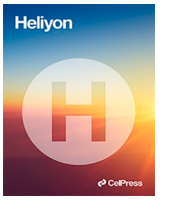
Heliyon 10 (2024) e30239

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Continuous distribution of cancer cells in the cell cycle unveiled by AI-segmented imaging of 37,000 HeLa FUCCI cells

Hamid Cheraghi [a,b], Kinga Dóra Kovács [a,c], Inna Székács [c], Robert Horvath [c], Bálint Szabó [a,b,*]

[a] Department of Biological Physics, Eötvös University (ELTE), H-1117, Budapest, Hungary
[b] CellSorter Scientific Company for Innovations, Prielle Kornélia utca 4A, 1117, Budapest, Hungary
[c] Nanobiosensorics Laboratory, HUN-REN, Institute of Technical Physics and Materials Science, Centre for Energy Research, Budapest, Hungary

ARTICLE INFO

Keywords:
Cell segmentation
Fluorescence imaging
Cell cycle analysis
Ground truth datasets

ABSTRACT

Classification of live or fixed cells based on their unlabeled microscopic images would be a powerful tool for cell biology and pathology. For such software, the first step is the generation of a ground truth database that can be used for training and testing AI classification algorithms. The Application of cells expressing fluorescent reporter proteins allows the building of ground truth datasets in a straightforward way. In this study, we present an automated imaging pipeline utilizing the Cellpose algorithm for the precise cell segmentation and measurement of fluorescent cellular intensities across multiple channels. We analyzed the cell cycle of HeLa–FUCCI cells expressing fluorescent red and green reporter proteins at various levels depending on the cell cycle state. To build the dataset, 37,000 fixed cells were automatically scanned using a standard motorized microscope, capturing phase contrast and fluorescent red/green images. The fluorescent pixel intensity of each cell was integrated to calculate the total fluorescence of cells based on cell segmentation in the phase contrast channel. It resulted in a precise intensity value for each cell in both channels. Furthermore, we conducted a comparative analysis of Cellpose 1.0 and Cellpose 2.0 in cell segmentation performance. Cellpose 2.0 demonstrated notable improvements, achieving a significantly reduced false positive rate of 2.7 % and 1.4 % false negative. The cellular fluorescence was visualized in a 2D plot (map) based on the red and green intensities of the FUCCI construct revealing the continuous distribution of cells in the cell cycle. This 2D map enables the selection and potential isolation of single cells in a specific phase. In the corresponding heatmap, two clusters appeared representing cells in the red and green states. Our pipeline allows the high-throughput and accurate measurement of cellular fluorescence providing extensive statistical information on thousands of cells with potential applications in developmental and cancer biology. Furthermore, our method can be used to build ground truth datasets automatically for training and testing AI cell classification. Our automated pipeline can be used to analyze thousands of cells within 2 h after putting the sample onto the microscope.

1. Introduction

The regulatory network controlling the eukaryotic cell cycle is now well-established. Previous research predominantly focused on

* Corresponding author.)
E-mail address: balint.szabo@cellsorter-scientific.com (B. Szabó).

https://doi.org/10.1016/j.heliyon.2024.e30239
Received 12 December 2023; Received in revised form 22 April 2024; Accepted 22 April 2024
Available online 23 April 2024

single-celled organisms or immortalized cultured cells capable of indefinite proliferation with adequate nutrients and growth factors [1]. However, in multicellular organisms like animals and plants, a cell's destiny - whether it divides, remains quiescent, or exits the cell cycle for differentiation - is intricately influenced by interactions with neighboring cells and physiological signals from diverse body regions. Understanding how cellular proliferation is governed by the cellular microenvironment is pivotal for addressing fundamental challenges in the development, regeneration, and transformation of normal cells into malignant ones.

Traditional cell cycle markers, such as nucleotide analogs (BrdU, EdU) or replication proteins (PCNA, Ki-67), rely on immunofluorescent detection that necessitates sample fixation [2,3]. Consequently, analyzing proliferating cells in intact organisms poses challenges. However, observing individual cell behavior during the cell cycle in live organisms offers insights into various developmental processes like pattern formation, morphogenesis, cell differentiation, growth, migration, and apoptosis. The introduction of the Fluorescent Ubiquitination-Based Cell Cycle Indicator (FUCCI) by Sakaue-Sawano et al., in 2008 revolutionized the visualization of

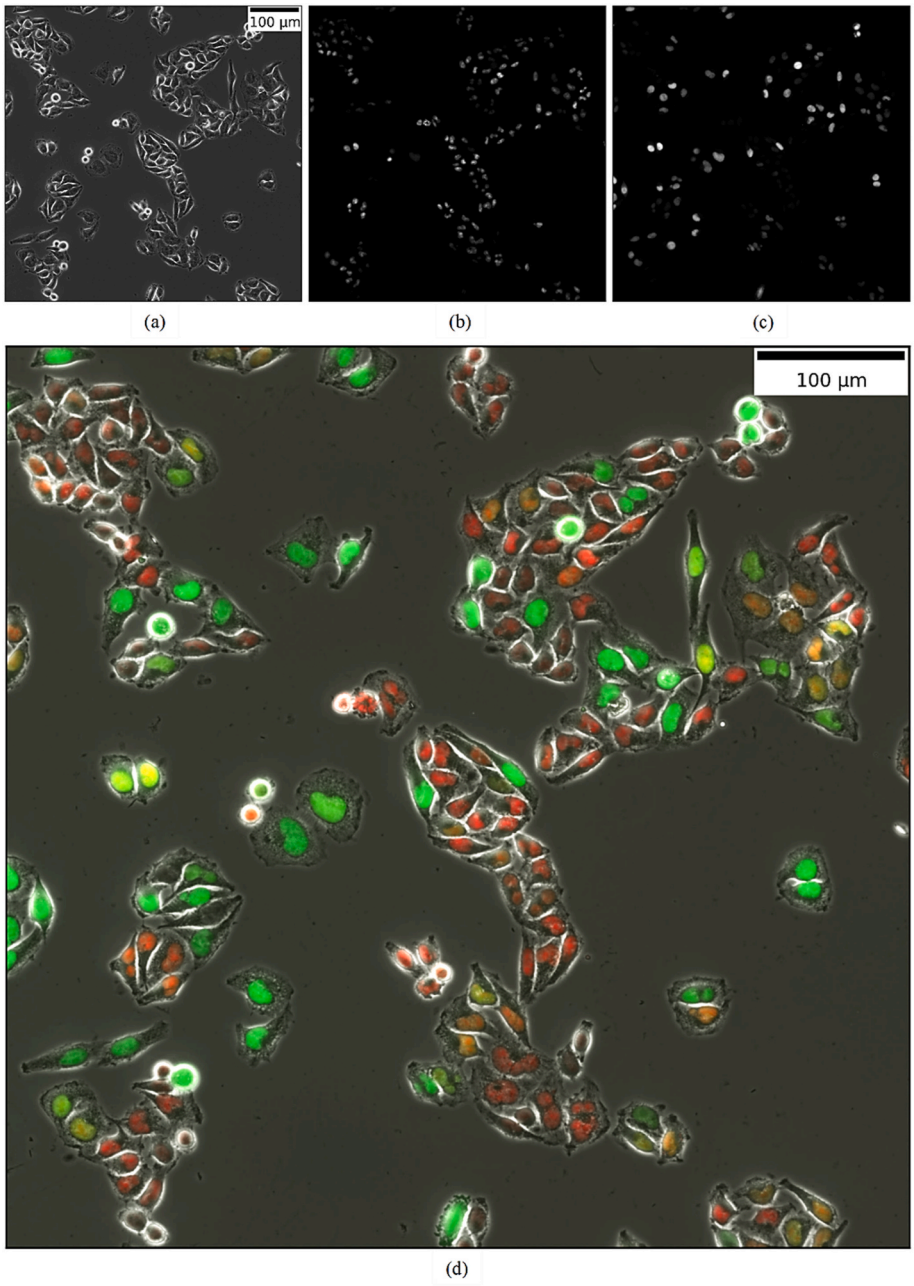


Fig. 1. Composite visualization of HeLa cells captured using a standard motorized microscope. (a) Phase contrast image, (b) Fluorescent Red image, (c) Fluorescent Green image. The composite image (d) combines all three channels for enhanced visualization. Scale bar: 100 µm. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

cell cycle progression in live cells, subsequently adopted by researchers [4–9].

The FUCCI method leverages Cdt1 and Geminin's distinct characteristics at different cell cycle phases. For the G1 phase, a fusion protein of Cdt1 (amino acids 30–120) and mKO2 fluorescent protein is utilized. To visualize the S, G2, and M phases, a fusion protein of Geminin (amino acids 1–110 or 1–60) and mAG1 fluorescent protein is employed [4]. By exploiting the ubiquitin-proteasome system to degrade replication licensing factors selectively and rapidly, FUCCI enables effective cell cycle visualization. Traditionally, fluorescence microscopy tracked DNA replication and repaired protein activity like Proliferating Cell Nuclear Antigen (PCNA) to distinguish between cell cycle phases [8,10].

Various image processing techniques such as nearest neighbor, level-set segmentation, support vector machine (SVM), intensity histogram, and intensity surface curvature have been used for fluorescence intensity image segmentation [11,12]. Recent advancements in deep learning and convolutional neural networks have facilitated cell cycle tracking in individual cell fluorescence images, albeit requiring labeled datasets for training [13]. Calculating cellular fluorescence intensity is crucial for generating ground truth datasets involving microscopic image segmentation and cellular mask creation.

Fluorescence microscopy image segmentation commonly employs Voronoi and seeded watershed techniques [14–16]. While the watershed method is versatile, modifications have been made to address splitting or merging issues [17]. Its efficacy heavily relies on initial seeds; without proper background seeding, accurate cell outlines are challenging to generate (similar to the Voronoi algorithm). Region-based segmentation techniques are well-suited due to relatively uniform statistical features of foreground and background in fluorescence microscopy images [18]. Although region-growing segmentation techniques exist in image processing and computer vision fields, their direct application to fluorescent microscopy imaging is limited [19]. Recent studies have displayed deep convolutional neural networks' effectiveness in addressing segmentation challenges unresolvable by conventional methods [20–22].

2. Methods

In this study, we propose an automated approach for segmenting cells in phase contrast images, generating masks and outlines, and measuring cell fluorescence intensity in both the red and green channels. Classification of live or fixed cells based on their unlabeled

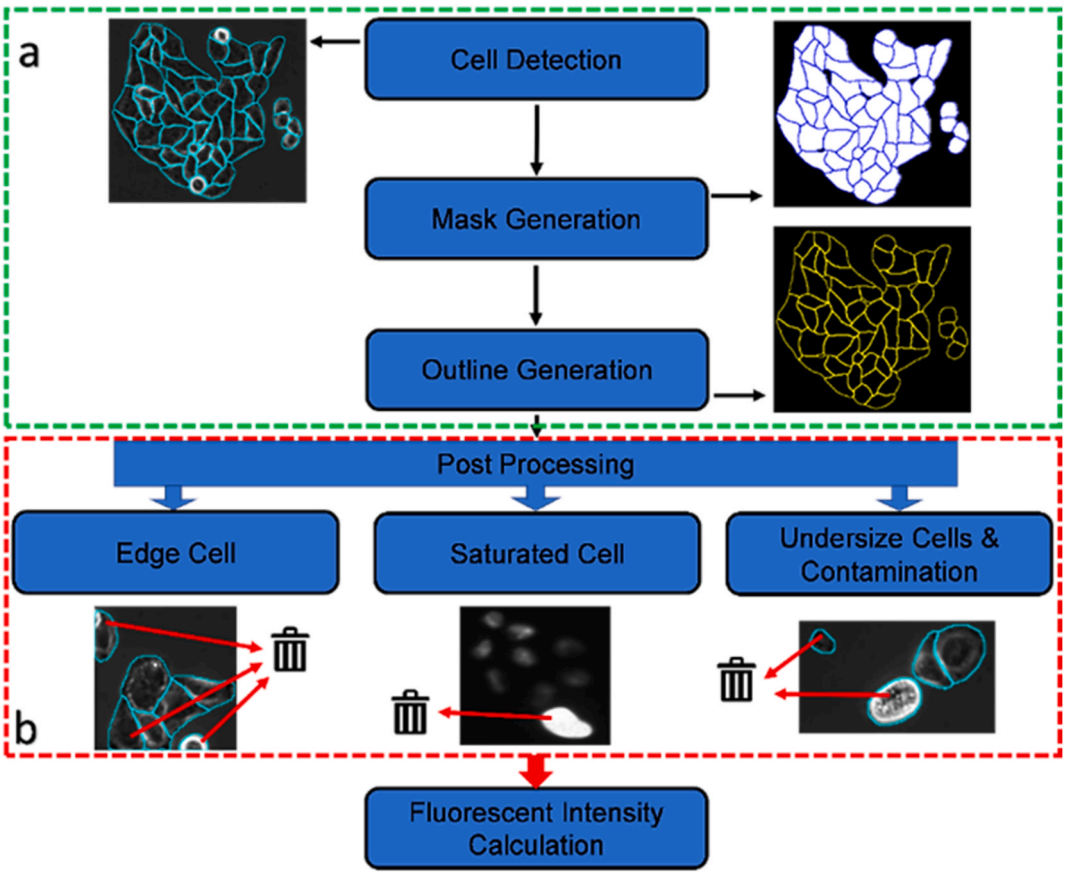


Fig. 2. Illustrating the automated pipeline employed for measuring cellular fluorescent intensities. The process consists of two primary stages: (a) Segmentation and outline generation; and (b) Post-Processing involving the removal of contaminants and non-fluorescing cells to ensure accurate quantification.

microscopic images would be a powerful tool for cell biology and pathology. Fig. 1 provides an example of the image data analyzed in this study.

We present an automated pipeline for cell segmentation utilizing the Cellpose algorithm [23]. This algorithm proves to be a versatile tool capable of accurately segmenting cells and generating cell outlines. Our study focuses on analyzing the cell cycle of HeLa–FUCCI cells, which express fluorescent red and green reporter proteins. The automated pipeline consists of several key steps. Following the preparation of HeLa–FUCCI cells, we acquired a dataset by scanning 37,000 cells using a standard motorized microscope to capture images in phase contrast, as well as fluorescent red and green channels. To segment the cells and create outlines and masks, we leverage the Cellpose algorithm, as shown in Fig. 2a. We refined cell segmentation and purified the image dataset in several postprocessing steps. Ultimately, we integrate the fluorescent pixel intensity of each cell using the masks derived from the phase contrast channel, as depicted in Fig. 2b.

2.1. HeLa FUCCI cell preparation

HeLa–FUCCI cells (RCB2812, RIKEN BRC) [4] were cultured in Dulbecco's Modified Eagle's Medium (DMEM, 21885025 Gibco) supplemented with 10 % Fetal Bovine Serum (Biowest SAS, France) and a mixture solution of 100 U/mL penicillin and 100 $(\mu g/mL)$ streptomycin (Merck, Germany). The cells were maintained in a humidified incubator at 37 °C with 5 % CO2. For imaging purposes, the cells were seeded in 33 mm glass bottom Petri dishes (Greiner, Hungary) and incubated overnight. Once the cells adhered to the dish, the medium was replaced with 20 mM 2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid (HEPES, Merck, Germany) in Hank's balanced salt solution (HBSS, Merck, Germany) with a pH of 7.4. Alternatively, cells were fixed with 4 % paraformaldehyde following the overnight incubation.

2.2. Image collection of HeLa FUCCI cells

HeLa–FUCCI cells were employed to create a comprehensive image dataset specifically designed for our analysis. This dataset comprises 324 images, each including phase contrast images along with corresponding fluorescent red and green channel images. The images were automatically captured using a standard motorized ZEISS microscope (Axio Observer) with a 20x objective magnification to ensure consistency and efficiency in data acquisition. A PCO Edge 4.2 camera was utilized for image acquisition, capturing images with a resolution of 3152 pixels/mm. To enhance image quality and contrast, particular attention was given to this dataset's generation, with an exposure time of 3 s for each fluorescent image. We utilized this high-quality imaging setup to ensure that our analysis is based on accurate and reliable data.

2.3. Cell segmentation and outline generation using cellpose

In our research, we focused on accurately segmenting cells and generating mask images and outlines from unlabeled phase-contrast images. To achieve this, we employed the powerful Cellpose algorithm, known for its precise cell segmentation capabilities and ability to produce accurate outlines. The algorithm utilized topological maps derived from ground truth masks that were manually annotated by a human expert. Through training a U-Net-based neural network [24], we successfully predicted the horizontal and vertical gradients of these topological maps, along with a binary map indicating whether a pixel belonged to a region of interest (ROI) or not. During testing, the neural network accurately predicted the vector fields representing these gradients. By merging pixels that converged to the same location, we were able to distinguish individual cells and capture their unique morphologies.

To further refine the cell shapes, we eliminated pixels projected by the network to be outside of cells. Utilizing the Cellpose algorithm, we generated masks and outlines for all the images in our dataset. For a visual demonstration of the cell segmentation and outline generation process, please refer to Fig. 2.

2.4. Post processing stage

The post-processing stage consists of three sublevels aimed at refining the accuracy and reliability of our results. These steps are Edge Cell Identification, Handling Saturated Cells, and Contamination Identification. Each sublevel addresses specific challenges encountered during the cell identification process, ensuring the generation of clean data. Through the implementation of these three sublevels, we achieved more accurate results in fluorescence intensity analysis. This comprehensive post-processing approach is vital in producing high-quality data for further analysis and interpretation, enabling us to draw meaningful insights and conclusions in our research.

1. Edge Cell Identification

In the image processing pipeline, the accurate calculation of fluorescence intensity values is crucial. However, one challenge arises from cells located at the image edges, which provide inaccurate intensity readings due to their incomplete morphology. We devised a strategy to identify and handle these edge cells. By investigating the overlap between cell coordinates and edge coordinates, we excluded cells with more than three common pixels from the analysis. This approach allowed us to avoid incorporating intensity data from edge cells, thereby preserving the accuracy and integrity of the calculations.

2. Handling Saturated Cells

In microscopic imaging, prolonged fluorescence exposure can lead to cells with saturated pixel values equal to 255 in the case of 8-bit images. Such high-intensity cells introduce artifacts in the analysis. To address this issue, we implemented a strategy during the post-processing stage to handle saturated cells. By excluding cells with pixel values equal to 255 from the calculation process, we ensure accurate fluorescence intensity measurements. Fig. 2 (b) provides an example of a saturated cell.

3. Small Particle Removal

We encountered images with Petri dish contaminations including small particles with only a few detected pixels. These contaminations were mistakenly identified as cells by the Cellpose 1.0 algorithm. To address this issue, we took a proactive approach by training the Cellpose 2.0 algorithm using 52 randomly selected images. This training process allowed us to teach the network to differentiate between actual cells and contaminations, improving the accuracy of cell identification.

To further refine our dataset, we implemented a filtering mechanism based on a set threshold. To determine this threshold, we conducted an analysis of the entire cell sizes. The process involved sorting them from the smallest to the largest. During this analysis, we observed that cells with an area of less than 130 $(\mu m)^2$ or 1300 pixels were categorized as contaminated.

By effectively managing the impact of contamination in our dataset, we can confidently draw meaningful conclusions and make informed decisions based on reliable data. The integration of Cellpose 2.0 and the exclusion of small particles strengthens the robustness of our analysis and enhances the quality of the results. Small particles were excluded from further analysis using the formula for cell area calculation, which is given by equation (1) in this study.

$$Cell\ area(\mu m)^2 = \frac{Cell\ area\ in\ pixels}{(Resolution)^2} \tag{1}$$

where the resolution depends on the microscope objective and camera.

2.5. Calculation of total cellular intensity and cellular brightness

The calculation of total cellular intensity and cellular brightness is a widespread method in the quantification of fluorescence levels within cells and the analysis of fluorescent signal distribution across cell populations [25–27]. This critical procedure commences with cell detection in phase images, yielding precise spatial coordinates, cellular masks, and outlines that serve as indispensable tools for identifying cells in the fluorescent red and green channels. By employing a pixel-wise multiplication technique, pixel values from fluorescent red and green signals are multiplied with corresponding masks derived from phase contrast images, effectively isolating and quantifying specific fluorescent intensity unique to each cell. The total cellular intensity is then determined by summing pixel values in the original image weighted by the corresponding mask value identified by Cellpose, ensuring accurate quantification across the dimensions of the cellular mask. The formulation of our approach can be summarized as follows:

- **Total Cellular Intensity (TCI):** The crux of this analysis lies in the computation of Total Cellular Intensity, which stands as the summation of pixel values (P_{ij}) in the original image, meticulously weighted by the corresponding mask value (M_{ij}) identified by Cellpose. The summation is carried out across the height (h) and width (w) of the bounding box of the cellular mask. The formula adds up pixel values in the image, adjusted by mask values, to measure cell fluorescence intensity. Total Cellular Intensity (TCI) calculation, as defined by equation (2) in this study, plays a crucial role in quantifying cell fluorescence intensity based on pixel values and corresponding mask values identified by Cellpose.

$$TCI = \sum_{i=1}^{h} \sum_{j=1}^{w} P_{ij} \times M_{ij} \tag{2}$$

- **Cell Area (CA):** The concept of Cell Area emerges as a pivotal metric, encapsulating the number of pixels nestled within each segmented cell, thereby portraying the spatial expanse of the cell under scrutiny:

$$CA = Number\ of\ Pixels\ in\ Segmented\ Cell$$

- **Cellular Brightness (CB):** The concept of Cellular Brightness (CB) is defined as the ratio of Total Cellular Intensity (TCI) to Cell Area (CA), providing a measure of the average intensity of fluorescent signals across individual cells. This metric is useful in understanding the levels of fluorescence in cells and can be calculated using equation (3) in the document. It is important to note that the accuracy of CB depends on the accurate measurement of TCI and CA, which in turn relies on the quality of the image and the effectiveness of the image processing techniques used. The use of appropriate image processing techniques and software can help ensure the accuracy of CB measurements, which can be valuable in various applications, such as cellular assays, live-cell imaging, and high-content screening.

$$CB = \frac{TCI}{CA} \tag{3}$$

- **Background Noise Reduction:** We conducted data visualization and identified background noise, which led to an increase in cellular brightness equivalent to the pixel size of a cell. To address this issue, we computed the background noise for the darkest cells per pixel. Subsequently, we determined the slope of the line across both axes (axis-x for the red channel and axis-y for the green channel) using a linear equation formed by two points on the outer side of the data. Following this, we applied equation (4) to mitigate the background noise effect. The experimentally calculated slope was 6.81 for the red channel and 9.9 for the green channel. Fig. 6 illustrates the process of reducing background noise in both the red and green fluorescent intensity signals.

$$CBWBN\ (Cellular\ Brightness\ Without\ Background\ Noise) = \frac{TCI - (slope\ value * CA)}{CA} \tag{4}$$

- **Normalization of Cellular Brightness:** In the final step, we applied equation (5) to normalize cellular brightness (CB) values for each channel, ensuring they are constrained within the range of zero to one. The maximum cellular brightness (CB_{max}) observed for the brightest cell in the red channel was 145.81 and 159.48 for the green channel. Similarly, the minimum cellular brightness (CB_{min}) was 6.95 for the darkest cell in the red channel and 9.97 for the green channel.

$$NCBWBN\ (Normalized\ Cellular\ Brightness\ Without\ Background\ Noise) = \frac{CBWBN - CBWBN_{min}}{CBWBN_{max} - CBWBN_{min}} \tag{5}$$

3. Results and discussion

We aimed to devise an automated pipeline for calculating the precise fluorescent intensity of HeLa FUCCI cells. This pipeline encompassed cell segmentation, data refinement, and integration of fluorescent signals within individual cells. Our dataset comprised 972 images captured across three channels: phase contrast, fluorescent red, and fluorescent green. Within this dataset, 37,000 individual cells were meticulously segmented, yielding cell masks and outlines. Rigorous data purification stages were employed to eliminate false positive detections and cells unsuitable for the measurement. Ultimately, we successfully derived cellular brightness values for the 30,505 remaining cells.

As shown in Fig. 3, to visually elucidate and dissect the outcomes, we provided a 2D scatter plot in conjunction with a heatmap

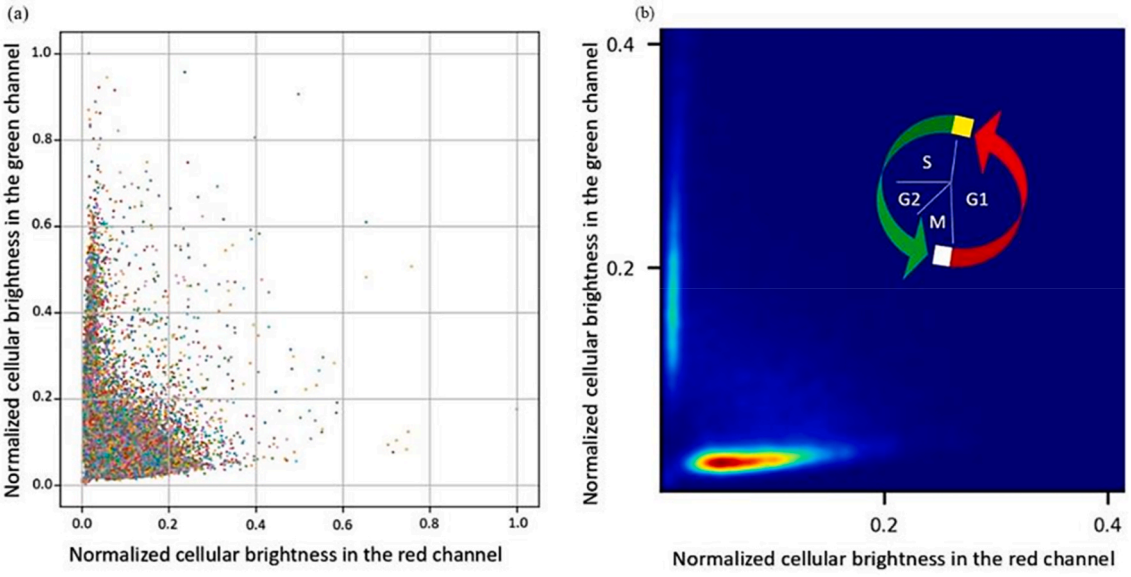


Fig. 3a. 2-dimensional phase space illustrating cells in various stages of the cell cycle. The depiction showcases the normalized cellular brightness of two distinct cell cycle reporter proteins across a dataset encompassing 30,505 individual cells, elucidating the specific phase of each cell within the cell cycle. This visualization offers a comprehensive insight into the distribution of cells within the cell cycle phases. The accompanying heatmap (Fig. 3b) provides a visual representation of cell density within different regions of this 2D phase space, aiding in the identification of cell clusters and enhancing the clarity of results.

generated with a smoothing rate (α) set to 8. The scatter plot presents the cellular brightness of each cell in the red and green channels, showing a continuous distribution of cells across the cell cycle. In the heatmap, two clusters of cells can be observed. Cells that predominantly align close to the x-axis can be categorized as red, indicating high intensity in the red and low in the green channel. Conversely, cells positioned predominantly along the y-axis can be classified as green (Fig. 3a and b).

The heatmap's classification scheme predominantly divides cells into red and green classes based on their respective intensity values. To further elucidate the classification process and deepen comprehension, we present a selection of 10 random images from each class, as shown in Fig. 4. Panel (a) displays cells in the green state of their life, panel (b) shows cells in the red phase, and panel (c) exhibits cells from the dark stage of the cell cycle. This depiction unveils the different morphology of cells in 3 different stages of the cell cycle, thereby contributing to a more in-depth analysis of our findings.

We employed the density heatmap as an alternative visualization method to represent the distribution of data throughout the cell cycle. In Fig. 5, we introduced three distinct rectangles, each representing the count of classes at different cell cycle stages. The red rectangle represents cells in the red class, the green rectangle signifies cells in the green stage of their life cycle, and the gray rectangle denotes dark cells. Any remaining cells were categorized as yellow cells.

The pursuit of minimizing background noise from detected cells constitutes a multistep process. Commencing this process involves constructing a graph to visualize cell data, where the x-axis represents cell area in pixels, and the y-axis signifies total cellular intensity. Fig. 6 demonstrates the graph. Subsequently, we determined the slope of the line fitted to the darkest cells with a certain area. This line corresponds to the fluorescent background intensity, which can be subtracted from the cellular brightness values.

Background noise reduction hinges on integrating pixel counts associated with identified cells. By multiplying pixel counts by the respective slope values of the *x* and *y* axes, an estimation of background noise contribution is derived. This estimation is then subtracted from the overall detected cell value, as shown in Fig. 6 (a) for the red fluorescent intensity signal and Fig. 6 (b) for the green fluorescent intensity signal. This method corrects the fluorescent brightness of cells by subtracting background noise. The x-axis in both visualizations represents the pixel count for both channels.

Red triangles highlight the slope corresponding to the darkest cells with a specific pixel area. We consider their brightness as the fluorescent background noise.

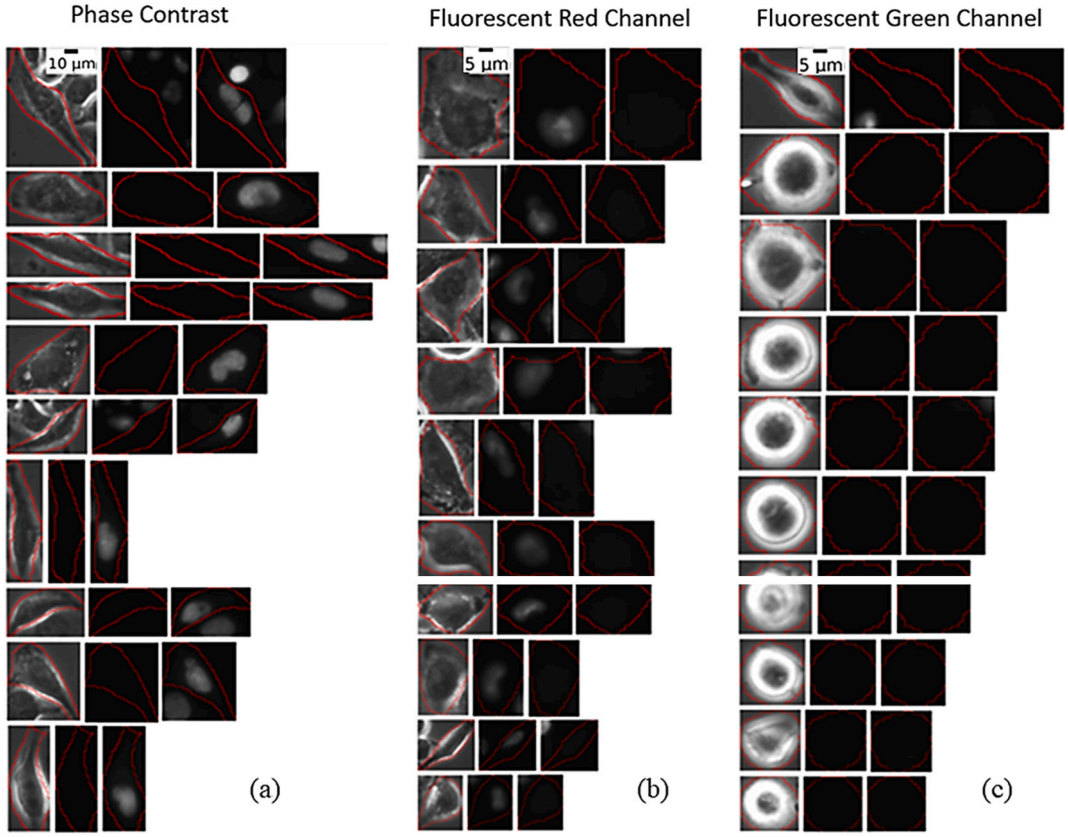


Fig. 4. An example of 10 randomly selected cells from the three classes. (a) cells in the green state of their life, (b) cells in the red phase, and (c) cells from dark stage of cell cycle. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

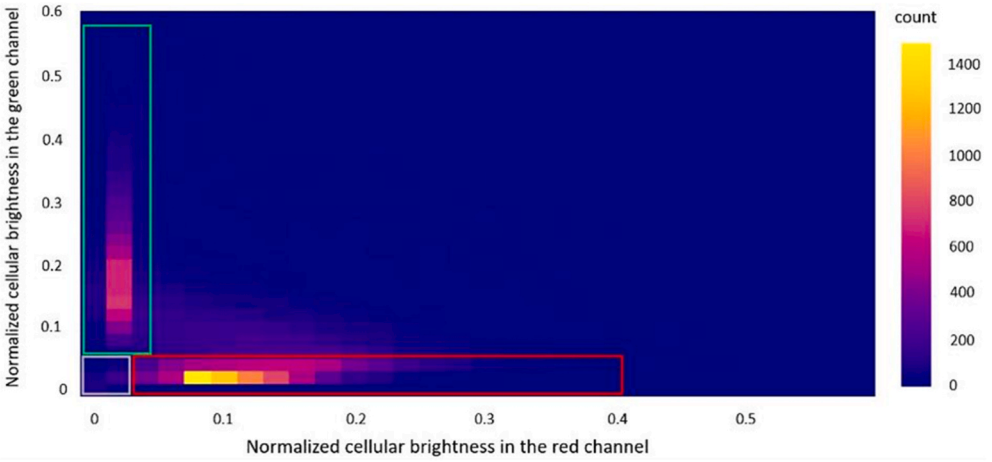


Fig. 5. Density heatmap depicting data distribution in the cell cycle. The figure displays three rectangles representing the count of classes at different cell cycle stages. The red rectangle signifies cells in the red class, the green rectangle represents cells in the green stage, and the gray rectangle shows dark cells. The definition of these 3 classes includes a manual step, i.e., drawing the rectangles. According to the continuous distribution of cells in Fig. 3a, classifying cells into red, green, yellow, and dark phases is not straightforward. However, previous studies used these classes [28–30], and our heatmap enables us to define them. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.1. Comparative analysis of cellpose 1.0 and cellpose 2.0 in cell detection performance

We employed Cellpose 2.0 to conduct a comparative analysis with its predecessor, Cellpose 1.0. Our research commenced with the critical task of preparing the ground truth dataset. To achieve this, we selected 52 images at random and leveraged Cellpose 1.0 for initial cell detection. Subsequently, a manual examination was performed to identify and annotate any contaminations and minute particles within the detected cells. Following this, we carefully refined the generated mask files, ensuring their accuracy. These refined masks were then provided to Cellpose 2.0, which served as the foundation for training our advanced model.

In this comparative analysis, we meticulously assess the performance of two versions of the same cutting-edge cell detection model: Cellpose 1.0 and its successor, Cellpose 2.0. The evaluation focuses on their capacity to accurately identify cells in 52 randomly selected images, considering the presence of contaminations that should not be detected as cells, as well as the background regions that do not represent cellular structures, and undetected cells, i.e., false positive and false negative detections. Fig. 7 illustrates instances of the segmentation performance using Cellpose 1.0 (depicted in (a)) and Cellpose 2.0 (shown in (b)). Notably, within the image below, Cellpose 2.0 displays remarkable accuracy by identifying only one falsely identified cell. In contrast, using Cellpose 1.0, there are instances where several falsely identified cells are present. These inaccuracies are highlighted by the yellow rectangles in Fig. 7 and quantified in Fig. 8.

By comparing these two models, we aim to discern the advancements and improvements in cell detection capabilities while addressing potential challenges posed by contaminants and non-cellular regions. As shown in Table 1, Cellpose 1.0 exhibits a false positive rate of 10 %, with 127 instances where non-cellular structures or contaminations were incorrectly classified as cells. This highlights a tendency of Cellpose 1.0 to generate false positive detections. Nonetheless, it demonstrates a commendable true positive rate of 97.6 % successfully identifying a substantial number of cells within the images. Conversely, Cellpose 2.0 exhibits remarkable improvements in performance. It highlights a significantly reduced false positive rate of 2.7 %, indicating a substantial reduction in the misclassification of contaminations and non-cellular regions. Furthermore, Cellpose 2.0 achieves a higher true positive rate of 98.4 % precisely detecting a greater number of cells while effectively managing the presence of contaminations and background regions.

It is noteworthy that despite the advancements, both models still exhibit false negatives. Cellpose 1.0 has a 10 % false negative error rate, whereas Cellpose 2.0 shows a reduced false negative error rate equal to 2.7 %. These false negatives represent instances where actual cells were missed in the detection process. As we focus on enhancing the accuracy of cell detection, the false negative error rate for Cellpose 1.0 is 2.4 %, and for Cellpose 2.0, it is further reduced to 1.4 %. To summarize the quantitative assessment, Cellpose 1.0 successfully detected 1275 cells, while Cellpose 2.0 also identified 1275 cells, highlighting comparable performance in terms of the total number of cells. However, Cellpose 2.0 outperforms Cellpose 1.0 in terms of false positive and false negative rates, demonstrating its superior accuracy in cell detection.

The table presents a comprehensive comparison of Cellpose 1.0 and Cellpose 2.0, providing detailed information on the counts of false positive, true positive, and false negative for each model, facilitating a deeper understanding of their respective detection efficiencies. This analysis reinforces the significance of employing Cellpose 2.0 for precise and reliable cell detection tasks, offering valuable insights for research and analysis in various fields of study.

To visually depict the comparative performance of Cellpose 1.0 and Cellpose 2.0, we have prepared a bar chart Fig. 7 displaying key metrics. The bar chart presents the false positive, true positive, and false negative counts for both versions of the model. This visual

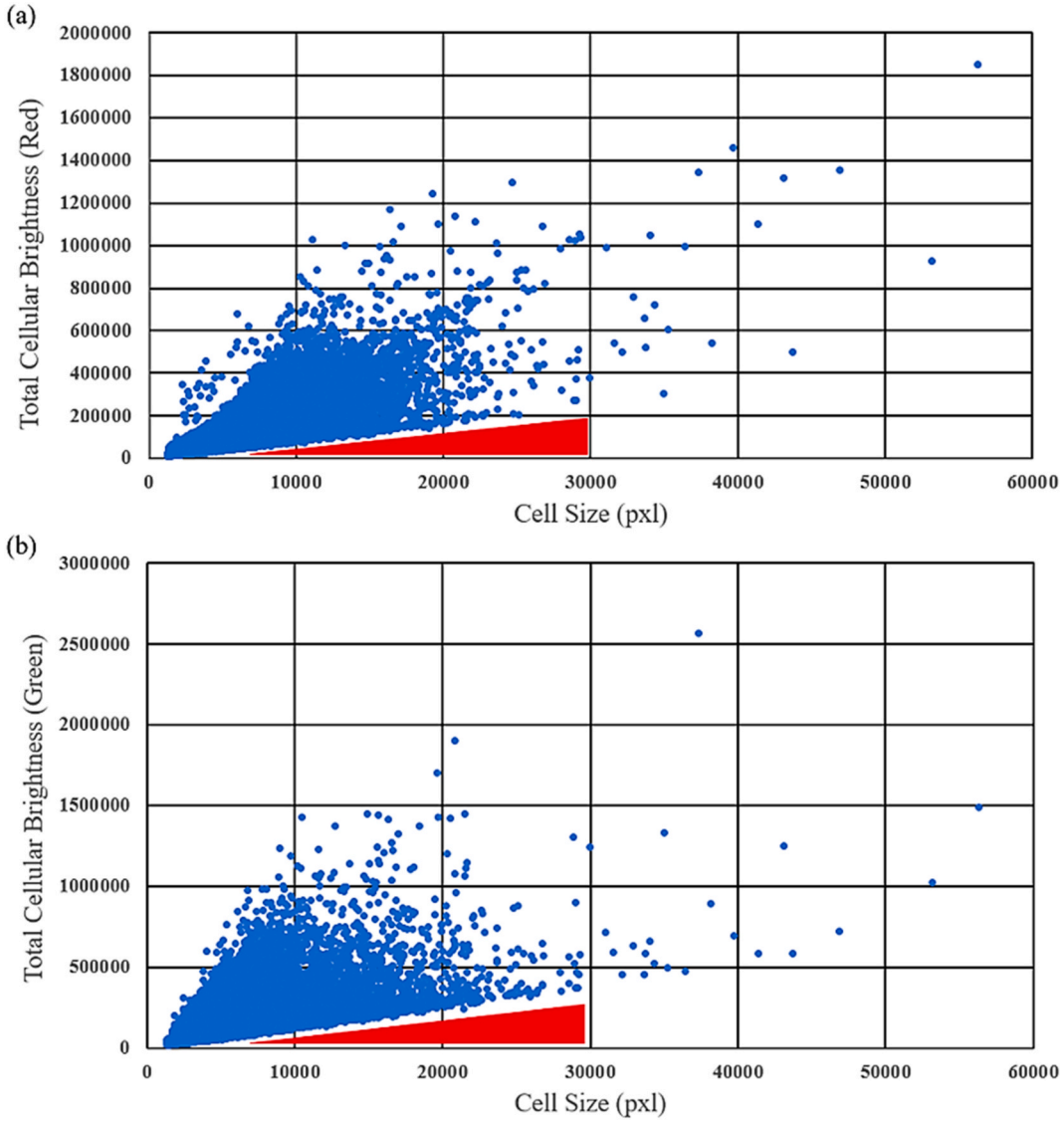


Fig. 6. Background noise reduction process _visualizing total cellular intensity as a function of cell area in pixels and line fitting. (a) red fluorescent channel and (b) green fluorescent channel. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

representation offers an easily interpretable overview of their detection results, enabling a direct comparison of their strengths and weaknesses. In the chart, the height of each bar corresponds to the respective count, making it straightforward to assess the differences in false positive and true positive rates between the two versions. By leveraging this bar chart, we gain valuable insights into the models' abilities to manage contaminations and non-cellular background regions effectively.

4. Conclusion

In conclusion, our study presents a comprehensive and automated pipeline for cell segmentation and fluorescent intensity measurement, addressing crucial challenges in quantitative cell biology. By leveraging the powerful Cellpose algorithm, we achieved precise cell segmentation and outline generation, enabling accurate measurement of fluorescence intensity in multiple channels for individual cells. Our focus on HeLa–FUCCI cells expressing red and green fluorescent reporter proteins allowed us to gain insights into the cell cycle. We found that the distribution of the cells is continuous in the cell cycle. The classification of cells into four classes (red, green, yellow, and dark) is rather arbitrary.

Through the integration of deep learning techniques, convolutional neural networks, and region-based segmentation, our approach outperforms traditional methods and streamlines the analysis process. The use of automated segmentation not only enhances efficiency

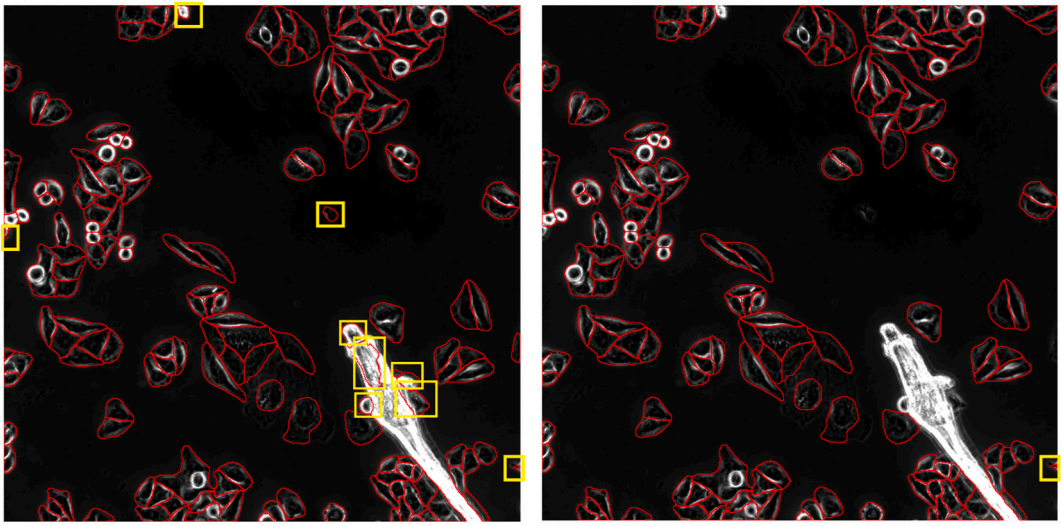


Fig. 7. Example image demonstrating the segmentation performance of Cellpose 1.0 (a) and Cellpose 2.0 (b) in identifying cellular structures. Yellow rectangles highlight instances of falsely identified cells (false positive detections). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

but also eliminates potential biases introduced by manual methods, ensuring robust and reliable results. Furthermore, our post-processing stage, consisting of three sublevels – Edge Cell Identification, Handling Saturated Cells, and Small Cell Removal, significantly improved the accuracy and reliability of our analysis. By effectively managing edge cells, saturated cells, and contaminants, we eliminated potential distortions in the data and produced high-quality results for further analysis and interpretation.

A key highlight of our study lies in the comparative analysis of Cellpose 1.0 and Cellpose 2.0 in cell detection performance. We demonstrated that Cellpose 2.0 exhibits notable improvements over Cellpose 1.0, with a significantly reduced false positive rate and a higher true positive rate. This higher efficiency of Cellpose 2.0 ensures more accurate identification of cells, effectively managing the presence of contaminations and background regions, and making it an ideal choice for cell detection tasks.

In summary, our automated pipeline, along with the comparative analysis of Cellpose 1.0 and Cellpose 2.0, contributes to advancing cell biology research by providing a robust, reliable, and high-throughput approach for cell segmentation and fluorescent intensity measurement. We believe that our findings will not only enhance current methodologies but also inspire future investigations in the field of quantitative cell biology, ultimately contributing to the advancement of scientific knowledge.

We observed distinct clusters in the 2D heatmap based on the red and green channels of the FUCCI construct. Cells in the red, and green states formed separate clusters, with yellow cells co-expressing both reporter proteins positioned between the red and green states.

Overall, our method offers a reliable and efficient solution for quantitatively analyzing cellular fluorescence and investigating the cell cycle dynamics of HeLa–FUCCI cells. By automating the segmentation process and integrating precise intensity measurements, we enable comprehensive analysis on a large scale, paving the way for further studies in quantitative cell biology and facilitating research in areas such as development, regeneration, and cancer biology.

Data availability

The data that support the findings of this study are available from the Zenodo at https://zenodo.org/. These data include raw image files of HeLa FUCCI cells. The dataset can be accessed and downloaded, by using the DOI provided https://doi.org/10.5281/zenodo.8289362.

CRediT authorship contribution statement

Hamid Cheraghi: Writing – original draft, Visualization, Software, Investigation. **Kinga Dóra Kovács:** Resources. **Inna Székács:** Writing – review & editing, Resources. **Robert Horvath:** Writing – review & editing, Conceptualization. **Bálint Szabó:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

We declare no competing interests.

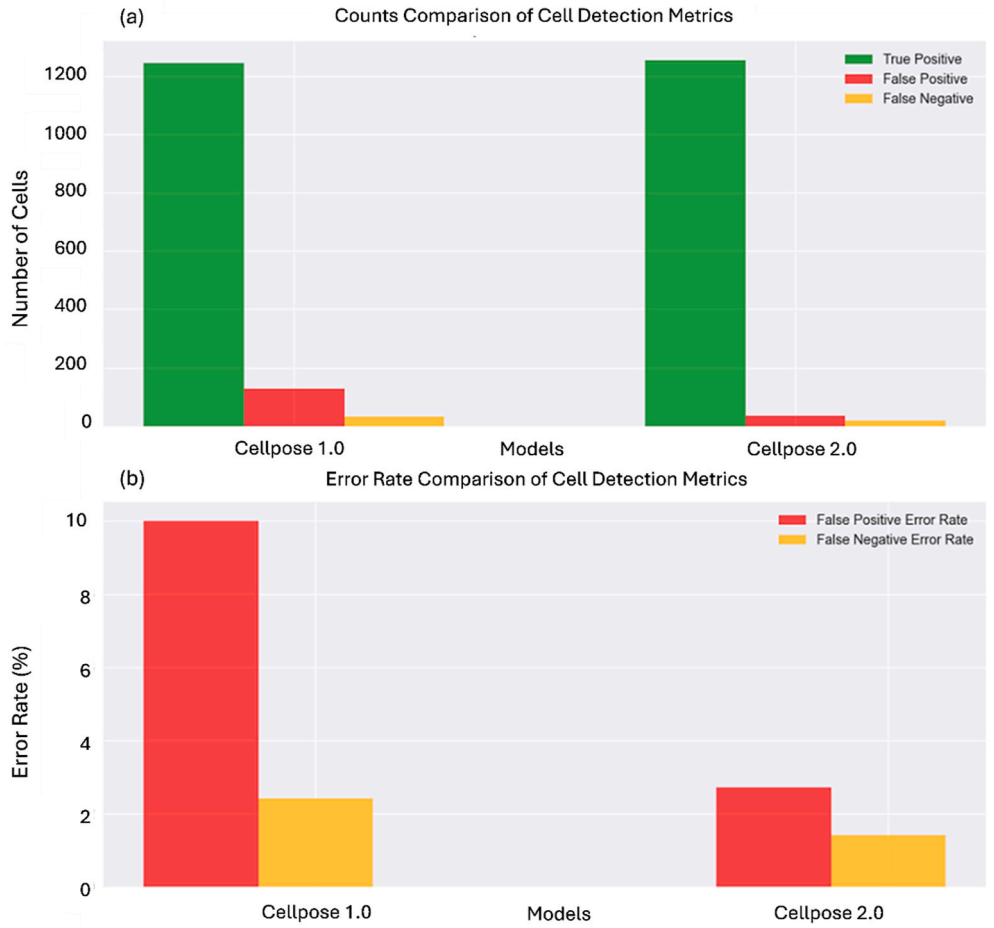


Fig. 8. Performance Comparison of Cellpose 1.0 and Cellpose 2.0 in Cell Detection. The bar chart (a) illustrates the performance metrics of Cellpose 1.0 and Cellpose 2.0, displaying their respective true positive, false positive, and false negative rates in a visual format. The chart (b) provides an insightful comparison of the two models' detection capabilities, highlighting the improvements achieved by Cellpose 2.0 in reducing false positive and false negative errors while accurately identifying cells.Overall, the bar chart enhances the comprehensibility of the comparison and provides a concise yet informative visualization of the performance differences between Cellpose 1.0 and its successor, Cellpose 2.0. This visual aid contributes to a deeper understanding of their respective strengths and serves as a valuable tool in assessing their performance in cell detection.

Table 1
Comparative analysis of cellpose 1.0 and cellpose 2.0 cell detection performance.

Model	True positive	False positive	False negative	False positive error rate	False negative error rate	Total number of cells
Cellpose 1.0	1245	127	30	10 %	2.4 %	1275
Cellpose 2.0	**1255**	**35**	**20**	**2.7 %**	**1.4 %**	1275

Acknowledgments

We extend our sincere appreciation to the numerous sources that have generously supported our research, and we would like to acknowledge the following grants and funding sources, along with the individuals and affiliations associated with them:

1. Marie Skłodowska-Curie grant numbered 860675 by the European Union's Horizon 2020 Research and Innovation program funded the salary of Hamid Cheraghi at CellSorter.
2. Cell2Cell network: Supporting our research at CellSorter.
3. GPU resources: Generously provided by Lénárd Szántó and Prof. Gergely Szöllősi.
4. Valuable suggestions in image processing and analysis: Dr. Imre Lajtos and Dr. Dávid Selmeczi.
5. Hungarian Academy of Sciences [Lendület (Momentum) Program]: Supporting our research endeavors.
6. National Research, Development, and Innovation Office (NKFIH) [KKP 129936 and TKP2021-EGA-04 Programs]: Supported Robert Horvath.

7. Ministry of Innovation and Technology of Hungary: Funding under the TKP2021 funding scheme and the KDP-2021 program funded the salary of Kinga Dóra Kovács.

We deeply appreciate the contributions of these sources, and their financial support has been instrumental in the success of our research.

References

[1] K. Nasmyth, A prize for proliferation, Cell 107 (6) (2001) 689–701.
[2] N. Zielke, B.A. Edgar, FUCCI sensors: powerful new tools for analysis of cell proliferation: FUCCI sensors, WIREs Dev. Biol. 4 (5) (Sep. 2015) 469–487, https://doi.org/10.1002/wdev.189.
[3] K.A. Schafer, The cell cycle: a review, Vet. Pathol. 35 (6) (Nov. 1998) 461–478, https://doi.org/10.1177/030098589803500601.
[4] A. Sakaue-Sawano, et al., Visualizing spatiotemporal dynamics of multicellular cell-cycle progression, Cell 132 (3) (Feb. 2008) 487–498, https://doi.org/10.1016/j.cell.2007.12.033.
[5] R. Ungai-Salánki, et al., Single-cell adhesion strength and contact density drops in the M phase of cancer cells, Sci. Rep. 11 (1) (Dec. 2021) 18500, https://doi.org/10.1038/s41598-021-97734-1.
[6] L. Rappez, A. Rakhlin, A. Rigopoulos, S. Nikolenko, T. Alexandrov, DeepCycle reconstructs a cyclic cell cycle trajectory from unsegmented cell images using convolutional neural networks, Mol. Syst. Biol. 16 (10) (Oct. 2020), https://doi.org/10.15252/msb.20209474.
[7] M. Malumbres, M. Barbacid, Cell cycle, CDKs and cancer: a changing paradigm, Nat. Rev. Cancer 9 (3) (Mar. 2009) 153–166, https://doi.org/10.1038/nrc2602.
[8] Y.R. He, et al., Cell cycle stage classification using phase imaging with computational specificity, ACS Photonics 9 (4) (Apr. 2022) 1264–1273, https://doi.org/10.1021/acsphotonics.1c01779.
[9] Á.G. Nagy, N. Kanyó, A. Vörös, I. Székács, A. Bonyár, R. Horvath, Population distributions of single-cell adhesion parameters during the cell cycle from high-throughput robotic fluidic force microscopy, Sci. Rep. 12 (1) (2022) 7747.
[10] G. Maga, U. Hübscher, Proliferating cell nuclear antigen (PCNA): a dancer with many partners, J. Cell Sci. 116 (15) (Aug. 2003) 3051–3060, https://doi.org/10.1242/jcs.00653.
[11] M. Wang, X. Zhou, F. Li, J. Huckins, R.W. King, S.T.C. Wong, Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy, Bioinformatics 24 (1) (Jan. 2008) 94–101, https://doi.org/10.1093/bioinformatics/btm530.
[12] D. Padfield, J. Rittscher, N. Thomas, B. Roysam, Spatio-temporal cell cycle phase analysis using level sets and fast marching methods, Med. Image Anal. 13 (1) (Feb. 2009) 143–155, https://doi.org/10.1016/j.media.2008.06.018.
[13] A. Zargari, G.A. Lodewijk, N. Mashhadi, N. Cook, C.W. Neudorf, K. Araghbidikashani, R. Hays, S. Kozuki, S. Rubio, E. Hrabeta-Robinson, A. Brooks, DeepSea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy, Cell Reports Methods 3 (6) (2023) 100500.
[14] G. Srinivasa, M.C. Fickus, Yusong Guo, A.D. Linstedt, J. Kovacevic, Active mask segmentation of fluorescence microscope images, IEEE Trans. Image Process. 18 (8) (Aug. 2009) 1817–1829, https://doi.org/10.1109/TIP.2009.2021081.
[15] T.R. Jones, A. Carpenter, P. Golland, Voronoi-based segmentation of cells on image manifolds, in: Y. Liu, T. Jiang, C. Zhang (Eds.), Computer Vision for Biomedical Image Applications, vol. 3765, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 535–543, https://doi.org/10.1007/11569541_54.
[16] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, IEEE Trans. Pattern Anal. Mach. Intell. 13 (6) (Jun. 1991) 583–598, https://doi.org/10.1109/34.87344.
[17] M. Gamarra, E. Zurek, H.J. Escalante, L. Hurtado, H. San-Juan-Vergara, Split and merge watershed: a two-step method for cell segmentation in fluorescence microscopy images, Biomed. Signal Process Control 53 (2019) 101575.
[18] Jadwiga Rogowska, Overview and fundamentals of medical image segmentation. Handbook of Medical Imaging, Processing and Analysis, 2000, pp. 69–85.
[19] Z. Liu, L. Jin, J. Chen, Q. Fang, S. Ablameyko, Z. Yin, Y. Xu, A survey on applications of deep learning in microscopy image analysis, Comput. Biol. Med. 134 (2021) 104523.
[20] Y. Al-Kofahi, A. Zaltsman, R. Graves, W. Marshall, M. Rusu, A deep learning-based algorithm for 2-D cell segmentation in microscopy images, BMC Bioinf. 19 (1) (2018) 1–11.
[21] A. Wang, Q. Zhang, Y. Han, S. Megason, S. Hormoz, K.R. Mosaliganti, J.C. Lam, V.O. Li, A novel deep learning-based 3D cell segmentation framework for future image-based disease detection, Sci. Rep. 12 (1) (2022) 342.
[22] P. Shrestha, N. Kuang, J. Yu, Efficient end-to-end learning for cell segmentation with machine generated weak annotations, Commun. Biol. 6 (1) (2023) 232.
[23] C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: a generalist algorithm for cellular segmentation, Bioinformatics (preprint, Feb. 2020), https://doi.org/10.1101/2020.02.02.931238.
[24] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, October, pp. 234–241.
[25] E.M. Smith, J.D. Mueller, The statistics of protein expression ratios for cellular fluorescence studies, Eur. Biophys. J. 41 (2012) 341–352.
[26] M.A. Digman, R. Dalal, A.F. Horwitz, E. Gratton, Mapping the number of molecules and brightness in the laser scanning microscope, Biophys. J. 94 (6) (2008) 2320–2332.
[27] K.H. Hur, P.J. Macdonald, S. Berk, C.I. Angert, Y. Chen, J.D. Mueller, Quantitative measurement of brightness from living cells in the presence of photodepletion, PLoS One 9 (5) (2014) e97440.
[28] L. Kurzawa, M.C. Morris, Cell-cycle markers and biosensors, Chembiochem 11 (8) (2010) 1037–1047.
[29] A. Sakaue-Sawano, H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, S. Kashiwagi, K. Fukami, T. Miyata, H. Miyoshi, T. Imamura, Visualizing spatiotemporal dynamics of multicellular cell-cycle progression, Cell 132 (3) (2008) 487–498.
[30] A. Sakaue-Sawano, M. Yo, N. Komatsu, T. Hiratsuka, T. Kogure, T. Hoshida, N. Goshima, M. Matsuda, H. Miyoshi, A. Miyawaki, Genetically encoded tools for optical dissection of the mammalian cell cycle, Mol. Cell 68 (3) (2017) 626–640.