SPEHEATAL: A Cluster-Enhanced Segmentation Method for Sperm Morphology Analysis

Yi Shi[1,2*†], **Yun-Kai Wang**[1,2*], **Xu-Peng Tian**[1,2], **Tie-Yi Zhang**[1,2], **Bing Yao**[3,4], **Hui Wang**[3,4],
Yong Shao[3,4], **Cen-Cen Wang**[3,4], **Rong Zeng**[3,4]

[1] School of Artificial Intelligence, Nanjing University, Nanjing, China
[2] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[3] Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China
[4] Jiangsu Provincial Medical Key Discipline Cultivation Unit, Nanjing, China
shiy@lamda.nju.edu.cn

Abstract

The accurate assessment of sperm morphology is crucial in andrological diagnostics, where the segmentation of sperm images presents significant challenges. Existing approaches frequently rely on large annotated datasets and often struggle with the segmentation of overlapping sperm and the presence of dye impurities. To address these challenges, this paper first analyzes the issue of overlapping sperm tails from a geometric perspective and introduces a novel clustering algorithm, CON2DIS, which effectively segments overlapping tails by considering three essential factors: **Con**nectivity, **Con**formity, and **Dis**tance. Building on this foundation, we propose an unsupervised method, SPEHEATAL, designed for the comprehensive segmentation of the **SPE**rm **HEA**d and **TA**iL. SPEHEATAL employs the Segment Anything Model (SAM) to generate masks for sperm heads while filtering out dye impurities, utilizes CON2DIS to segment tails, and then applies a tailored mask splicing technique to produce complete sperm masks. Experimental results underscore the superior performance of SPEHEATAL, particularly in handling images with overlapping sperm.

Introduction

The analysis of sperm morphology constitutes a critical component within the framework of andrological diagnostics, serving as a key indicator of sperm viability and male reproductive capacity (Moretti et al. 2022; Menkveld, Holleboom, and Rhemrev 2011; Oehninger and Kruger 2021; Gatimel et al. 2017). Traditionally, the evaluation of sperm morphology has relied heavily on the expertise of andrologists, who conduct detailed examinations of semen samples under microscopy to distinguish between normal and abnormal sperm structures. This conventional approach not only imposes significant time demands on medical professionals but also fails to meet the growing need for high-throughput, efficient analysis. Compounded by the inherent subjectivity and the absence of uniform evaluative benchmarks, sperm morphology assessment emerges as a particularly challenging domain within reproductive diagnostics. Consequently,

---

*These authors contributed equally.

†Corresponding author.

(a) Raw image

(b) Labeled by experts

(c) Segmented by SAM
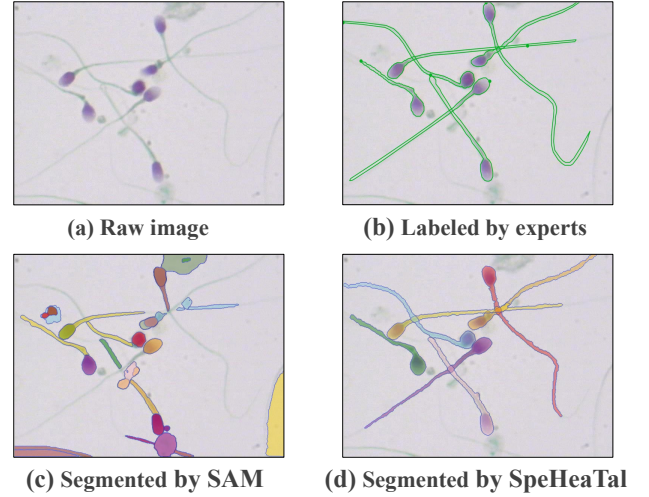
(d) Segmented by SpeHeaTal

Figure 1: Significant sperm overlap and dye impurities are evident in clinical practice. Compared with previous methods, such as SAM, our proposed SPEHEATAL effectively addresses these two challenges, achieving complete segmentation of multiple independent sperm.

there is an increasing demand for automated detection systems capable of identifying sperm abnormalities. A primary challenge in this context is the accurate segmentation of individual spermatozoa within imaging, a task that is complicated by frequent sperm overlap, low contrast, and the presence of dye impurities in microscopic imagery. These factors exacerbate the segmentation challenge for current imaging models and complicate manual annotation, resulting in a marked scarcity of labeled sperm image datasets.

Recent advancements in image segmentation technologies have been widely applied across various domains(Fang et al. 2024; He et al. 2024), including medical imaging(Bi et al. 2024; Dai et al. 2024; Ma et al. 2024). In the specific context of sperm morphology segmentation, existing models generally require labeled data for training. However, due to the limited availability of labeled sperm image datasets, these models often simplify the objects to be segmented, such as by only segmenting sperm heads or ignoring over-

lapping sperm. The ability to accurately segment *overlapping* spermatozoa *without* relying on labeled data is recognized as a crucial step in applying artificial intelligence to sperm morphology assessment.

Sperm image segmentation, as a task of instance segmentation, requires the segmentation of individual sperm masks without the necessity of identifying semantic differences among them. While recent advancements have shown significant progress in unsupervised instance segmentation (Niu et al. 2023; Wang et al. 2023; Kirillov et al. 2023), their effectiveness is primarily confined to common instances in everyday scenes, with limited applicability to specialized targets such as sperm. For instance, our experimental results indicate that SAM (Kirillov et al. 2023) struggles to accurately identify the elongated structure of sperm tails, particularly in cases involving overlap and dye impurities, as depicted in Figure 1(c). Additionally, some approaches have explored the use of clustering methods, such as K-means (MacQueen et al. 1967) and Spectual Clutering (SC) (Shi and Malik 2000), for unsupervised instance segmentation, yet these methods also fail to adequately address the issue of overlap. While tubular structure segmentation (Hu 2022; Qi et al. 2023) can deal with sperm tails, these methods treat overlapping sperm as a single entity rather than segmenting each sperm individually.

In response to these challenges, we first conduct a geometric analysis of sperm tail overlap, concluding that effective distinction of tails requires consideration of three key factors: **Con**nectivity, **Con**formity, and **Dis**tance. Based on these insights, we propose a novel clustering method, CON2DIS, designed to achieve effective segmentation of sperm tails, with particular efficacy in overlapping regions. To the best of our knowledge, CON2DIS is the first clustering algorithm explicitly developed to address the segmentation challenges posed by intersecting slender structures in sperm imagery. This approach may also offer valuable insights for the segmentation of images with similar structural characteristics, such as blood vessels and neural pathways.

Building on the foundation of CON2DIS, we present SPEHEATAL, an end-to-end segmentation algorithm designed to accurately segment complete sperm. SPEHEATAL functions through a "decomposition-combination" approach, dividing the sperm segmentation task into distinct head and tail segmentation processes, which are subsequently integrated. Specifically, SPEHEATAL employs SAM to generate masks for sperm heads while filtering out dye impurities, leverages CON2DIS to segment tails, and finally applies a tailored mask splicing technique to produce complete sperm masks. As an unsupervised algorithm, SPEHEATAL effectively mitigates the challenges posed by sperm overlap and dye impurities, delivering multiple independent and complete sperm masks without the need for any annotations.

Given that existing sperm datasets do not adequately capture the characteristics of sperm overlap and dye impurities observed in clinical practice, we collaborated with several prominent hospitals to compile a dataset comprising approximately 2,000 unlabeled sperm images for model calibration, along with an additional 240 expert-annotated images for rigorous model evaluation, as depicted in Figure 1(b).

Experimental results demonstrate the superior performance of SPEHEATAL, particularly in the context of overlapping sperm, as illustrated in Figure 1(d). This advancement not only lays a foundation for future work in sperm morphology classification but also signals a significant step towards fully automated sperm morphology analysis using AI methodologies. Our contributions are summarized as follows:

- We introduce a novel clustering method, CON2DIS, which effectively segments overlapping sperm tails by considering Connectivity, Conformity, and Distance.

- We propose SPEHEATAL, a method capable of achieving effective and complete segmentation of overlapping sperm images under unsupervised conditions.

- Experimental results on our clinical sperm dataset highlight the effectiveness of SPEHEATAL, especially in images with overlapping structures.

Related Work

Sperm image segmentation is a pivotal task in automating sperm morphology analysis. Existing studies, such as (Lv et al. 2022; Iqbal, Mustafa, and Ma 2020), primarily focus on the segmentation of sperm heads, often neglecting the tails. In contrast, research efforts in (Ilhan et al. 2020; Fraczek et al. 2022) have concentrated on segmenting independent, non-overlapping sperm. Other approaches, such as those in (Kheirkhah, Mohammadi, and Shahverdi 2019; Lewandowska et al. 2023), treat overlapping sperm as a single entity, thereby neglecting the need to segment individual spermatozoa. These methodologies often simplify the segmentation task, leading to suboptimal results when applied to real-world clinical specimens. Moreover, these approaches typically require labeled datasets for training, which are limited in availability.

Unsupervised instance segmentation has seen significant advancements in recent years. The Segment Anything Model (SAM) (Kirillov et al. 2023) demonstrates high-quality segmentation across a wide range of objects using diverse prompts without requiring specific fine-tuning. U2Seg (Niu et al. 2023) employs a hierarchical feature extraction framework to capture multi-scale information, thereby improving segmentation accuracy. Similarly, CutLER (Wang et al. 2023) utilizes contrastive learning and local feature representations to distinguish individual instances. However, these methods encounter challenges when applied to medical imaging due to variations in color spectrum, texture, and contrast specific to this domain. Their performance is particularly inadequate when addressing sperm overlap and dye impurities.

Clustering methods for segmentation have been widely integrated with advanced techniques to enhance image segmentation tasks. K-Means (MacQueen et al. 1967) optimizes segmentation by iteratively minimizing within-cluster variance, primarily focusing on the distance between pixels. Agglomerative Hierarchical Clustering (AHC) (Sneath, Sokal et al. 1973) constructs a hierarchy by iteratively merging
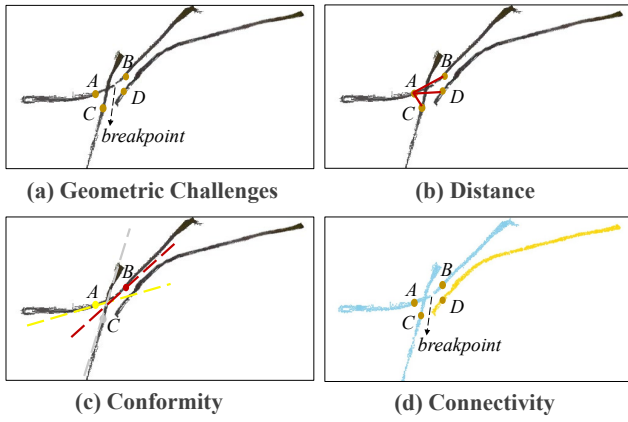
Figure 2: (a) A pre-processed sperm tail image. The goal is to ensure that point A is similar to point B on the same tail, but not similar to points C and D on different tails. (b) Considering only distance, if B is similar to A, then C and D are also similar to A. (c) Considering conformity, point B is more similar to point A than point C based on the similarity of tangent angles. (d) Considering connectivity, among B, C, and D, only D is not in the same connected interval as A. Notably, small breakpoints can interfere with connectivity.

the closest pairs of clusters from bottom to up, primarily focusing on distance and connectivity. Some clustering methods incorporate both distance and conformity, such as Local Subspace Affinity (LSA) (Yan and Pollefeys 2006) and Spectral Clustering (SC) (Shi and Malik 2000). LSA computes an affinity between pairs of points based on principal subspace angles across different linear manifolds, while SC leverages the eigenvalues of a similarity matrix to reduce dimensionality before clustering in the reduced space. Despite their strengths, these methods often overlook the simultaneous consideration of all three crucial aspects—distance, conformity, and connectivity—limiting their effectiveness in segmenting overlapped sperm tails.

Method

Sperm image segmentation presents two principal challenges. The first challenge arises from the scarcity of labeled samples, often resulting in the undertraining of supervised learning models. The second challenge is the frequent occurrence of tail overlap and dye impurities, which complicates the independent segmentation of sperm and subsequently degrades the overall segmentation performance. In this section, we first examine how CON2DIS effectively addresses the difficulty of overlapping tail segmentation. We then introduce the other components of our unsupervised method, SPEHEATAL, detailing how SAM is utilized for head segmentation, how dye interference is mitigated, and how the head and tail segments are subsequently spliced together.

Tail Segmentation with CON2DIS

Geometric Challenges in Clustering. Viewing all sperm tail pixels as independent scattered points, clustering methods can be employed to group these points into differ-

ent clusters, each representing an independent tail, thus achieving unsupervised instance segmentation of overlapping sperm tails. The clustering objective is to maximize intra-cluster similarity, ensuring that data points from the same sperm tail are grouped together, while minimizing inter-cluster similarity to segregate points from different tails. This section explores the complexities associated with achieving this objective. Referring to Figure 2(a), with point A as the reference, the aim is to increase the similarity between points A and B while reducing the similarity between point A and points C and D. Some clustering methods, such as K-Means, rely solely on Euclidean distance as the measure of similarity between points, which can lead to problems. As illustrated in Figure 2(b), if point B is considered similar to point A based on distance, then points C and D would also be similar to point A. Manifold clustering methods, such as LSA (Yan and Pollefeys 2006), consider additional factors like conformity, such as the angle of the tangent line at the point. As shown in Figure 2(c), based on this consideration, point B is closer to point A than point C. However, for a sperm tail with variable width and uneven thickness, calculating tangent angles is challenging, particularly in overlapping regions. To exclude point D from being similar to point A, connectivity must also be considered. As shown in Figure 2(d), different connected components are marked with different colors, allowing point D to be excluded from A's similar points. However, sperm images often contain breakpoints in the tail due to noise or impurities, which can interfere with connectivity-based judgments, leading to misclassification.

In summary, a clustering method that simultaneously considers *distance*, *conformity*, and *connectivity* is required for effective segmentation of overlapping tails. The assessment of conformity and connectivity needs to be specially designed to handle the unique characteristics of sperm tails.

Clustering Methodology. For two sample points x_i and x_j in sperm tails, Our clustering method CON2DIS defines *distance* p_{ij} as:

$$p_{ij} = \begin{cases} 1 & \text{if } x_i \in \text{Knn}(x_j) \text{ or } x_j \in \text{Knn}(x_i) \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $\text{Knn}(x)$ denotes K nearest neighbors of x.
The *connectivity* r_{ij} is defined as:

$$r_{ij} = \begin{cases} 1 & \text{if } x_i, x_j \text{ are in the same connected component} \\ 0 & \text{otherwise} \end{cases}. \quad (2)$$

To mitigate the interference caused by breakpoints, SPEHEATAL employs the DBSCAN (Ester et al. 1996) algorithm for connected component filtration, setting appropriate thresholds to prevent excessive segmentation while ensuring distinct tails are properly separated, as shown in Figure 2(d).

In general, the calculation of conformity relies on the construction of an undirected graph or affinity matrix. However, clusters near the intersection point can easily connect within the undirected graph, potentially corrupting the affinity matrix with poor pairwise affinity values, propagating mislead-
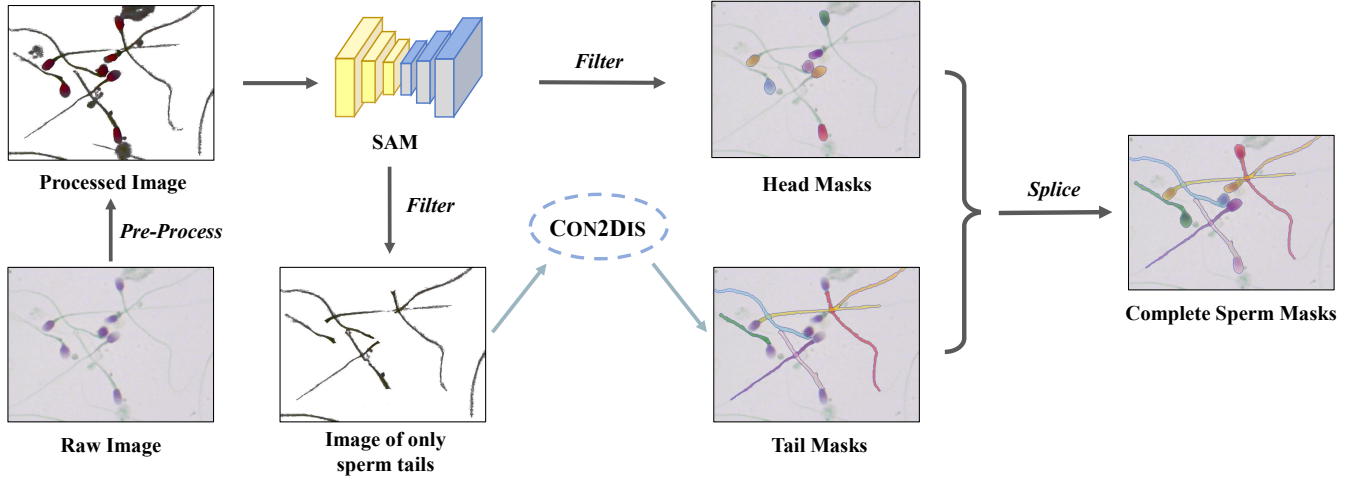
Figure 3: The pipeline of our SPEHEATAL. SPEHEATAL works in a "decomposition-combination" manner, utilizing SAM and CON2DIS for segmenting heads and tails, respectively, and subsequently splicing them into complete masks.

ing information among different clusters. Inspired by Spectral Multi-Manifold Clustering method (SMMC) (Wang et al. 2011), the *conformity* q_{ij} is defined as:

$$q_{ij} = q(\Theta_i, \Theta_j) = \left(\prod_{l=1}^{d} \cos(\theta_l)\right)^o , \quad (3)$$

where Θ_i is the tangent space at x_i, and $o \in \mathrm{N}^+$ is an adjustable parameter. $0 \leq \theta_1 \leq, \cdots, \leq \theta_d \leq \pi/2$ are a series of principal angles (Golub and Van Loan 1996) between two tangent spaces Θ_i and Θ_j, defined recursively as:

$$\cos(\theta_1) = \max_{\substack{u_1 \in \Theta_i, v_1 \in \Theta_j \\ \|u_1\| = \|v_1\| = 1}} u_1^T v_1 , \quad (4)$$

and

$$\cos(\theta_l) = \max_{\substack{u_l \in \Theta_i, v_l \in \Theta_j \\ \|u_l\| = \|v_l\| = 1}} u_l^T v_l, \quad l = 2, \cdots, d . \quad (5)$$

The tangent vectors u_i and v_i, $i = 1, \cdots, l$, are derived by training mixtures of probabilistic principal component analyzers using the EM (Archambeau, Delannay, and Verleysen 2008) algorithm. Further details are provided in the supp. These multiple local tangent subspaces enhance the reliability of affinity values between pairwise points, particularly those located in overlapping areas of sperm tails.

The amalgamation of Eq. 1, Eq. 2, and Eq. 3 yields the similarity metric w_{ij} defined as:

$$w_{ij} = p_{ij} \cdot q_{ij} \cdot r_{ij} . \quad (6)$$

This metric satisfies the prerequisite clustering criteria, positing that a substantial w_{ij} value is indicative of x_i and x_j being in close distance, possessing analogous tangent angles, and residing within the same connected component. With Eq. 6, the affinity matrix W is calculated as $W_{ij} = w_{ij}$ and the diagonal matrix E is defined as $E_{ii} = \sum_j w_{ij}$. The clustering outcome is obtained by solving the first k generalized eigenvectors $u_1, \cdots, u_k$ and the corresponding k smallest eigenvalues of the generalized eigenproblem:

$$(E - W)u = \lambda E u . \quad (7)$$

Finally, the K-means method is applied on the row vectors $U = [u_1, \cdots, u_k]$ to obtain the clusters.

Skeletonization and Restoration of Tails. The calculation of *conformity* q_{ij} is complicated by variations in sperm tail thickness and noise interference. To address this, we apply skeletonization, reducing sperm tails to single-pixel-width lines, thereby improving the precision of angular assessments. Therefore, the clustering process produces a skeletal representation of the sperm tails, requiring a subsequent restoration phase to reconstruct the tails to their full form. This involves mapping each original tail point x_i to its nearest skeletal counterpart x_j, facilitated by the distance similarity p'_{ij}, calculated using Euclidean metrics as follows:

$$p'_{ij} = \|x_i - x_j\|^2 . \quad (8)$$

A threshold γ determines the allocation of point x_i to the cluster associated with point x_j if $p'_{ij} < \gamma$. It is important to note that overlapping sperm tails may result in a single point being classified into multiple tails, accurately reflecting the biological complexity of sperm samples. Additionally, skeletonization can simplify the computation of the distance p_{ij} and the connectivity r_{ij}, while also facilitating the mask splicing of sperm heads and tails, as detailed in the following section.

Sperm Segmentation with SPEHEATAL

Following the segmentation of overlapping sperm tails using CON2DIS, we propose SPEHEATAL for the comprehensive segmentation of complete sperm. As illustrated in Figure 3, SPEHEATAL employs a "decomposition-combination" approach, comprising the following three main steps:

1. The Initial segmentation of pre-processed images using SAM to obtain masks for different sperm heads;

2. The segmentation of sperm tails with significant overlap using our clustering algorithm CON2DIS;

3. The combination of head and tail masks with a tailored mask splicing approach to produce complete masks.

Initial segmentation with SAM. In sperm images, sperm heads generally appear as purple ovals, sperm tails as green tubular structures, and dye impurities as green irregular blocks. We utilize SAM for preliminary segmentation of sperm images, distinguishing these components based on color and shape. To enhance SAM's segmentation efficiency, we initiate a pre-processing phase for the input images, which includes adjustments to brightness, contrast, saturation, and resolution, along with edge sharpening and background whitening. SAM is then applied to these pre-processed images to generate various segmentation masks.

We use shape and color indices, denoted as SI and CI respectively, to classify the mask types. For each mask m segmented by SAM, we calculate its true area $S(m)$ and the area of its smallest circumscribed circle $S_{scc}(m)$. The shape index SI is defined by their ratio:

$$SI(m) = \frac{S(m)}{S_{scc}(m)} . \qquad (9)$$

A larger $SI(m)$ indicates a clump-shaped mask, while a smaller $SI(m)$ suggests an elongated tubular shape. We set a threshold α ($0 < \alpha < 1$) for shape filtering. If $SI(m) < \alpha$, the mask m is considered as an elongated tubular, indicative of a sperm tail; otherwise, it is deemed closer to a clump shape, suggestive of either a dye block or a sperm head.

Additionally, we use $S_{pa}(m)$ to denote the "purple area" within the mask m, where "purple area" corresponds to the region in the raw image classified as purple by the RGB or HSV value. The color index $CI(m)$ is defined as:

$$CI(m) = \frac{S_{pa}(m)}{S(m)} . \qquad (10)$$

In raw sperm images, only sperm heads are dyed purple, while the rest is dyed green. We set another threshold β ($0 < \beta < 1$) for color filtering, where a larger β value indicates a greater presence of purple areas in the mask. The complete filtering method is defined as follows:

$$m = \begin{cases} \text{sperm head} & \text{if } SI(m) > \alpha \text{ and } CI(m) > \beta \\ \text{dye block} & \text{if } SI(m) > \alpha \text{ and } CI(m) < \beta \\ \text{sperm tail} & \text{if } SI(m) \leq \alpha \end{cases} .$$

$$(11)$$

With Eq. 11, we effectively categorize all masks into three distinct types: heads, tails, and dye impurities. Following this classification, we isolate and store all masks identified as sperm heads separately. Subsequently, we remove the sperm heads and dye blocks from the *raw* images, leaving *only* the sperm tails as the foreground. Notably, sperm tails often overlap and intersect, a scenario where SAM typically struggles, often identifying these overlapping sections as a single entity rather than distinguishing them as separate structures. Given these limitations, we confine SAM's segmentation capabilities to sperm heads.

Tail Segmentation with CON2DIS**.** We then apply CON2DIS to further segment these images, which contain only overlapping tails, to obtain multiple independent masks for different sperm tails. This segmentation process may result in some sperm tails being fragmented into multiple sections due to occlusions by larger entities in raw images, such as heads or dye blocks, necessitating the subsequent detection and splicing of the masks corresponding to the same sperm tail. Fortunately, this is addressed in the head and tail splicing step.

Splicing Masks of Heads and Tails. In the previous steps, SAM is employed to segment sperm heads, followed by CON2DIS for sperm tail segmentation. Finally, we implement a speicialized mask splicing approach to combine the corresponding head and tail masks associated with the same sperm. We fit each head mask into an ellipse and extract the line segment where the major axis of the ellipse lies, while the tail lines are obtained by skeletonization in CON2DIS. The endpoints of these skeletonized tails and the major axes of the heads are identified, with particular emphasis on the terminal slopes at these endpoints. Our approach is based on the hypothesis that a minimal distance between endpoints and a similarity in terminal slopes suggest a higher likelihood of the endpoints belonging to the same sperm entity. For any two distinct endpoints x_i and x_j that do not originate from the same line, we first evaluate their distance similarity p'_{ij} using Eq. 8. We then assess their angular similarity q'_{ij}, derived from the tangent slopes at these endpoints, as follows:

$$q'_{ij} = |s(x_i) - s(x_j)| , \qquad (12)$$

where $s(\cdot)$ denotes the slope value, and $|\cdot|$ denotes the absolute value. By instituting distance threshold λ_1 and angle threshold λ_2, we proceed to connect x_i and x_j with a linear path, if $p'_{ij} < \lambda_1$ and $q'_{ij} < \lambda_2$ simultaneously. If multiple endpoints meet this criterion for a single anchor endpoint, the endpoint closest in *angle* is preferred. In addition to head-tail splicing, our mask splicing approach also matches endpoints from different tails, as sperm tails may be broken into several segments due to various disturbances. Successful matching results in the assembly of complete sperm masks, effectively integrating the head and tail components.

Experiment

We evaluate the effectiveness of our proposed method, SPE-HEATAL, on a sperm dataset obtained from clinical practice. The visual results demonstrate that SPEHEATAL provides superior segmentation of overlapping sperm tails compared to existing methods. Additional experimental settings and results are provided in the supplementary materials.

Experimental Setups

Datasets. The sperm datasets used in previous works (Lv et al. 2022; Meléndez et al. 2021; Iqbal, Mustafa, and Ma 2020) predominantly feature images with independent sperm, which do not adequately represent the cross-overlapping phenomena commonly encountered in clinical practice. To address this gap, we collaborated with several leading hospitals to obtain approximately 2,000 unlabeled sperm images from clinical settings for model parameter

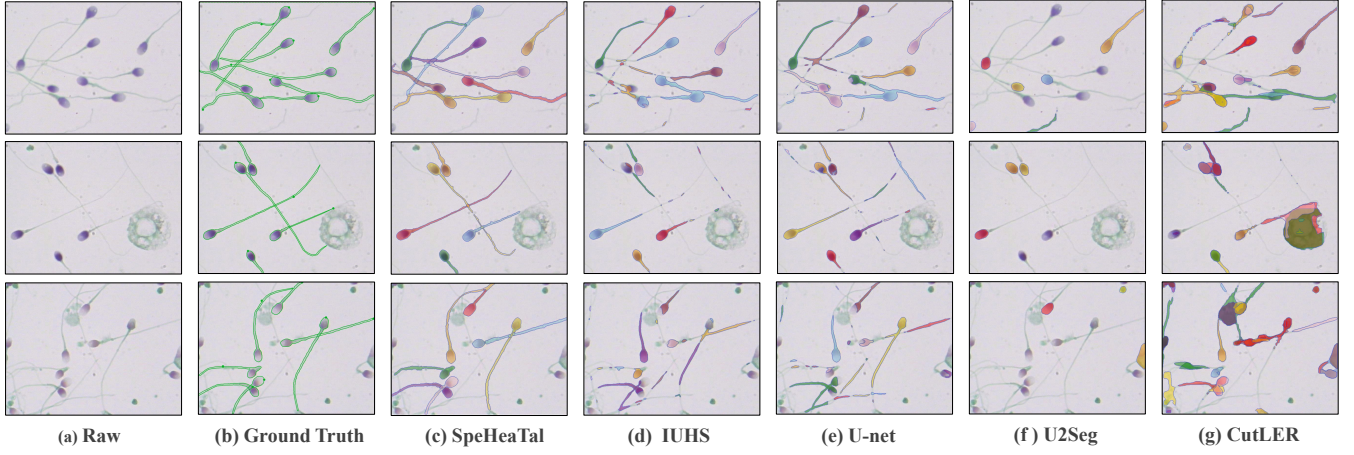| (a) Raw | (b) Ground Truth | (c) SpeHeaTal | (d) IUHS | (e) U-net | (f) U2Seg | (g) CutLER |

Figure 4: Visualization of segmentation results obtained by different methods. More results can be found in the supp.

adjustment. Additionally, 240 sperm images were annotated by andrology experts, with each annotation verified and approved by at least three senior experts to ensure accuracy for model performance testing. The original images are all 720×540 pixels in size. For those models requiring labeled samples for training, we divided the 240 labeled images into a 3:1 ratio, with 180 images used for their training and the remaining 60 images reserved for performance evaluation.

Evaluation Metrics. Following the methodologies outlined in (Müller, Rey, and Kramer 2022; Costa et al. 2019; Baccouche et al. 2021; Shamir et al. 2019), we employ mean Intersection over Union (mIoU) and mean Dice index (mDice), two standard metrics in instance segmentation, to evaluate model performance. Both mIoU and mDice assess performance by measuring the degree of overlap between the ground truth areas ($A_i, i = 1, \cdots, N$) and the predicted areas ($B_i, i = 1, \cdots, N$). A higher degree of overlap corresponds to a larger index value and better model performance. The specific definitions are as follows:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{|A_i \cap B_i|}{|A_i \cup B_i|} \right), \quad (13)$$

$$\text{mDice} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{2 \times |A_i \cap B_i|}{|A_i| + |B_i|} \right). \quad (14)$$

Prior to calculating mIoU and mDice, we establish the optimal pairing relationship by comparing the IoU ratios between each instance in Ground Truth and the predicted results. Notably, the segmentation outcomes of Ground Truth and certain methods may not correspond on a one-to-one basis, potentially leading to a disparity in quantities. In such cases, we apply a filtering criterion based on selecting pairs with an optimal pairing IoU ratio for calculation.

Comparison Methods. We categorize our comparison methods into two groups. The first group consists of supervised learning methods based on frameworks such as U-Net (Ronneberger et al. 2015). Previous sperm segmentation methods fall into this category, including IUHS (Lv

Criteria	mIOU	mDice
U-net (Ronneberger et al. 2015)	49.21	62.17
IUHS (Lv et al. 2022)	47.56	58.46
CN2UA (Meléndez et al. 2021)	19.26	32.94
SAM (Kirillov et al. 2023)	34.17	46.75
U2Seg (Niu et al. 2023)	16.44	28.86
CutLER (Wang et al. 2023)	31.21	44.17
SPEHEATAL	**72.21**	**81.07**

Table 1: mIOU and mDice of different segmentation methods on our sperm dataset. The best results are in bold.

et al. 2022) and CN2UA (Meléndez et al. 2021). Since these methods do not provide source codes, we reproduced them based on the original papers. The second group comprises unsupervised instance segmentation methods, including SAM (Kirillov et al. 2023), U2Seg (Niu et al. 2023), and CutLER (Wang et al. 2023), for which we directly used the authors' provided codes.

Implementation details. Our method utilizes the SAM model in "everything mode" to segment microscopic sperm images. The images are standardized to a resolution of 720×540 pixels to ensure consistency across the dataset. According to the parameter settings described in the methodology section, the shape filtering threshold $\alpha = 0.25$, color filtering threshold $\beta = 0.4$, and point allocation threshold $\gamma = 5$ pixels. The distance threshold $\lambda_1 = 30$ pixels and angle threshold $\lambda_2 = 35°$ are determined to effectively splice different parts of each sperm. Codes are available at *https://www.github.com/shiy19/SpeHeaTal*.

Performance Evaluation

The experimental results on our sperm dataset are presented in Table 1. All comparison methods and our proposed SPE-HEATAL are tested on pre-processed images. Supervised methods, such as U-Net, IUHS, and CN2UA, require a large amount of labeled data for training and exhibit limited effec-

Method	Distance	Conformity	Connectivity	mIOU	mDice
K-Means	✓	✗	✗	14.63	26.94
AHC	✓	✗	✓	27.76	38.17
LSA	✓	✓	✗	37.61	45.92
SC	✓	✓	✗	46.86	56.33
CON2DIS	✓	✓	✓	**72.21**	**81.07**

Table 2: mIOU and mDice of different clustering methods on our sperm dataset. The best results are in bold.

tiveness on sperm datasets lacking large-scale labeled data. Consequently, these methods often simplify the sperm segmentation task, rendering them ineffective on our real clinical sperm dataset. In contrast, unsupervised instance segmentation methods, including SAM, U2Seg, and CutLER, often rely on texture and structural patterns in images. However, sperm imagery significantly deviates from typical natural images, characterized by high contrast, minimal texture, and disruptions due to dye blocks. These unique attributes complicate the learning process for these models. Although some methods utilize robust data augmentation to induce variability, such enhancements risk obliterating critical features in sperm imagery, such as tail position and morphology, thus impairing model performance. Furthermore, the frequent presence of numerous elongated tails and their irregular distribution across the image exacerbates the difficulty of characterization. All comparison methods struggle with processing these overlapping structures.

In contrast, our method SPEHEATAL exhibits a remarkable ability to accurately segment sperm tails, even in scenarios with significant overlap and dye impurities, and effectively identifies masks for multiple independent sperm. This capability results in superior performance across the evaluated metrics. Visualizations of segmentation results obtained by different methods are shown in Figure 4.

Ablation Study

We further investigate the impact of using alternative clustering techniques, such as K-Means, Agglomerative Hierarchical Clustering (AHC), Local Subspace Affinity (LSA), and Spectral Clustering (SC), in place of CON2DIS. The experimental results are detailed in Table 2. All methods are applied solely to sperm tail segmentation, with SAM used for sperm head segmentation. Sperm tail imagery presents intricate geometric challenges, necessitating a multi-faceted approach to clustering for precise segmentation of overlapping sperms. This approach encompasses three critical perspectives: distance, conformity, and connectivity. The comparison methods inadequately address all these factors. While some methods consider certain aspects, such as SC considering conformity, they still fall short. For instance, the cross-overlap at the tail significantly disrupts the calculation of the affinity matrix in SC. Consequently, our tailored CON2DIS achieves the best performance.

We also conduct an additional ablation study to explore the impact of two hyperparameters, the distance threshold λ_1 and the angle threshold λ_2, in our mask splicing ap-
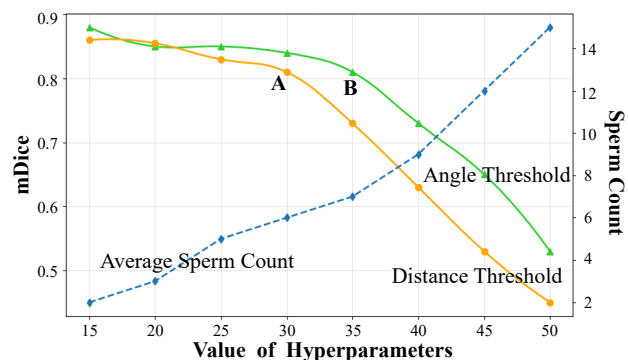

Figure 5: The impact of distance threshold λ_1 and angle threshold λ_2 on SPEHEATAL.

proach. The results are shown in Figure 5. We also visualize the average *sperm count* detected in each image for different *distance thresholds*. As λ_1 and λ_2 increase, the model relaxes the screening conditions, leading to more sperm with head-to-tail mismatches and incomplete head-to-tail matches being counted, resulting in a continued decline in mDice. Additionally, due to the looser restrictions, the number of successfully detected sperm in each image gradually increases. Based on the graph, we select points A and B, where the "elbow" inflection point of the mDice parameter curve appears, as appropriate hyperparameter values. These hyperparameter values represent an optimal balance between detection accuracy and detection completeness.

Conclusion

The challenges of tail overlap and dye impurities pose significant obstacles to the unsupervised segmentation of sperm images. In this work, we first conduct a geometric analysis, revealing that the effective segmentation of overlapping tails requires the simultaneous consideration of Conformity, Connectivity, and Distance. Leveraging these insights, we develop a clustering algorithm, CON2DIS, specifically tailored to address the segmentation of overlapping tails. Building on this foundation, we introduce SPEHEATAL, a comprehensive sperm segmentation method that employs a "decomposition-combination" approach. SPEHEATAL utilizes SAM for head segmentation and impurity removal, while CON2DIS is used for tail segmentation. These segments are then integrated into complete masks through a tailored mask splicing technique. Experimental results on an in-house clinical sperm dataset demonstrate that SPEHEATAL outperforms existing methods, particularly in images with significant overlap.

Limitations. SPEHEATAL encounters difficulties when processing images with extremely complex tail overlaps, particularly in scenarios where more than ten sperm are entangled—situations that even human experts struggle to resolve. Therefore, adequate dilution of semen samples remains necessary in clinical practice to ensure manageable analysis conditions.

Acknowledgments

References

Archambeau, C.; Delannay, N.; and Verleysen, M. 2008. Mixtures of robust probabilistic principal component analyzers. *Neurocomputing*, 71(7-9): 1274–1282.

Baccouche, A.; Garcia-Zapirain, B.; Castillo Olea, C.; and Elmaghraby, A. S. 2021. Connected-UNets: a deep learning architecture for breast mass segmentation. *NPJ Breast Cancer*, 7(1): 151.

Bi, Q.; Yi, J.; Zheng, H.; Ji, W.; Huang, Y.; Li, Y.; and Zheng, Y. 2024. Learning Generalized Medical Image Segmentation from Decoupled Feature Queries. In *AAAI*, 810–818.

Costa, M. G. F.; Campos, J. P. M.; de Aquino e Aquino, G.; de Albuquerque Pereira, W. C.; and Filho, C. F. F. C. 2019. Evaluating the performance of convolutional neural networks with direct acyclic graph architectures in automatic segmentation of breast lesion in US images. *BMC Medical Imaging*, 19(1): 38.

Dai, Q.; Wei, D.; Liu, H.; Sun, J.; Wang, L.; and Zheng, Y. 2024. Federated Modality-Specific Encoders and Multimodal Anchors for Personalized Brain Tumor Segmentation. In *AAAI*, 1445–1453.

Ester, M.; Kriegel, H.; Sander, J.; and Xu, X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD*, 226–231.

Fang, C.; Zhou, Z.; Chen, J.; Su, H.; Wu, Q.; and Li, G. 2024. Variance-Insensitive and Target-Preserving Mask Refinement for Interactive Image Segmentation. In *AAAI*, 1698–1706.

Fraczek, A.; Karwowska, G.; Miler, M.; Lis, J.; Jezierska, A.; and Mazur-Milecka, M. 2022. Sperm segmentation and abnormalities detection during the ICSI procedure using machine learning algorithms. In *HSI*, 1–6.

Gatimel, N.; Moreau, J.; Parinaud, J.; and Léandri, R. 2017. Sperm morphology: assessment, pathophysiology, clinical relevance, and state of the art in 2017. *Andrology*, 5(5): 845–862.

Golub, G. H.; and Van Loan, C. F. 1996. *Matrix Computations*. Baltimore, MD: John Hopkins University Press.

He, J.; Cheng, L.; Fang, C.; Feng, Z.; Mu, T.; and Song, M. 2024. Progressive Feature Self-Reinforcement for Weakly Supervised Semantic Segmentation. In *AAAI*, 2085–2093.

Hu, X. 2022. Structure-Aware Image Segmentation with Homotopy Warping. In *NeurIPS*, 24046–24059.

Ilhan, H. O.; Sigirci, I. O.; Serbes, G.; and Aydin, N. 2020. A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods. *Medical Biol. Eng. Comput.*, 58(5): 1047–1068.

Iqbal, I.; Mustafa, G.; and Ma, J. 2020. Deep Learning-Based Morphological Classification of Human Sperm Heads. *Diagnostics*, 10(5): 325.

Kheirkhah, F. M.; Mohammadi, H. R. S.; and Shahverdi, A. 2019. Efficient and robust segmentation and tracking of sperm cells in microscopic image sequences. *IET Comput. Vis.*, 13(5): 489–499.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.; Dollár, P.; and Girshick, R. B. 2023. Segment Anything. In *ICCV*, 3992–4003.

Lewandowska, E.; Wesierski, D.; Mazur-Milecka, M.; Lis, J.; and Jezierska, A. 2023. Ensembling noisy segmentation masks of blurred sperm images. *Comput. Biol. Medicine*, 166: 107520.

Lv, Q.; Yuan, X.; Qian, J.; Li, X.; Zhang, H.; and Zhan, S. 2022. An Improved U-Net for Human Sperm Head Segmentation. *Neural Process. Lett.*, 54(1): 537–557.

Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; and Wang, B. 2024. Segment anything in medical images. *Nature Communications*, 15(1): 654.

MacQueen, J.; et al. 1967. Some methods for classification and analysis of multivariate observations. *Berkeley Symposium*, 1: 281–297.

Meléndez, R.; Castañón, C. A. B.; Rodríguez, R. M.; and Rodríguez, R. M. 2021. Sperm Cell Segmentation in Digital Micrographs based on Convolutional Neural Networks using U-Net Architecture. In *CBMS*, 91–96.

Menkveld, R.; Holleboom, C. A.; and Rhemrev, J. P. 2011. Measurement and significance of sperm morphology. *Asian journal of andrology*, 13(1): 59.

Moretti, E.; Signorini, C.; Noto, D.; Corsaro, R.; and Collodel, G. 2022. The relevance of sperm morphology in male infertility. *Frontiers in Reproductive Health*, 4.

Müller, D.; Rey, I. S.; and Kramer, F. 2022. Towards a Guideline for Evaluation Metrics in Medical Image Segmentation. *CoRR*, abs/2202.05273.

Niu, D.; Wang, X.; Han, X.; Lian, L.; Herzig, R.; and Darrell, T. 2023. Unsupervised Universal Image Segmentation. *CoRR*, abs/2312.17243.

Oehninger, S.; and Kruger, T. F. 2021. Sperm morphology and its disorders in the context of infertility. *F&S Reviews*, 2(1): 75–92.

Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; and Yang, G. 2023. Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation. In *ICCV*, 6047–6056.

Ronneberger, O.; Fischer, P.; Brox, T.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, 234–241.

Shamir, R. R.; Duchin, Y.; Kim, J.; Sapiro, G.; and Harel, N. 2019. Continuous dice coefficient: a method for evaluating probabilistic segmentations. *CoRR*, abs/1906.11031.

Shi, J.; and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8): 888–905.

Sneath, P. H.; Sokal, R. R.; et al. 1973. *Numerical taxonomy. The principles and practice of numerical classification.* W. H. Freeman and Company.

Wang, X.; Girdhar, R.; Yu, S. X.; and Misra, I. 2023. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. In *CVPR*, 3124–3134.

Wang, Y.; Jiang, Y.; Wu, Y.; and Zhou, Z. 2011. Spectral Clustering on Multiple Manifolds. *IEEE Trans. Neural Networks*, 22(7): 1149–1161.

Yan, J.; and Pollefeys, M. 2006. A General Framework for Motion Segmentation: Independent, Articulated, Rigid, Non-rigid, Degenerate and Non-degenerate. In *ECCV*, 94–106.