Electronic Supplementary Materials

For https://doi.org/10.1631/jzus.A2500360

Deep-learning-enabled automatic gene abnormality detection via fluorescence in situ hybridization

Lemin SHI[1], Yuqiang ZHANG[2], Haoyu QI[3], Chengyue LU[4], Menglei HU[5], Mingye LI[6], Dianxin SONG[1], Hao ZHANG[1], Xin FENG[1]✉, Ping GONG[1]✉, Shan JIANG[3]✉

[1]School of Computer Science and Technology, School of Life Science and Technology, Changchun University of Science and Technology, Changchun 130022, China
[2]Modern Industry College, Jilin Jianzhu University, Changchun 130119, China
[3]Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China
[4]Department of Electromechanical Engineering, Centre for Artificial Intelligence and Robotics, University of Macau, Macau 999078, China
[5]Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver V6T 1Z4, Canada
[6]School of Computing and Information Systems, The University of Melbourne, Parkville 3010, Australia

Section S1: Materials and Methods

Sample Processing: Peripheral blood samples (2-5 ml, typically 3 ml) were collected into a centrifuge tube and mixed with 10 ml of hypotonic solution (0.075 M). This concentration helps swell the cell nucleus and separate the chromosomes (Döhner et al., 2000; Ouillette et al., 2008), making them easier to stain and analyze later). The mixture was incubated in a 37 °C water bath for 25 minutes. Following incubation, 2 ml of fixative (methanol/acetic acid in a 3:1 ratio) was added, and the sample was centrifuged at 2000 rpm for 8 minutes. The supernatant was quickly discarded, and 10 ml of fixative was added to resuspend the cell nucleus pellet, followed by another centrifugation at 2000 rpm for 10 minutes. This process was repeated with 5 ml of fixative for nuclei fixation. The supernatant was then discarded, leaving approximately 1 ml of cell nucleus suspension, which was then thoroughly mixed and placed onto a pre-cooled slide.

Hybridization and Staining: To detect prognostically significant chromosomal abnormality among chromosome 13 for the early diagnosis of CLL, multiplexed FISH treatment was performed on the peripheral blood lymphocytes from the collected human blood samples using XL RB1/DLEU/LAMP1 (Jelinek et al., 2024) probes (MetaSystems, Altlussheim). (Fig. S1b) demonstrates the genetic mapping of the corresponding fluorescence probes on chromosome 13 (Kushwaha et al., 2024; Zhang et al., 2024). The sample treatment procedure involved three primary steps: (i) sample collection and preprocessing, (ii) denaturation and hybridization, and (iii) washing and counterstaining. Firstly, whole blood samples were stained with the tricolor FISH probes and centrifuged at 2000 rpm for 20 minutes. The leukocyte layer was then mixed with 37-℃-preheated 5 mL 75mM KCl in a tube, followed by resuspension and incubation in a 37 ℃ water bath for 25 minutes. After incubation, cells were mixed using methanol/acetic acid (3:1 vol/vol) fixative solution three times and centrifuged at 2000 rpm for 10 minutes. Following supernatant removal and resuspension, the cell suspension was air-dried on pre-cooled glass slides at room temperature for density observation. Slides were further treated with 2x SSC and subsequently in 100% ethanol for 2 minutes and air-dried thoroughly at room temperature. In the (ii) step, the hybridization probe solution was prepared for 5μL per cell sample and applied to the cell region of slides, followed by covering with a cover glass and sealing. The slide was then denatured at 75 ℃ for 2 minutes and hybridized at 37 ℃ overnight. Furthermore, in a dark or low-light environment, the slides were washed in 2x SSC to remove the coverslip, followed by washing in 0.4x SSC at 72 ℃ for 2 minutes, 2x SSC with 0.05% Tween-20 for 30 s, and DI water subsequently. After air-drying in the dark room, slides were stained with 5 μL DAPI for 10 minutes and stored at -20 ℃ for further microscopy.

Microscope Presetting: ET-DAPI filter set (49000 ET-DAPI, Chroma Technology) was used for fluorescent observation. Upon initializing the fluorescent microscope (Olympus BX61), the stage moved to the far-right and bottom positions to facilitate sample placement. After placing the sample in the designated slot on the stage, the corresponding slot is selected in the software to start the acquisition process. The stage then moved the sample under the objective lens and raised it to the preset focusing height of 13,835.0μm. The stage was moved to align the objective lens with the center of the coverslip before image acquisition.

Automated Image Acquisition: With the objective lens fixed, the focusing and acquisition were operated through stage movement in vertical and horizontal directions. During the pre-acquisition process, the stage was moved downward in 0.8μm increments, capturing 300 preliminary images. Notably, the short working distance of the 60× oil objective (0.1 mm to 0.2 mm) required precise focusing accordingly to avoid out-of-focus images. An energy gradient algorithm was then employed to automatically identify the clearest preliminary images, and the corresponding height is chosen as the pre-focusing height. Subsequently, the stage was adjusted to five intervals (0.44 μm each) above the pre-focusing height to process the main acquisition process. During this process, image capturing was commenced at 11 focal planes while moving downward to complete the acquisition for one field of view. Images were captured using an Olympus BX61 motorized microscope equipped with a grayscale camera (FLIR GS3-U3-51S5M-C) and a motorized color filter set. including an ET-DAPI UV filter (#49000, Chroma Technology) and a FISH trichrome filter set including Orange (#49305, Chroma Technology), Aqua (#49302, Chroma Technology), and Green (#49303, Chroma Technology) filters. To minimize ambient light interference, images were captured at 4 units of camera exposure in a light-sealed environment.

Section S2: Working Principle of Automatic Acquisition

The automatic acquisition function primarily relies on a smart focusing process, guided by an algorithm that evaluates the clarity of each image. This system, in addition to a commercial motorized microscope equipped with an objective stage controller and oil immersion lens, also includes a fluorescence light source and associated accessories. Notably, a grayscale camera is used instead of a color camera due to its superior sensitivity and quantum efficiency. The key principles and procedures are outlined below:

Presetting and Sample Placement: To reduce the coarse focusing time for each sample, the microscope slides are pre-aligned in the center of the field of view and marked. And the preset height is adjusted according to the focal distance required for nuclear size. Specifically, in this work, the preset height for the Olympus BX61 microscope is set to 13,835.0μm, which is slightly higher than the average focusing height determined from pretests. The prepared sample slide is then positioned at the marked point.

Automatic Focusing and Image Selection: Starting from the preset height, the focal distance is increased incrementally by the object stage controller. For the 60× oil immersion lens used in this work, the interval between each layer is set at 0.44μm, based on the lens's working distance. The FAST system captures 11 images per field of view, and each image's gradient energy (EOG) value is calculated to evaluate image clarity and information richness. The EOG is given as follows:

$$EOG = \sum^{x}\sum^{y}\left(f_x^2 + f_y^2\right) \tag{S1.1}$$

$$f_x = f(x+1, y) - f(x, y) \tag{S1.2}$$

$$f_y = f(x, y+1) - f(x, y) \tag{S1.3}$$

where $f(x, y)$ is pixel intensity at the coordinates (x, y) within the image. Since a higher EOG value indicates clearer edges, the image with the highest EOG value is automatically selected for further enhancement, ensuring optimal image clarity. This refined process ensures that the system efficiently captures the clearest possible image in each field of view, improving the accuracy and consistency of automated analysis.

Section S3: Working Principle of Image Enhancement

Image enhancement was conducted using the multi-scale Retinex32 algorithm and an image fusion model. An optimized watershed-based algorithm was employed for nuclear segmentation.

Image Decomposition and Difference of Gaussians: A microscopic image consists of two primary components (Sun et al., 2022): irradiation and reflection components. The irradiation component corresponds to the low-frequency part of the image, while the reflection component represents the high-frequency part. Both components accurately capture the essential information of the image. Thus, the mathematical expression of the highest violet clarity image $I(x,y)$ can be expressed as:

$$I(x,y) = L(x,y)R(x,y) \tag{S2.1}$$

where $L(x,y)$ represents the irradiation component, which includes ambient light, and $R(x,y)$ represents the reflection component, which contains the key information such as nucleus features within the image.

By transforming both sides of **Eq. (2.1)** into the logarithmic domain, the impact of incident light is isolated and mitigated, thus revealing the inherent attributes of the target nucleus objects, as follows:

$$Log[I(x,y)] = Log[L(x,y)] + Log[R(x,y)] \tag{S2.2}$$

Through convolving a Gaussian kernel function with $I(x,y)$, $L(x,y)$ is approximated. Consequently, $R(x,y)$ can be expressed as:

$$Log[L(x,y)] = Log[I(x,y) * G(x,y)] \tag{S2.3}$$

$$Log[R(x,y)] = Log[I(x,y)] - Log[I(x,y) * G(x,y)] \tag{S2.4}$$

$$G(x,y) = Ke^{-(x^2+y^2)/\sigma^2} \tag{S2.5}$$

where $G(x,y)$ is the Gaussian kernel function, * represents the convolution operation, and σ represents the Gaussian surround scale parameter, controlling the range of the kernel function. The extreme values of $Log[R(x,y)]$ are individually calculated, followed by linear mapping and quantization to pixel values within [0, 255] to obtain $R(x,y)$.

Additionally, we used a tailored multiscale difference of Gaussians (DoGs) to enhance nucleus details in low-frequency subimages and mitigate the imbalance of incident light (Kim et al., 2015). The globally enhanced image first undergoes Gaussian kernel transformations G_1, G_2, and G_3 to produce three differently blurred images:

$$L'_1 = G_1 * L(x,y),\ L'_2 = G_2 * L(x,y),\ L'_3 = G_3 * L(x,y) \tag{S2.6}$$

where G_1, G_2, and G_3represent the Gaussian kernels with σ=1.0, 2.0, and 3.0, respectively. Image contour details at different scales D_1, D_2, and D_3 can be correspondingly calculated as:

$$D_1 = L'_1,\ D_2 = L'_1 - L'_2,\ D_3 = L'_2 - L'_3 \tag{S2.7}$$

Multiscale image clarity is calculated using the Laplacian sharpness evaluation algorithm, and the image

sharpness ω can be expressed as:

$$\omega = \sum_x \sum_y |G(x,y)| * L \tag{S2.8}$$

$$L = \frac{1}{6}\begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \tag{S2.9}$$

where $G(x,y)$ represents the convolution with the Laplace operator at a specific pixel point. During multiscale calculations, segmentation functions are required to avoid oversaturation and maintain the average grayscale level difference within the image:

$$sgn(D_1) = \begin{cases} -1, & D_1 < 0 \\ 1, & D_1 \geq 0 \end{cases} \tag{S2.10}$$

$$D^* = \left(1 - \omega_1 * \mathrm{sig}(D_1)\right) * D_1 + \omega_2 * D_2 + \omega_3 * D_3 + alpha * Src \tag{S2.11}$$

Here, ω_1, ω_2, and ω_3 represent the multiscale image sharpness values calculated using **Eq. (2.10),** corresponding to D_1, D_2, and D_3, respectively. D^* represents the overall contour details of the merged image at different scales. Based on multiple experimental results, an alpha parameter was introduced and set to 0.1, which retains 10% of the original image information. This balance between detail enhancement and original image preservation significantly enhances details without over-processing, resulting in a natural appearance. This method is well-suited for analyzing FISH and other microscopic images, leading to better image quality. by superimposing the multi-scale detail image with the original image, a multi-scale enhanced image is obtained

$$R'(x,y) = I(x,y) + D^* \tag{S2.12}$$

Cut-off Frequency: A Gaussian high-pass filter is applied to the enhanced high-frequency component $R'(x,y)$, followed by mask construction with $R'(i,j)$. The conventional Gaussian high-pass filter is defined as:

$$H(u,v) = 1 - e^{-D^2(u,v)/2D_0^2} \tag{S2.13}$$

where $D(u,v)$ indicates the distance to the frequency center. High-frequency filtering removes zero-frequency components, causing the average intensity of the background to become zero.

The position corresponding to $R'(i,j)$ is set to zero, and the image center is used as the origin to create rectangular blocking masks of various lengths and widths. The SNR (mean/variance) (Bhutto et al., 2024) versus cutoff frequency is calculated for images with rectangular blocking masks of various sizes:

$$Q = \sum_{m=0}^{m=50} (R'(x,y) * mask(h-m:h+m, w-m:w+m)) \tag{S2.14}$$

$$\bar{X} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} Q_{(i,j)}}{MN} \tag{S2.15}$$

$$SD = \sqrt{\frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}(R'(i,j) - \bar{Q}^2)} \quad \text{(S2.16)}$$

$$SNR = \sum_{n=0}^{m=50}\frac{M}{SD} \quad \text{(S2.17)}$$

where Q represents the rectangular blocking mask of different sizes, $w, h,$ and m denote the image width, height, and blocking window of different sizes, respectively. $R'(i,j)$ denotes the pixel value at the *i*-th row and *j*-th column, $\bar{Q}$ denotes the multiscale mean value, and the optimal cutoff frequency D_0 is determined according to the peak SNR applied to the multiscale mask.

Local contrast factor: Gaussian filtering enhances details in dark areas of nuclei and reduces interference from noises and vignetting, but it lacks sufficient edge sharpening and clarity in high-illumination areas. To address this, an image fusion model is employed to derive a Local contrast factor for the $R(x,y)$ component in the frequency domain. The average image brightness $avgG$ and Local contrast factor Coe (Wang et al., 2018) are given as:

$$avgG = G_r/G_{Num} \quad \text{(S2.18)}$$

$$Coe = \frac{255}{avgG} \quad \text{(S2.19)}$$

where G_r is the grayscale value and G_{Num} is the number of pixels. the number of pixels.

To achieve more precise control over highlight adjustments while minimizing the impact on other image areas, ensuring detail retention and balanced exposure, and preventing overexposure, the top 0.1% of brightness values are selectively extracted. These values represent the most critical parts of the image, allowing for fine-tuning in the brightest highlight areas. By selectively processing these brightness values, image quality can be enhanced while preserving key information to the greatest extent.

The Local contrast factor, along with the optimal cutoff frequency, is fused with the enhanced high-frequency subimage, thereby improving edge information and enhancing high-frequency details and overall image quality. The image fusion model is given as:

$$H'(u,v) = Coe \cdot H(u,v) + L(x,y) \quad \text{(S2.20)}$$

Fourier transformation is then applied to the resultant image, followed by the combination of the inverse Fourier transform with the incident component $H'(u,v)$. The gain difference is further adjusted using C_{oe} to prevent overexposure, resulting in the final enhanced image output.

Section S4: Working Principle of Image Segmentation

Local maxima: The Euclidean distance metric is applied to both the original and the enhanced images, calculating the distance between each non-zero pixel $p(x_1, y_1)$ and its nearest background pixel $q(x_2, y_2)$ as follows (Shen et al., 2019):

$$dis(p,q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{S3.1}$$

where $p(x_1, y_1)$ represents the coordinates of an object pixel and $q(x_2, y_2)$ denotes the coordinates of a background pixel. The distance set $D[dis(p,q)]$ is thereby derived, with its maximum and minimum values denoted as D_{max} and D_{min}, respectively.

The varying locations of nucleus pixel points are transformed into corresponding grayscale information, defined as the distance map $G(x,y)$. The converted grayscale values are calculated as follows (Sharma et al., 2022):

$$G(x,y) = 255 \cdot \frac{|dis(p,q) - D_{min}|}{|D_{max} - D_{min}|} \tag{S3.2}$$

The local extrema are then extracted from $G(x,y)$, peaks and their coordinates across all distance maps are marked, and the distances between neighboring peak points are calculated as (Jardim et al., 2022):

$$dis_{\min}\big(p(A), q(B)\big) = \sum_{i,j}^{m,n} \sqrt{(a_i - c_j)^2 + (b_i - d_j)^2} \tag{S3.3}$$

where $A(a_i, b_i)$, $B(c_j, d_j)$ denote peak points obtained from the enhanced image.

Gradient mapping: This algorithm estimates the gradient of the image by using the Sobel operator and calculates the magnitude and direction of the gradient, thereby generating an expression similar to an HSV (hue, saturation, value) image.

The Sobel operator is employed to calculate the gradient of the image:

$$G_x = Sobel_x(Image) \tag{S3.4}$$

$$G_y = Sobel_y(Image) \tag{S3.5}$$

where G_x and G_y are the gradients of the image in the x and y directions, respectively. Therefore,

$$v = \sqrt{G_x^2 + G_y^2} \tag{S3.6}$$

$$\theta = arccos\left(\frac{G_x}{v}\right) \tag{S3.7}$$

where G and θ are the magnitude and direction of the image gradient, respectively.

Calculate the direction angle θ of the optical flow based on G_y.

$$\theta = \theta * \left(\frac{sing(G_y)}{2\pi}\right) \quad \text{(S3.8)}$$

To adjust the angle from possible negative values to positive values, we consider that in some cases, when G_y is negative, the angle θ will be mapped to the negative semi-axis. By adding 0.5 and taking the modulo 1 (%1), we ensure that the angle θ can be properly mapped to the [0, 1) interval, thereby correctly expressing the hue in the HSV color space.

$$\theta = \theta + 0.5 \quad \text{(S3.9)}$$

$$v = \frac{v}{\max(v)} \quad \text{(S3.10)}$$

HSV mapping and data processing: The saturation S of all elements is initialized to 1, and the horizontal angle is mapped to the integer range [0, 1], where a is the integer part of h.

$$a_{ij=}\lfloor h_{ij} \rfloor \quad \text{(S3.11)}$$

$a_{ij}$is the value of a at position (i, j), and h_{ij} is the value of h at the same position.

Calculate the decimal part of the horizontal angle as:

$$b = h \quad \text{(S3.12)}$$

$$b *= a \quad \text{(S3.13)}$$

Define a zero array p of the same size as the amplitude, and the interpolation parameters based on the horizontal angle are t, q.

$$t = v * b \quad \text{(S3.14)}$$

$$q = v - t \quad \text{(S3.15)}$$

Assume $sdata$ is a flattened four-channel data array containing the optical flow information calculated from the original image data. Each pixel consists of four values (amplitude v, interpolation parameter t, occupancy parameter p, interpolation parameter q), which are scaled and converted to 8-bit integer format and scaled to the range [0,255].

$$\text{sdata} = \begin{bmatrix} v_{11} & t_{11} & p_{11} & q_{11} \\ v_{12} & t_{12} & p_{12} & q_{12} \\ \vdots & \vdots & \vdots & \vdots \\ v_{nm} & t_{nm} & p_{nm} & q_{nm} \end{bmatrix} \quad \text{(S3.16)}$$

Create an index array: According to the integer part a of the optical flow direction angle θ, select an index array to rearrange the color channels. idx is an index array calculated based on the optical flow direction, which is used to reorder the data in sdata according to the optical flow direction of each pixel to correctly map the hue, saturation, and brightness information, and reshape the data to generate an image with three color channels.

$$HSV\_image = sdata[idx] \quad \text{(S3.17)}$$

Convex Hull Boundary: Converting grayscale images to binary images highlights the structural features of cell nuclei and removes noise through morphological opening operations, enabling clearer feature extraction. Subsequently, we utilize two contour detection algorithms: Simple Contour Approximation and C89-L1 Contour Approximation, to detect contours in the image. For accurate identification of convex defects on the contours, we conduct two detection phases: the first detection is for the preliminary identification of contours and convex defect points, and the second detection involves detailed analysis of deeper defect points, defined as the pixel distance from the convex hull boundary to the farthest point on the contour, including iterating and classifying defect points.

Additionally, we calculate the minimum distance between peak points and connect these peak points with common convex hull points. Finally, the distances between these connections and the intersection points on the common convex hull are calculated to further analyze contour features.

Optimized Watershed: After two passes, the distance from the convex hull boundary to the farthest point on the contour is detected and marked in pixels, classifying and marking points with larger defect depths. In the connecting line of the common convex hull points $C(x_c, y_c)$ of adjacent peak points $P_i(x_i, y_i)$ and $P_{i+1}(x_{i+1}, y_{i+1})$, the distance between the intersection points of the lines connecting the peak points to C, represented as D, is given by:

$$D = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + (x_{i+1} - x_c)^2 + (y_{i+1} - y_c)^2} \quad \text{(S3.18)}$$

When the local maximum equals the global maximum (Jardim, et al., 2022), the distance between peaks $D = \frac{dis_{min}}{2}$ meets the optimal geodesic distance for most nuclei segmentation. This effectively segments each nucleus while maintaining reasonable spacing between segmented regions. However, for overlapping nuclei, D does not equal $\frac{dis_{min}}{2}$ and fails to meet the optimal geodesic distance for segmentation. To ensure the horizontal line passes through both low and high peak regions, the distance D' should be closer to the small peak region, with $D' = D/2$ to achieve optimal segmentation. Note that if the distance is too small, over-segmentation of nuclei may occur.

Section S5: Performance comparison of different image enhancement methods

To evaluate the enhancement performance of FAST, enhancing results of FAST are compared with traditional image enhancement methods, including contrast limited adaptive histogram (CLAHE), single-scale Retinex (SSR), and Homomorphic filtering (HF).

Evaluation Metrics for Image Enhancement: The evaluation is standardized through several parameters:

1) Tenengrad: it evaluates sharpness by calculating the image gradient, with higher values indicating sharper images. Tenengrad is calculated as:

$$Tenengrad = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(S\text{obel}I(i,j))^2} \tag{S4.1}$$

where $I(i,j)$ represents the pixel value of the image at position (i,j). M,N denotes the height and width of the image. T indicates the sharpness, $S\text{obel}I(i,j)$ shows the gradient value after applying the operator.

2) Contrast: it is determined by calculating the standard deviation of pixel values. The greater contrast means more differences between light and dark areas. It can be expressed as:

$$Contrast = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2} \tag{S4.2}$$

where X_i is the second grayscale value in the image, i is the total number of pixels, and μ is the average value of all grayscale pixels.

3) Spatial Frequency: it is measured using Fourier transform, reflecting the frequency of details in the image. The higher frequencies indicate more details. The grayscale variation in the horizontal and vertical directions can be expressed as row frequency (RF) and column frequency (CF):

$$RF = \sqrt{\frac{1}{M*N}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}(I(i,j-1) - I(i,j))^2} \tag{S4.3}$$

$$CF = \sqrt{\frac{1}{M*N}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}(I(i-1,j) - I(i,j))^2} \tag{S4.4}$$

where $I(i,j)$ represents the pixel value of the image at position (i,j). M and N denote the height and width of the image respectively.

$$Spatial\ Frequency = \sqrt{RF^2 + CF^2} \tag{S4.5}$$

4) Standard Deviation: it reflects the distribution of image pixel values, with a larger standard deviation

relative to higher contrast. The standard deviation ($StdDev$) of pixel values is calculated by:

$$Mean = \frac{1}{M * N}\sum_{i=1}^{M}\sum_{j=1}^{N} I(i,j) \tag{S4.6}$$

$$StdDev = \sqrt{\frac{1}{M * N}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j) - M\text{ean})^2} \tag{S4.7}$$

where $I(i,j)$ represents the pixel value of the image at position (i,j). M,Ndenotes the height and width of the image.

Image Enhancement Evaluation Results: Compared with other methods, the enhancement function of FAST offers significant advantages in enhancing image quality. It excels particularly in improving image clarity, contrast, and detail, proving highly valuable for applications requiring high-quality image processing. The calculated results of the evaluation metrics are as follows:

Table SN4-1: performance comparison of the image enhancement methods

Sample	Method	Tenengrad	Contrast	Spatial frequency	Standard deviation
	Original	27.8930	2.2931	2.1359	26.6211
	CLAHE	34.8991	4.1387	2.8716	33.7330
1	SSR	39.6033	5.6423	3.3535	41.2158
	HF	27.3487	2.0399	2.0143	23.9109
	FAST	**66.8459**	**10.4073**	**4.5562**	**65.0439**
	Original	28.0469	2.4089	2.1896	24.5014
	CLAHE	40.6740	5.4394	3.2931	35.8694
2	SSR	40.4881	6.4834	3.5942	38.6020
	HF	27.3703	2.1190	2.0541	21.5866
	FAST	**69.3687**	**10.1580**	**4.5030**	**65.3544**
	Original	19.9402	1.4094	1.6747	17.5255
	CLAHE	30.2948	3.6170	2.6858	26.5583
3	SSR	37.9520	5.7193	3.3770	34.7232
	HF	19.5287	1.2416	1.5714	15.5825
	FAST	**83.5987**	**12.8715**	**5.0698**	**77.8721**
	Original	42.8403	4.5821	3.0174	37.1505
	CLAHE	50.7689	6.9770	3.7261	45.4387
4	SSR	48.0749	6.9986	3.7329	48.1122
	HF	41.9512	4.1924	2.8864	32.6971
	FAST	**73.2701**	**11.3854**	**4.7625**	**68.5246**

	Original	26.1232	2.1546	2.0710	24.0392
	CLAHE	37.9646	4.8489	3.1089	35.5464
5	SSR	44.7234	7.0554	3.7504	44.2779
	HF	25.4521	1.8568	1.9223	21.4386
	FAST	**75.1987**	**11.8355**	**4.8589**	**72.2549**
	Original	31.7055	2.8004	2.3636	28.8947
	CLAHE	41.5946	5.2664	3.2414	38.9005
6	SSR	43.8349	6.1161	3.4928	43.9859
	HF	31.0349	2.4890	2.2282	25.8147
	FAST	**70.7244**	**10.3056**	**4.5356**	**70.0438**
	Original	43.2200	4.8607	3.1149	41.9808
	CLAHE	49.8985	6.9707	3.7306	50.4672
7	SSR	52.8680	11.4847	4.7883	61.7537
	HF	42.2665	4.4584	2.9832	36.5472
	FAST	**82.2189**	**11.7771**	**4.8498**	**91.2843**
	Original	15.0337	1.0795	1.4590	12.2170
	CLAHE	25.9858	3.3578	2.5844	21.0458
8	SSR	23.5356	3.7987	2.7492	19.4702
	HF	14.7276	0.9460	1.3653	11.1038
	FAST	**61.9287**	**8.1351**	**4.0284**	**58.4464**
	Original	26.0783	2.1264	2.0603	22.5198
	CLAHE	38.6798	5.0197	3.1654	33.8910
9	SSR	15.1416	1.1536	1.5110	14.1682
	HF	25.4542	1.8574	1.9258	19.9069
	FAST	**105.4847**	**19.1251**	**6.1805**	**94.7853**
	Original	42.6883	4.7402	3.0746	39.8610
	CLAHE	50.5938	7.1107	3.7657	49.4942
10	SSR	51.8703	10.8186	4.6458	58.1845
	HF	41.7269	4.3486	2.9448	34.5193
	FAST	**78.9599**	**13.3348**	**5.2290**	**78.0832**

Section S6: Performance Comparison of Different Nuclear Segmentation Methods

For a comprehensive evaluation of FAST's segmentation abilities, its segmentation results are compared with traditional segmentation methods and deep learning methods. The traditional methods include marker-based watershed (MW) and morphological segmentation (MS), and the deep learning method is selected as common U-Net-based semantic segmentation. Specifically, the semantic segmentation model is trained by the Kaggle 2018 Data Science Bowl nucleus segmentation dataset (Caicedo et al., 2019).

Evaluation Metrics for Image Segmentation: The evaluation of segmentation is based on the number of identifications and misidentifications within the enhanced images from FAST. Through manual detection, the delineated "nuclei" are classified into true positives (TP) and false positives (FP), and the "backgrounds" are sorted as true negatives (TN) and false negatives (FN). They can be defined as:

a) True positives (TP): the number of "nuclei" that are correctly identified and segmented.

b) False positives (FP): the number of "nuclei" that are mistakenly identified, including over-segmentation that recognizes non-cellular areas as cellular ones.

c) True negatives (TN): the number of "backgrounds" that are correctly identified as non-cellular backgrounds.

d) False negatives (FN): the number of "backgrounds" that are mistakenly identified, which means actual cell nucleus areas are missed by the algorithm.

And the evaluation includes four metrics:

i) Accuracy measures the proportion of cell nucleus and backgrounds correctly identified:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{S5.1}$$

ii) Precision reflects the accuracy within the regions predicted as cell nucleus:

$$Percision = \frac{TP}{TP + FP} \tag{S5.2}$$

iii) Recall indicates the proportion of actual cell nucleus that are correctly identified:

$$Recall = \frac{TP}{TP + FN} \tag{S5.3}$$

iv) The F1 score, a harmonic mean of precision and recall, is used to comprehensively assess the performance of the algorithm:

$$F1(Score) = \frac{2 \times (Recall \times Percision)}{Recall + Percision} \tag{S5.4}$$

Nuclear Segmentation Evaluation Results: The counted number and the calculated evaluation metrics for segmentation are shown as:

Table SN5-1: performance comparison of the segmentation methods

Subject (Count)	Method	TP	FP	FN	TN	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
1 (91)	MW	86	2	3	0	94.51%	97.73%	96.63%	97.18%
	MS	77	14	0	0	84.62%	84.62%	100.00%	91.67%
	U-net	53	38	0	0	58.24%	58.24%	100.00%	73.61%
	FAST	**91**	**0**	**0**	**0**	100.00%	100.00%	100.00%	100.00%
2 (166)	MW	103	19	44	0	62.05%	84.43%	70.07%	76.58%
	MS	153	3	10	0	92.17%	98.08%	93.87%	95.92%
	U-net	87	79	0	0	52.41%	52.41%	100.00%	68.77%
	FAST	**163**	**2**	**1**	**0**	98.19%	98.79%	99.39%	99.09%
3 (131)	MW	57	1	73	0	43.51%	98.28%	43.85%	60.64%
	MS	105	5	21	0	80.15%	95.45%	83.33%	88.98%
	U-net	69	60	2	0	52.67%	53.49%	97.18%	69.00%
	FAST	**126**	**3**	**2**	**0**	96.18%	97.67%	98.44%	98.05%
4 (146)	MW	112	33	1	0	76.71%	77.24%	99.12%	86.82%
	MS	128	2	16	0	87.67%	98.46%	88.89%	93.43%
	U-net	89	57	0	0	60.96%	60.96%	100.00%	75.74%
	FAST	**142**	**3**	**1**	**0**	97.26%	97.93%	99.30%	98.61%
5 (167)	MW	125	40	2	0	74.85%	75.76%	98.43%	85.62%
	MS	152	4	11	0	91.02%	97.44%	93.25%	95.30%
	U-net	96	71	0	0	57.49%	57.49%	100.00%	73.00%
	FAST	**164**	**1**	**2**	**0**	98.20%	99.39%	98.80%	99.09%
6 (147)	MW	110	35	2	0	74.83%	75.86%	98.21%	85.60%
	MS	135	3	9	0	91.84%	97.83%	93.75%	95.74%
	U-net	90	57	0	0	61.22%	61.22%	100.00%	75.95%
	FAST	**146**	**0**	**1**	**0**	99.32%	100.00%	99.32%	99.66%
7 (137)	MW	130	4	3	0	94.89%	97.01%	97.74%	97.38%
	MS	63	71	3	0	45.99%	47.01%	95.45%	63.00%
	U-net	69	67	1	0	50.36%	50.74%	98.57%	66.99%
	FAST	**137**	**0**	**0**	**0**	100.00%	100.00%	100.00%	100.00%
8 (181)	MW	171	1	9	0	94.48%	99.42%	95.00%	97.16%
	MS	144	28	9	0	79.56%	83.72%	94.12%	88.62%
	U-net	77	95	9	0	42.54%	44.77%	89.53%	59.69%
	FAST	**176**	**5**	**0**	**0**	97.24%	97.24%	100.00%	98.60%
9	MW	165	5	2	0	95.93%	97.06%	98.80%	97.92%
	MS	95	75	2	0	55.23%	55.88%	97.94%	71.16%

(172)	U-net	98	74	0	0	56.98%	56.98%	100.00%	72.59%
	FAST	**168**	**1**	**3**	**0**	97.67%	99.41%	98.25%	98.82%
	MW	148	5	3	0	94.87%	96.73%	98.01%	97.37%
10	MS	102	52	2	0	65.38%	66.23%	98.08%	79.07%
(156)	U-net	84	69	3	0	53.85%	54.90%	96.55%	70.00%
	FAST	**154**	**2**	**0**	**0**	98.72%	98.72%	100.00%	99.35%

The FAST's segmentation method demonstrated exceptional performance in the task of cell nucleus segmentation, with an average accuracy of 98.3%. It excelled particularly in handling complex scenarios such as closely packed or overlapping nuclei. This indicates that FAST not only offers high stability and efficiency across various conditions but also has significant advantages in computational complexity. In comparison, the MW method performed well on certain datasets, especially in precision and F1 score, but exhibited some variability in performance across other datasets. The MS method achieved high precision and recall on specific datasets but had a lower overall accuracy. The U-net method performed better in recall but showed lower precision and F1 scores. All methods faced limitations in terms of over-segmentation and missed nuclei.

Evaluation from Medical Laboratory Center: In evaluating ten validation image sets, FAST's results closely matched the high standards set by medical experts, indicating its ability to achieve expert-level accuracy in handling complex cell nucleus images. This consistency demonstrates not only FAST's efficiency in cell nucleus segmentation but also its capability to accurately identify nuclei, meeting stringent expert assessment criteria.

Description of the accuracy evaluation of cell segmentation and counting of 10 images including 00221834

Objective: To compare and analyze the accuracy of manual segmentation versus algorithmic segmentation of fluorescence in situ hybridization (FISH) images, with a focus on the consistency of cell shape, size, and alignment with actual observed cells.

1. Image Data for Evaluation: The evaluation images were provided by Professor Shi Lemin from Changchun University of Science and Technology. These images were collected in 2022 using the Olympus BX61 fluorescence microscope. The dataset includes fluorescence images of peripheral blood B lymphocytes from patients with chronic lymphocytic leukemia, specifically DAPI channel images with the following IDs:00221834、00378449、00282649、00292393、00292397、00292530、00292619、00292620、00316557、00337484、

2. Evaluation Criteria:

Shape Consistency: The similarity between the shape of cells post-segmentation and their original shape.

Size Consistency: The degree to which the size of segmented cell nuclei matches the original cell nuclei size.

Segmentation Integrity: The preservation of cell structures post-segmentation, avoiding over-segmentation or under-segmentation.

3. Segmentation Algorithm: The segmentation algorithm used is an optimized and improved watershed segmentation method (0TS algorithm) developed by Professor Shi Lemin. This method performs automatic segmentation of all cells in the fluorescence images, with algorithm parameters adjusted to accommodate various image types.

4. Results Analysis:

Cell Count Consistency:

The analysis will include a detailed comparison of cell counts between manual and algorithmic segmentations to assess their consistency and accuracy.:

Image ID	Manual Analysis (Count)	Algorithmic Segmentation (Count)
00221834	91	91
00378449	166	166
00282649	131	129
00292393	146	144
00292397	167	166
00292530	147	146
00292619	137	137
00292620	180	178
00316557	172	170
00337484	156	154

The missing cells after algorithm segmentation are mainly reflected in the cells at the edge of the image, and the number of cells after the overall segmentation of the image is consistent with the number of manual segmentation.

Shape consistency: In most cases, the OWS algorithm segmentation can accurately identify and track the cell shape, and the cell morphology is consistent with the real cell morphology.

Size consistency: The OWS algorithm can accurately calculate the area of the cell. In most cases, the algorithm segmentation can accurately identify and track the cell shape when the cells are dense or overlapping, and the cell size is consistent with the real cell morphology.

Segmentation integrity: The cell nucleus after OWS algorithm segmentation has no over-segmentation or missed segmentation phenomenon. Some cells that have not been segmented are in the edge area of the image, which does not affect the evaluation of cell integrity.

5. Conclusion:

The OWS algorithm segmentation shows a high degree of accuracy in maintaining the consistency of cell shape and size, especially in identifying and tracking cell morphology. It also performs well in maintaining segmentation integrity. Although the cell nucleus at the edge of the image is not fully identified, this does not affect the overall cell integrity evaluation, ensuring consistency with manual cell counting.。

Evaluator's unit (department): **Jilin Provincial People's Hospital Medical Laboratory Center**

Assessor:

Date:2022.2.2

Figure SN 5-1: The evaluation report from the Medical Laboratory Center.

Section S7: Resource Usage of Different Network Models

The resource consumption of different models (Somu et al., 2021; Menghani, 2023) is compared during the training or inference processes, including time, GPU utilization, CPU utilization, and memory usage. The FAST system is configured with an Intel Core i7-12700 processor, an NVIDIA 3090Ti graphics card, 16 GB of memory, and a Windows 10 operating system. The focusing time is calculated as (500 ×1.55)/60 = 12.5 minutes, with a CPU usage of 10% and memory usage between 1.5-3 GB. The full automated multi-sampling time is (36 ×11 ×1.8)/60 = 11.88 minutes, with a CPU usage of 30% and memory usage between 2-4 GB. Image enhancement takes 2 minutes, with a CPU usage of 10% and memory usage between 2-3 GB. The CPU usage during the image enhancement phase is 10%. Cell nucleus segmentation takes 2 minutes, with a CPU usage of 1% and memory usage of 1 GB. When tasks are executed sequentially, the total CPU utilization is 51% and total memory usage ranges from 1.5 to 4 GB. FAST is better suited for applications requiring rapid processing of large volumes of images, such as clinical research and real-time analysis. For tasks that need a balance between precision and efficiency, this indicates that the system maintains a relatively balanced resource demand across different stages and effectively utilizes CPU and memory resources to complete tasks.

Table SN6-1: Practical Performance Comparison

Model	Sample for Model Building (Number)	Time Consumption (Hours)	GPU Utilization (%)	CPU Utilization (%)	Memory (GB)
U-Net	> 1000	2-4	25-33	10-20	10-14
ENET	> 1000	1.5-3	13-21	10-15	6-10
SegNet	> 1000	3-5	33-42	10-20	12-16
ICNet	> 1000	2.4-4.5	17-25	10-15	8-12
LEDnet	> 1000	2.4-4.5	21-29	10-15	10-14
DeepLabV3+ResNet18	> 1000	3-5	29-38	10-15	10-14
DeepLabV3+ResNet34	> 1000	3.5-5.5	38-46	10-20	12-16
DeepLabV3+ResNet50	> 1000	4.5-6.5	42-50	15-20	14-18
DeepLabV3+ResNet101	> 1000	5.5-7.5	50-58	15-25	16-20
FAST	**1**	**0.5**	**None**	**51%**	**4.3-8.3**

Datasets: Two datasets were employed in this study: one for training the ResNet152 network for cell abnormality classification and the other for training commercial networks in the comparative analysis of systematic performance in our dual-step approach.

The first ground-truth dataset included 4600 images of cells with stained nuclei and fluorescence signals from tricolor FISH channels. The most-focused nuclei images in the DAPI channel were selected and annotated using Labelme to obtain nucleus contour annotations, followed by nuclei segmentation on the n-layer-fused images across ten samples (n=1, 3, 5, 7, 9). Each n-layer-fused image in the three channels was then manually labeled as either "good" or "bad".

The second dataset consisted of 26,733 images with 18,713 designated for the training set and 8,020 for the validation set. In this dataset, the most-focused images across ten samples were segmented through our segmentation method. 9-layer-fused images per sample underwent nucleus contour identification accordingly, establishing a ground truth dataset with 3819 images of nuclei. It was then enlarged to 26, 733 images through classical data augmentation techniques such as translation, rotation, and scaling. Wherein, for each FISH channel, images featuring fewer than two fluorescence signals were labeled as bad-truth for chromosomal abnormality, and good-truth for chromosomal normality otherwise.

Training: The ResNet152 network was trained for 20 epochs on datasets using n-layer fusion (n=1, 3, 5, 7, 9) with a weight decay of 0.1 and an initial learning rate of 0.001. To mitigate overfitting, dropout regularization was employed to randomly disable individual neurons after each fully connected layer. After training, the corresponding model, including structures and parameters was loaded and trained on the 9-layer-fused dataset for parameter optimization. 30% of images from the dataset were then randomly selected as a test set and input into the trained model for prediction.

The training set is input into the ResNet152 model for automatic learning training, with an initial learning rate of 1e-3, 40 epochs set during training, batch_size=64, and 300 training times. In order to avoid overfitting of the model, global average pooling is used instead of the fully connected layer to prevent overfitting. The classification results are output through the classifier. After the training is completed, the corresponding model, including the structure and parameters, is read, and the red, green, and blue 9-layer fluorescence fusion images are imported into the ResNet152 model as a data set; by optimizing the model parameters, the prediction results are obtained.

The precision, recall rate and F1-score of the cell abnormality analysis results are calculated. TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Table SN6-2: Performance comparison of average diagnostic performance using different networks across ten samples

Method	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
AlexNet (Battleday et al., 2020)	6665	890	357	108	94.20	94.92	98.41	96.63
DenseNet121 (Singh et al., 2025)	6720	850	300	150	94.39	95.73	97.82	96.76
DenseNet169 (Biancalani et al., 2021)	6780	820	250	170	94.76	96.44	97.55	97.00
VGG16 (Yang et al., 2021)	6864	797	158	201	95.52	97.75	97.15	97.45
VGG19 (Le Goallec et al., 2022)	6857	821	165	177	95.74	97.65	97.48	97.57
InceptionV4 (Li et al., 2022)	6900	800	140	180	96.01	98.01	97.46	97.73
ResNet50 (Hossain et al., 2022)	6943	900	79	98	97.79	98.87	98.61	98.74
ResNet101(Rolfe et al., 2024)	6943	901	79	97	97.81	98.87	98.62	98.75
FAST	6951	897	71	101	97.86	98.57	98.99	98.78

The ResNet50, ResNet101, and FAST models exhibit the highest performance across accuracy, precision, recall, and F1-score when screening cells with abnormal fluorescence characteristics. These models, in particular, show significant improvements compared to AlexNet, VGG16, VGG19, and DenseNet169. For instance, the ResNet50 (Hossain, et al., 2022) model achieves 97.79% accuracy, 98.87% precision, and 98.61% recall, while the ResNet101 (Rolfe et al., 2024) model shows similar performance with 97.81% accuracy, 98.87% precision, and 98.62% recall. The FAST model outperforms them with 97.86% accuracy, 98.57% precision, and 98.99% recall. These models also have a lower missed detection rate, indicating their reliability in detecting abnormal fluorescence characteristics. The DenseNet121 (Singh, et al., 2025) and DenseNet169 (Biancalani, et al., 2021) models also perform well, with DenseNet121 showing 94.39% accuracy, 95.73% precision, and 97.82% recall,

while DenseNet169 performs slightly better with 94.76% accuracy, 96.44% precision, and 97.55% recall. The InceptionV4 (Li, et al., 2022) model performs strongly with 96.01% accuracy and 98.01% precision, but still lags slightly behind the top performers like ResNet50 and FAST in recall and F1-score.

Section S8: Comparative Evaluation with Commercial Tools

To comprehensively assess the performance of FAST, the comparative analysis includes qualitative and quantitative comparisons with commercial tools.

Selection of Counterpart Segmentation Tool: Based on publicly available information from the official websites of these tools and relevant literature, a comparative summary is shown in this table:

Tool	Cost	Automation	Flexibility	Difficulty of Learning	Adaptability Scenario
V7	Paid	Automatic labeling pre-labeling	High	Medium	Autonomous driving, medical imaging, surveillance video
Labelbox	Paid	Automatic labeling	Wide	Medium	Various annotation tasks
Scale AI	Paid	Automatic labeling	Limited	Low	Autonomous driving scenarios, satellite image analysis, complex target detection
SuperAnnotate	Paid	Automatic labeling	High	Low	Complex image segmentation, target detection
DataLoop	Paid	Automatic labeling	Wide	Medium	Industrial detection, medical imaging
Playment	Paid	Automatic labeling pre-labeling	Customizable	Low	Target detection and segmentation in autonomous driving
Supervise.ly	Paid	Automatic labeling	Customizable	Medium	Image segmentation, target detection
Hive	Paid	Automatic labeling	Basic	Medium	Video surveillance, advanced image analysis
LabelMe	Free	Manual labeling	High	Low	Basic image annotation tasks

Table SN7-1: Commercial Tools for Labeling

Since FAST is designed to achieve accurate and efficient nuclear segmentation while minimizing costs, LabelMe (Khan and Al-Habsi, 2020; Kipke et al., 2022) was selected as the counterpart product for further quantitative analysis for several reasons:

First, in nuclear identification for FISH images, ensuring precise cell nucleus recognition is paramount. Although many paid tools, such as V7, DataLoop and SuperAnnotate, excel in automation and processing efficiency, they are generally designed for broader applications rather than targeting low-contrast tasks such as cell nucleus segmentation tasks. As a result, they often lack the flexibility needed to handle complex images that

involve challenges like cell nucleus overlapping, uneven staining, and impurities. In contrast, LabelMe allows for flexible manual adjustments tailored to experimental needs, effectively addressing these challenges.

Furthermore, as an open-source tool, LabelMe offers a cost-effective solution for research projects with limited budgets, aligning well with the original goal of FAST to reduce the costs associated with expensive high-resolution microscopy. Additionally, LabelMe is user-friendly and does not require complex learning or operation processes, such as extensive parameter adjustments, making it the ideal counterpart for FAST.

Dataset Generation of Different Segmentation Tools: Using fifteen microscopy images that include FISH-labeled nuclei, the contour labels are obtained through LabelMe and FAST respectively.

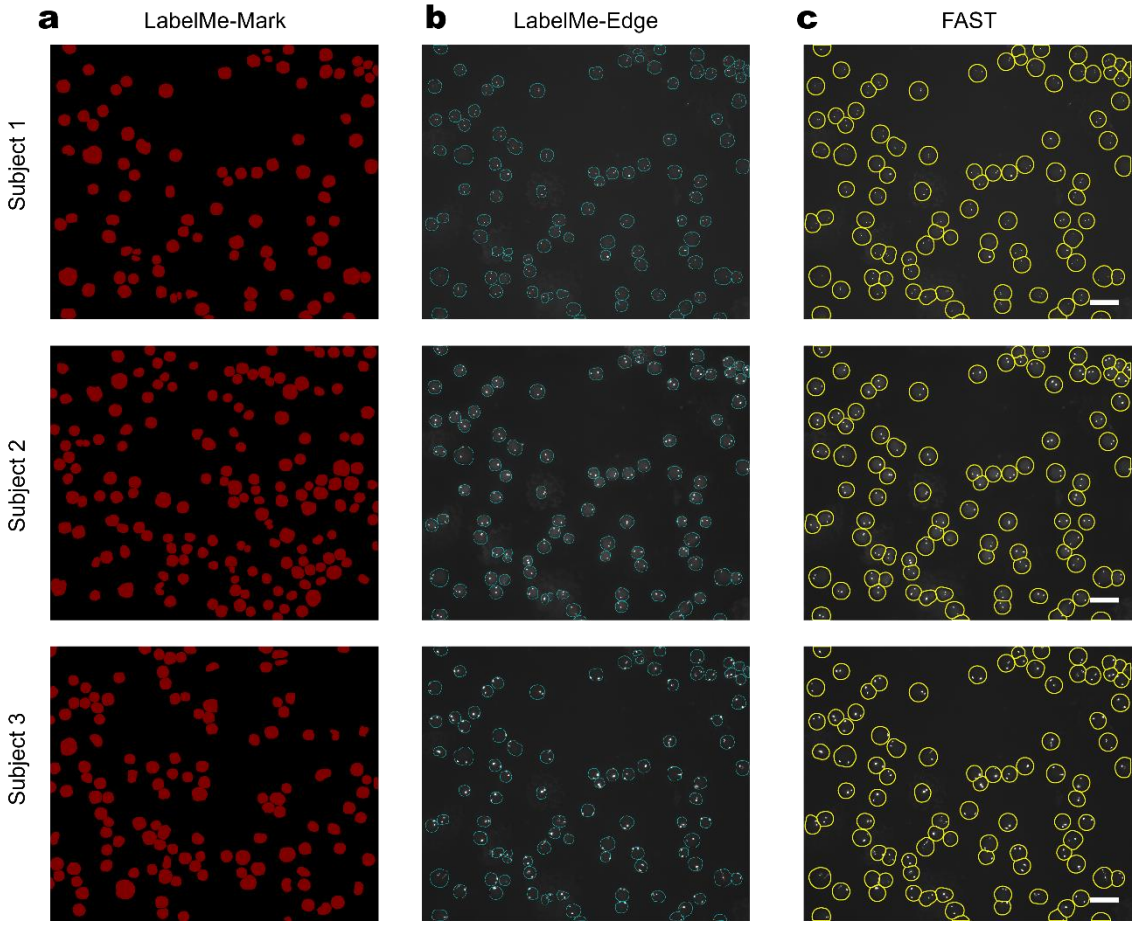


Figure SN7-1: The contour labels from (**a**) LabelMe-Mark, (**b**) LabelMe-Edge, and (**c**) FAST.

The nucleus image datasets, each containing 80 FISH images, are generated based on these labels respectively: (1) LabelMe-based dataset contains cell nucleus labels through LabelMe's annotation; (2) FAST-based dataset includes cell nucleus labels through FAST's nuclear segmentation function.

Selection of Evaluation Networks*:* To quantitively compare the performance of the FAST method with LabelMe in image segmentation tasks, the optimal evaluation network is selected based on LabelMe-based dataset.

The models are trained for the same number of epochs, with Dice loss as the loss function, and the adaptive moment estimation method as the optimizer, under a stepwise learning rate decay. These models represent the cutting edge of image segmentation technology, which encompass different network architectures and characteristics. And the evaluation metrics include Dice, Jaccard, Recall and Accuracy:

(1) Dice and Jaccard: they quantify the similarity between the predicted results and the ground truth labels.

(2) Recall: it is defined as the proportion of true positive pixels to total number of pixels that should have been segmented.

(3) Accuracy: it is the ratio of correctly classified pixels to the total number of pixels in the image.

Table SN7-2: Comparison of Commercial Networks Based on LabelMe Dataset

Networks	Dice	Jaccard	Recall	Accuracy (%)
U-Net (Ronneberger et al., 2015)	0.7859	0.6715	0.7828	98.02
ENET (Paszke et al., 2016)	0.6945	0.5686	0.6983	97.44
SegNet (Badrinarayanan et al., 2017)	0.6722	0.5387	0.6414	96.93
ICNet (Zhao et al., 2018)	0.6250	0.5042	0.6075	97.22
LEDnet (Wang et al., 2019)	0.6870	0.5602	0.6495	97.59
DeepLabV3+ResNet18 (Chen et al., 2018)	0.7708	0.6577	0.7530	98.12
DeepLabV3+ResNet34 (Chen, et al., 2018)	0.7740	0.6617	0.7612	98.10
DeepLabV3+ResNet50 (Chen, et al., 2018)	0.8077	0.7006	0.7926	98.19
DeepLabV3+ResNet101 (Chen, et al., 2018)	0.8142	0.7061	0.8050	98.32
FAST	0.8157	0.7125	0.8137	98.28

The results showed that the DeepLabV3+ResNet152 method significantly outperformed other segmentation networks, demonstrating exceptional performance in segmentation tasks. This provides strong evidence for selecting the optimal image segmentation strategy.

Quantitative Comparison of FAST and LabelMe*:* In the previous comparison, the dataset annotated by LabelMe demonstrates superior performance, especially when combined with the advanced segmentation model DeepLabV3+ResNet152. Therefore, the comparison of FAST-based datasets and LabelMe-based datasets is conducted with the model DeepLabV3+ResNet152. For the comprehensively comparable analysis of commercial LabelMe annotation software and FAST, the evaluation includes crucial factors in this table:

Table SN7-3: Performance Comparison of LabelMe and FAST

Method	Recall	Accuracy (%)	Contour resolution (Segment/nucleus)	Processing Speed (Image/hour)	Automation	Image pre-processing
LabelMe	0.8261	98.51	30	0.25	No	No
FAST	0.8086	96.66	130	0.1	Yes	Yes

The nucleus detection and annotation accuracy of LabelMe and FAST are comparable in terms of recall value (0.83 and 0.81 respectively) and accuracy (99% and 97% respectively) mentioned above. However, FAST performs LabelMe in key aspects such as contour resolution and processing speed. Specifically, the contour resolution is defined by the average segments of each nuclear contour, with higher number of segments contains finer details of nuclear edges. The contour resolution of FAST is nearly ten folds of that in LabelMe, which means more robustness in complex nuclear shapes.

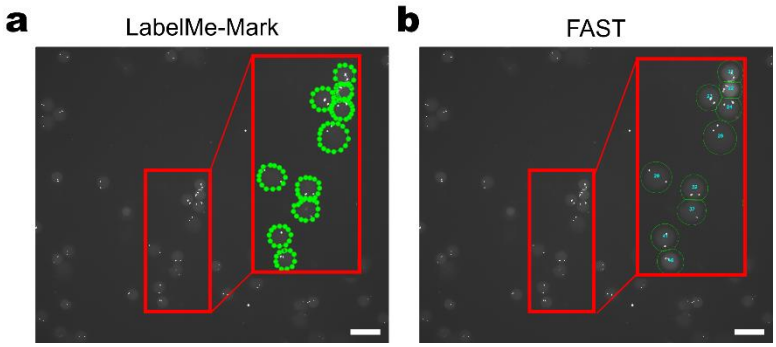


Figure SN7-2: Contour resolution of LabelMe and FAST.

Additionally, FAST excels in automation and image preprocessing functions. The processing speed of FAST, which includes the image acquisition and analysis process, is twice that of LabelMe, making it a more efficient and effective tool for segmentation tasks. Furthermore, FAST offers high automation and a user-friendly interface, positioning it as a competitive tool in the market. These results demonstrate a multi-faceted evaluation of FAST and LabelMe, showing that FAST produces superior outcomes overall.

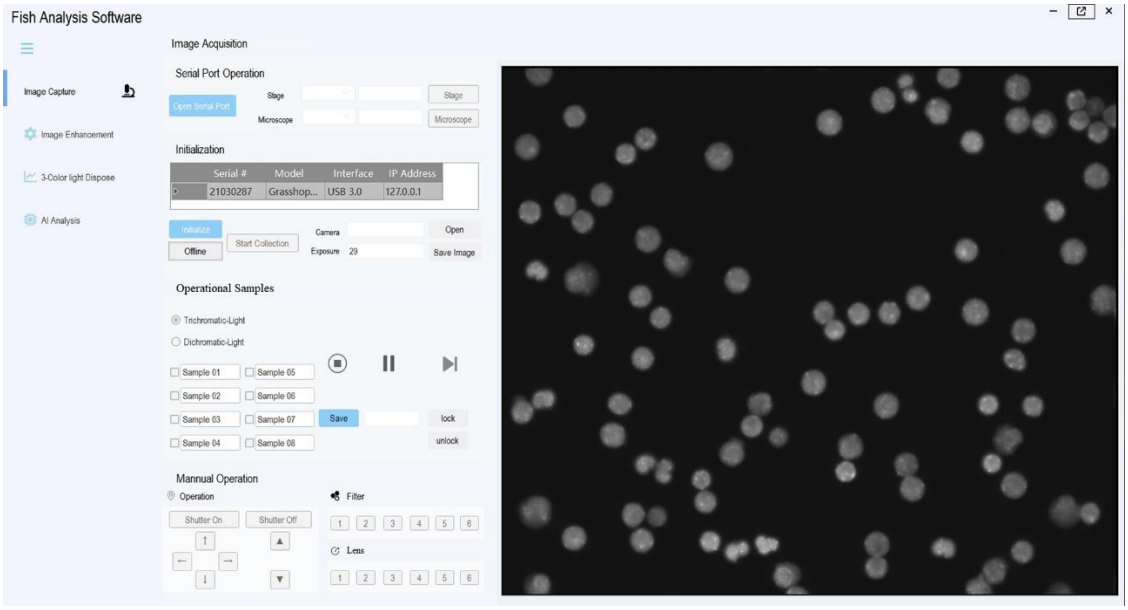
Conclusion: A comprehensive evaluation is conducted to compare the performance of the FAST and LabelMe. From the perspective of algorithm, the accuracy and recall are evaluated through DeepLabV3+ResNet152 model, showing comparable results. And taking the practical use into consideration, the contour resolution, speed, and other useful functions are also assessed, in which FAST outperforms. The FAST method, with its strong

interpretability, multi-scale filtering, and improved cell nucleus segmentation techniques, makes it particularly effective for large-scale image processing tasks, real-time analysis, and clinical research. Despite not showing a clear advantage in segmentation accuracy in this evaluation, FAST's efficiency and stability make it a valuable tool in practical applications, such as cancer detection, where it meets the high demands for processing large volumes of image data quickly.

Section S9: The best balance point for base layer and multi-layer feature fusion

With the increase in n, the quantity and clarity of detected fluorescence signals within each nucleus improve, ultimately reaching a relatively saturated state. The qualitative and quantitative evaluation of the composite output through multilayer feature fusion, respectively, we employ average EOG in each channel to assess image clarity of the fused images, while using average ratio of clear signal at $n = i$, defined as $\frac{number\ of\ clear\ signal\ at\ n=i}{number\ of\ clear\ signal\ at\ n=11}$, to evaluate our method's efficacy in restoring missing or obscured signals with the increase in n. Both image clarity and clear signal ratio progressively increase until n=9, beyond which a plateau is reached, indicating n=9 as an optimal balance between signal compensation and processing cost for FISH analysis in this work.

Section S10: Operation Guidance of Graphical User Interface



1) **System Startup and Initialization:** Double-click to run the Fish Analysis Software.exe to start the application. After the system completes initialization, the "Start Collection" button will become inactive until initialization is complete. Click the "System Offline" button to put the camera into a suspended state.

2) **Saving Address Selection:** Click the "Save Photo" button to save the current image from the camera's field of view. The image will be saved by default in the root directory of the C drive, and the system will notify you when the save is successful.

3) **Camera Settings:** Manually open the serial port and enter the exposure time in the exposure settings area. Click the "Open Camera" button to configure the camera's exposure time based on the settings. Camera Information Parameters are as below:

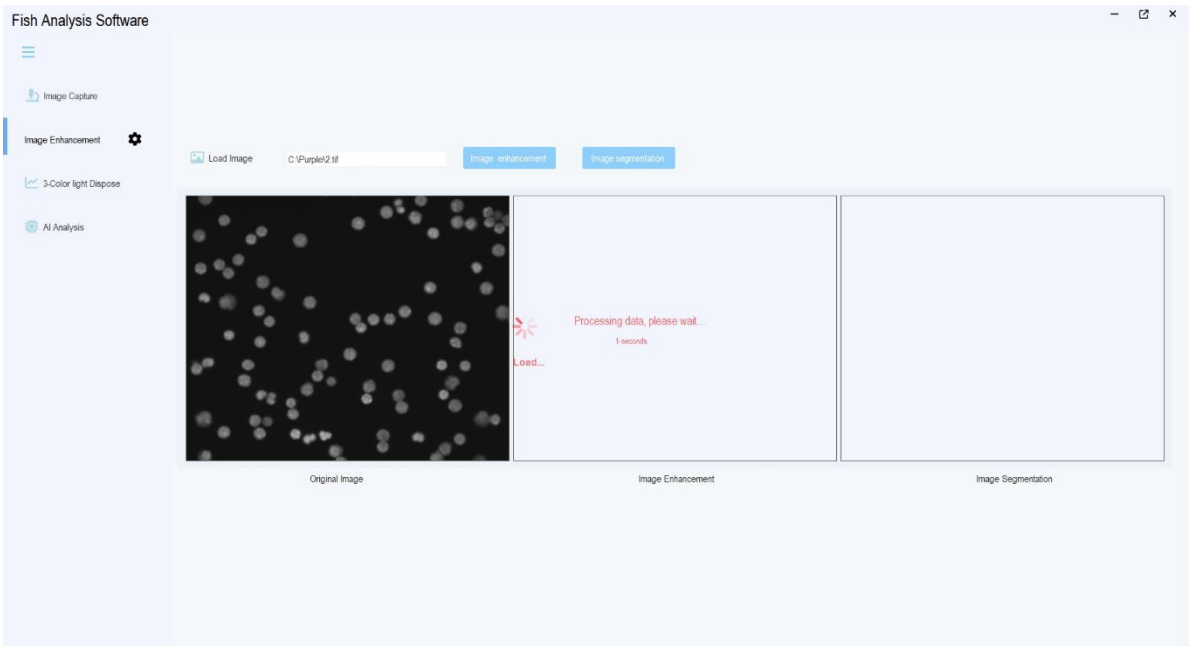
Table SN9-1: Camera parameters

Serial	The camera's serial number
Model	The camera's name and model
Interface	The camera's data interface
IP Address	The camera's IP address

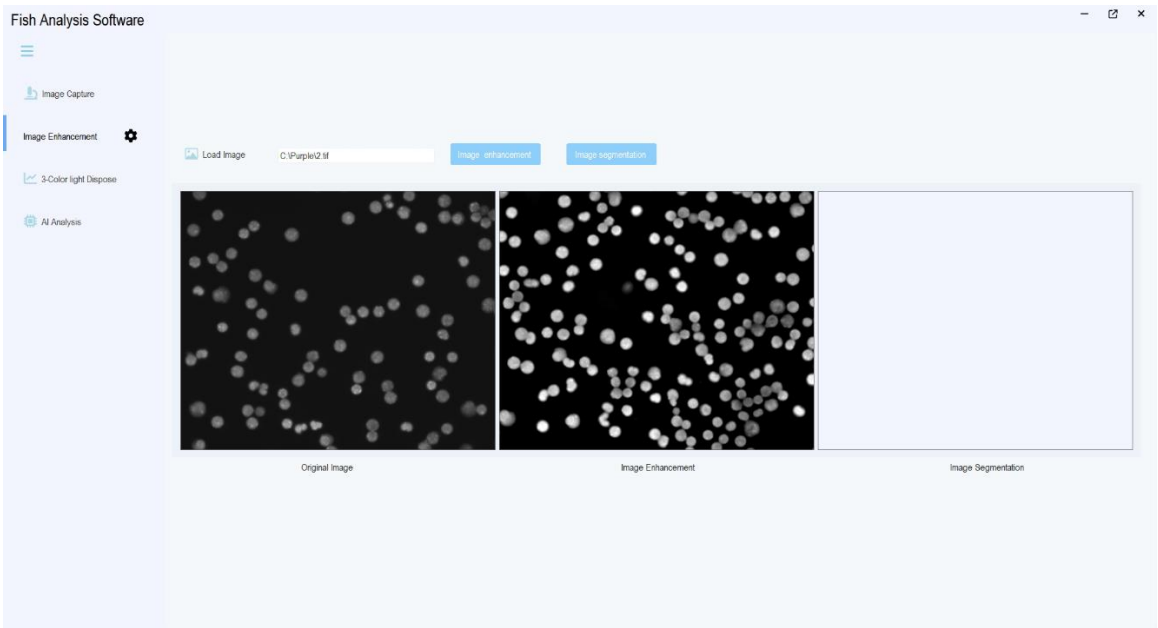
4) **Mode Selection and Image Collection:** Choose different modes to switch between different filters; this system supports 2-color and 3-color filters. After placing the sample in the corresponding position on the stage, rename the sample and check the checkbox to confirm the sample. Once confirmed, click the "Start Collection" button to collect the sample data.

5) **Microscope and Component Control:** Use the buttons on the software interface to operate the microscope, stage, filters, and light shields, simulating the physical control of the actual microscope and its components.

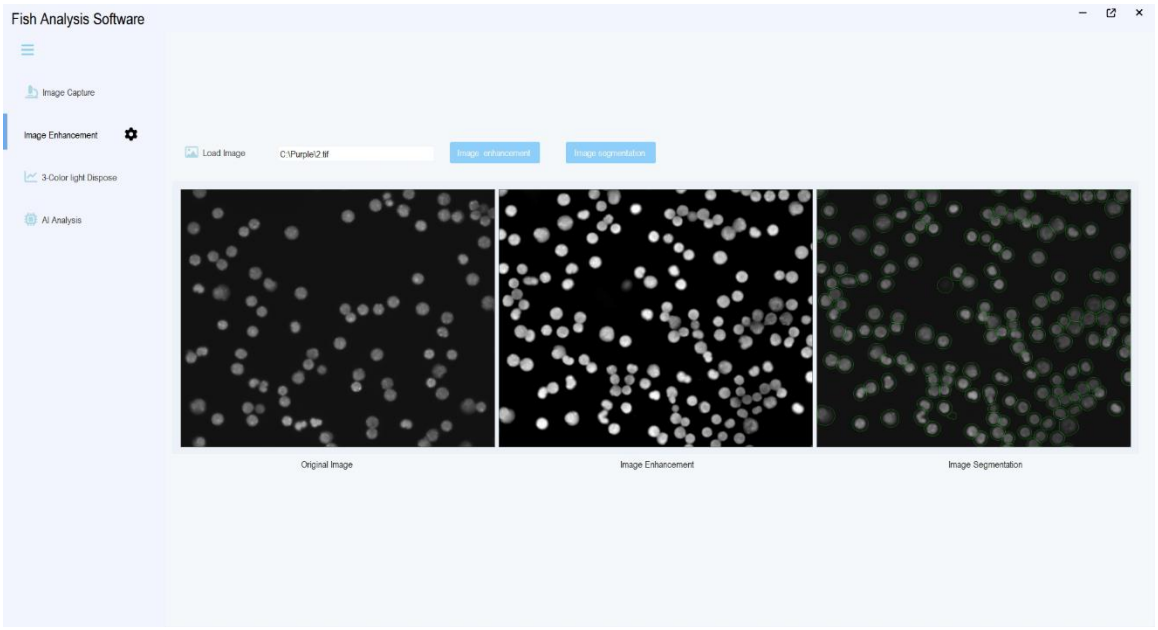
6) **Automatic Acquisition:** Click the "Load FISH Image" button to select and load the saved images. Click the "Load Image" button, and the system will automatically calculate and display the image with the highest clarity among the current 11 layers.



7) **Automatic Enhancement:** Click the "Image Enhancement" button to complete the image enhancement process.



8) **Automatic Segmentation:** Click the "Image Segmentation" button to complete the image segmentation and counting process.



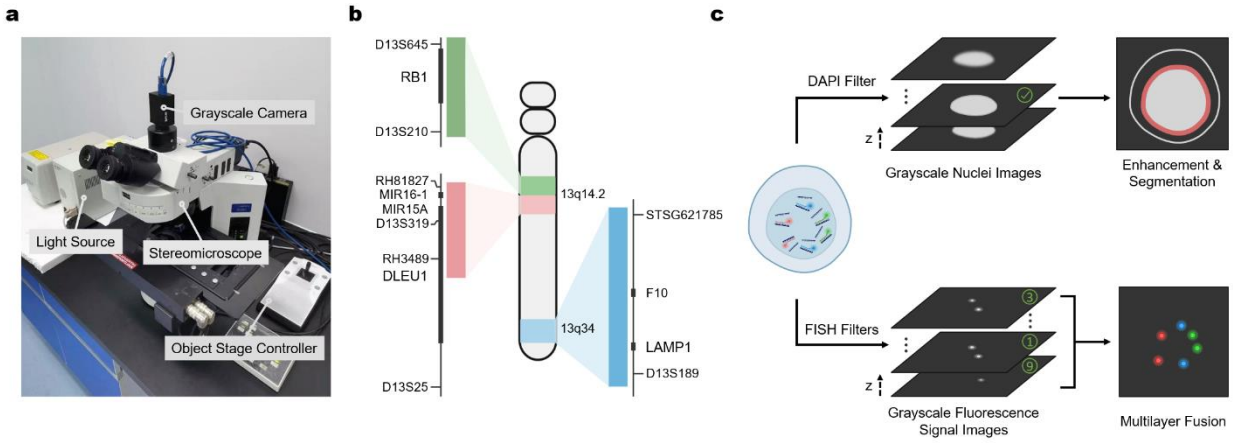


Fig. S1: Setup and working framework of the integrated FISH imaging and analysis system. (**a**) Overview of the integrated system. (**b**) Schematic of the tricolor FISH fluorescence probes mapping on chromosome 13. (**c**) Overall workflow of imaging and preprocessing for nuclei segmentation and multilayer fusion using a grayscale camera.

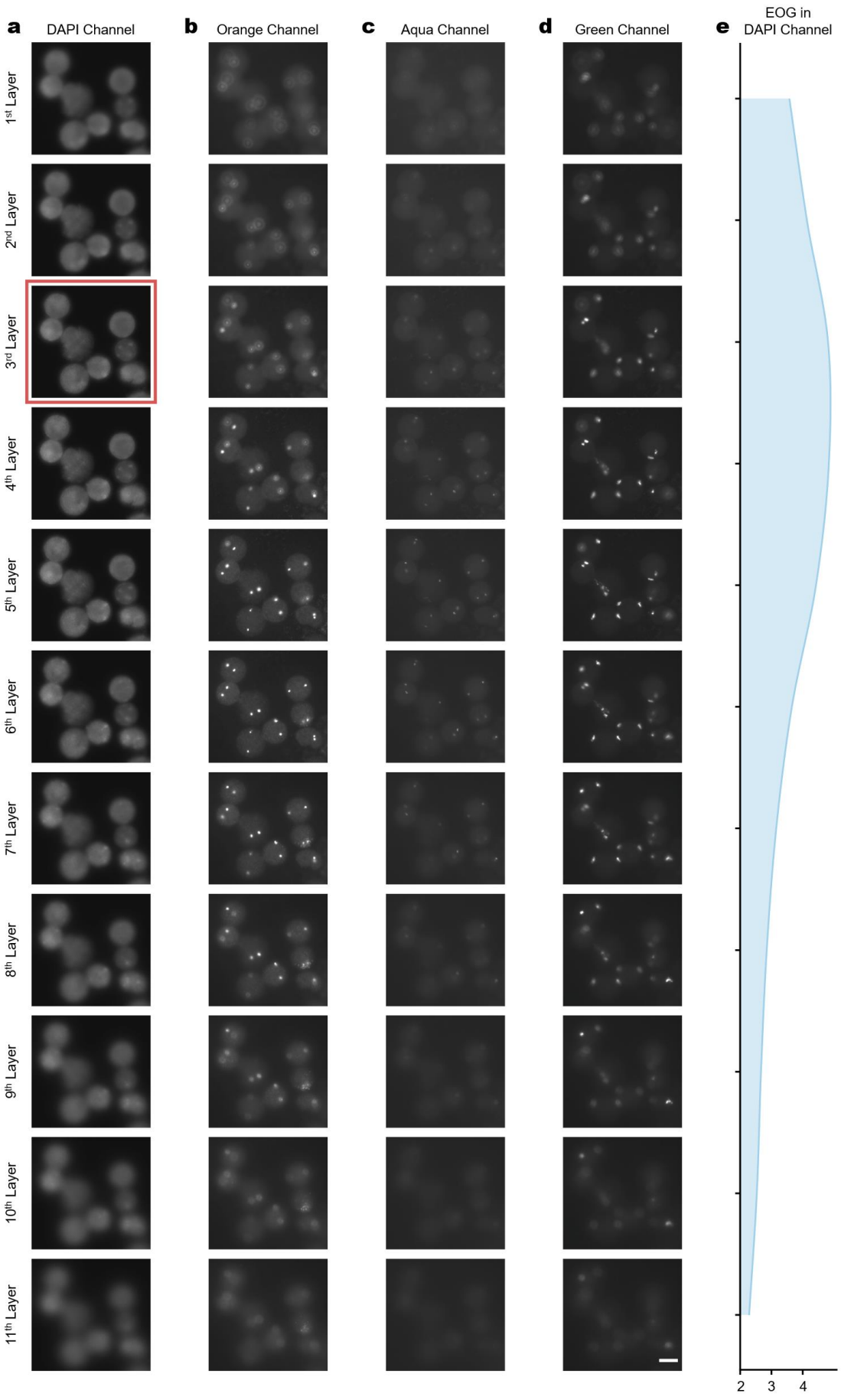

a
DAPI Channel
b
Orange Channel
c
Aqua Channel
d
Green Channel
e
EOG in
DAPI Channel
1st Layer
2nd Layer
3rd Layer
4th Layer
5th Layer
6th Layer
7th Layer
8th Layer
9th Layer
10th Layer
11th Layer
2
3
4

Fig. S2: Representative microscopic images in tetrachromatic channels using 11-layer scanning, including (**a**) nuclei components in the DAPI channel and (**b**)–(**d**) genetic insights in tricolor fluorescence channels. (**e**) EOG of images in the DAPI channel to identify the most focused layer for enhancement and nuclei segmentation processing. Interval between adjacent layers is 0.44μm. Scale bar: 5μm.

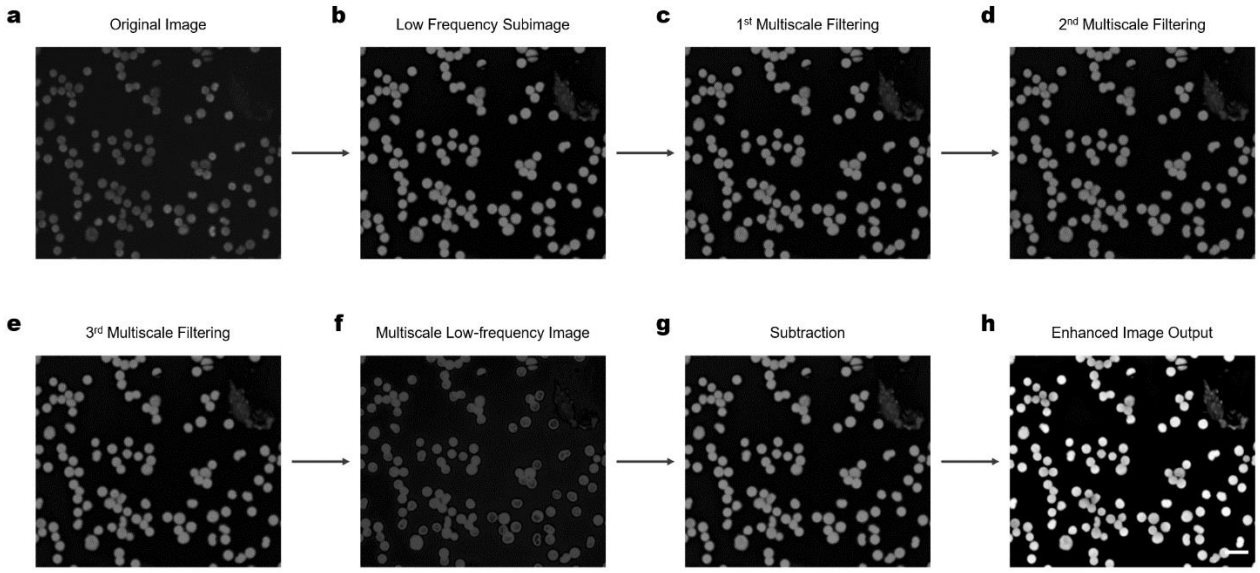


Fig. S3: Sequential steps of image enhancement. (**a**) Original image selected for optimal focus among 11 layers in the DAPI channel. (**b**) Extraction of the low frequency components from the input image. (**c**)–(**f**) Steps of multiscale DoGs to blur the extracted low frequency subimage. (**g**) Subtraction from original image to highlight high frequency components with enhanced edge features. (**h**) Enhanced image output through the combination of multiscale DoGs and the image fusion model to significantly improve image clarity and details. Scale bar: 20μm.

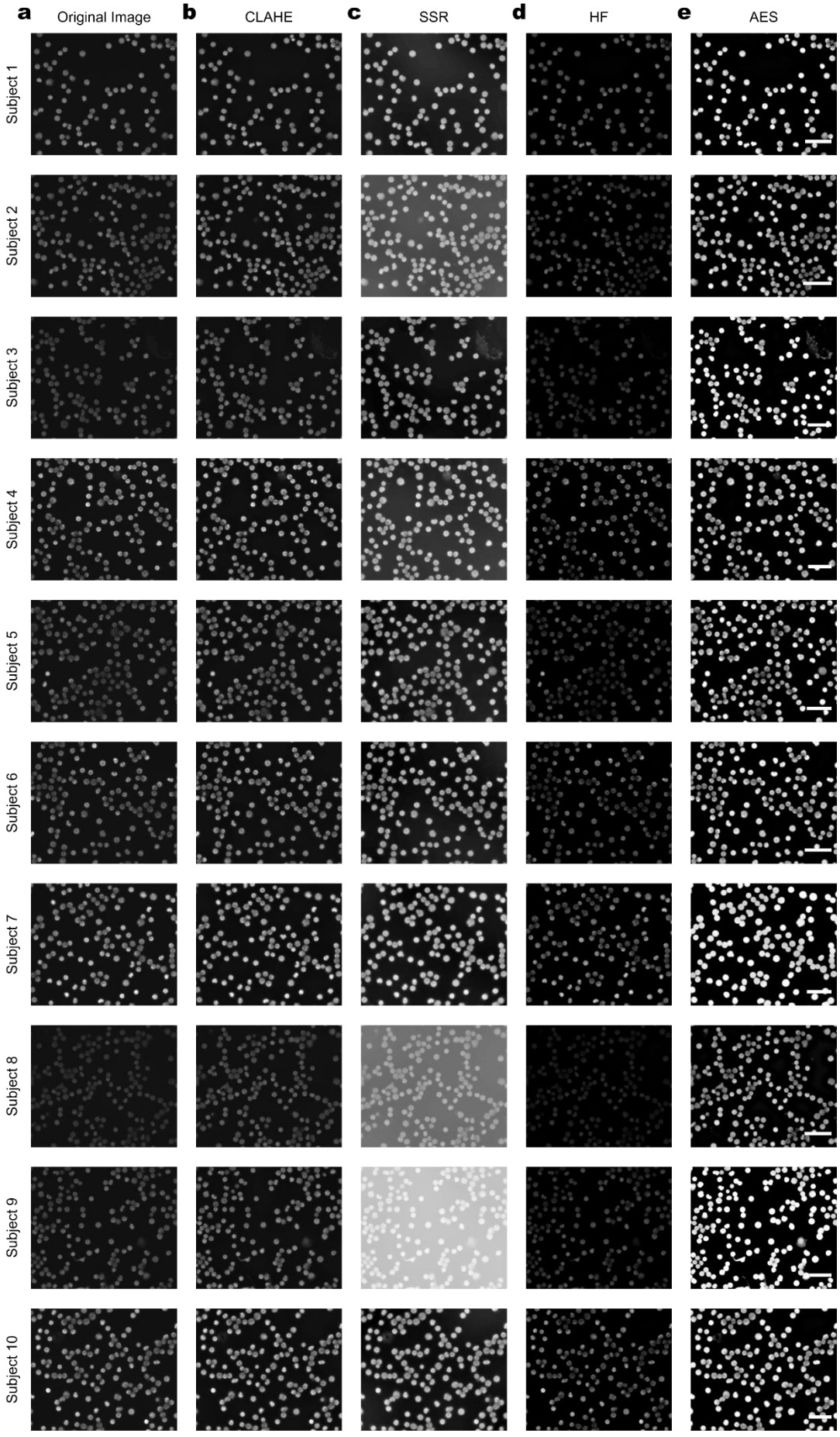

a
Original Image
b
CLAHE
c
SSR
d
HF
e
AES
Subject 1
Subject 2
Subject 3
Subject 4
Subject 5
Subject 6
Subject 7
Subject 8
Subject 9
Subject 10

Fig. S4: Performance comparison of different image enhancement methods across ten representative samples. (**a**) Original images presenting baseline data. (**b**)–(**e**) Enhancement results using (**b**) CLAHE, (**c**) SSR, (**d**) HF, and (**e**) FAST, respectively. The comparative analysis suggests a significant improvement of our method in both overall and local contrast in comparison to others. Scale bar: 30μm.

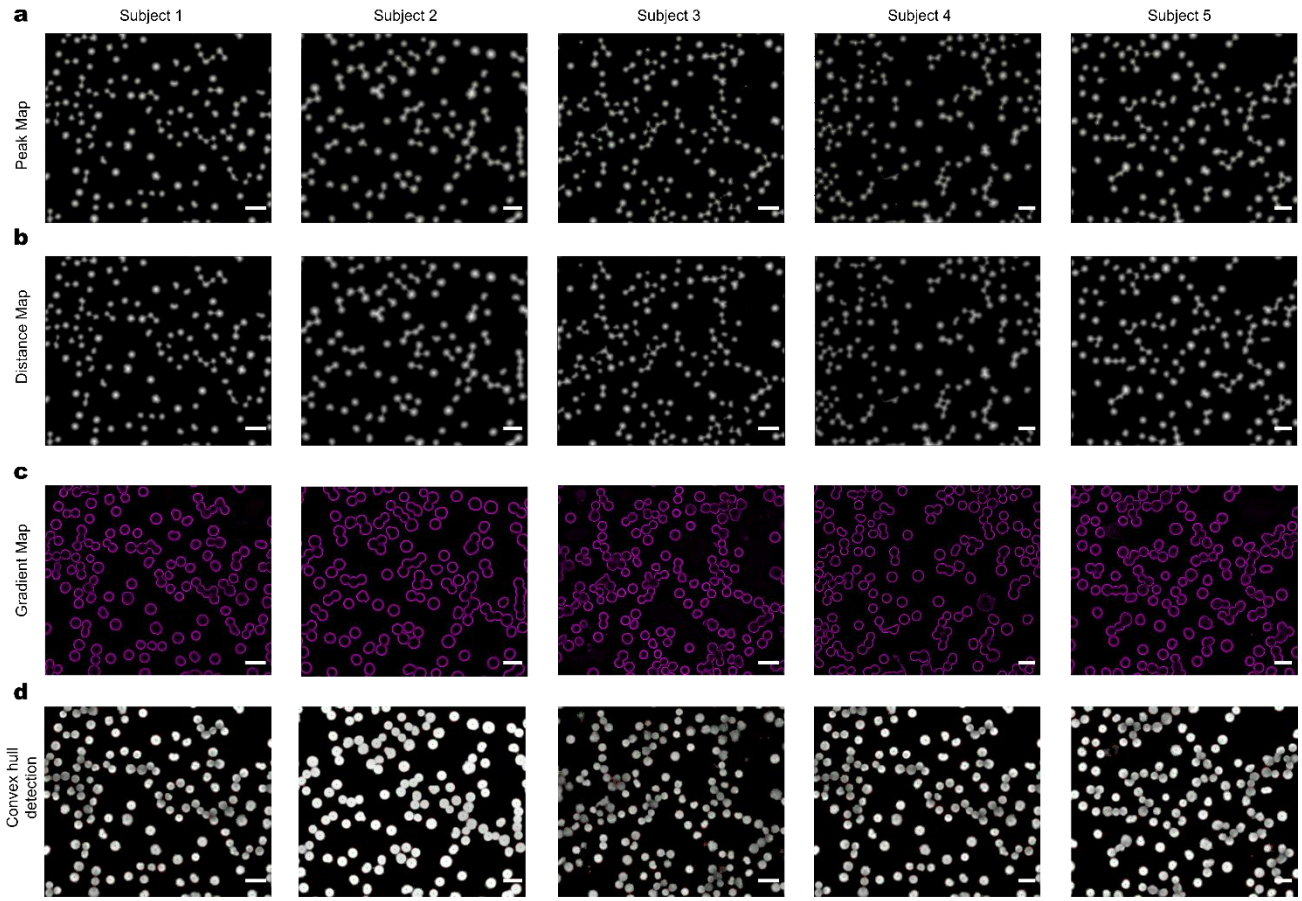


Fig. S5: Cell nucleus segmentation process. (**a**) Peak map localizes the presence of cell nuclei in the image. (**b**) Distance map provides reference information for segmentation. (**c**) Gradient map makes the boundary of the cell nucleus more obvious. (**d**) Convex hull detection ensures that each nucleus region is accurately segmented. Scale bar: 20μm.

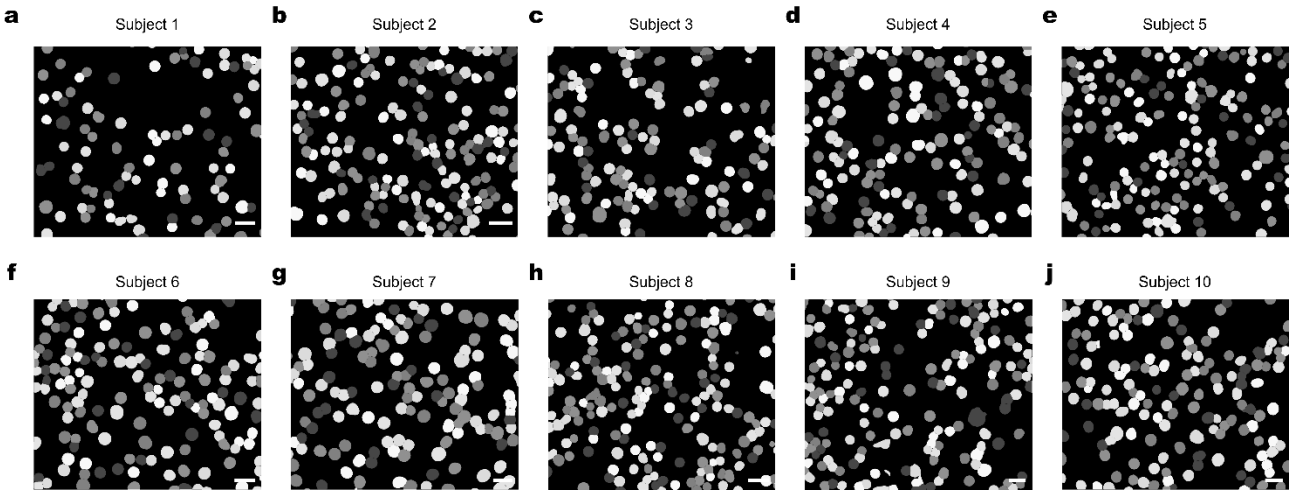


Fig. S6: Mask images across ten samples. Scale bar: 30μm.

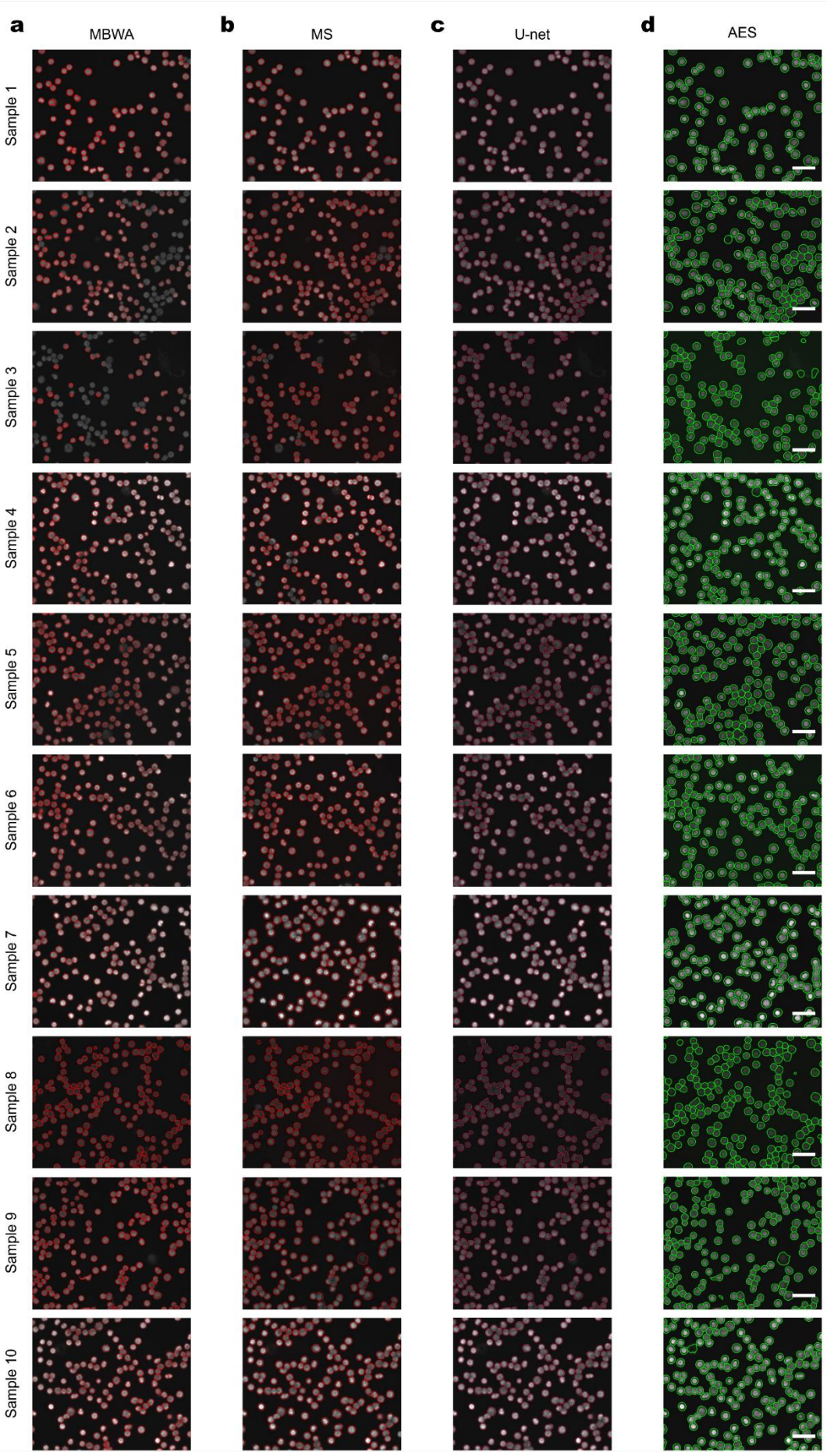

a
MBWA
b
MS
c
U-net
d
AES
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Sample 7
Sample 8
Sample 9
Sample 10

Fig. S7: Comparison of different image segmentation methods across ten samples, including (**a**) MW, (**b**) MS, (**c**) U-net, and (**d**) our segmentation method. The comparative analysis suggests the outperformance of our method against others in terms of segmentation precision and robustness. Scale bar: 30μm.

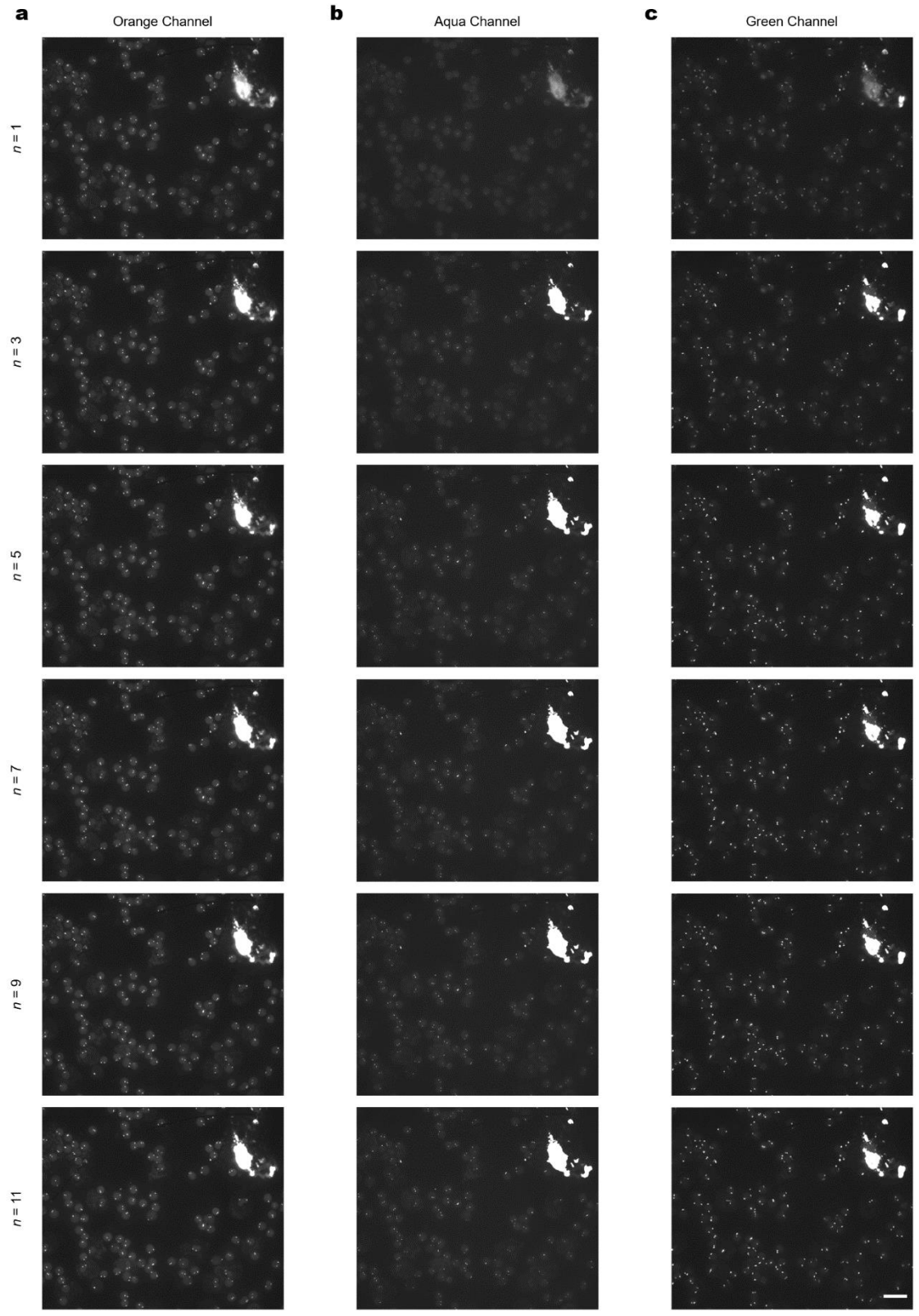


Fig. S8: Representative results of multilayer fluorescence feature fusion across tricolor FISH channels. Number of fused layers *n*=1, 3, 5, 7, 9, 11. Scale bar: 2μm.

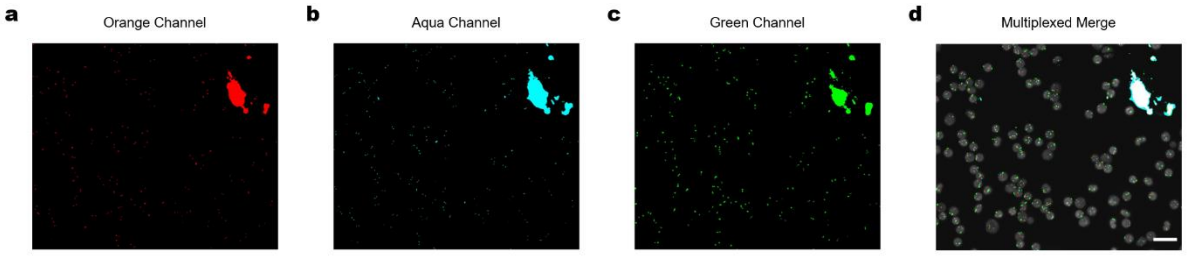


Fig. S9: **Representative results of pseudocolor processing and merging of the 9-layer fused image across tricolor FISH channels**, including (**a**) orange, (**b**) aqua, and (**c**) green channels. (**d**) Multiplexed merging of all three channels, visualizing the comprehensive genetic information for FISH analysis. Scale bar: 20μm.

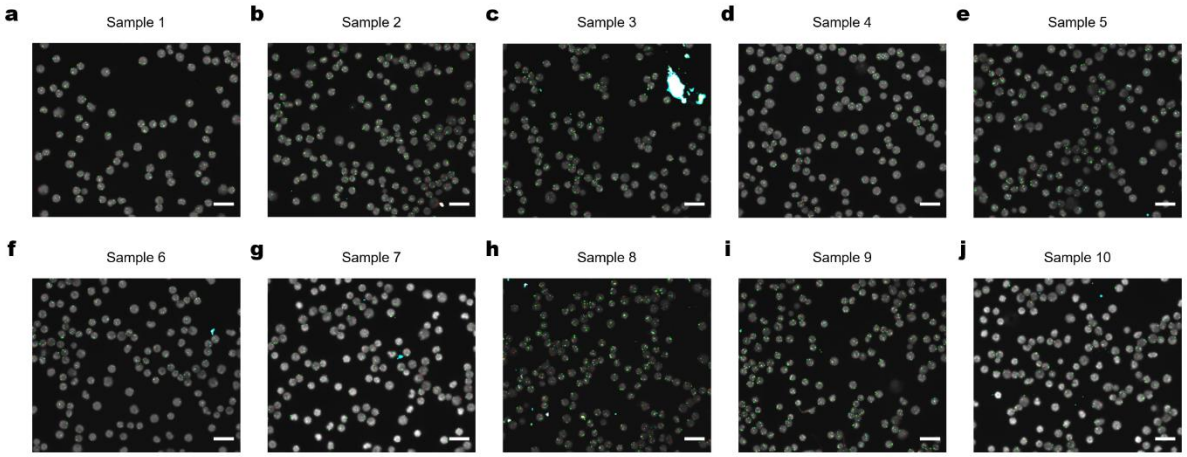


Fig. S10: **Visualization of 9-layer-fused multiplexed merged images across ten samples**, illustrating the comprehensive integration of fluorescence signals through the multilayer fluorescence feature fusion method for chromosomal abnormality classification. Scale bar: 20μm.

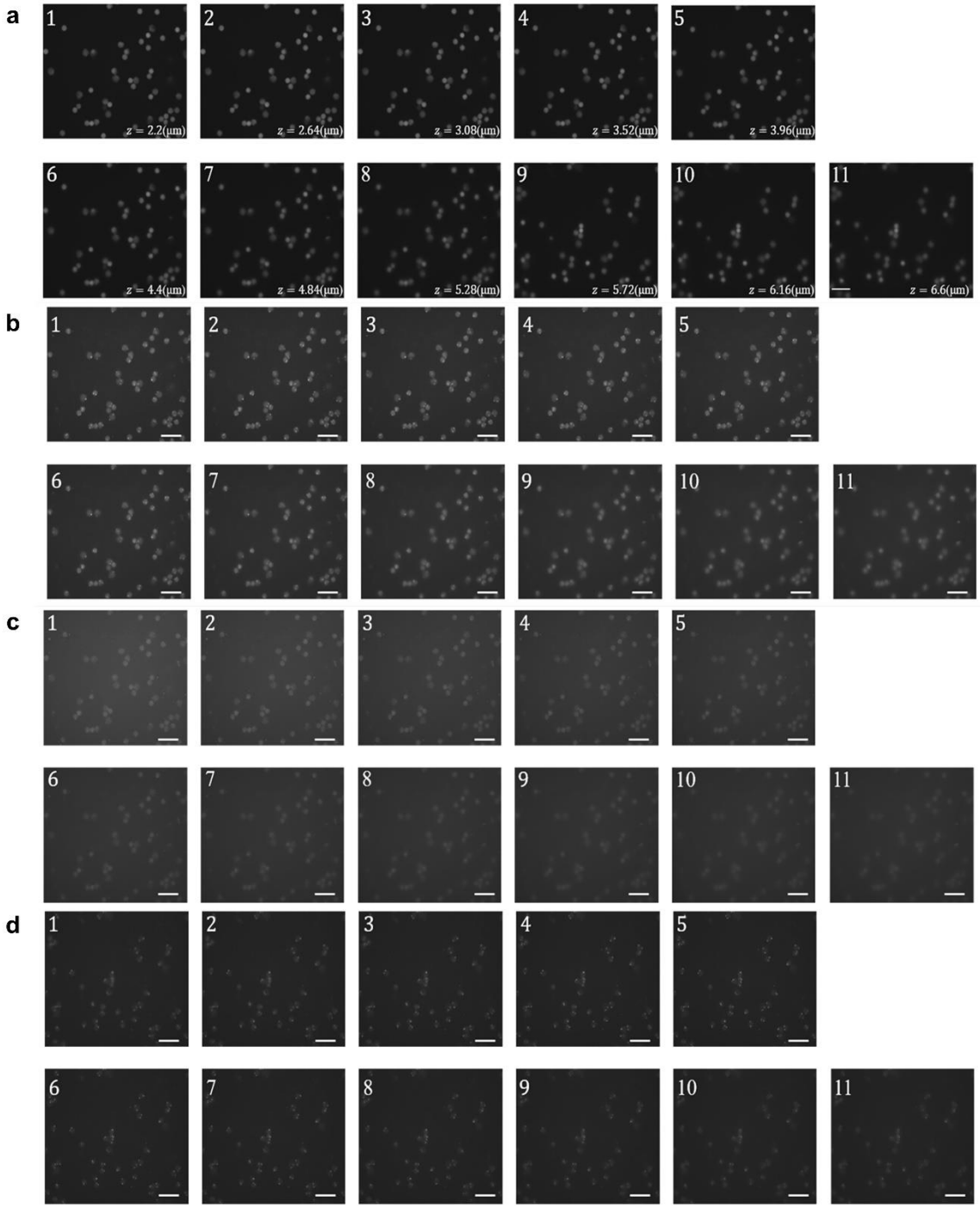


Fig. S11: Fluorescence characteristic images under different filters. (**a**) Fluorescence image under DAPI staining. (**b**) Fluorescence signal under aqua light. (**c**) Fluorescence signal under orange light. (**d**) Fluorescence signal under green light.

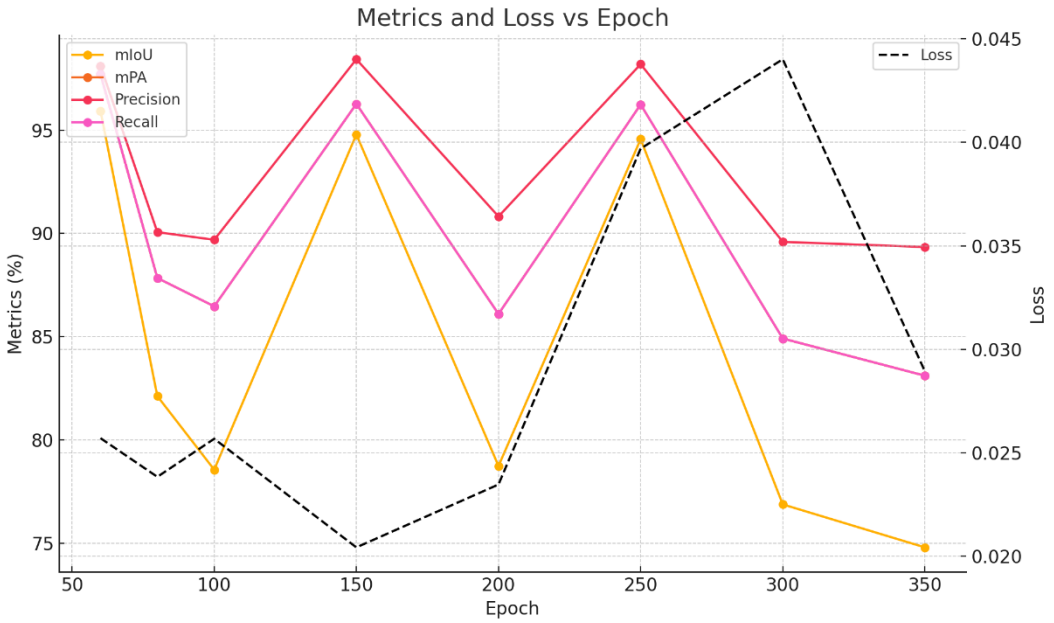


Fig. S12: Trends in model performance indicators and loss values at different Epochs. Four important evaluation indicators were used for the training of the fine segmentation network: mIoU (mean Intersection over Union), mPA (mean Pixel Accuracy), Precision and Recall, which comprehensively consider the various performance dimensions of the model and fully reflect the segmentation accuracy, pixel accuracy, etc. of the model. The indicators mIoU (94.79%), mPA (96.27%), Precision (98.43%) and Recall (96.27%) all showed the best results, with high accuracy in distinguishing the "nucleus" and "background" categories. The recognition effect of the background class was particularly outstanding, and the model had the lowest Loss, indicating that the model had the best fitting effect on the training data and the smallest error at this time. In contrast, the recognition of the cell nucleus class was slightly behind. Although the precision and recall were 0.99 and 0.93 respectively, it still showed strong performance, but the recall was slightly low, indicating that there were some missed cell pixels. There are almost no errors in background identification, while cell identification can still be further optimized, especially improving the recall rate to reduce missed cells.

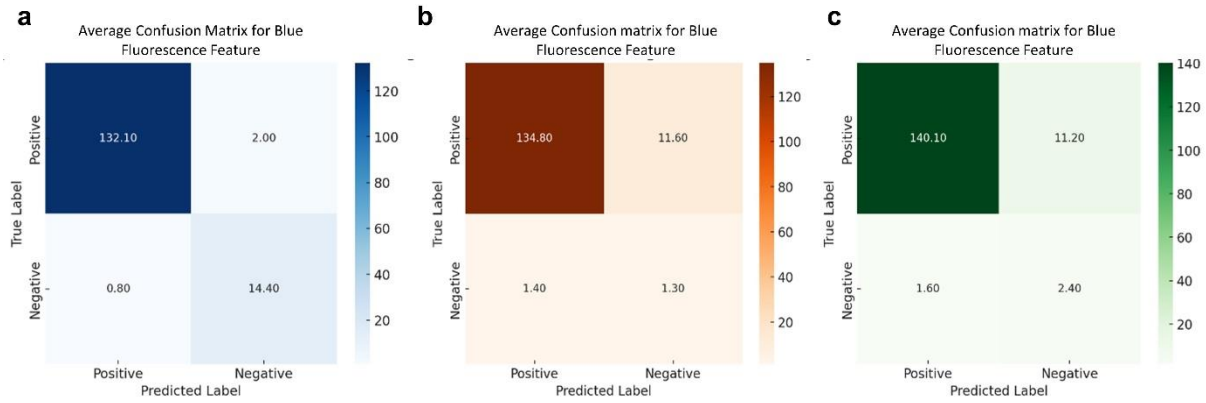


Fig. S13: Confusion matrix of classification tasks under different fluorescence. (**a**) Average confusion matrix heat map of blue fluorescence features. (**b**) Average confusion matrix heat map of orange fluorescence features. (**c**) Average confusion matrix heat map of green fluorescence features. In order to verify the generalization performance of the classification network model, this paper conducted statistics on indicators such as TN, FN, TP, and FP for single fluorescence images under the fusion of 9-layer fluorescence features, and comprehensively evaluated the classification performance of the model on different fluorescence features. The statistical results show that the classification accuracy of the water blue fluorescence feature is 98.10%, the classification accuracy of the orange fluorescence feature is 98.08%, the classification accuracy of the green fluorescence feature is 97.41%, and the classification accuracy of the fusion feature is 97.86%. This shows that the fusion strategy can effectively improve the classification performance of the model, making its prediction ability on unknown data more robust, and also verifies the generalization ability of the model under the fusion of multi-layer fluorescence features.

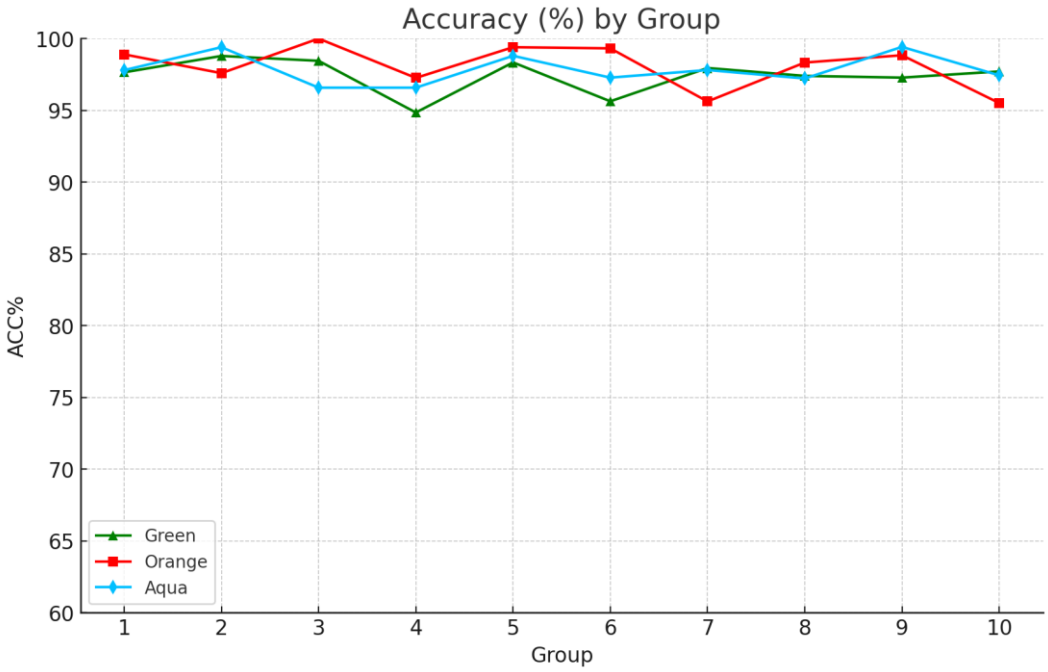


Fig. S14: Cross-validation accuracy curve under different fluorescence.

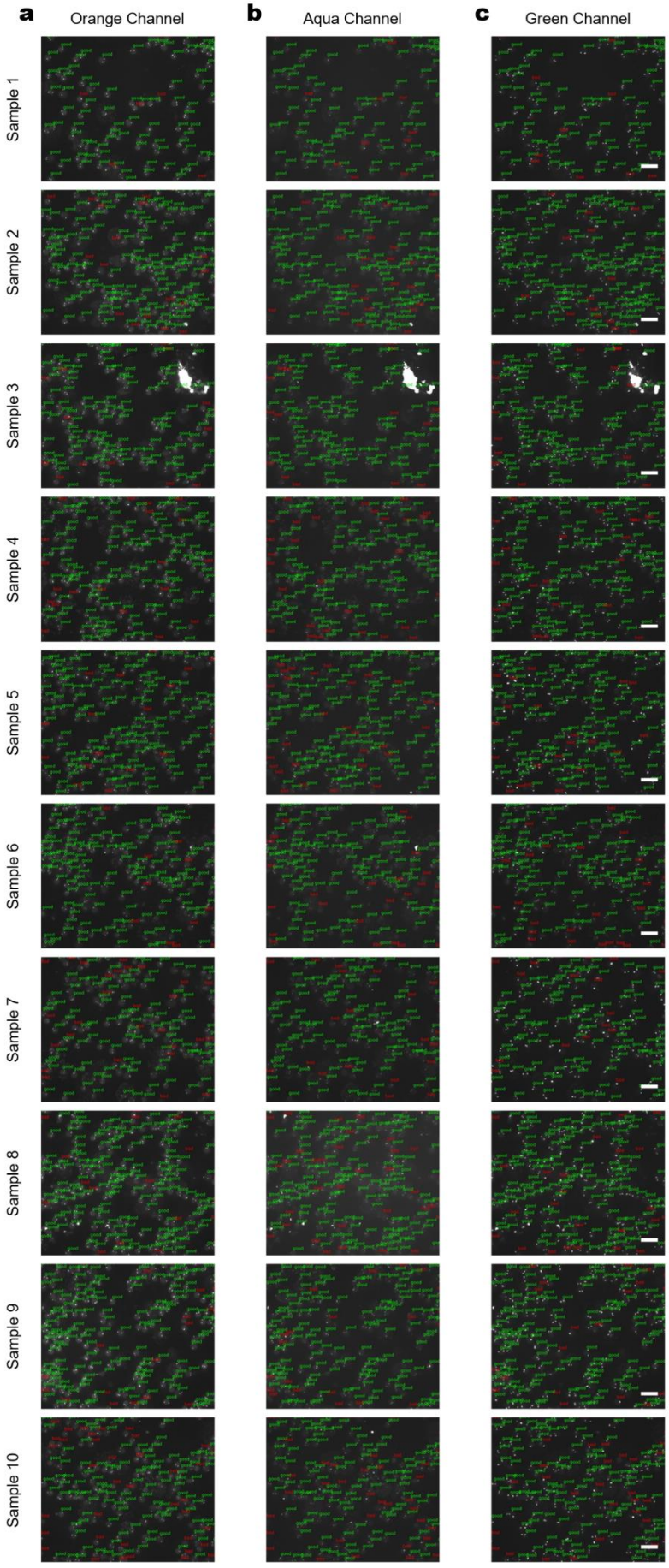

a
Orange Channel
b
Aqua Channel
c
Green Channel
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Sample 7
Sample 8
Sample 9
Sample 10

Fig. S15: **Automated classification of chromosomal abnormality using the ResNet152 network on the 9-layer-fused dataset**. Result images in (**a**) Orange channel, (**b**) aqua channel, (**c**) green channel. Green and red annotations in each image indicate normal and abnormal signal patterns as classified, respectively. Scale bar: 20μm.

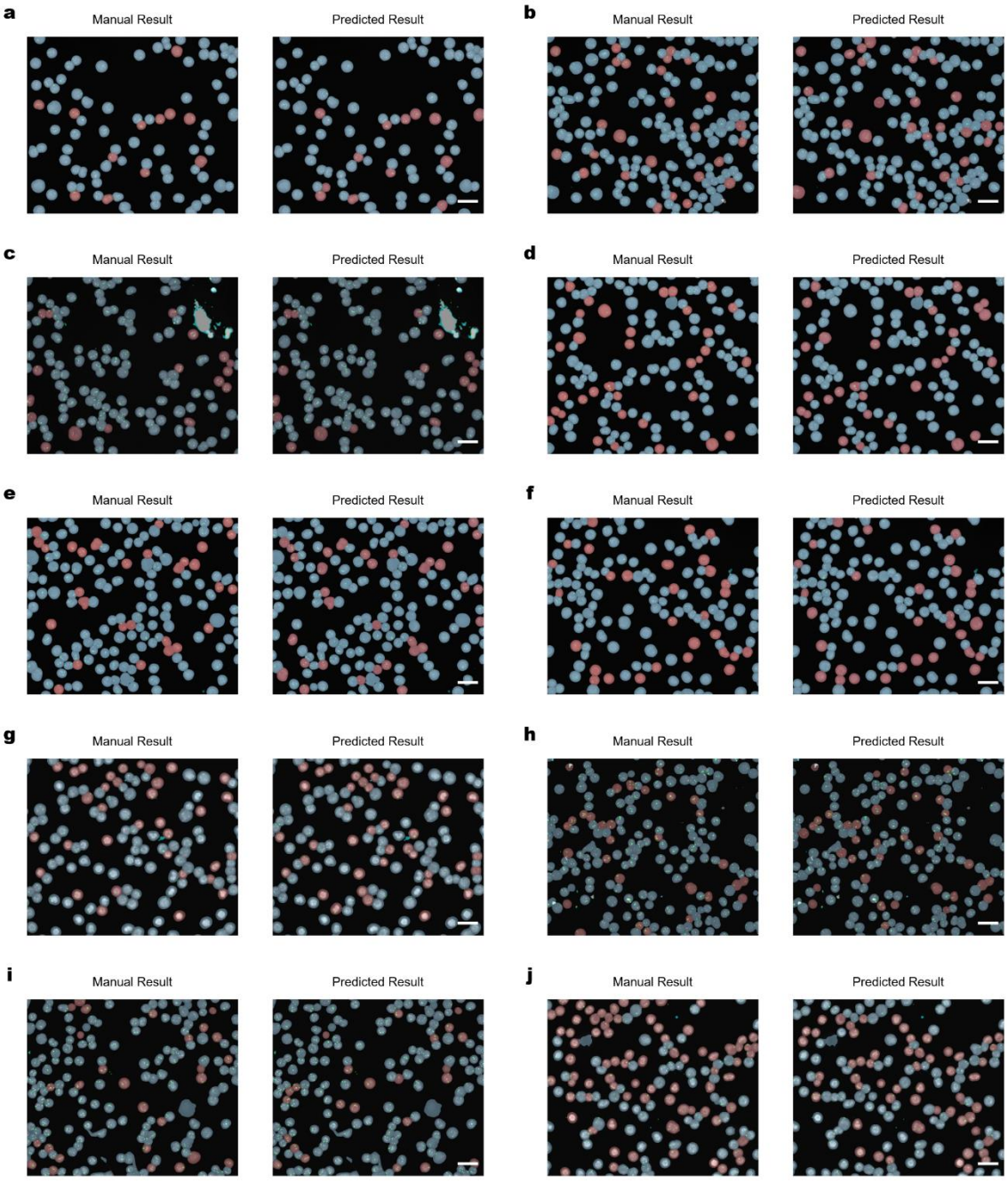


Fig. S16: Comparison of cell abnormality classification across ten samples using manual and deep-learning-based methods. Scale bar: 20μm.

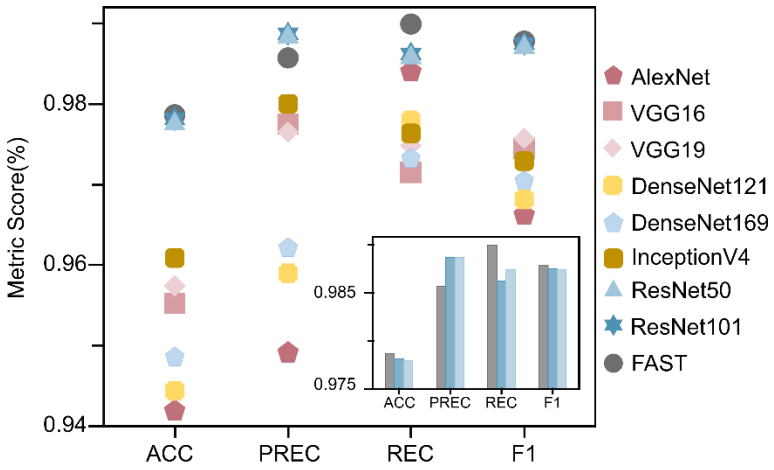


Fig. S17: Comparative evaluation of diagnostic performance across various established commercial networks and the proposed model. Inset plot illustrates a zoomed-in view for a clearer comparison among ResNet50, ResNet101, and our model, indicating the outperformance of our model against the other networks on automated cancer detection. ACC, accuracy; PREC, precision; REC, recall; and F1, F1 scores.

Table S1. Approximate cost of the proposed FISH imaging and analysis system

Item	Type	Cost ($)
Olympus fully intelligent motorized microscope	BX61	50,000
Wide-field fluorescence light source	X-Cite® 120Q	6,500
Computer workstation	Precision 3660	2,200
Industrial Camera	FLIR GS3-U3-51S5M-C	1,882
Total Cost		**60,582**

Table S2. Comparison of the costs of microscopy for FISH images

Microscope Model	Price (Thousand USD)	Maintenance Cost	Scanning Type	Space	Learning Curve
Thermo Fisher Helios	1,500-3,000	High	Point-by-point	Large	Gentle
Hitachi TM4000	100-150	Low	Point-by-point	Flexible	Gentle
Nikon A1R HD25	400-600	Medium	Point-by-point	Medium	Gentle
Leica DM4 B	30-60	Medium	Widefield	Medium	Steep
Zeiss Axio lmager 2	50-100	High	Widefield	Large	Steep
Olympus BX63	40-90	Medium	Widefield	Medium	Steep
Nikon Eclipse Ni-E	30-70	Low	Widefield	Flexible	Gentle
Olympus BX61	50	Medium	Widefield	Medium	Low

Table S3. Sample profiles used in this work

Samples	ID	Gender	Age	Clinical Diagnosis	Sample Type
1	00221834	M	61	CLL	Peripheral blood-EDTA anticoagulation
2	00378449	M	83	CLL	Peripheral blood-EDTA anticoagulation
3	00282649	M	56	CLL	Peripheral blood-EDTA anticoagulation
4	00292393	M	67	CLL	Peripheral blood-EDTA anticoagulation
5	00292397	F	55	CLL	Peripheral blood-EDTA anticoagulation
6	00292530	F	71	CLL	Peripheral blood-EDTA anticoagulation
7	00292619	F	59	CLL	Peripheral blood-EDTA anticoagulation
8	00292620	M	63	CLL	Peripheral blood-EDTA anticoagulation
9	00316557	M	61	CLL	Peripheral blood-EDTA anticoagulation
10	00337484	F	52	CLL	Peripheral blood-EDTA anticoagulation

Table S4. Energy gradient of 10 sets of DAPI images

images / layers	1	2	3	4	5	6	7	8	9	10
1	3.130	3.574	2.601	6.116	3.578	0.376	3.772	**2.220**	3.421	3.660
2	3.331	4.137	**2.817**	7.750	4.003	4.471	3.682	2.157	3.802	3.961
3	3.874	**4.841**	2.746	8.480	4.047	5.234	4.083	1.909	**4.249**	3.935
4	**4.583**	4.814	2.453	**9.158**	**4.305**	**5.596**	4.434	1.940	3.597	4.343
5	4.341	4.427	2.251	7.515	3.924	4.906	5.110	1.699	3.479	4.573
6	3.619	3.653	2.052	6.386	3.612	4.124	5.601	1.625	2.966	5.572
7	3.048	3.150	1.858	5.615	3.205	3.645	6.223	1.520	2.622	6.329
8	2.635	2.846	1.742	4.459	2.943	3.415	7.915	1.492	2.399	7.320
9	2.459	2.662	1.648	3.910	2.789	2.992	9.289	1.452	2.226	8.790
10	2.193	2.527	1.587	3.528	2.541	2.816	**9.714**	1.419	2.056	**9.473**
11	2.120	2.281	1.528	3.540	2.489	2.597	8.702	1.358	2.047	8.719

Table S5. Accumulated EOG in tricolor FISH channels across ten samples

Sample	Fused layers *n*	Orange channel	Aqua channel	Green channel
1	1	6.8111	0.5459	4.8605
	3	12.0431	7.1378	14.6885
	5	13.3810	9.5862	17.0186
	7	13.7334	9.8022	17.3268
	9	13.8763	10.0028	17.5253
	11	13.8791	10.0075	17.5275
2	1	11.1707	0.8033	6.7987
	3	16.9583	10.8042	15.4365
	5	18.3182	12.4693	21.2062
	7	18.8320	13.5947	22.8560
	9	19.0109	13.8188	23.2028
	11	19.0109	14.0005	23.4800
3	1	13.6578	0.7555	4.7667
	3	25.2906	14.6671	19.7257
	5	28.1512	23.7202	29.1148
	7	30.1688	23.9645	32.7816
	9	32.3322	23.9937	32.8691
	11	30.6715	24.0281	32.9049
4	1	7.8331	0.7212	4.6256
	3	13.3410	7.3640	16.7654
	5	17.8250	9.5689	21.8090
	7	18.5470	10.3869	22.6080
	9	18.6799	11.7099	26.8531
	11	18.6799	10.6539	27.1524
5	1	5.0410	0.3972	2.4114
	3	12.9769	7.2585	11.6973
	5	18.1129	9.1470	19.6466
	7	19.1678	11.1092	22.1015
	9	19.3125	18.6248	32.3313
	11	20.1562	18.6249	32.4572
6	1	6.7779	0.5576	3.4612
	3	14.0130	10.0300	16.5957
	5	17.0307	13.7158	18.5824
	7	17.6254	13.9761	18.9112

	9	18.3669	14.2199	19.0107
	11	19.3049	14.2374	19.0370
7	1	8.1623	15.4893	36.6859
	3	11.5243	20.6143	39.9787
	5	11.5459	20.5850	41.5214
	7	11.5494	20.7538	41.5513
	9	11.5600	20.7620	41.5577
	11	11.5600	20.7620	41.5577
8	1	18.9248	14.7443	28.0902
	3	25.5517	24.1849	37.1118
	5	27.9846	25.7692	46.6073
	7	30.2424	26.5994	47.7702
	9	31.5132	26.9330	48.5740
	11	31.5132	26.9330	48.5740
9	1	22.3213	19.4766	23.0264
	3	27.5800	25.7798	39.2013
	5	29.6857	28.2928	42.9576
	7	31.1715	28.9129	47.3661
	9	31.3987	28.9301	47.8729
	11	31.3987	28.9301	47.8729
10	1	13.6881	17.0131	31.6022
	3	15.4147	21.6060	38.6522
	5	15.4166	22.4429	38.6722
	7	15.4315	22.4541	38.6773
	9	15.4409	22.4608	38.6845
	11	15.4409	22.4608	38.6845

Table S6. Clear signal counts in each channel using *n*-layer feature fusion

Sample	Fused layers *n*	Number of clear signals		
		DLEU	LAMP1	RB1
1	1	35	12	33
	3	27	6	5
	5	3	1	3
	7	3	1	3
	9	2	1	3
	11	1	1	3
2	1	39	20	32
	3	17	4	28
	5	12	3	5
	7	9	3	4
	9	5	2	3
	11	5	2	3
3	1	43	14	33
	3	28	9	33
	5	6	3	22
	7	3	2	5
	9	1	1	2
	11	1	1	2
4	1	14	14	10
	3	13	11	9
	5	10	5	6
	7	8	5	6
	9	6	5	6
	11	6	5	6
5	1	23	13	17
	3	13	9	13
	5	12	3	9
	7	11	2	6
	9	11	2	5
	11	11	2	5
6	1	24	9	14
	3	14	4	11
	5	10	3	10

	7	6	3	10
	9	5	3	10
	11	5	3	10
7	1	9	5	4
	3	7	4	1
	5	4	4	1
	7	3	4	1
	9	3	4	1
	11	3	4	1
8	1	63	23	30
	3	22	10	23
	5	14	8	7
	7	12	7	7
	9	7	2	7
	11	7	2	7
9	1	27	22	57
	3	9	7	23
	5	4	2	7
	7	1	1	6
	9	1	1	6
	11	1	1	6
10	1	15	23	12
	3	6	22	9
	5	4	22	9
	7	4	22	9
	9	4	22	9
	11	4	22	9

Table S7. Manual classification results of chromosomal and cell abnormality using *n*-layer feature fusion

Sample	Fused layers *n*	Cell number	Refined threshold (%)	Standard missing signal threshold	R_{al} (Count)			Result	Cell abnormality rate (%)
					DLEU	LAMP1	RB1		
1	1	91	5	5	8	6	24	Positive	-
	3			5	7	5	8	Positive	-
	5			5	6	5	5	Positive	-
	7			**5**	**5**	**5**	**5**	**Negative**	5.49
	9			**5**	**5**	**5**	**5**	**Negative**	5.49
	11			**5**	**5**	**5**	**5**	**Negative**	5.49
2	1	166	5	8	19	20	12	Positive	-
	3			8	13	15	12	Positive	-
	5			8	10	14	7	Positive	-
	7			8	8	14	6	Positive	-
	9			8	8	14	6	Positive	-
	11			8	8	14	6	Positive	-
3	1	131	5	7	13	11	25	Positive	-
	3			7	12	10	25	Positive	-
	5			7	9	7	13	Positive	-
	7			**7**	**7**	**6**	**5**	**Negative**	4.58
	9			**7**	**6**	**6**	**5**	**Negative**	4.33
	11			**7**	**6**	**6**	**5**	**Negative**	4.33
4	1	146	5	7	37	22	17	Positive	-
	3			7	31	21	16	Positive	-
	5			7	22	13	14	Positive	-
	7			7	20	13	13	Positive	-
	9			7	20	13	9	Positive	-
	11			7	20	13	9	Positive	-
5	1	167	5	8	38	22	20	Positive	-
	3			8	20	17	16	Positive	-
	5			8	15	9	10	Positive	-
	7			8	15	9	10	Positive	-
	9			8	14	9	10	Positive	-
	11			8	14	9	10	Positive	-
6	1	147	5	7	19	13	22	Positive	-

	3			7	10	12	19	Positive	-
	5			7	9	10	16	Positive	-
	7			7	9	9	16	Positive	-
	9			7	9	9	16	Positive	-
	11			7	9	9	16	Positive	-
7	1	137	5	7	19	15	11	Positive	-
	3			7	16	15	11	Positive	-
	5			7	16	15	11	Positive	-
	7			7	16	12	11	Positive	-
	9			7	16	12	11	Positive	-
	11			7	16	12	11	Positive	-
8	1	180	5	9	27	8	12	Positive	-
	3			9	21	8	11	Positive	-
	5			9	21	8	11	Positive	-
	7			9	21	8	11	Positive	-
	9			9	20	8	11	Positive	-
	11			9	20	8	10	Positive	-
9	1	172	5	9	23	21	48	Positive	-
	3			9	10	10	18	Positive	-
	5			9	10	9	12	Positive	-
	7			9	10	9	12	Positive	-
	9			9	10	9	12	Positive	-
	11			9	10	9	12	Positive	-
10	1	156	5	8	30	54	25	Positive	-
	3			8	24	54	24	Positive	-
	5			8	24	54	24	Positive	-
	7			8	24	54	24	Positive	-
	9			8	24	54	24	Positive	-
	11			8	24	54	24	Positive	-

Table S8. Automated classification results of DLEU-labelled chromosomal abnormality using the CNN network

Sample	TP	TN	FP	FN	Accuracy (%)
1	86	4	1	0	98.90
2	148	14	4	0	97.59
3	124	7	0	0	100.00
4	130	12	1	3	97.26
5	154	12	0	1	99.40
6	138	8	1	0	99.32
7	117	14	0	6	95.62
8	168	9	2	1	98.33
9	162	8	0	2	98.84
10	121	28	5	2	95.51

Table S9. Automated classification results of LAMP1-labelled chromosomal abnormality using the CNN network

Sample	TP	TN	FP	FN	Accuracy (%)
1	83	6	0	2	97.80
2	148	17	0	1	99.40
3	123	7	1	0	99.24
4	125	16	1	4	96.58
5	145	20	2	0	98.80
6	133	10	0	4	97.28
7	120	14	2	1	97.81
8	155	20	0	5	97.22
9	161	10	0	1	99.42
10	128	24	2	2	97.44

Table S10. Automated classification results of RB1-labelled chromosomal abnormality using the CNN network

Sample	TP	TN	FP	FN	Accuracy (%)
1	83	6	1	1	97.80
2	153	11	1	1	98.80
3	122	7	2	0	98.47
4	130	9	1	6	95.21
5	150	14	0	3	98.20
6	127	13	4	3	95.24
7	125	9	1	2	97.81
8	166	12	1	4	97.27
9	155	12	1	4	97.09
10	133	19	4	0	97.44

Table S11. Automated classification results of chromosomal and cell abnormality using the CNN network on the 9-layer-fused dataset

Sample	Cell number	Refined threshold (%)	Standard missing signal threshold	R_{al} (%)						Result	Cell abnormality rate (%)
				DLEU		LAMP1		RB1			
1	91	5	5	4	4.40	6	6.59	6	6.59	Positive	-
2	166	5	8	14	8.43	17	10.24	11	6.63	Positive	-
3	131	5	7	7	5.34	7	5.34	7	5.34	**Negative**	**5.34**
4	146	5	7	12	8.22	16	10.96	9	6.16	Positive	-
5	167	5	8	12	7.19	20	11.98	14	8.38	Positive	-
6	147	5	7	8	5.44	10	6.80	13	8.84	Positive	-
7	137	5	7	14	10.22	14	10.22	9	6.57	Positive	-
8	180	5	9	9	5.00	20	11.11	12	6.67	Positive	-
9	172	5	9	8	4.65	10	5.81	12	6.98	Positive	
10	156	5	8	28	17.95	24	15.38	20	12.8	Positive	-

Table S12. Total system response time

Image preprocessing time (seconds)	Segmentation algorithm time (seconds)	Total time (seconds)
51.77	21.75	73.52
51.09	38.12	89.21
50.36	28.64	79.00
48.05	33.94	81.99
48.67	53.56	102.23
63.29	53.65	116.94
61.05	58.75	119.80
67.90	65.34	133.24
63.03	65.96	128.99
77.98	55.12	133.10

In the experimental setup, the total processing time was approximately 45 minutes, with an average preprocessing time of 105 seconds for 10 sets of images. This includes both algorithm execution and the time required for microscope control and image acquisition. The workflow involves tasks such as image acquisition, stage movement, and automated computational processes including image enhancement, segmentation, and multi-layer feature fusion. Preprocessing, which encompasses enhancement and segmentation, typically requires around 5 minutes but may extend up to this duration depending on variations in the images. The microscope control operations account for approximately 40 minutes. It is important to note that these timings reflect operations performed on a standard personal computer rather than a dedicated workstation.

Supplementary References

Badrinarayanan V, Kendall A, Cipolla R, 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*, 39(12):2481-2495. https://doi.org/10.1109/tpami.2016.2644615

Battleday RM, Peterson JC, Griffiths TL, 2020. Capturing human categorization of natural images by combining deep networks and cognitive models. *Nat Commun*, 11(1):5418. https://doi.org/10.1038/s41467-020-18946-z

Bhutto DF, Zhu B, Liu JZ, et al., 2024. Uncertainty estimation and out-of-distribution detection for deep learning-based image reconstruction using the local lipschitz. *IEEE J Biomed Health Inform*, 28(9):5422-5434. https://doi.org/10.1109/jbhi.2024.3404883

Biancalani T, Scalia G, Buffoni L, et al., 2021. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods*, 18(11):1352-1362. https://doi.org/10.1038/s41592-021-01264-7

Caicedo JC, Goodman A, Karhohs KW, et al., 2019. Nucleus segmentation across imaging experiments: The 2018 data science bowl. *Nat Methods*, 16(12):1247-1253. https://doi.org/10.1038/s41592-019-0612-7

Döhner H, Stilgenbauer S, Benner A, et al., 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*, 343(26):1910-1916. https://doi.org/10.1056/nejm200012283432602

Hossain MB, Iqbal S, Islam MM, et al., 2022. Transfer learning with fine-tuned deep cnn resnet50 model for classifying covid-19 from chest x-ray images. *Inform Med Unlocked*, 30:100916. https://doi.org/10.1016/j.imu.2022.100916

Jardim S, António J, Mora C, 2022. Graphical image region extraction with k-means clustering and watershed. *J Imaging*, 8(6) https://doi.org/10.3390/jimaging8060163

Jelinek T, Zihala D, Sevcikova T, et al., 2024. Beyond the marrow: Insights from comprehensive next-generation sequencing of extramedullary multiple myeloma tumors. *Leukemia*, 38(6):1323-1333. https://doi.org/10.1038/s41375-024-02206-w

Khan AI, Al-Habsi S, 2020. Machine learning in computer vision. *Procedia Computer Science*, 167:1444-1451.

Kushwaha S, Stinnett V, Zou YS, et al., 2024. Live-born autosomal ring chromosomes at the johns hopkins hospital cytogenomics laboratory: Case series-spanning 52 years of experience in a single center. *Am J Med Genet A*, 194(2):253-267. https://doi.org/10.1002/ajmg.a.63429

Le Goallec A, Diai S, Collin S, et al., 2022. Using deep learning to predict abdominal age from liver and pancreas magnetic resonance images. *Nat Commun*, 13(1):1979. https://doi.org/10.1038/s41467-022-29525-9

Li F, Yang Y, Wei Y, et al., 2022. Predicting neoadjuvant chemotherapy benefit using deep learning from stromal histology in breast cancer. *NPJ Breast Cancer*, 8(1):124. https://doi.org/10.1038/s41523-022-00491-1

Menghani G, 2023. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, 55(12):1-37.

Ouillette P, Erba H, Kujawski L, et al., 2008. Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14. *Cancer Res*, 68(4):1012-1021. https://doi.org/10.1158/0008-5472.Can-07-3105

Paszke A, Chaurasia A, Kim S, et al., 2016. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*,

Rolfe M, Hayes S, Smith M, et al., 2024. An ai based smart-phone system for asbestos identification. *J Hazard Mater*, 463:132853. https://doi.org/10.1016/j.jhazmat.2023.132853

Sharma AK, Nandal A, Dhaka A, et al., 2022. Enhanced watershed segmentation algorithm-based modified resnet50 model for brain tumor detection. *Biomed Res Int*, 2022:7348344. https://doi.org/10.1155/2022/7348344

Shen X, Qi Y, Ma T, et al., 2019. A dicentric chromosome identification method based on clustering and watershed algorithm. *Sci Rep*, 9(1):2285. https://doi.org/10.1038/s41598-019-38614-7

Singh Y, Eaton JE, Venkatesh SK, et al., 2025. Deep learning analysis of mri accurately detects early-stage perihilar cholangiocarcinoma in patients with primary sclerosing cholangitis. *Hepatology*, https://doi.org/10.1097/hep.0000000000001314

Somu N, Mr GR, Ramamritham K, 2021. A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews*, 137:110591.

Sun Y, Zhao Z, Jiang D, et al., 2022. Low-illumination image enhancement algorithm based on improved multi-scale retinex and abc algorithm optimization. *Front Bioeng Biotechnol*,

10:865820. https://doi.org/10.3389/fbioe.2022.865820

Wang J-G, Tai S-C, Lin C-J, 2018. The application of an interactively recurrent self-evolving fuzzy cmac classifier on face detection in color images. *Neural Computing and Applications*, 29:201-213.

Yang H, Ni J, Gao J, et al., 2021. A novel method for peanut variety identification and classification by improved vgg16. *Sci Rep*, 11(1):15756. https://doi.org/10.1038/s41598-021-95240-y

Zhang Z, Rong X, Xie T, et al., 2024. Fluorogenic crispr for genomic DNA imaging. *Nat Commun*, 15(1):934. https://doi.org/10.1038/s41467-024-45163-9

Zhao H, Qi X, Shen X, et al., 2018. Semantic image segmentation algorithms. *Proc. Eur. Conf. Comput. Vis*, :405–420.