai, China
[e] Shanghai Human Sperm Bank, Shanghai, China
[f] Shanghai Ji Ai Genetics and IVF Institute, Obstetrics and Gynecology Hospital, Fudan University, Shanghai, China
[g] Suzhou Dushu Lake Hospital (Dushu Lake Hospital Affiliated to Soochow University), Suzhou, Jiangsu, China
[h] State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Dataset link: https://jz.basecare.cn/

Dataset link: https://lw.basecare.cn/

Keywords:
Non-staining
Live sperm
Morphological analysis
Multiple sperm tracking
Motility
Deep learning

ABSTRACT

Manual semen evaluation methods are subjective and time-consuming. In this study, a deep learning algorithmic framework was designed to enable non-invasive multidimensional morphological analysis of live sperm in motion, improve current clinical sperm morphology testing methods, and significantly contribute to the advancement of assisted reproductive technologies. We improved the FairMOT tracking algorithm by incorporating the distance and angle of the same sperm head movement in adjacent frames, as well as the head target detection frame IOU value, into the cost function of the Hungarian matching algorithm. For sperm morphology, we used the BlendMask segmentation method to segment individual sperm. SegNet was used to separate the head, midpiece, and principal piece comments from each sperm. Experienced in vivo sperm physicians confirmed a morphological accuracy percentage of 90.82%. A total of 1272 samples were collected from multiple tertiary hospitals for validation of the system, which were also evaluated by physicians. The results of our system were highly consistent with those of manual microscopy. This study realized the automated detection of progressive motility and morphology of sperm simultaneously, which is crucial for selection of morphologically normal and motile sperm for intracytoplasmic sperm injection.

1. Introduction

Assessment of semen quality is critical for diagnosing male infertility. The sixth edition of the World Health Organization's reference manual for routine semen analysis includes criteria such as liquefaction, consistency, volume, color, and pH Organization and others [19]. Sperm morphological abnormalities are categorized into three types: defective head, midpiece, and principal piece. Currently, clinical methods of sperm morphology detection involve manual assessment of the size and shape of all three sperm substructures by trained technicians after staining the sperm to determine whether the sperm meets normal conditions, which is time-consuming and highly dependent on the expertise of clinical technicians.

Therefore, automated algorithms have been developed for the morphological analysis of sperms. Early computer-aided sperm morphometric analyses processed images to detect and locate sperm using threshold-, edge-, and region-based segmentation (techniques [20,23, 18,4]). These techniques require various preset thresholds for the ac-

* Corresponding authors.
E-mail addresses: gy_guoyi@tongji.edu.cn (Y. Guo), boliang@basecare.cn (B. Liang), tengxiaoming@51mch.com (X. Teng).
[1] These authors contributed equally.

https://doi.org/10.1016/j.csbj.2024.02.025
Received 8 December 2023; Received in revised form 27 February 2024; Accepted 27 February 2024
Available online 1 March 2024
2001-0370/© 2024 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

curate segmentation and extraction of regions of interest. Furthermore, the sperm images captured by the microscope's optical system are affected by the light and color of the sample, contaminants, and cells in the semen. The preset thresholds cannot be adaptively adjusted to changes in the target image. Therefore, segmentation techniques are often unsatisfactory [15].

Several techniques are available for the automated assessment of stained animal sperm. One such method was presented by Zhang [27], who used fluorescence microscopy to capture images of stained ram sperm. The morphometric analysis of these sperms was conducted using ImageJ software. In another study, Alegre et al. [1] used a combination of digital image processing and learning vector quantization (LVQ) to automatically classify the acrosome of boar sperm. This involved acquiring images of sperm heads using a phase-contrast microscope and evaluating the acrosome status based on staining color. The findings from their experiments demonstrated a classification error of 6.8% for sperm heads.

The application of deep learning to sperm image categorization and segmentation has been reported in several studies ([24], [17]). Riordon et al. [21] used the VGG16 neural network to categorize sperm according to the WHO standards for sperm shape, and validated the algorithm using two publicly available head datasets (HuSHeM and SCIAN) [10]. On both datasets, their system outperformed existing methods in terms of classification accuracy. However, this method could not distinguish between different parts of the same sperm. Javadi and Mirroshandel [16] used deep learning algorithms to classify sperm morphology with relatively high accuracy. Their model could detect morphological abnormalities in different parts of the human sperm and was trained and tested on the MHSMA dataset, demonstrating relatively good accuracy.

Several studies have explored the potential of advanced techniques for evaluating and classifying sperm cells according to various criteria. Butola et al. [9] used high spatial phase sensitivity to capture the phase map of entire sperm cells, including the head, neck, and tail, and applied an end-to-end deep learning approach to classify normal sperm, H2O2-stressed, ethanol-stressed, and cryopreserved cells. Similarly, Sato et al. [22] developed and evaluated a YOLOv3-based machine learning model for rapid morphological evaluation and tracking, with sensitivity and positive predictive value (PPV) for abnormal sperm at 0.881 and 0.853, respectively, and for normal sperm at 0.794 and 0.689, respectively. Moreover, their model successfully tracked 78.4% of the objects completely, 21.6% partially, and no objects were lost. Fraczek et al. [14] introduced a comprehensive sperm classification method using Mask R-CNN, trained on two publicly available and one specially created sperm database, and proposed a 14-element feature vector for classifying four typical head defects (amorphous, normal, tapered, and pyriform) using a Support Vector Machine. Moreover, Dobrovolny et al. [12] demonstrated the potential of the YOLOv5 Architecture to detect sperm in images, achieving the best not-super-tuned MAP of 72.15.

Despite the numerous studies on sperm morphological analysis, implementing such a high-standard analysis remains challenging because of factors such as

1) sperm death and unavailability for in vitro fertilization treatment;

2) staining increasing the variability of sperm morphology and reducing the consistency of sperm analysis;

3) failure to subdivide the morphology of sperm into the head and tail (midpiece and principal piece) regions, motility assessment decoupled from morphological analysis, and absence of a full-dimensional analysis of motile sperm multi-morphology.

Recently developed algorithms have been used for sperm classification. FairMOT is a tracker that dominates multi-target tracking [28], and its structure can be used to accurately track sperm. BlendMask is an efficient, high-precision instance segmentation algorithm with good scalability and segmentation effects on crossed sperms [11]. The number of network layers of the SegNet model can be used in a larger context to predict the three parts of sperm [3]. Finally, in 2019, Google proposed the EfficientNet model, a new generation of high-performance convolutional neural networks [25] designed to distinguish between the normal and pathological morphology of the head, midpiece, and principal piece of sperm.

This study aimed to develop a novel algorithm based on FairMOT, BlendMask, SegNet, and EfficientNet to accurately classify microscopic sperm images with minimal computational effort. These findings could help improve the management of infertility in men.

2. Methods

2.1. Data collection and experimental methods

All data used in the algorithm development were obtained from the Male Department Laboratory of the Shanghai First Maternity and Infant Hospital between March and May 2022. Semen samples were collected from the Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine (FMAI), affiliated with Shanghai Jiaotong University School of Medicine (Renji), and Shanghai JIAI Genetics & Institute, affiliated with the Obstetrics and Gynecology Hospital of Fudan University (JIAI). Renji is recognized as a leading sperm bank in Shanghai [26]. All the participants provided informed consent. The study was approved by the ethics committees of the Shanghai First Maternal and Child Health Hospital (approval KS2201), Renji Hospital, Shanghai Jiaotong University School of Medicine (approval KY2022-032-B), and Fudan University Obstetrics and Gynecology Hospital (approval 2022-38).

For each sample, 30–50 photographs and two or three videos were collected (see Fig. 1). In March, 250 samples, 10,700 images, and 750 videos were collected at the Shanghai First Maternity and Infant Hospital. In April, 200 samples were collected at the JIAI, with 8600 images and 500 videos. In May, 250 samples were collected from Renji with 10,700 images and 750 videos. Semen was liquefied, placed onto a sperm counting plate, and left to stand for 2 min. Images were captured using a light microscope equipped with a camera at 1000× magnification. The microscope was equipped with an X-Y motorized stage and the acquisition route was customized for specialized slides to ensure a homogeneous and effective field of view. Videos of the semen smear were captured at 25 frames/s with a frame size of 1536×1024 pixels, corresponding to an area of 240×160 μm on the slide. A total of 2000 1-s sperm motion videos were captured, of which 1800 were used for training and 200 for testing. Additionally, among the captured 30,000 static sperm images, 27,000 were used for training and 3000 for testing. Three experienced male laboratory professional technicians independently labeled all morphological data as normal or abnormal according to WHO standards. The final labels were determined using the majority-vote method.

Morphological analyses were performed on the same semen sample using our algorithm and a traditional manual microscopic examination to evaluate the accuracy of the dynamic sperm morphology recognition algorithm. A total of 1272 samples were collected from the three hospitals for analysis, which were independent of the 2000 videos and 30,000 images described earlier. Hospital technicians performed the staining process according to the manufacturer's instructions (Sperm Morphology Stain Kit - Papanicolaou Method, ANHUI ANKE BIOTECHNOLOGY). Subsequently, they manually analyzed the sperm morphology using an oil lens at 100 × magnification for artificial sperm morphology analysis. They observed and recorded at least 200 sperms and categorized them as either normal or abnormal. For each specimen, the developed morphological analysis algorithm was used to directly analyze the sample, with a cumulative total of 200 thoroughly segmented sperm or the first 20 complete fields of view extracted from the recorded videos.

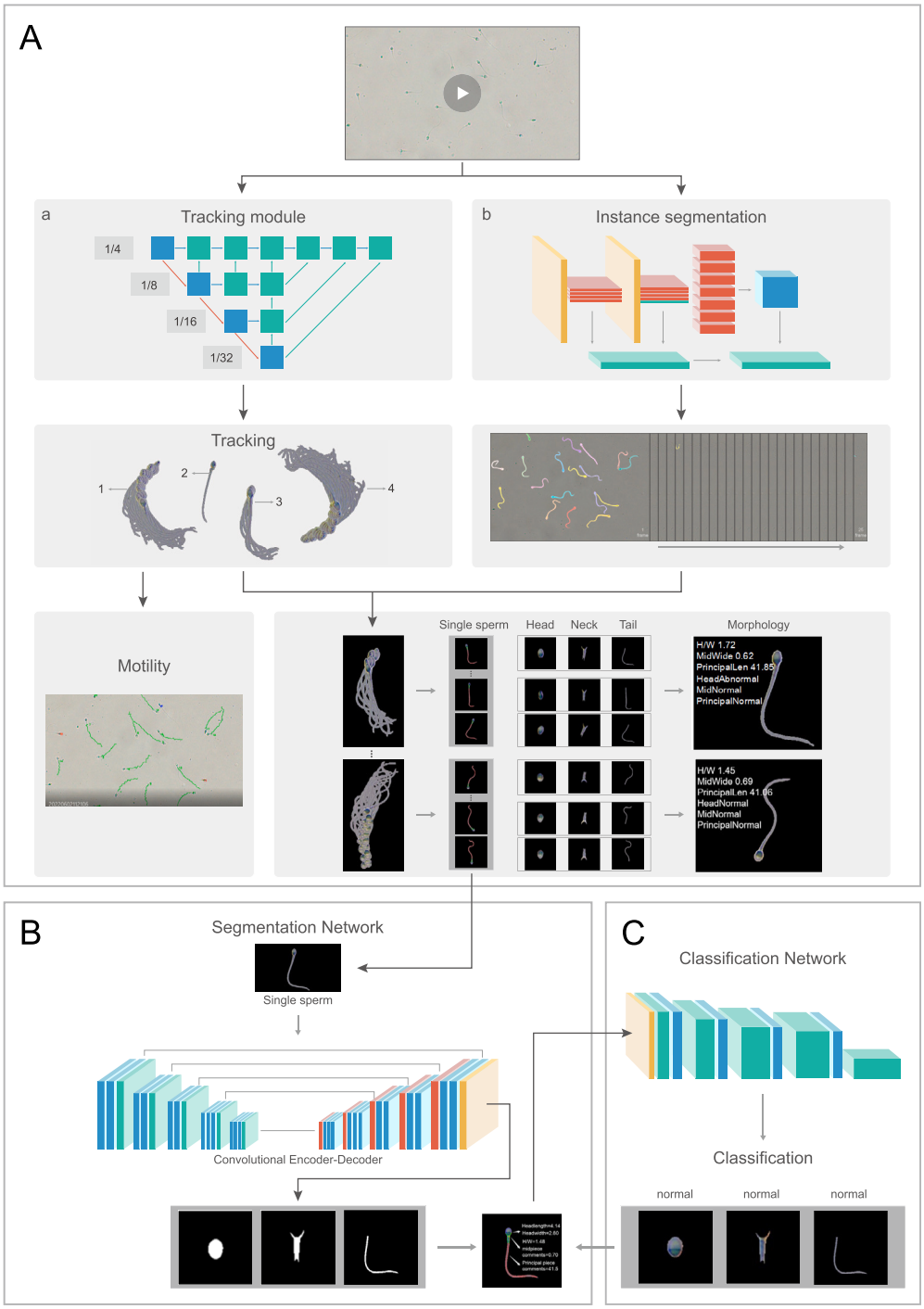

Fig. 1. Structure of the algorithm framework. (A.a) Modified FairMOT network for multiple sperm tracking to obtain the viability results. (A.b) BlendMask instance segmentation network for multi-frame single sperm segmentation. (B) SegNet semantic segmentation network. (C) EfficientNet-b0 classification network for morphological analysis of sperm. (Sperm dimension: μm).

2.2. Development of morphological recognition algorithms

2.2.1. Multiple sperm tracking methods

The FairMOT algorithm allows multi-object tracking and consists of two homogeneous branches that can predict pixel-wise objectness scores and re-ID features, resulting in a high detection and tracking accuracy. The detection branch uses an anchor-free method to estimate the centers and sizes of objects, which are represented as position-aware measurement maps. Similarly, the re-ID branch estimates a re-ID feature for each pixel to characterize the object centered at the pixel. Unlike previous anchor-based methods, which use feature maps of stride 32, FairMOT uses high-resolution feature maps of stride four. The elimination of anchors and the use of high-resolution feature maps better align Re-ID features with object centers, which significantly improves tracking accuracy [28].

Sperm morphology changes during movement, necessitating resolution of the sperm tracking issue. To achieve sperm tracking, the FairMOT multi-object tracking algorithm was improved by incorporating sperm motion characteristics to increase the accuracy rate of sperm tracking. FairMOT is a highly advanced tracking algorithm that uses

an anchor-free target detection method, the Centernet algorithm [29], to estimate the target center on a high-resolution characteristic graph. By adding an output branch for head keypoint detection using the Centernet algorithm, we were able to predict the central point and detect more keypoints without changing the network structure. The head keypoint was connected to the target detection center point to represent the direction of the head movement. FairMOT was used to determine the appearance distance of the Hungarian algorithm match using the ID feature, which considered that the appearance distance of the same sperm in adjacent frames was greater than that of different sperm. Based on the motion characteristics of the sperm, the limited motion of sperm in adjacent frames, and the $IOU > 0$ of its head target detection frame, we integrated the head frame IOU and the motion angle of the sperm. The combination of the head frame IOU, angle, and distance of the head central point was the cost function for the first Hungarian algorithm match, and the cost function calculated by the sperm appearance distance was used for the remaining sperm matches that did not match for the first time. The FairMOT algorithm was trained using sperm videos captured using a microscope, whereas other models were trained using static images.

We performed Hungarian algorithm matching by combining the head frame IOU, angle, central point distance, and sperm appearance distance as the cost function to address the problem of sperm motion crossover and the similar appearance of identical sperm exchanging IDs. The algorithm successfully achieved accurate tracking of each sperm even in dense distributions. After reliable sperm tracking results were obtained, the motion trajectory of the same sperm was determined. The viability findings were then acquired by counting according to WHO standards [19]. Progressive motility (PR) was defined as the active movement of sperm in a straight line or a large circle, regardless of speed. Non-progressive motility (NP) encompassed additional types of non-forward motion, such as swimming in a small circle with little or no tail momentum driving the head to move or only a tail wobble. Immotility (IM) was characterized by the absence of movement.

2.2.2. Extraction of single sperm segmentation

BlendMask is an instance segmentation framework based on the FCOS object detector. The bottom module of the framework predicts bases using either backbone or FPN features. A single convolution layer is then added to generate attention masks and bounding box predictions. For each predicted instance, the blender crops the bases with its bounding box and linearly combines them according to the learned attention maps [11].

We used the BlendMask framework to obtain a single sperm from each frame. The algorithm learns different levels of feature maps by introducing the fading module before Roi pooling, which allows for better capture of instance boundaries and improves the segmentation accuracy. Additionally, the feature fusion module fuses the feature maps learned by the fading module with the backbone network features of the Mask R-CNN to further improve the segmentation accuracy. BlendMask also uses feature fusion of multiple positive and negative samples to improve the algorithm's segmentation of small- and high-density targets by suppressing category-independent information. We trained the instance segmentation model using BlendMask on 27,000 sperm images. The trained model accurately separated and obtained a single sperm, and then correlated the head position coordinates with the sperm tracking findings to obtain a single sperm with the same ID in all video frames.

2.2.3. Extraction of head, midpiece, and principal piece comments

After the multiple sperm images underwent the instance segmentation step, we used the thresholding technique to exclude overly large, connected regions (some sperm were heavily adhered or heavily covered by impurities) that were difficult for the instance segmentation algorithm to segment correctly. To eliminate background interference, we placed each sperm individually in the background image with 380×380-pixel values of zero, and then extracted different parts of each sperm, including the head, midpiece, and principal piece, for individual morphological analysis.

SegNet is a core trainable segmentation architecture that consists of an encoder network, corresponding decoder network, and pixel-wise classification layer. The architecture of the encoder network is topologically identical to that of the 13 convolutional layers in the VGG16 network. The decoder network converts the low-resolution encoder feature maps into full input-resolution feature maps for pixel-wise classification. The novelty of SegNet lies in the manner in which the decoder upsamples its lower-resolution input feature maps. Specifically, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling [3].

We used the SegNet semantic segmentation model to segment the head, midpiece, and principal piece comments, instead of continuing with the instance segmentation method, because the three sperm parts are positioned independently of each other. Accurate segmentation results could be obtained using only semantic segmentation algorithms. This technique allowed SegNet to achieve high segmentation accuracy using fewer computational resources. We annotated and screened 27,000 sperm images used previously to obtain 80,000 single-sperm images for training the semantic segmentation model. Using the semantic segmentation approach, a single sperm could be divided into three independent parts (head, midpiece, and principal piece comments).

2.2.4. Classification of normal morphological abnormalities

We then used EfficientNet to perform a morphological classification of the three sperm components. EfficientNet is a convolutional neural network architecture specifically designed for image classification tasks and uses a compound scaling method. This method simultaneously adjusts the depth, width, and resolution in the network, allowing it to adapt better to images of different sizes and achieve optimal outcomes across multiple image classification tasks. Its faster training speed and updated model size can be efficiently applied for sperm morphology classification [25].

The classification and semantic segmentation algorithms used the same data batch for model training. In addition to using the classification algorithm, we estimated the head length, width, and aspect ratio based on the head segmentation findings in accordance with the WHO 6th edition standards. For individual sperms, the morphology of the head, midpiece, and principal piece was examined separately to determine their normality. Any sperm with abnormalities in any of these regions was considered to have abnormal morphology. Due to the limited field of view during examination, only sperms with complete and intact morphology of the head, midpiece, and principal piece were included in the analysis.

Based on the tracking results of multiple sperm, morphological analysis was performed on each sperm in multiple consecutive frames. The final morphological analysis results for each sperm was obtained by examining the morphological effects of the sperm in multiple frames. This approach, which fully considered the morphological characteristics of different frames of the same sperm, enhanced the accuracy and objectivity of the final morphological analysis.

2.2.5. Algorithm deployment

We developed our system using C++ and accelerated image processing using an Nvidia RTX A5000 32 GB graphics card and multi-threaded processing. Depending on the semen concentration, 10 to 15 videos were captured, and the forward inference of the network model was accelerated using TensorRt36. The process achieved a speed of up to 22 s/video, resulting in a complete morphological analysis within 3 min.

2.3. Statistical analysis

MedCalc software was used for statistical analyses. Least-squares regression analysis was used to compare the manual and automated

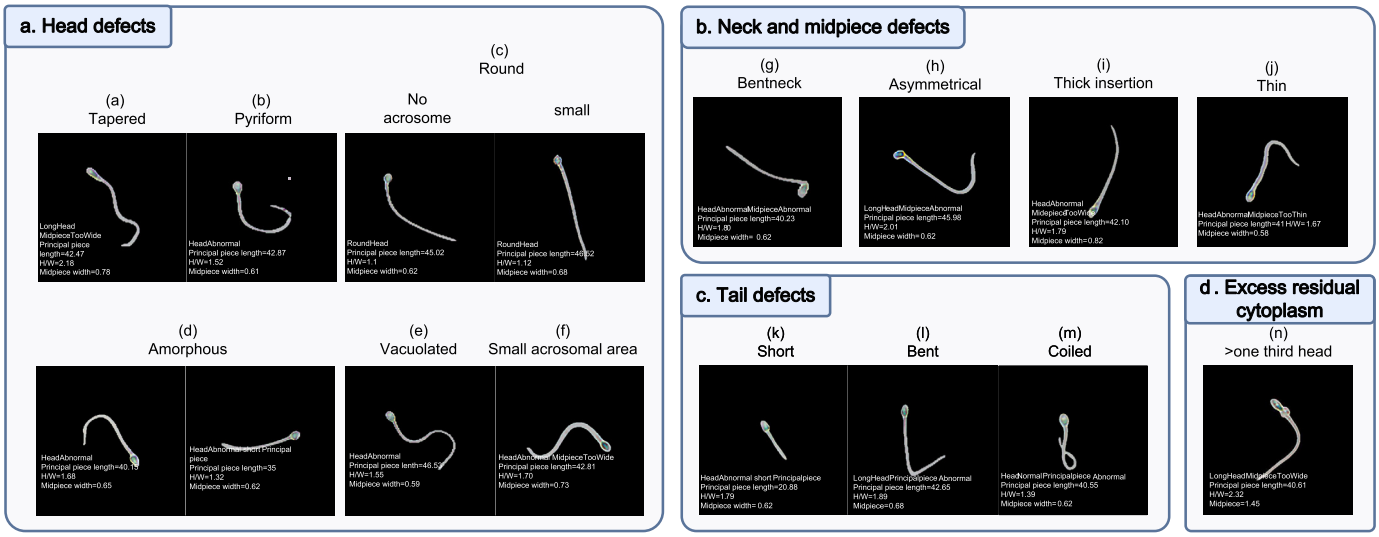

Fig. 2. Algorithm implemented for sperm defect type identification and morphological findings. (A) Head defect. (B) Midpiece defect comments. (C) Principal piece defect comments. (D) Excess residual cytoplasm. (Dimensions of sperm: μm).

methods and to determine the Pearson correlation coefficients. Additionally, Bland-Altman plots and Passing-Bablok regression methods were used to analyze the consistency of the two methods and detect any systematic errors. The Bland-Altman analysis [5–7] is based on the calculation of the average difference between the results obtained by the two methods, and enables the identification of systematic bias. It is widely used to assess consistency between two different instruments or measurement techniques. Passing-Bablok regression is a linear regression approach that makes no specific assumptions regarding the sample distribution and measurement error [2]. The results were independent of the distribution of X and Y using the method (or instrument). The slope B and intercept A were calculated based on 95% confidence intervals. These confidence intervals were used to determine whether the differences between B and one and between A and zero were only random.

3. Results

3.1. Achieving accurate judgment of unstained non-invasive sperm morphology

Fig. 2 displays the single sperm morphologies obtained by the algorithm, which accurately detected the four categories of abnormal sperm morphology according to WHO criteria. Our algorithm used an unstained, non-invasive live sperm morphology calculation and analyzed each sperm in the video. Each sperm was classified using deep learning, and the shape parameters of the head, midpiece, and principal piece of the sperm were measured. Highly accurate morphological results were obtained for unstained sperm. Our algorithm accurately classified the 11 abnormal sperm morphologies listed by WHO (Head: Tapered, No-acrosome, and Small; Neck and midpiece defects: Bent neck, Asymmetrical, Thick insertion, Thin, and Excess residual cytoplasm; Tail defects: Short, Bent, and Coiled).

3.1.1. Tracking accuracy

We used the improved FairMOT multi-object tracking algorithm in videos to track sperm [28]. The performance of multi-sperm tracking was evaluated using multi-object tracking accuracy (MOTA), a standard evaluation metric for multi-object tracking algorithms. Finally, 200 test sets yielded an average MOTA value of 93.47%, which was calculated as follows:

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \quad (1)$$

where m_t is the number of targets missed in t-frames, fp_t is the number of misidentified targets in t-frames, mme_t is the number of incorrectly matched targets in t-frames, and g_t is the number of real targets in t-frames.

Table 1 presents the multi-sperm tracking performance data for the 200 test sets. These 200 sets were divided into six groups based on sperm concentration, and the average results for each group were calculated separately. When sperm distribution was relatively dispersed, each sperm could be accurately tracked. However, when sperm were close or intersecting, false matches or sperm misses could have occurred. Among the six groups of videos, the first and second groups had dense sperm distribution and a very high probability of sperm proximity or intersection, resulting in an MOTA of approximately 90%. The third and fourth group videos had moderate sperm distributions, and the MOTA reached 95%. The fifth and sixth group videos had sparser sperm distributions, and the MOTA reached 97%. Hence, good tracking results could be obtained regardless of sperm concentration; however, more precise results could be achieved with lower sperm concentrations.

3.1.2. Segmentation accuracy

We used BlendMask segmentation to extract a single sperm from static images, and then used SegNet semantic segmentation to identify its three parts: head, midpiece comments, and principal piece comments. Subsequently, we used the EfficientNet classification algorithm for the triple classification of sperm to obtain the morphological classification of the head, midpiece comments, and principal piece comments. Finally, we incorporated constraints for each sperm part to evaluate the same frame morphological result for that sperm. We used the BOX Average Precision (BOX AP) and MASK Average Precision (MASK AP) to evaluate the instance segmentation results. The AP is calculated by integrating the PR curve, a curve graph with the recall value on the horizontal axis, and the precision value on the vertical axis. The recall and precision are explained below. BlendMask instance segmentation achieved an integrated BOX AP of 0.9437 and MASK AP of 0.9290 after testing with 3000 test sets. For the semantic segmentation, we used the Dice method:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2)$$

where X denotes the Ground Truth, and Y denotes the Prediction. The Dice values of all 3000 test sets were 0.996, 0.976, and 0.998 for the head, midpiece comments, and principal piece comments, respectively.

Table 1
Evaluation of multi-sperm tracking performance.

Teams	1	2	3	4	5	6
Frames T	25	25	25	25	25	25
$\sum_t m_t$	97.15 ± 13.64	77.09 ± 14.02	20.9 ± 6.39	8.9 ± 6.44	7.33 ± 2.22	6.24 ± 2.18
$\sum_t fp_t$	44.94 ± 11.88	25.09 ± 11.84	32.21 ± 8.23	19.42 ± 6.24	10.06 ± 7.2	4.9 ± 2.75
$\sum_t mme_t$	14.03 ± 3.84	5.97 ± 5.36	1.30 ± 1.12	0.0	0.0	0.0
$\sum_t g_t$	1379.42 ± 16.18	1080.72 ± 14.32	809 ± 10.73	658.42 ± 16.37	469.42 ± 16.91	367.47 ± 12.76
MOTA (%)	88.68 ± 3.75	89.99 ± 2.65	93.27 ± 2.62	95.70 ± 2.04	96.23 ± 1.61	96.97 ± 0.86

Table 2
Performance recognition results for head, midpiece comments, and principal piece comments.

	Precision	Recall	Sensitivity	Specificity
Head	0.589	0.918	0.918	0.885
Midpiece comments	0.936	0.927	0.927	0.931
Principal piece comments	0.937	0.976	0.976	0.976

3.1.3. Accuracy of morphological analysis

A total of 63,092 single spermatozoa were evaluated for morphological recognition accuracy using 3000 test sets. Accuracy was calculated as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

True positive (TP) signifies that the sperm labeled as abnormal by the physician was also predicted as abnormal by the algorithm. True negative (TN) indicates that the sperm labeled as normal by the physician was also predicted as normal by the algorithm. False positive (FP) signifies that the sperm labeled as normal by the physician was predicted as abnormal by the algorithm. False negative (FN) indicates that the sperm labeled as abnormal by the physician was predicted to be normal by the algorithm. The morphological recognition accuracy of whole sperm was 90.82%.

Because of the unbalanced distribution of positive and negative data in the test sets, we extracted head normal and abnormal data for 882 and 1026 sets, respectively, midpiece normal and abnormal data for 927 and 821 sets, respectively, and principal piece normal and abnormal data for 1658 and 1669 sets, respectively. Physician-determined outcomes were considered to be the gold standard. Accuracy, sensitivity, specificity, recall, and precision were used to determine the accuracy of the algorithm. The formula for accuracy is presented above. The formulas for the remaining four indicators are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

The test results for the head, midpiece comments, and principal piece comments are presented in Table 2. We used the EfficientNet classification algorithm [25] in addition to the head aspect ratio, midpiece width, and principal piece length to assess the normal and abnormal conditions of the three parts of sperm. Because of important influencing factors, such as the vacuole of the head and the acrosome, the head was accurately classified at 90.6%. The accuracy rate for the midpiece recognition was 92%. However, errors in the recognition process mainly occurred because of the asymmetric access of the midpiece to the head and incorrect measurement of the midpiece width. The length of the principal piece was computed using the principal piece comments obtained through the segmentation. Our primary segment recognition algorithm recognized the acute angle bend and curl with high accuracy, resulting in an overall recognition rate of 97% for the principal piece.

3.2. Calculating the percentage of progressively motile and morphologically normal sperm

Fig. 3 shows that most samples obtained from the three hospitals had higher values of Sperms × (PR & Normal)% than Sperms × PR% × Normal%. The conventional method can only provide an approximate estimate of the number of sperm with progressive motility and normal morphology by multiplying the regular total number of sperm by the percentage of progressive motility and the percentage of morphologically normal sperm. However, this method may not be able to detect sperm with progressive motility and normal morphology in individuals with low motility or low morphological normality, resulting in an almost zero rate that cannot be detected using the conventional method. To overcome this limitation, our system could accurately calculate the value of progressively motile and morphologically normal sperm by conducting simultaneous morphological and motility analyses on the same motile sperm. This novel method addresses the shortcomings of existing methods for calculating the results for this specific set of sperms with low motility or normal morphology.

3.2.1. High consistency with multicenter morphological test results

To validate the accuracy of our algorithm in assessing sperm morphology, we analyzed the same semen sample separately using both manual and algorithmic approaches, and reported the morphological normality results for each sample. This study was conducted in three Grade A tertiary hospitals in Shanghai. A total of 1272 samples were collected from the three hospitals for validation. These 1272 samples were completely independent of those used to test the models.

Table 3 shows a comparison between the morphological results after staining in the three hospitals and the results of our algorithm for analyzing unstained semen samples. The data are presented as means ± standard deviation, slope, 95% consistency range of the regression line intercept, and Pearson correlation coefficient. Morphological data were expressed as the percentage of normal sperm to the total number of sperm.

Table 3 and Fig. 4 A-C show that the Pearson's r-values of our algorithm were all greater than 0.9 when compared to the stained sperm morphology from Shanghai First Maternity and Infant Hospital and Shanghai JIAI Genetics & Institute. The range of morphological differences in the Bland-Altman plots was less than 4, which was consistent with the WHO sampling error range. Fig. 4 D-E shows that the 95% consistency range of the slope of the stained morphology between our

Table 3
Slope of the Passing-Bablok plot and 95% consistency range of the regression line intercept.

	FMAI (algorithm)	RenJi (algorithm)	JIAI (algorithm)
$Mean \pm SD$	5.22 ± 4.02 (4.71 ± 3.98)	7.48 ± 4.37 (7.40 ± 4.64)	3.9 ± 4.0 (5.1 ± 4.9)
	FMAI	RenJi	JIAI
Pearson	0.916	0.922	0.903
Slop	1.000 to 1.111	0.875 to 1.000	0.667 to 0.778
intercept	-0.111 to 1.000	1.000 to1.250	0.111 to 0.500

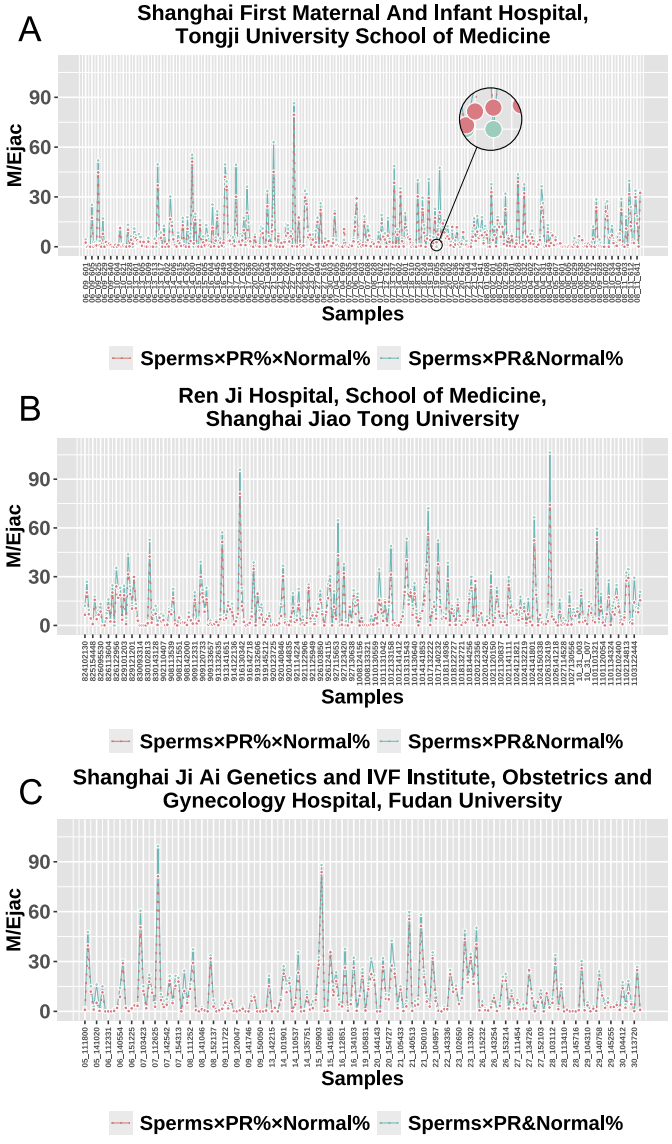

Fig. 3. Comparison of the two methods for calculating sperm with progressive motility and normal morphology. (A-C) Comparative results of the two methods in the three hospitals (ICH, Renji, and Ji'ai). respectively). Sperms × PR% × Normal% indicates the multiplication of the total number of conventional sperm by the percentage of progressive motility and the percentage of morphologically normal sperm. Sperms × (PR & Normal)% indicates the number of progressively motile and morphologically normal sperm, where the horizontal coordinate is the number of sperm samples, and the vertical coordinate is the total number of progressively motile and morphologically normal sperm in millions.

algorithm and manual microscopic examination was 1, suggesting that the difference between the unstained and stained morphology of the algorithm was small and the consistency was high.

4. Discussion

We used the BlendMask instance segmentation algorithm [11] to segment all single sperm morphologies in the video frames. We integrated the tracking algorithm with sperm ID identification to achieve multi-frame morphology detection of the same sperm. For a single sperm, the SegNet semantic segmentation algorithm was used to extract the head, midpiece, and principal piece comments [3]. The EfficientNet model [25] was used to perform a three-part classification and combine the results to provide more accurate and objective sperm morphology analysis results. Unstained sperm morphology analysis ensures that the activity and viability of the detected sperm are maintained, allowing for non-destructive, accurate, and rapid detection of live sperm morphology. The system uses a deep learning algorithm to perform detailed morphological analysis of each sperm, allowing the selection of sperm with good forward motility and normal morphology for assisted reproduction technologies. Sperm selection is of great significance and has applications in assisted reproduction [13,8].

Several algorithms have been proposed to analyze live sperm, but they are frequently restricted to specific criteria, such as motility or morphology. However, these earlier algorithms achieved a relatively high accuracy for specific tasks. The issue was that the motility tracking algorithms were unable to determine morphology, whereas the morphology algorithms were unable to compensate for sperm movement, necessitating the use of fixed specimens. Compared with previous algorithms, we developed an innovative non-invasive sperm morphology analysis algorithm that can dynamically detect the morphological and motility information of live sperm, including the head, midpiece, and principal piece, and accurately measure the viability of individual sperm, consistent with the standard semen analysis method recommended by the WHO. The method accurately tracks all sperm in the field of view and generates multi-frame morphological data for each sperm by matching the tracked position information, which can reflect sperm quality more comprehensively and objectively through multidimensional analysis. The tracking algorithm solves the issues of sperm motion crossover, similar appearance of similar sperm exchanging ID, and lost and missed frames and achieves an average MOTA value of 93.47%. The sperm morphological analysis method accurately segments and extracts data to obtain the head, midpiece, and principal piece comments and then uses the deep learning classification algorithm to classify and evaluate each part of the segmented sperm according to the WHO standards. Our algorithm accurately classified normal sperm and provided various explanations for the abnormalities. The morphological accuracy was verified by multicenter clinical verification to be $> 90\%$, consistent with the standard of clinical detection of sperm morphology, and demonstrating high clinical application value.

This study detected progressively motile and morphologically normal sperms. The percentage of progressively motile and morphologically normal sperm in traditional semen analysis can only be determined by multiplying the percentage of progressively motile sperm by the percentage of morphologically normal sperm. However, this method of calculation may not provide objective results for patients who do not possess any progressively motile or morphologically normal sperm. The integrated viability and morphology analysis method for live sperm can accurately locate each sperm with progressive motility and normal morphology, providing precise measurement of the quantity of high-quality

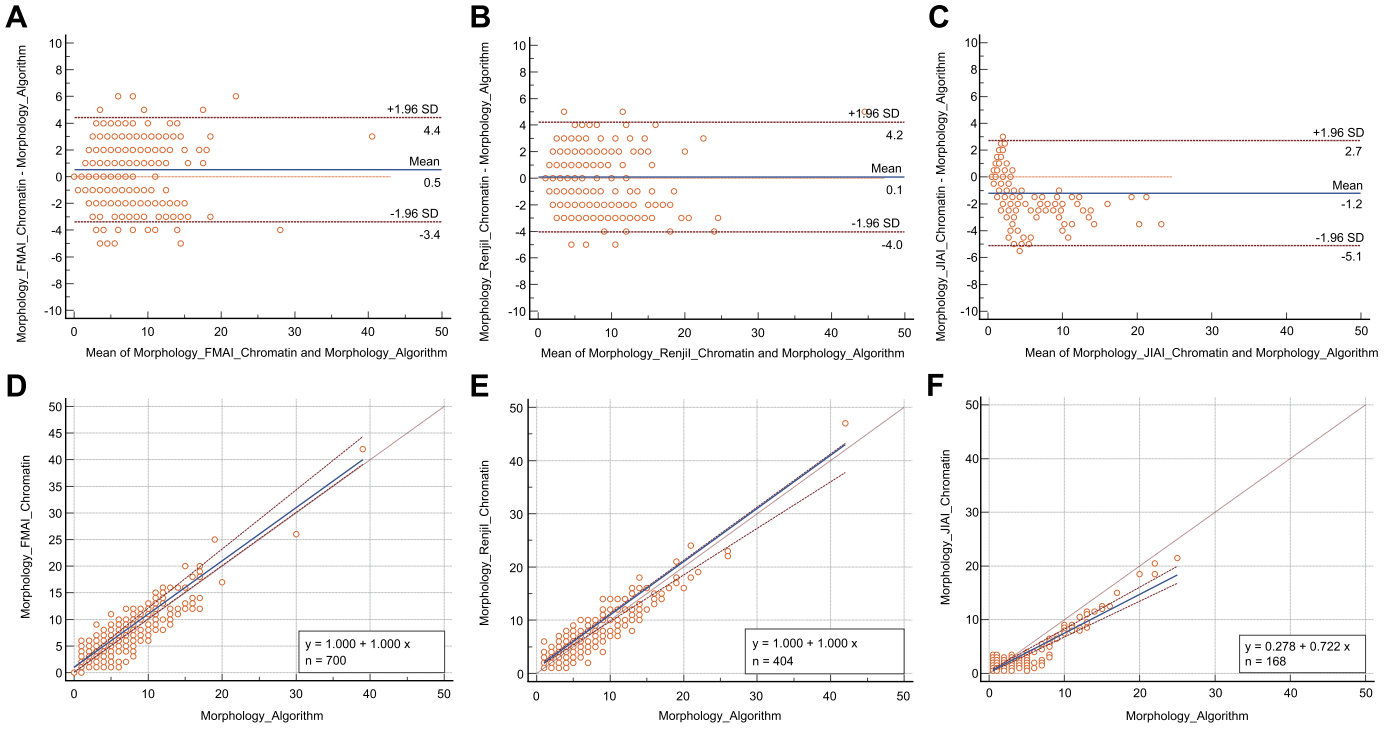

Fig. 4. Comparison of stained morphology with algorithmic morphology using Bland-Altman and Passing-Bablok plots for the three hospitals. (A-C) Bland-Altman plots for morphological results detected at the three hospitals (One Woman Infant, Renji, and Ji'ai, respectively). The solid line in the Bland-Altman plot indicates the mean difference between the two methods, and the dashed line indicates the 95% limits of agreement (1.96 standard deviations of the mean difference). (D-F) Passing-Bablok plots for the three hospitals. The solid line indicates the regression line and the dashed line indicates the regression line confidence intervals.

sperm. This non-invasive method can locate progressively motile and morphologically normal sperm for intracytoplasmic sperm injection to improve the success rate of second-generation IVF, with far-reaching implications for the future of assisted reproductive medicine.

Compared with existing methods, the innovations of this study are notable in the following three areas. (1) The algorithm can directly analyze live sperm from liquefied semen smears without the need for staining, which significantly reduces the workload of laboratory technicians and enables multidimensional analysis of active sperm with different morphologies, resulting in more accurate and objective results for the assessment of sperm morphology. (2) We propose a new method for sperm morphology and multi-sperm tracking, which improves the accuracy of sperm morphology assessment and enables precise tracking of each sperm, even in cases where the heads intersect, thereby providing a basis for multi-frame multidimensional analysis of sperm morphology. (3) The algorithm can dynamically monitor sperm and accurately identify all progressively motile and morphologically normal sperm in the field of view, addressing the limitations of existing methods in accurately calculating minority groups of non-existent progressive motility and morphologically normal sperm, and accurately selecting high-quality sperm with normal morphology and excellent motility for intracytoplasmic sperm injection (ICSI).

Throughout our study, we strictly followed the manufacturer's recommendations and WHO manual guidelines for semen analysis. As a result, near-perfect experimental results were obtained. Our system demonstrated high consistency with the stained morphology obtained from several hospitals that operated manually, based on the WHO manual. Morphological results obtained from these hospitals were stained using a manual microscope. The slope of the regression line of the Passing-Bablok plot between stained and unstained morphology at Shanghai First Maternity and Infant Hospital and Renji Hospital was 1, indicating that the results of the system's analysis of unstained morphology were consistent with those of the WHO recommendations, and the unstained morphology could be used for clinical semen analysis.

This study had some limitations. All the semen samples used in this study were obtained from Shanghai. Because of the possible regional differences in semen parameters, further collaborative studies should be performed in hospitals across China to validate our algorithm.

In conclusion, this study achieved automated detection of progressively motile and morphologically normal sperm, which is essential for the subsequent selection of morphologically normal and motile sperm for ICSI. This technology holds promise for streamlining and enhancing the process of sperm analysis. We reached an agreement with our partners to jointly develop this technology into a fully automated semen analysis device that integrates the assessment of concentration, motility, and morphology. This significant improvement will boost the precision and consistency of analyses, aiding physicians in making more scientifically informed diagnostic decisions. Additionally, this technology will reduce the procedural steps required by andrology laboratory physicians during semen analysis, offering quicker and more efficient feedback to those being tested.

CRediT authorship contribution statement

Hao Yang: Project administration, Resources, Supervision. **Mengmeng Ma:** Project administration, Resources, Supervision. **Xiangfeng Chen:** Project administration, Resources, Supervision. **Guowu Chen:** Project administration, Resources, Supervision. **Yi Shen:** Formal analysis, Software. **Lijun Zhao:** Formal analysis, Software. **Jianfeng Wang:** Formal analysis, Software. **Feifei Yan:** Data curation, Investigation, Validation. **Difeng Huang:** Formal analysis, Software. **Huijie Gao:** Data curation, Investigation, Validation. **Hao Jiang:** Formal analysis, Software. **Yuqian Zheng:** Writing – original draft, Writing – review & editing. **Yu Wang:** Data curation, Investigation, Validation. **Qian Xiao:** Data curation, Investigation, Validation. **Ying Chen:** Data curation, Investigation, Validation. **Jian Zhou:** Data curation, Investigation, Validation. **Jie Shi:** Investigation, Validation. **Yi Guo:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Bo**

Liang: Conceptualization, Methodology. **Xiaoming Teng:** Conceptualization, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Yi Guo reports financial support was provided by National Natural Science Foundation of China. Xiaoming Teng reports financial support was provided by Joint Research Project on Emerging Frontier Technologies for Municipal Hospitals. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Since the system will be commercialized, we only disclose a part of the sperm morphology identification data and the statistical results of sperm morphology, sperm forward motion and the percentage of normal morphology, all labeled by collaborating physicians from three hospitals, open at https://jz.basecare.cn/ for interested personnel to study. For those who require an account and password, please get in touch with the corresponding author to obtain the necessary credentials. We hope that more experts in related fields will join in to calibrate these unstained sperm morphology results to help our system to obtain more accurate results. In addition, we have established a dedicated web server at https://lw.basecare.cn/, offering a platform for live demonstrations of our model. Users can input sperm videos to obtain results related to sperm morphology and motility.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (no. 82271632) and the Joint Research Project on Emerging Frontier Technologies for Municipal Hospitals (no. SHDC12021115).

References

[1] Alegre E, Biehl M, Petkov N, Sánchez L. Automatic classification of the acrosome status of boar spermatozoa using digital image processing and lvq. Comput Biol Med 2008;38:461–8.
[2] Bablok W, Passing H. Comparison of several regression procedures for method comparison studies and determination of sample sizes. J Clin Chem Clin Biochem 1984;22:431–45. Publisher: Citeseer.
[3] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–95. Publisher: IEEE.
[4] Bhargavi K, Jyothi S. A survey on threshold based segmentation technique in image processing. Int J Innov Res Dev 2014;3:234–9.
[5] Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327:307–10. Publisher: Elsevier.
[6] Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995;346:1085–7. Publisher: Elsevier.
[7] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60. Publisher: Sage Publications Sage CA: Thousand Oaks, CA.
[8] Boulet SL, Mehta A, Kissin DM, Warner L, Kawwass JF, Jamieson DJ. Trends in use of and reproductive outcomes associated with intracytoplasmic sperm injection. JAMA 2015;313:255–63. Publisher: American Medical Association.
[9] Butola A, Popova D, Prasad DK, Ahmad A, Habib A, Tinguely JC, et al. High spatially sensitive quantitative phase imaging assisted with deep neural network for classification of human spermatozoa under stressed condition. Sci Rep 2020;10:13118.
[10] Chang V, Garcia A, Hitschfeld N, Härtel S. Gold-standard for computer-assisted morphological sperm analysis. Comput Biol Med 2017;83:143–50. Publisher: Elsevier.
[11] Chen H, Sun K, Tian Z, Shen C, Huang Y, Yan Y. Blendmask: top-down meets bottom-up for instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 8573–81.
[12] Dobrovolny M, Benes J, Langer J, Krejcar O, Selamat A. Study on sperm-cell detection using yolov5 architecture with labaled dataset. Genes 2023;14:451.
[13] Esteves SC, Roque M, Bedoschi G, Haahr T, Humaidan P. Intracytoplasmic sperm injection for male infertility and consequences for offspring. Nat Rev Urol 2018;15:535–62. Publisher: Nature Publishing Group UK London.
[14] Fraczek A, Karwowska G, Miler M, Lis J, Jezierska A, Mazur-Milecka M. Sperm segmentation and abnormalities detection during the icsi procedure using machine learning algorithms. In: 2022 15th international conference on human system interaction (HSI). IEEE; 2022. p. 1–6.
[15] Ghasemian F, Mirroshandel SA, Monji-Azad S, Azarnia M, Zahiri Z. An efficient method for automatic morphological abnormality detection from human sperm images. Comput Methods Programs Biomed 2015;122:409–20. Publisher: Elsevier.
[16] Javadi S, Mirroshandel SA. A novel deep learning method for automatic assessment of human sperm images. Comput Biol Med 2019;109:182–94. Publisher: Elsevier.
[17] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60:84–90. Publisher: ACM New York, NY, USA.
[18] Maree L, Du Plessis S, Menkveld R, Van der Horst G. Morphometric dimensions of the human sperm head depend on the staining method used. Hum Reprod 2010;25:1369–82. Publisher: Oxford University Press.
[19] Organization WH, et al. WHO laboratory manual for the examination and processing of human semen publisher. World Health Organization; 2021.
[20] Pal NR, Pal SK. A review on image segmentation techniques. Pattern Recognit 1993;26:1277–94. Publisher: Elsevier.
[21] Riordon J, McCallum C, Sinton D. Deep learning for the classification of human sperm. Comput Biol Med 2019;111:103342. Publisher: Elsevier.
[22] Sato T, Kishi H, Murakata S, Hayashi Y, Hattori T, Nakazawa S, et al. A new deep-learning model using yolov3 to support sperm selection during intracytoplasmic sperm injection procedure. Reprod Med Biol 2022;21:e12454.
[23] Senthilkumaran N, Rajesh R. Image segmentation-a survey of soft computing approaches. In: 2009 international conference on advances in recent technologies in communication and computing. IEEE; 2009. p. 844–6.
[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint. arXiv:1409.1556, 2014.
[25] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019. p. 6105–14.
[26] Wang X, Chen MX, Zhang F, Liang GQ, Zhu H, Feng BL, et al. Sperm donors in Shanghai, China: a study of motivations, characteristics, and semen parameters of actual sperm donors. Reprod Dev Med 2021;5:213–9. Publisher: Chinese Medical Journals Publishing House Co., Ltd. 42 Dongsi Xidajie ....
[27] Zhang Y. Animal sperm morphology analysis system based on computer vision. In: 2017 eighth international conference on intelligent control and information processing (ICICIP). IEEE; 2017. p. 338–41.
[28] Zhang Y, Wang C, Wang X, Zeng W, Liu W. Fairmot: on the fairness of detection and re-identification in multiple object tracking. Int J Comput Vis 2021;129:3069–87. Publisher: Springer.
[29] Zhou X, Wang D, Krähenbühl P. Objects as points. arXiv preprint. arXiv:1904.07850, 2019.