

ORIGINAL RESEARCH ARTICLE

Dynamic U-Net: A lightweight hierarchical network for accurate and efficient sperm morphology segmentation

Qingxin Zhang¹, Shuai Lei², Xiaoyi Gu³, Shaofeng Liang⁴, Tong Wu⁵,
Ming Yuan^{6*}, Yanhua Fei^{6*}, and Runwei Guan^{3,4*}¹Glasgow College Hainan, University of Electronic Science and Technology of China, Lingshui, Hainan, China²School of Medicine, Shanghai Jiao Tong University, Shanghai, China³Department of Research and Development, FertiTech AI, Shanghai, China⁴Thrust of Artificial Intelligence, Information Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China⁵Department of People Experience and Technology, Amazon.com Services LLC, Seattle, Washington, United States of America⁶Department of Oncology, Jiangyin People's Hospital, Jiangyin, Jiangsu, China(This article belongs to the *Special Issue: Artificial intelligence in reproductive health and pathology*)***Corresponding authors:**Ming Yuan
(yuan__ming@163.com)
Yanhua Fei
(ferteichai@gmail.com)
Runwei Guan
(runwayrwguan@hkust-gz.edu.cn)**Citation:** Zhang Q, Lei S, Gu X, *et al.* Dynamic U-Net: A lightweight hierarchical network for accurate and efficient sperm morphology segmentation. *Artif Intell Health*. 2026;3(3):025460101.
doi: 10.36922/AIH025460101**Received:** November 16, 2025**Revised:** December 26, 2025**Accepted:** January 15, 2026**Published online:** February 13, 2026**Copyright:** © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.**Publisher's Note:** AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.**Abstract**

Accurate sperm segmentation is essential for automated semen analysis and male fertility assessment, yet existing methods struggle to balance segmentation accuracy with computational efficiency. We propose Dynamic U-Net (Dy-UNet), a novel lightweight deep learning architecture that achieves state-of-the-art performance while maintaining exceptional efficiency. The architecture integrates four key innovations: a cross-resolution dual-branch network that preserves high-resolution spatial details, a dynamic snake convolution that adapts to curved sperm morphology, a grouped multi-axis Hadamard product attention that efficiently captures long-range dependencies, and a fast cross-resolution aggregation that enables multi-scale feature fusion. Comprehensive evaluation on the SegSperm dataset demonstrates that Dy-UNet substantially outperforms existing methods. Dy-UNet achieves a state-of-the-art mean Intersection over Union of 0.44260 and Dice similarity coefficient of 0.61362, which surpasses the baseline U-Net by 25.5% and 17.7%, respectively. Remarkably, Dy-UNet achieves this superior performance with only 184 K parameters, especially suitable for deploying mobile platforms and embedded systems in clinical settings. The lightweight architecture enables accurate, objective, and accessible automated sperm morphology assessment with significant implications for clinical fertility diagnosis and point-of-care testing.

Keywords: Sperm segmentation; U-Net; Lightweight model; Computer-assisted semen analysis

1. Introduction

Human infertility refers to the inability to achieve pregnancy after a defined period of unprotected intercourse, and it remains a major global health challenge. This inability affects approximately 17.5% of adults worldwide, with male factors contributing to 30% of infertility cases.^{1,2} According to the World Health Organization guidelines, semen analysis serves as the primary diagnostic method for evaluating male infertility through the assessment of sperm concentration, motility, and morphological characteristics.³ Among these parameters, the morphology of the sperm cell is considered a significant indicator of male-factor infertility,⁴ making morphological analysis crucial in determining fertility potential and selecting high-quality sperm for assisted reproductive technologies, such as *in vitro* fertilization.^{5,6} Traditional morphological assessment requires trained clinical personnel to follow World Health Organization guidelines meticulously, involving pixel-level annotation of sperm structures, including the acrosome, vacuole, nucleus, midpiece, and tail.⁷ The World Health Organization mandates the analysis of at least 200 sperm per semen sample, which translates to a manual annotation of more than 1,000 contours per patient, rendering the process extraordinarily time-consuming and subject to considerable inter-observer variability.³

The challenges of manual sperm analysis are further compounded by the limitations of the current microscopy techniques. While high-magnification microscopy at a 100× objective enables detailed subcellular analysis,⁸ it restricts observation to individual sperm due to the limited field of view. Lower-magnification microscopy with a 20× objective is essential for comprehensive population-level assessment within semen samples, yet individual sperm occupy less than 1% of the visual field under these conditions, making manual identification and selection extremely challenging and expertise dependent.^{9,10} These limitations have driven the development of computer-assisted semen analysis systems. The systems aim to systematically quantify shape and movement and perform statistical analysis after counting sperm according to specific selection criteria.¹¹ Over the years, the analysis process evolved from traditional handcrafted feature-based methods to contemporary deep learning approaches that demonstrate superior performance and reproducibility.^{12–19}

Deep learning techniques have gained prominence and have proven effective at automatically extracting meaningful image features. However, model performances depend heavily on the quality and representativeness of the training dataset, which can lead to degraded detection on low-quality images and limited generalizability in

real-world applications. Current sperm analysis datasets, such as the Human Sperm Head Morphology dataset²⁰ and the Sperm Morphology Image Data Set,²¹ primarily focus on abnormality classification of individual sperm cells captured at high magnification. This approach fails to address the clinical need for comprehensive morphological and motility assessment across entire sperm populations in lower-magnification images containing multiple cells. Factors such as unclear neck and overlapping sperm heads can complicate the segmentation process.¹¹ Accurate segmentation of individual sperm from multi-cell microscopy images is therefore increasingly critical for automated sperm health evaluation. However, existing segmentation approaches face substantial challenges due to variations in image background, illumination conditions, and inconsistent annotation standards across different laboratories.^{22,23} Although these data challenges persist, several segmentation models have reported good performance; however, most focus primarily on sperm head detection,^{13,14,23,24} neglecting the complete morphological structure that is essential for comprehensive fertility assessment.

In terms of deep learning methods for medical segmentation tasks, the U-Net architecture has emerged as a cornerstone in medical image segmentation. Its encoder-decoder structure with skip connections effectively preserves fine-grained spatial information while capturing high-level semantic features, making it particularly suitable for segmenting small biological structures.²⁵ Building on the U-Net's success, several architectural innovations have further enhanced segmentation performance in medical imaging.^{26–28} However, since the sperm size is relatively small, lacking sufficient pixels for straightforward local information extraction, current models are still not well adapted to the task.²⁹ The linear morphology of sperm also presents new challenges for feature extraction in deep learning models, which is essential to segment the sperm flagellum.

To address the above challenges in sperm segmentation, we propose a novel Dynamic U-Net (Dy-UNet) framework. Our aims are as follows:

- (i) To introduce a dynamic attention mechanism specifically designed to capture the elongated and serpentine morphological characteristics of sperm cells.
- (ii) To enhance information interaction across resolution by extending the U-Net to a smart multi-layer variant, and to foster the low-resolution feature through a convolution stream.
- (iii) To create a Dy-UNet model that features high inference speed and small model size, which could be easily

deployed on edge systems for sperm segmentation, and also demonstrates state-of-the-art performance in competing with other models.

2. Related work

The evolution of medical image segmentation architectures traces back to the seminal U-Net architecture introduced by Ronneberger *et al.*²⁵ in 2015. This work fundamentally established the encoder-decoder paradigm with skip connections for precise biomedical image segmentation, effectively addressing the challenge of preserving fine-grained spatial information during feature extraction. Building on this architecture, subsequent research has pursued two primary directions: enhancing feature representation and improving computational efficiency.

In the realm of feature enhancement, Oktay *et al.*³⁰ proposed Attention U-Net, which incorporated spatial attention gates to selectively emphasize salient features. Zhou *et al.*³¹ introduced UNet++, which employed nested and dense skip connections to reduce the semantic gap between encoder and decoder feature maps, achieving superior performance across multiple medical imaging datasets.

The integration of transformer architectures then shifted the paradigm in capturing the global context. Chen *et al.*³² pioneered TransUNet, synergistically combining convolutional neural networks with Vision Transformers³³ to model long-range dependencies, establishing new performance benchmarks for multi-organ segmentation. Subsequently, Cao *et al.*³⁴ adopted the shifted window mechanism from Swin Transformer³⁵ to Swin-UNet, efficiently processing high-resolution medical images while maintaining computational tractability.

Alongside the development in feature extraction, recent studies have aimed to optimize models for computational efficiency in resource-constrained medical environments, particularly for mobile health applications and point-of-care diagnostics.

Early lightweight architectures focused on efficient feature extraction for classification tasks. MobileNetV1³⁶ first proposed depth-wise separable convolutions to decompose standard convolutions, substantially reducing computational complexity. Subsequently, MobileNetV2³⁷ enhanced this approach with inverted residuals and linear bottlenecks, and ShuffleNetV2³⁸ introduced channel shuffling to minimize memory access costs during group convolutions.

Adapting these principles to segmentation posed unique challenges due to high-resolution output requirements. Bilateral Segmentation Network³⁹ addressed this through

bilateral paths separating spatial detail and context, while the Image Cascade Network⁴⁰ employed multi-resolution cascade processing to balance speed and accuracy for real-time applications. In medical segmentation, the multi-attention and lightweight U-Net²⁷ utilized multi-axis attention for efficient feature fusion in skin lesion segmentation. Efficient group enhanced U-Net²⁸ further introduced linear-complexity Hadamard product attention and mask-guided feature aggregation, achieving state-of-the-art skin lesion segmentation performance with a model size of only 50 kB. The Lightweight and Vanilla Model for Medical Image Segmentation⁴¹ demonstrated that leveraging pre-trained MobileNetV3-Large⁴² with structural re-parametrization could also realize low parameters without compromise in performance.

3. Methods

The Dy-UNet introduces a hierarchical multi-resolution framework specifically tailored for sperm morphology analysis. Unlike conventional encoder-decoder architectures that compress spatial information through sequential downsampling, our approach maintains parallel processing streams at multiple resolutions throughout the network. This design philosophy draws inspiration from the hierarchical nature of biological vision, where the human visual system simultaneously processes information at different scales to achieve comprehensive scene understanding.

The architecture consists of five progressive stages, each maintaining an increasing number of parallel branches that operate at different spatial resolutions. Information flows both horizontally within each resolution level and vertically across different scales, creating a dense information exchange network that preserves fine-grained spatial details while capturing high-level semantic context. To adaptively weight cross-resolution information transfer, we strategically incorporated conditional channel weighting (CCW) modules at stages 4 and 5, where the increased number of branches necessitates more sophisticated fusion mechanisms. Similarly, fast cross-resolution aggregation (FCRA) modules were positioned at the transition between adjacent resolution levels to facilitate efficient multi-scale feature integration.

As illustrated in [Figure 1](#), the network architecture can be decomposed into three fundamental components: (i) a cross-resolution interaction mechanism that enables bidirectional information flow between parallel branches, (ii) specialized feature extraction modules that adapt to sperm-specific morphological characteristics, and (iii) multi-scale perception mechanisms that integrate features across different resolution levels. The subsequent sections

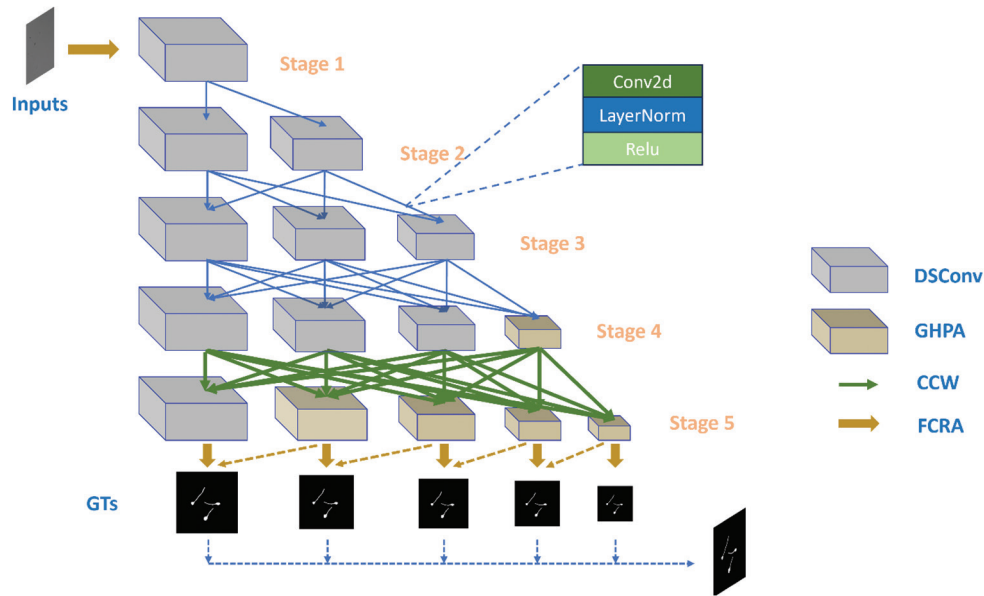


Figure 1. The network architecture of the proposed Dynamic U-Net

Abbreviations: CCW: Conditional channel weighting; Conv2d: Two-dimensional convolution; DSConv: Dynamic snake convolution; FCRA: Fast cross-resolution aggregation; GHPA: Grouped multi-axis Hadamard product attention; GT: Ground truth; ReLU: Rectified linear unit.

provide a detailed exposition of each component and its contribution to the overall segmentation performance.

3.1. Cross-resolution interaction

The accurate segmentation of sperm morphological structures demands both fine-grained detail preservation and comprehensive contextual understanding. Traditional U-Net architectures and their variants typically follow an encoder-decoder paradigm where high-resolution representations are first compressed into low-resolution, semantically rich features, then gradually recovered through upsampling operations. However, this sequential downsampling-upsampling process inevitably leads to spatial information loss, which is particularly detrimental for detecting subtle morphological abnormalities in sperm analysis. Inspired by the success of high-resolution networks⁴³ in maintaining multi-resolution representations throughout the network, we propose a novel cross-resolution enhancement strategy that fundamentally reimagines how multi-scale information should be integrated within medical image segmentation frameworks.

3.1.1. Parallel multi-resolution architecture

The network maintains parallel multi-resolution streams throughout its entire depth, departing from conventional single-path designs. The processing begins with a stem layer that extracts initial features from the input image $\{I \in \mathbb{R}^{3 \times H \times W}\}$, as shown in Equation 1:

$$F_0 = \text{ReLU}(\text{LN}(3 \times 3_{s=2}^{\text{Conv}}(I))) \quad (1)$$

where $3 \times 3_{s=2}^{\text{Conv}}$ denotes a 3×3 convolution with a stride of 2, reducing the spatial resolution to $\frac{H}{2} \times \frac{W}{2}$, and LN represents layer normalization applied in the channel dimension.

The network progressively expands from a single high-resolution branch to five parallel branches across five stages. At stage t , the network maintains t parallel branches with resolutions $r_1, r_2, \dots, r_t$, where $r_i = \frac{H}{2^i} \times \frac{W}{2^i}$. Here, the index i indicates the resolution level, with $i = 1$, representing the highest resolution branch (half of the input resolution), and larger values of i correspond to progressively lower resolution branches. To manage the increased complexity of information fusion in later stages, CCW modules were integrated at stages 4 and 5, operating before the multi-resolution fusion process to adaptively recalibrate channel-wise features based on cross-resolution dependencies. FCRA modules were positioned at the network's decoder phase, where they create efficient bidirectional information flow between adjacent resolution levels. Features from lower-resolution branches (with larger receptive fields and richer semantic context) were upsampled and aggregated into higher-resolution branches to refine spatial details.

The progressive expansion strategy allows the network to gradually build multi-scale representations while

maintaining computational efficiency in early stages. For example, the stage-wise expansion for the highest resolution layer ($i = 1$) can be described as Equation 2:

$$S_t : \{X_i^{t-1}\}_{i=1}^{t-1} \rightarrow \{X_i^t\}_{i=1}^t \quad (2)$$

where S_t represents the transformation at stage t , including both intra-stage processing and inter-stage fusion.

The information preservation can be quantified through the mutual information between the input and features at stage t , as shown in Equation 3:

$$I\left(\mathbf{I}; \{X_i^t\}_{i=1}^t\right) \geq I\left(\mathbf{I}; X_{enc}^t\right) \quad (3)$$

where X_{enc}^t represents the encoder features at the corresponding depth in traditional U-Net architectures.

3.1.2. Stage transition

For the stage transition from s to $s + 1$, the fusion process for branch i is formulated as Equation 4:

$$i_{s+1}^Y = ReLU\left(\sum_{j=1}^{N_s} T_{ij}\left(j_s^X\right)\right) \quad (4)$$

where N_s denotes the number of branches in stage s , and the transformation function T_{ij} is defined as Equation 5:

$$T_{ij}(X) = \begin{cases} X, & \text{if } i = j, \\ \mathcal{H}_{ij}\left(\mathcal{U}_{2^{j-i}}(X)\right), & \text{if } i < j, \\ \mathcal{H}_{ij}\left(\mathcal{D}_{2^{i-j}}(X)\right), & \text{if } i > j. \end{cases} \quad (5)$$

where, $\mathcal{U}_r$ and $\mathcal{D}_r$ represent upsampling and downsampling operations with ratio r , while $\mathcal{H}_{ij}$ denotes channel harmonization to align feature dimensions between branches. For upsampling operations, where $i < j$ (lower to higher resolution), we employed a bilinear interpolation followed by 1×1 convolution, as shown in Equation 6:

$$\mathcal{U}_{2^k}(X) = \text{Conv}_{1 \times 1}\left(\text{Bilinear}\left(X, \text{scale} = 2^k\right)\right) \quad (6)$$

The design preserves spatial smoothness while adapting channel dimensions. For downsampling operations where $i > j$ (higher to lower resolution), we utilized strided convolutions to simultaneously reduce spatial dimensions and learn task-specific downsampling kernels, as shown in Equation 7:

$$\mathcal{D}_{2^k}(X) = \begin{cases} \text{LN}\left(\text{Conv}_{3 \times 3}^{s=2}(X)\right), & \text{if } k = 1, \\ \mathcal{D}_{2^k}(\mathcal{D}_{2^{k-1}}(X)), & \text{if } k > 1. \end{cases} \quad (7)$$

The recursive formulation for $k > 1$ ensures that large downsampling ratios are decomposed into sequential operations, preventing excessive information loss.

3.1.3. Conditional channel weighting

While the full-connectivity fusion pattern enables comprehensive information exchange, not all cross-resolution connections contribute equally to the final representation. To address this, we introduced the CCW module in stages 4 and 5, where the increased number of branches demands more sophisticated fusion strategies.

As depicted in Figure 2, the CCW module learns to adaptively recalibrate channel-wise features based on cross-resolution dependencies. The CCW module first performs spatial-channel decoupling by splitting each branch's features into two halves along the channel dimension, as shown in Equation 8:

$$X_i = \left[X_i^{(1)}, X_i^{(2)} \right], X_i^{(k)} \in \mathbb{R}^{C_i/2 \times H_i \times W_i} \quad (8)$$

The first half $X_i^{(1)}$ maintains the original features as a residual connection, while the second half $X_i^{(2)}$ undergoes cross-resolution weighting, as shown in Equation 9:

$$W = \text{Sigmoid}\left(\mathcal{F}_{cw}\left(\left\{\text{AvgPool}\left(X_i^{(2)}\right)\right\}_{i=1}^N\right)\right) \quad (9)$$

where $\mathcal{F}_{cw}$ represents the cross-resolution weighting function implemented as Equation 10:

$$\mathcal{F}_{cw}\left(\left\{z_i\right\}_{i=1}^N\right) = \text{Conv}_{1 \times 1}^{(2)}\left(\text{ReLU}\left(\text{Conv}_{1 \times 1}^{(1)}\left(\text{Concat}_{i=1}^N\left[\mathcal{R}_s\left(z_i\right)\right]\right)\right)\right) \quad (10)$$

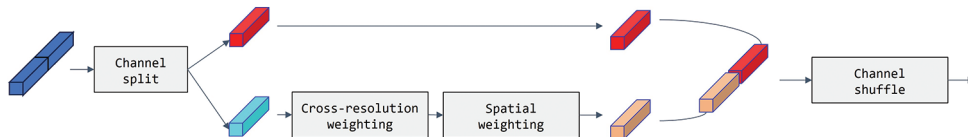


Figure 2. The pipeline of conditional channel weighting

where R_s denotes resolution alignment to the smallest resolution through adaptive pooling, and the two 1×1 convolutions formed a bottleneck structure with a reduction ratio $r = 16$ for computational efficiency.

The weighted features were then processed through depth-wise separable convolutions for spatial modeling, followed by spatial attention, as shown in Equation 11:

$$Y_i^{(2)} = \text{SpatialAtt}\left(\text{DSConv}\left(W_i \odot X_i^{(2)}\right)\right) \quad (11)$$

The final output combines both pathways through channel shuffling to enhance information flow, as shown in Equation 12:

$$Y_i = \text{ChannelShuffle}\left(\left[X_i^{(1)}, X_i^{(2)}\right], \text{groups} = 2\right) \quad (12)$$

The parallel branches naturally provide diverse receptive fields without requiring specialized operations such as atrous spatial pyramid pooling. Adding channel-wise attention, the effective receptive field for branch i at stage t can be approximated as Equation 13:

$$\text{ERF}_i^t = \text{ERF}_i^{t-1} + \sum_{j \neq i} w_{ij} \cdot \text{ERF}_j^{t-1} \quad (13)$$

where w_{ij} represents the learned importance weight from branch j to branch i through the CCW module.

3.2. Linear feature extraction

The morphological analysis of sperm cells presents unique challenges due to their distinctive elongated structure and complex internal organization. Sperm cells exhibit characteristic linear features, including the acrosomal cap, nucleus boundary, midpiece, and flagellum, each requiring precise delineation for accurate morphological assessment. Traditional convolutional operations with fixed rectangular kernels often struggle to adapt to these curved and elongated structures, leading to suboptimal feature extraction, especially at boundary regions.

To address these challenges, we proposed a progressive linear feature extraction strategy that combines dynamic snake convolution (DSConv) for adaptive spatial sampling with the grouped multi-axis Hadamard product attention (GHPA) mechanism for efficient multi-dimensional feature modeling.

3.2.1. Deformable spatial convolution

While sperm orientations are random, individual sperm structures maintain local directional consistency. A sperm tail, once it begins curving in a particular direction, tends to continue that curvature smoothly rather than

exhibiting abrupt directional changes. The observation led to our cumulative offset mechanism, where deformations propagate smoothly along the kernel's sampling direction.

As illustrated in Figure 1, we employed DSConv⁴⁴ as the fundamental building block within each resolution branch. The operation further optimizes multi-resolution dense feature flow through adaptive receptive field adjustment based on local morphological features.

For an input feature map $X \in \mathbb{R}^{C \times H \times W}$ and a regular grid $R = (-1, -1), (-1, 0), \dots, (1, 1)$ defining the 3×3 kernel sampling locations, the deformable convolution augments each grid position with learnable offsets Equation 14:

$$Y(p) = \sum_{r \in R} w(r) \cdot X(p + r + \Delta r) \cdot m(r) \quad (14)$$

where p denotes the output position, $w(r)$ represents the convolution weights, Δr denotes the learnable offset, and $m(r) \in [0, 1]$ is the modulation scalar for each sampling point.

The offsets and modulation scalars were predicted through a parallel convolutional branch, as shown in Equation 15:

$$\{\Delta R, M\} = \text{Conv}_{3 \times 3}(X), \Delta R \in \mathbb{R}^{2K \times H \times W}, M \in \mathbb{R}^{K \times H \times W} \quad (15)$$

where $K = |R| = 9$ for a 3×3 kernel. To ensure stable training and prevent excessive deformation, we applied offset regularization $\Delta r_{\text{reg}} = \tanh(\Delta r) \cdot \alpha$, where α is a learnable scale parameter initialized to 1. The center position remained fixed as an anchor point while peripheral positions accumulated offsets. The modeling created a snake-like sampling pattern that naturally follows curved structures without fragmenting them. The offset accumulation ensured that adjacent sampling points maintain spatial coherence.

3.2.2. Grouped multi-axis Hadamard product attention

While DSConv excels at following local morphological contours, sperm analysis requires understanding spatial relationships between different cellular components. For instance, bent neck defects are visible as sharp angles between the head and midpiece, or coiled tails indicating motility issues. These patterns require receptive fields larger than what local convolutions can provide.

In deeper network stages where feature channels expand to capture increasingly abstract representations, we introduced the GHPA mechanism. As shown in Algorithm 1, GHPA divides input features into four equal groups and processes each through a specialized attention

pathway. Given input features $X \in \mathbb{R}^{C \times H \times W}$, after layer normalization and channel splitting into groups $G_{ij} = 1^4$, each group undergoes distinct processing.

The spatial attention pathway (XY-plane) learns to identify coherent sperm structures against the cluttered background, while the channel-height attention (ZX-plane) captures the longitudinal continuity of sperm structures. For the channel-width attention (ZY-plane), lateral features were modeled. In addition, when two sperm tails cross, they create a characteristic X-pattern in the XY-plane, while maintaining distinct signatures in the ZX and ZY planes. The multi-dimensional processing enables the network to disambiguate overlapping structures that would be inseparable using single-plane features alone.

The Hadamard product provides efficient, parameter-light attention computation while maintaining the ability to model complex feature dependencies. The use of interpolated learnable parameters from lower resolutions reduces memory footprint and introduces implicit regularization against high-frequency noise prevalent in phase-contrast microscopy.

3.3. Multi-scale perception

The human visual system processes visual information through a sophisticated hierarchical mechanism. When examining sperm morphology, clinicians first identify the overall cell structure and orientation (global perception), then focus on specific regions of each sperm (local perception). The cognitive process involves continuous information exchange between different scales of observation, where global context guides local attention and local findings refine global understanding.

Translating this biological insight to our computational framework, we implemented an FCRA mechanism that facilitates bidirectional information flow between feature maps at different resolutions. As illustrated in Figure 3, the FCRA module processes hierarchical features from adjacent resolution levels.

The residual connection was used to preserve low-resolution information, as shown in Equation 16:

$$Y = X_i + \mathcal{F}_{\text{fusion}} \left(\left[X_h^{\text{resized}}, X_i \right] \right) \quad (16)$$

In deeper network stages, we introduced a lightweight mask-guided modulation mechanism. Unlike attention-based approaches that require expensive matrix operations, our modulation simply scales features based on predicted confidence, as shown in Equation 17:

Algorithm 1: Group multi-axis Hadamard product attention

```

Require: Input feature  $X \in \mathbb{R}^{B \times C_m \times H \times W}$ 
Ensure: Output feature  $Y \in \mathbb{R}^{B \times C_{out} \times H \times W}$ 
1:  $X \leftarrow \text{LayerNorm}(X)$ 
2:  $[X_1, X_2, X_3, X_4] \leftarrow \text{Split}(X, 4)$  {Split along channel dimension}
3: //Spatial attention branch (XY-axis)
4:  $P_{xy} \in \mathbb{R}^{\frac{1}{4} \times \frac{C_m}{4} \times H_0 \times W_0}$  {Learnable parameters}
5:  $A_{xy} \leftarrow \text{Conv}_{xy}$  (Interpolate  $(P_{xy}, [H, W])$ )
6:  $X_1 \leftarrow X_1 \odot A_{xy}$  {Hadamard Product}
7: //Channel-height attention branch (ZX-axis)
8:  $P_{zx} \in \mathbb{R}^{\frac{1}{4} \times \frac{C_m}{4} \times H_0 \times H_0}$  {Learnable parameters}
9:  $X_2 \leftarrow \text{Permute}(X_2, [0, 3, 1, 2])$ 
10:  $A_{zx} \leftarrow \text{Conv}_{zx}$  (Interpolate  $(P_{zx}, [\frac{C_m}{4}, H])$ )
11:  $X_2 \leftarrow X_2 \odot A_{zx}$ 
12: //Channel-Width attention branch (ZY-axis)
13:  $P_{zy} \in \mathbb{R}^{\frac{1}{4} \times \frac{C_m}{4} \times W_0 \times W_0}$  {Learnable parameters}
14:  $X_3 \leftarrow \text{Permute}(X_3, [0, 2, 1, 3])$ 
15:  $A_{zy} \leftarrow \text{Conv}_{zy}$  (Interpolate  $(P_{zy}, [\frac{C_m}{4}, W])$ )
16:  $X_3 \leftarrow X_3 \odot A_{zy}$ 
17: //Depth-wise convolution branch
18:  $X_4 \leftarrow \text{DWConv}(X_4)$ 
19:  $X \leftarrow \text{Concat}([X_1, X_2, X_3, X_4])$ 
20:  $X \leftarrow \text{LayerNorm}(X)$ 
21:  $Y \leftarrow \text{Conv}_{1 \times 1}(\text{DWConv}_{3 \times 3}(X))$ 
22: return Y
    
```

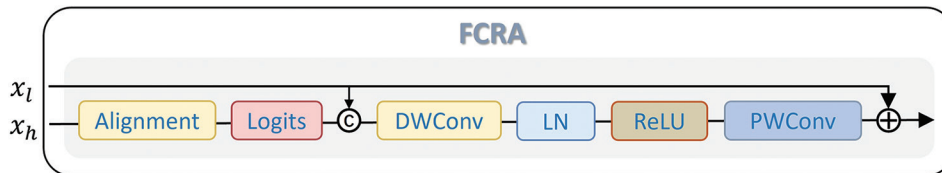


Figure 3. The pipeline of the FCRA module

Abbreviations: DWConv: Depth-wise convolution; FCRA: Fast cross-resolution aggregation; LN: Layer normalization; PWConv: Pointwise convolution; ReLU: Rectified linear unit.

$$X_h^{\text{modulated}} = X_h^{\text{resized}} \cdot (1 + 0.5 \cdot \text{Interpolate}(M)) \quad (17)$$

The modulation factor of 0.5 was empirically determined to provide optimal enhancement without over-suppression.

3.4. Experiments

We evaluated our proposed Dy-UNet on the SegSperm dataset, a specialized benchmark for sperm morphology segmentation.⁴⁵ The dataset presents the clinical scenario where multiple sperm cells appear within a single field of view, with each image containing pixel-level annotations for complete sperm structures, including the head, midpiece, and tail regions. All images were resized to 256×256 pixels to balance computational efficiency and morphological detail preservation. To comprehensively assess the effectiveness of our architecture, we compared Dy-UNet against several established methods representing different architectural paradigms: the original U-Net²⁵ as the fundamental baseline, R2U-Net and R2AttU-Net,⁴⁶ incorporating recurrent residual blocks and attention mechanisms, MobileNetV2-UNet,³⁷ employing efficient depth-wise separable convolutions, and DeepLab v3+⁴⁷ utilizing atrous spatial pyramid pooling for multi-scale contextual information. All baseline methods were implemented with identical channel configurations ($c_list = [8, 16, 24, 32, 48]$) to ensure a fair comparison under similar parameter budgets.

All models were trained for 300 epochs using the AdamW optimizer with an initial learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay of 0.01. The learning rate was scheduled using cosine annealing with $T_{max} = 50$ and $\eta_{min} = 0.00001$, allowing a gradual decrease following a cosine curve with periodic restarts. For Dy-UNet, we employed a compound loss function combining binary cross-entropy and Dice loss at multiple resolution levels, as shown in Equation 18:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{final}} + 0.4 \sum_{i=1}^5 w_i (\mathcal{L}_{\text{BCE}}(p_i, y) + \mathcal{L}_{\text{Dice}}(p_i, y)) \quad (18)$$

where w_i are weights that decrease with resolution depth. This multi-level supervision strategy encourages meaningful representations at different scales, proving effective for capturing both coarse-level structures and fine-grained morphological details. Data augmentation during training included random horizontal flipping ($p = 0.5$) and random vertical flipping ($p = 0.5$) to account for arbitrary sperm orientations, along with dataset-specific normalization using pre-computed training set statistics.

All experiments were conducted using PyTorch 1.12.0 with CUDA 12.6 on Python 3.10, utilizing a single NVIDIA

V100 GPU with 32 GB memory. Training was performed with a batch size of 8, and reproducibility was ensured by a fixed random seed of 42 across all random number generators. Model performance was assessed using four standard metrics: mean Intersection over Union (mIoU), Dice similarity coefficient (DSC), pixel-wise accuracy, and sensitivity, with all metrics computed at the default threshold of 0.5.

4. Results

4.1. Sample results

Figure 4 presents representative segmentation results comparing Dy-UNet against baseline U-Net on the SegSperm dataset. The qualitative comparison demonstrated that Dy-UNet produced more accurate and complete segmentation across diverse morphological variations and imaging conditions. Notably, in the second example, the baseline U-Net completely failed to detect one sperm cell in a multi-cell scenario, while Dy-UNet successfully identified and segmented all sperm structures present in the field of view. This superior performance can be attributed to our cross-resolution interaction mechanism, which maintains high-resolution representations throughout the network and enables simultaneous detection of both prominent and subtle sperm structures.

Across all examples, Dy-UNet demonstrated improved boundary precision and complete structure preservation, particularly for challenging regions, such as curved flagella and fine-grained morphological features. The DSConv module enabled adaptive spatial sampling that follows the serpentine morphology of sperm tails, while the GHPA mechanism captured long-range dependencies to maintain structural continuity. In contrast, U-Net exhibited characteristic limitations, including fragmented tail segmentation, boundary smoothing artifacts, and incomplete detection in complex scenarios. These qualitative results also complement the quantitative improvements presented in the following sections.

4.2. Comparative results

To evaluate the effectiveness of our proposed Dy-UNet architecture, we conducted comprehensive experiments on the SegSperm dataset. We compared our method against several established lightweight segmentation approaches, which included U-Net, R2U-Net, R2AttU-Net, DeepLab v3+, and MobileNet v2. Due to the lack of lightweight models on the sperm segmentation task, two state-of-the-art models in medical image segmentation were also included for comparison. These models all demonstrated comparable inference speeds in the order

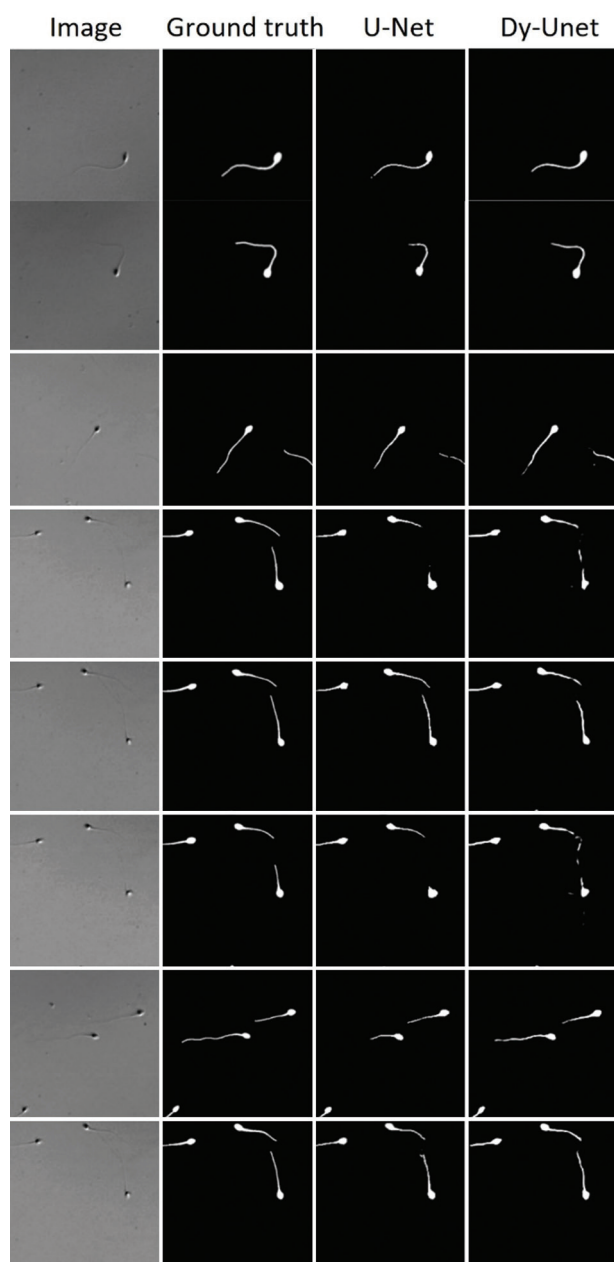


Figure 4. Sample results comparing our Dynamic U-Net (Dy-UNet) with the baseline U-Net

of tens of milliseconds. As presented in Table 1, Dy-UNet achieved superior performance across multiple evaluation metrics with a comparatively small parameter count and demonstrated the effectiveness of our architectural innovations for sperm segmentation tasks.

4.2.1. Performance baseline

The Dy-UNet achieved the best performance across all evaluation metrics, with substantial improvements over existing methods. Compared to the baseline U-Net,

our method achieved a 25.5% improvement in mIoU and a 17.7% improvement in DSC, demonstrating the effectiveness of the multi-resolution architecture for capturing sperm-specific morphological patterns. Notably, Dy-UNet also obtained the highest sensitivity among all compared methods and surpassed the second-best R2AttU-Net by 0.3%. This improvement in sensitivity is clinically significant, as it reflects the model's enhanced capability to detect true positive sperm structures, particularly the fine-grained components, such as flagella and midpiece regions, that are critical for comprehensive morphological assessment. These results collectively demonstrate that our approach effectively addresses the fundamental challenges in sperm segmentation, which include the accurate delineation of subtle morphological features from complex microscopy backgrounds and the precise identification of complete sperm structures even under challenging imaging conditions.

4.2.2. Comparison with attention-based methods

Compared to other attention-based architectures, Dy-UNet outperformed R2AttU-Net by 10.4% in mIoU and 7.2% in DSC, despite the fact that R2AttU-Net incorporates recurrent and attention mechanisms. This result indicates that our parallel multi-resolution processing strategy and linear feature-extraction modules are more effective than traditional attention gates at capturing sperm-specific morphological patterns. The performance advantage can be attributed to several key factors: (i) The cross-resolution interaction mechanism enables effective information exchange between high-resolution detail-preserving features and low-resolution semantic features; (ii) the DSConv adapts to the curved structures of sperm tails; and (iii) the GHPA efficiently captures both local and global dependencies.

4.2.3. Lightweight architecture comparison

Furthermore, when compared against the lightweight MobileNet v2 baseline, Dy-UNet achieved 6.9% higher mIoU and 4.8% higher DSC. Remarkably, Dy-UNet accomplishes this superior performance with only 184 K parameters, which makes it exceptionally suitable for deployment on resource-constrained devices, such as mobile platforms, edge computing systems, or embedded devices commonly used in clinical settings. This efficiency is achieved through strategic architectural design choices, which include the use of depth-wise separable convolutions, efficient grouped attention mechanisms, and lightweight multi-resolution feature fusion.

4.2.4. State-of-the-art model comparison

Compared to recent state-of-the-art lightweight medical image segmentation models, Dy-UNet demonstrates

Table 1. Comparative experimental results on the SegSperm dataset

Model	mIoU	DSC (F1)	Accuracy	Sensitivity	Parameters (K)
U-Net ²⁵	0.35265	0.52142	0.99021	0.53603	125
R2U-Net ⁴⁶	0.39853	0.56993	0.99263	0.49104	130
R2AttU-Net ⁴⁶	0.40091	0.57235	0.99173	0.55657	132
DeepLab v3+ ⁴⁷	0.32033	0.48522	0.99039	0.45515	1,460
MobileNet v2 ³⁷	0.41399	0.58557	0.99310	0.48971	100
Efficient group enhanced U-Net ²⁸	0.37707	0.54764	0.99166	0.50721	53
Multi-attention and lightweight U-Net ²⁷	0.30031	0.46190	0.98970	0.44439	177
Dynamic U-Net	0.44260	0.61362	0.99279	0.57546	184

Abbreviations: DSC: Dice similarity coefficient; mIoU: Mean Intersection over Union.

clear advantages. Efficient group enhanced U-Net, despite having only 53 K parameters, achieved an mIoU of 0.37707 and DSC of 0.54764, which are 17.4% and 12.0% lower than Dy-UNet, respectively. The multi-attention and lightweight U-Net, with 177 K parameters comparable to our model, showed the weakest performance with an mIoU of 0.30031 and DSC of 0.46190. These results indicate that architectural designs optimized for skin lesion segmentation do not transfer well to sperm segmentation tasks.

4.3. Ablation results

To systematically evaluate the contribution of each proposed component, we conducted comprehensive ablation studies with a cross-resolution U-Net (CR-UNet) as the baseline architecture. As shown in Table 2, we progressively incorporated our proposed modules on the standard U-Net: cross-resolution interaction, DSConv, GHPA, and FCRA. The results demonstrate that each component continuously improved performance under restricted computational overhead.

4.3.1. Baseline cross-resolution architecture

The baseline CR-UNet, which implements parallel multi-resolution branches without specialized feature extraction modules, achieved an mIoU of 0.37483 and DSC of 0.55478. This represents a 6.3% improvement in mIoU over the standard U-Net architecture (0.35265) and validates our hypothesis that the maintenance of high-resolution representations throughout the network preserves critical spatial information essential for accurate sperm segmentation. The improvement demonstrates the fundamental advantage of the dual-branch architecture in the simultaneous capture of both fine-grained spatial details and high-level semantic context, which are both crucial for the identification of sperm structures with various scales and complex morphologies.

Table 2. Ablation studies based on cross-resolution U-Net

Model	mIoU	DSC (F1)	Accuracy	Sensitivity
U-Net	0.35265	0.52142	0.99021	0.53603
CR-UNet	0.37483	0.55478	0.99131	0.52194
CR-UNet+DSConv	0.41328	0.56740	0.99203	0.55876
CR-UNet+DSConv+GHPA	0.42105	0.59259	0.99259	0.54171
CR-UNet+DSConv+GHPA+FCRA	0.44260	0.61362	0.99279	0.57546

Abbreviations: DSC: Dice similarity coefficient; DSConv: Dynamic snake convolution; FCRA: Fast cross-resolution aggregation; GHPA: Grouped multi-axis Hadamard product attention; mIoU: Mean Intersection over Union.

4.3.2. Dynamic snake convolution

The integration of DSConv (CR-UNet+DSConv) yielded substantial performance gains and elevated mIoU to 0.41328 and sensitivity to 0.55876. The 10.3% improvement in mIoU demonstrates the effectiveness of adaptive spatial sampling for the tracking of curved and elongated structures of sperm cells. The deformable convolution's ability to adjust receptive fields dynamically based on local morphological features proves particularly valuable for the capture of serpentine patterns of flagella and the precise boundaries of sperm heads. Unlike standard convolutions with fixed rectangular kernels, DSConv employs a cumulative offset mechanism where deformations propagate smoothly along the kernel's sampling direction and naturally follow curved structures without fragmentation. This capability is especially important for sperm segmentation, as individual sperm structures maintain local directional consistency.

4.3.3. Grouped multi-axis Hadamard product attention

Further incorporation of the GHPA mechanism (CR-UNet+DSConv+GHPA) enhanced the DSC to 0.59259. This improvement of 4.4% in DSC highlights

GHPA's capacity to model long-range spatial dependencies across different cellular components while it maintains computational efficiency through grouped processing and Hadamard product operations. The multi-axis attention's ability to process features along spatial (XY-plane), channel-height (ZX-plane), and channel-width (ZY-plane) dimensions enables the network to disambiguate overlapping structures and identify coherent sperm patterns against cluttered backgrounds.

The complete Dy-UNet architecture, which incorporates all proposed components (CR-UNet+DSConv+GHPA+FCRA), achieved the optimal overall performance. This result validates our design principle that bidirectional information flow between feature maps at different resolutions, which mimics the hierarchical processing observed in human visual perception, is crucial for comprehensive sperm morphology analysis. Notably, the progressive incorporation of modules consistently improved performance across all evaluation metrics. Each component addresses different aspects of the sperm segmentation challenge: CR-UNet establishes the multi-resolution foundation, DSConv handles curved structures, GHPA captures long-range dependencies, and FCRA integrates multi-scale information. Their combination created a comprehensive solution that significantly outperformed the baseline architecture and existing methods. Furthermore, Dy-UNet has a small parameter size of 184 K, making it one of the most compact models among the compared approaches, while still achieving superior performance.

5. Discussion

This study introduced Dy-UNet, a novel lightweight deep learning architecture specifically designed for automated sperm segmentation in phase-contrast microscopy images. The key contributions of this work may advance the field of automated sperm analysis through a synergistic network that addresses fundamental challenges in morphological assessment.

Our cross-resolution architecture maintains high-resolution representations throughout the network while simultaneously extracting multi-scale semantic features. This design fundamentally differs from traditional encoder-decoder networks and preserves fine-grained morphological information essential for accurate sperm segmentation. This finding aligns with advances in human pose estimation and object detection, where high-resolution networks have shown superior performance in the localization of small and thin structures. DSConv, as a specialized feature extraction module, enables adaptation to the curved and serpentine morphology characteristic

of sperm cells, while the GHPA mechanism efficiently models long-range spatial dependencies across multiple feature dimensions.

The high sensitivity achieved by our model (0.57546) is particularly valuable for clinical applications, as it ensures comprehensive detection of sperm cells, including those with subtle morphological abnormalities. In fertility assessment, the failure to detect abnormal sperm can lead to incomplete morphological analysis and potentially affect clinical decision-making. The superior sensitivity of Dy-UNet compared to existing methods suggests that our approach can provide more reliable morphological assessments, which may improve the accuracy of male fertility diagnosis and treatment planning. Moreover, the lightweight architecture with 184 K parameters enables deployment on portable and point-of-care devices, which democratizes access to advanced sperm analysis technology.

While our model achieved state-of-the-art performance on average metrics, analysis of failure cases revealed that Dy-UNet occasionally struggles with highly overlapping or tightly clustered sperm cells. In such cases, the boundaries between adjacent sperm can be ambiguous even for human experts. Future work could explore instance segmentation approaches that explicitly model individual sperm instances.

The dataset used in this study, while valuable, has a limited size and diversity compared to the full spectrum of cases encountered in clinical practice. Future work should focus on the collection of larger, more diverse datasets that capture this variability and enable more robust model training.

6. Conclusion

This study presents Dy-UNet, a novel lightweight deep learning architecture that achieves state-of-the-art performance in automated sperm segmentation while maintaining exceptional computational efficiency. Through the integration of cross-resolution processing, DSConv, grouped multi-axis attention, and FCRA, Dy-UNet effectively captured the unique morphological characteristics of sperm cells and provided accurate segmentation across diverse imaging conditions. With only 184 K parameters, Dy-UNet is well-suited for deployment in resource-constrained clinical settings and enables accessible, objective sperm morphology assessment. The comprehensive evaluation and ablation studies validated the effectiveness of each proposed component and demonstrated significant improvements over existing approaches. While limitations remain and future work is needed to expand the model's capabilities and validate its generalization across diverse datasets and

imaging modalities, Dy-UNet represents a meaningful advance toward automated, accurate, and accessible male fertility assessment.

Acknowledgments

None.

Funding

None.

Conflict of interest

Runwei Guan is the Guest Editor of this special issue, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization: Qingxin Zhang, Runwei Guan

Data curation: Qingxin Zhang, Shuai Lei, Shaofeng Liang

Formal analysis: Qingxin Zhang, Shuai Lei, Xiaoyi Gu, Runwei Guan

Funding acquisition: Ming Yuan, Yanhua Fei, Runwei Guan

Investigation: Ming Yuan, Yanhua Fei

Methodology: Qingxin Zhang, Shuai Lei, Shaofeng Liang

Project administration: Runwei Guan

Resources: Ming Yuan, Yanhua Fei

Software: Tong Wu

Supervision: Ming Yuan, Yanhua Fei, Runwei Guan

Validation: Qingxin Zhang, Shuai Lei, Xiaoyi Gu, Shaofeng Liang, Tong Wu, Ming Yuan, Yanhua Fei

Visualization: Tong Wu

Writing—original draft: Qingxin Zhang, Shuai Lei, Xiaoyi Gu, Shaofeng Liang

Writing—review & editing: Tong Wu

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The SegSperm dataset can be accessed online. (<https://www.kaggle.com/datasets/shubhamchauhan2222/segspem/data>)

References

1. World Health Organization. *Infertility Prevalence Estimates, 1990–2021*; 2023. Available from: <https://www.who.int/publications/i/item/978920068315> [Last accessed on 2025
2. Cox CM, Thoma ME, Tchangalova N, *et al.* Infertility prevalence and the methods of estimation from 1990 to 2021: A systematic review and meta-analysis. *Hum Reprod Open.* 2022;2022(4):hoac051. doi: 10.1093/hropen/hoac051
3. World Health Organisation. *WHO Laboratory Manual for the Examination and Processing of Human Semen*; 2021. Available from: <https://www.who.int/publications/i/item/9789240030787> [Last accessed on 2025 Sep 26].
4. Chang V, Saavedra JM, Castañeda V, Sarabia L, Hitschfeld N, Härtel S. Gold-standard and improved framework for sperm head segmentation. *Comput Methods Progr Biomed.* 2014;117(2):225–237. doi: 10.1016/j.cmpb.2014.06.018
5. Lee RKK, Hou JW, Ho HY, *et al.* Sperm morphology analysis using strict criteria as a prognostic factor in intrauterine insemination. *Int J Androl.* 2002;25(5):277–280. doi: 10.1046/j.1365-2605.2002.00355.x
6. Katz DE, Overstreet JW, Samuels SJ, Niswander PW, Bloom TD, Lewis EL. Morphometric analysis of spermatozoa in the assessment of human male fertility. *J Androl.* 1986;7(4):203–210. doi: 10.1002/j.1939-4640.1986.tb00913.x
7. Chen W, Song H, Dai C, *et al.* Automated Sperm Morphology Analysis Based on Instance-Aware Part Segmentation. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE; 2024:17743–17749. doi: 10.1109/icra57147.2024.10611339
8. Dai C, Zhang Z, Huang J, *et al.* Automated non-invasive measurement of single sperm's motility and morphology. *EEE Trans Med Imaging.* 2018;37(10):2257–2265. doi: 10.1109/tmi.2018.2840827
9. Dai W, Wu Z, Liu R, *et al.* Automated non-invasive analysis of motile sperms using sperm feature-correlated network. *IEEE Trans Autom Sci Eng.* 2024;22:3960–3970. doi: 10.1109/tase.2024.3404488
10. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ.* 1992;304(6840):1491–1494. doi: 10.1136/bmj.304.6840.1491
11. Lewandowska E, Węsierski D, Mazur-Milecka M, Liss J, Jezierska A. Ensembling noisy segmentation masks of blurred sperm images. *Comput Biol Med.* 2023;166:107520. doi: 10.1016/j.compbimed.2023.107520
12. İlhan HO, Serbes G, Aydın N. Dual Tree Complex Wavelet Transform Based Sperm Abnormality Classification. In: *2018 41st International Conference on Telecommunications*

- and Signal Processing (TSP)*; 2018:1-5.
doi: 10.1109/tsp.2018.8441431
13. Shaker F, Monadjemi SA, Alirezaie J, Naghsh-Nilchi AR. A dictionary learning approach for human sperm heads classification. *Comput Biol Med*. 2017;91:181-190.
doi: 10.1016/j.compbiomed.2017.10.009
14. Chang V, Garcia A, Hitschfeld N, Härtel S. Gold-standard for computer-assisted morphological sperm analysis. *Comput Biol Med*. 2017;83:143-150.
doi: 10.1016/j.compbiomed.2017.03.004
15. Riordon J, McCallum C, Sinton D. Deep learning for the classification of human sperm. *Comput Biol Med*. 2019;111:103342.
doi: 10.1016/j.compbiomed.2019.103342
16. Ilhan HO, Serbes G. Sperm morphology analysis by using the fusion of two-stage fine-tuned deep networks. *Biomed Sign Process Control*. 2021;71:103246.
doi: 10.1016/j.bspc.2021.103246
17. Yüzkat M, Ilhan HO, Aydin N. Multi-model CNN fusion for sperm morphology analysis. *Comput Biol Med*. 2021;137:104790.
doi: 10.1016/j.compbiomed.2021.104790
18. Liu R, Wang M, Wang M, Yin J, Yuan Y, Liu J. Automatic microscopy analysis with transfer learning for classification of human sperm. *Appl Sci*. 2021;11(12):5369.
doi: 10.3390/app11125369
19. Zhang Y, Zhang J, Zha X, Zhou Y, Cao Y, Chen D. Improving Human Sperm Head Morphology Classification With Unsupervised Anatomical Feature Distillation. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. United States: IEEE; 2022:1-5.
doi: 10.1109/isbi52829.2022.9761633
20. Shaker F. Human Sperm Head Morphology dataset (HuSHeM). Mendeley Data, V3; 2018.
doi: 10.17632/tt3yj2pf38.3
21. Ilhan HO, Serbes G, Aydin N. A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods. *Med Biol Eng Comput*. 2020;58(5):1047-1068.
doi: 10.1007/s11517-019-02101-y
22. Aktas A, Serbes G, Huner Yigit M, Aydin N, Uzun H, Osman Ilhan H. Hi-LabSpermMorpho: A novel expert-labeled dataset with extensive abnormality classes for deep learning-based sperm morphology analysis. *IEEE Access*. 2024;12:196070-196091.
doi: 10.1109/access.2024.3521643
23. Sapkota N, Zhang Y, Li S, *et al*. Shmc-Net: A Mask-Guided Feature Fusion Network for Sperm Head Morphology Classification. In: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*; 2024:1-5.
doi: 10.1109/isbi56570.2024.10635339
24. Yi WJ, Park KS, Paick JS. Parameterized characterization of elliptic sperm heads using fourier representation and wavelet transform. In: *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*. Vol. 2. United States: IEEE; 1998:974-977.
doi: 10.1109/iembs.1998.745610
25. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer Assisted Intervention*. Berlin: Springer; 2015:234-241.
doi: 10.1007/978-3-319-24574-4_28
26. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging*. 2020;39(6):1856-1867.
doi: 10.1109/tmi.2019.2959609
27. Ruan J, Xiang S, Xie M, Liu T, Fu Y. MALUNet: A multi-attention and light-weight UNet for skin lesion segmentation. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. United States: IEEE; 2022:1150-1156.
doi: 10.1109/bibm55620.2022.9995040
28. Ruan J, Xie M, Gao J, Liu T, Fu Y. EGE-UNet: An efficient group enhanced UNet for skin lesion segmentation. In: *Lecture Notes in Computer Science*. Berlin: Springer Nature Switzerland; 2023:481-490.
doi: 10.1007/978-3-031-43901-8_46
29. Dai W, Liu R, Wu Z, *et al*. Exploiting scale-variant attention for segmenting small medical objects. *IEEE Trans Neural Netw Learn Syst*. 2026:1-18.
doi: 10.1109/tnnls.2025.3645355
30. Oktay O, Schlemper J, Folgoc LL, *et al*. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv*. Preprint postd online 2018.
doi: 10.48550/arXiv.1804.03999
31. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Vol. 11045. Berlin: Springer; 2018:3-11.
doi: 10.1007/978-3-030-00889-5_1
32. Chen J, Lu Y, Yu Q, *et al*. TransUNet: Transformers Make

- Strong Encoders for Medical Image Segmentation. *arXiv*. Preprint posted online 2021.
doi: 10.48550/arXiv.2102.04306
33. Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv*. Preprint posted online 2020.
doi: 10.48550/arXiv.2010.11929
34. Cao H, Wang Y, Chen J, *et al.* Swin-unet: Unet-like pure transformer for medical image segmentation. In: *Lecture Notes in Computer Science*. Berlin: Springer Nature Switzerland; 2023:205-218.
doi: 10.1007/978-3-031-25066-8_9
35. Liu Z, Lin Y, Cao Y, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. United States: IEEE; 2021:9992-10002.
doi: 10.1109/iccv48922.2021.00986
36. Howard AG, Zhu M, Chen B, *et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv*. Preprint posted online 2017.
doi: 10.48550/arXiv.1704.04861
37. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. United States: IEEE; 2018:4510-4520.
doi: 10.1109/cvpr.2018.00474
38. Ma N, Zhang X, Zheng HT, Sun J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In: *Computer Vision - ECCV 2018 (Lecture Notes in Computer Science)*. Berlin: Springer International Publishing; 2018:122-138.
doi: 10.1007/978-3-030-01264-9_8
39. Yu C, Wang J, Peng C, Gao C, Yu G, Sang N. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. In: *Computer Vision - ECCV 2018 (Lecture Notes in Computer Science)*. Berlin: Springer International Publishing; 2018:334-349.
doi: 10.1007/978-3-030-01261-8_20
40. Zhao H, Qi X, Shen X, Shi J, Jia J. ICNet for real-time semantic segmentation on high-resolution images. In: *Computer Vision - ECCV 2018 (Lecture Notes in Computer Science)*. Berlin: Springer International Publishing; 2018:418-434.
doi: 10.1007/978-3-030-01219-9_25
41. Jiang J, Wang M, Tian H, Cheng L, Liu Y. LV-UNet: A lightweight and vanilla model for medical image segmentation. In: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. United States: IEEE; 2024:4240-4246.
doi: 10.1109/bibm62325.2024.10822465
42. Howard A, Sandler M, Chen B, *et al.* Searching for mobileNetV3. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. United States: IEEE; 2019:1314-1324.
doi: 10.1109/iccv.2019.00140
43. Yu C, Xiao B, Gao C, *et al.* Lite-HRNet: A Lightweight high-resolution network. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. United States: IEEE; 2021:10435-10445.
doi: 10.1109/cvpr46437.2021.01030
44. Qi Y, He Y, Qi X, Zhang Y, Yang G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. United States: IEEE; 2023:6047-6056.
doi: 10.1109/iccv51070.2023.00558
45. Lewandowska E, Węsierski D, Mazur-Milecka M, Liss J, Jezierska A. SegSperm - a dataset of sperm images for blurry and small object segmentation [Data set]. Gdańsk University of Technology; 2023.
doi: 10.34808/6wm7-1159
46. Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK. Recurrent residual U-Net for medical image segmentation. *J Med Imaging (Bellingham)*. 2019;6(01):014006.
doi: 10.1117/1.jmi.6.1.014006
47. Yang Z, Peng X, Yin Z, Yang Z. Deeplab_v3_plus-net for image semantic segmentation with channel compression. In: *2020 IEEE 20th International Conference on Communication Technology (ICCT)*. United States: IEEE; 2020:1320-1324.
doi: 10.1109/icct50939.2020.9295748