DOI: 10.2478/ijasitels-2025-0014

Deep Learning Segmentation for stem cells images

Constantin-Cristian DRĂGHICI[1], *Cătălina NEGHINĂ*[1]

[1]*Department of Computer Science and Electrical Engineering, Faculty of Engineering, University "Lucian Blaga" of Sibiu, Sibiu, Romania*

Cristian DRĂGHICI, e-mail: constantincristian.draghici@ulbsibiu.ro
Cătălina NEGHINĂ, e-mail: catalina.neghina@ulbsibiu.ro

Abstract

In this study, we investigate the application of deep learning techniques for automatic segmentation of the Fluo-N2DH-GOWT1 dataset, which consists of time-lapse grayscale images of rat neural stem cells. We employ the DeepLabV3+ semantic segmentation framework with a ResNet-18 backbone, chosen for its balance between accuracy and computational efficiency on relatively small biomedical datasets. To improve generalization and robustness, we apply data augmentation strategies including rotation, scaling, shear, and reflection. The performance of the proposed model is evaluated using standard metrics such as F1-score, Intersection of Union, Precision and Recall. Experimental results demonstrate that the ResNet-18–based DeepLabV3+ achieves reliable segmentation of stem cells, effectively distinguishing cells from the background.

Keywords: Stem Cells, Semantic Segmentation, Instance Segmentation, Deep Learning, ResNet

---

1 Introduction

Accurate detection and segmentation of stem cells represent a critical step in the quantitative analysis of cellular dynamics within biomedical research. Advances in computer science have profoundly transformed biological research, particularly through computational image analysis. In microscopy, automatic cell segmentation enables quantitative and reproducible study of morphology and dynamics, overcoming the limitations of manual annotation like time inefficiency, subjectivity, and variability [1], [2]. Recent progress in deep learning, especially convolutional neural networks, has revolutionized image understanding. CNNs learn hierarchical visual representations directly from data, achieving high accuracy in biomedical image segmentation and enabling the precise delineation of complex cellular structures [3], [4].

The focus of this study was the segmentation of fluorescence microscopy images from dataset Fluo-N2DH-GOWT1 [5]. Previous studies using this dataset employed architectures such as U-Net [6], [7], Motion UMiTNet [8], and hybrid or recurrent variants like C-LSTM [9], combining spatial and temporal modeling for improved segmentation accuracy. This work highlights the potential of deep learning techniques in automating the analysis of microscopy images, providing accurate and efficient tools for stem cell research and other biomedical applications [10], [11].

Convolutional Neural Networks (CNNs) extract hierarchical spatial features using convolution and pooling operations, achieving high performance in tasks such as classification and segmentation. Residual Networks (ResNet) introduce skip connections to ease the training of deeper models by learning residual mappings. For semantic segmentation, encoder-decoder architectures predict pixel-level class labels while preserving spatial context. DeepLabV3+ extends this design by integrating Atrous Spatial Pyramid Pooling (ASPP) for multi-scale context extraction and a refinement decoder for boundary accuracy [3], [12], [13].

The primary goal of this study is to set up and evaluate a context-aware deep learning framework for automatic segmentation of stem cells in fluorescence microscopy images. Specifically, the objectives are:

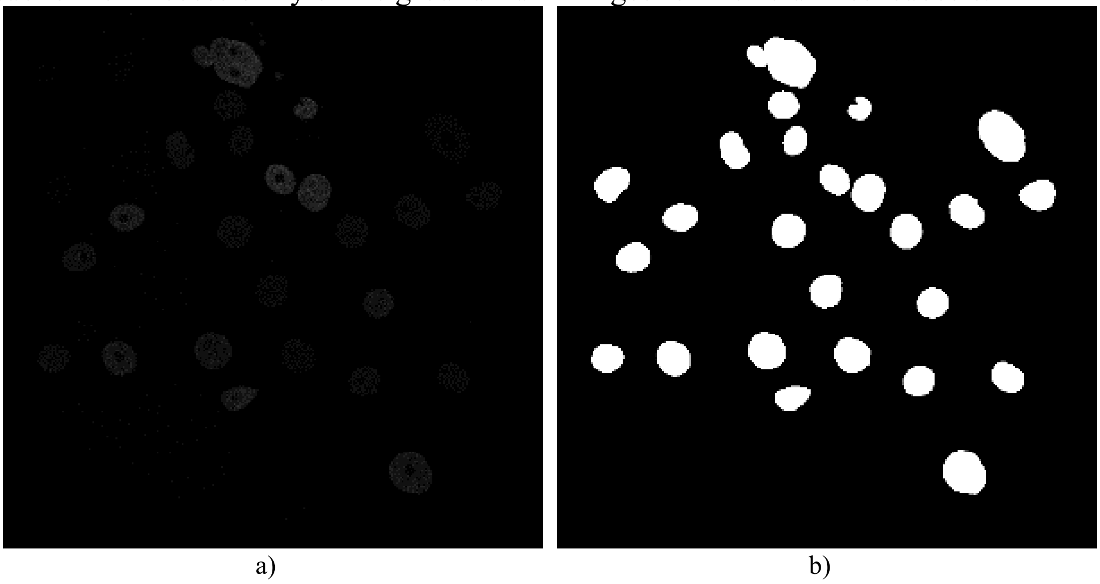
- To implement an image segmentation framework based on the DeepLabV3+ architecture with a ResNet-18 backbone [12], [13], fit for small biomedical datasets
- To train the network on the Fluo-N2DH-GOWT1 dataset with appropriate data augmentation techniques to improve generalization.
- To evaluate the semantic segmentation performance using standard quantitative metrics such as *Intersection over Union* (IoU) and *F1*-score
- To evaluate the instance segmentation using metrics such as *Precision* and *Recall*

By addressing these objectives, this work highlights the potential of contextual deep learning in automating the analysis of microscopy images, providing accurate and efficient tools for stem cell research and other biomedical applications.

2 Dataset

The Fluo-N2DH-GOWT1 dataset, part of the *Cell Tracking Challenge* [14], contains 2 subsets with 2D fluorescence microscopy sequences (*dataset1* with 92 frames and *dataset2* also with 92 frames) of GOWT1 stem cells. Each image is accompanied by instance-level annotations, such as in *Figure 1*. The image resolution is 1024x1024 px. Each dataset consists of successive frames of the same cell culture.

The dataset was split into training (*dataset2*) and validation subsets (*dataset1*), due to the inaccessibility of the ground truth images for the actual test subsets.



a) b)

Figure 1. a) Original image from *dataset2,* b) ground truth instance

3 Methods

A flowchart of the proposed methods can be seen in *Figure 2*.

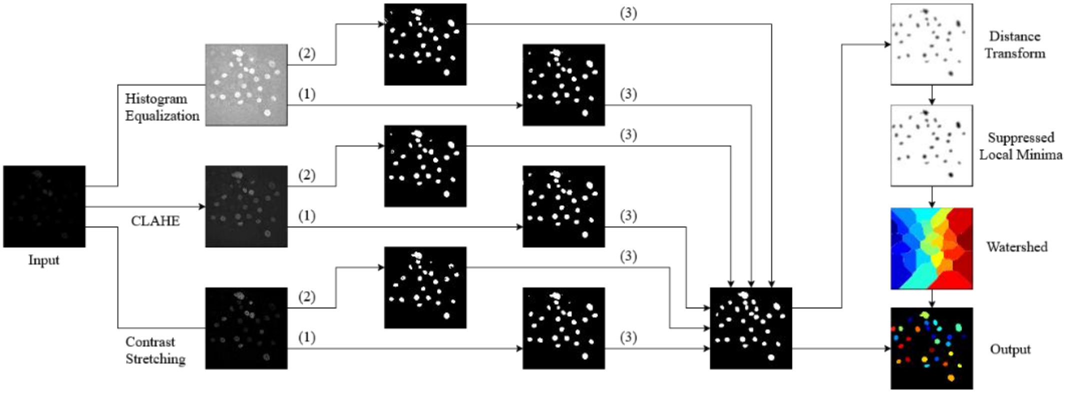


Figure 2. Scenario scheme: (1) Segmentation using a model trained on images without brightness, contrast modifications, or added noise; (2) Segmentation using a model trained on images with brightness and contrast adjustments and with added noise; (3) Series of morphological operations: removal of objects smaller than a given threshold, dilation, hole filling, and erosion.

3.1 Preprocessing

To enhance feature extraction, three grayscale preprocessing techniques were applied, which can be seen in *Figure 3*:

a) *Contrast Stretching*. The method consists of linearly adjusting the image contrast so that the range of pixel values is stretched across the entire available domain [0, 255] using eq.1, where I_{min} and I_{max} represents the minimum respectively the maximum intensity of the image $Imag_{orig}$.

$$Imag_{New} = \frac{Imag_{orig} - I_{min}}{I_{max} - I_{min}} \cdot 255 \qquad (1)$$

b) *Histogram Equalization*. This method improves contrast by redistributing pixel intensities so that the histogram becomes more uniformly distributed [15]. However, this technique may lead to amplification of the noise present in the image.

c) *Contrast Limited Adaptive Histogram Equalization (CLAHE)* [16]. This method divides the image into blocks (in this case of dimension 8x8 px) and applies local histogram equalization to each of them. In this way the contrast oversaturation in homogeneous regions is prevented and the noise is reduced.

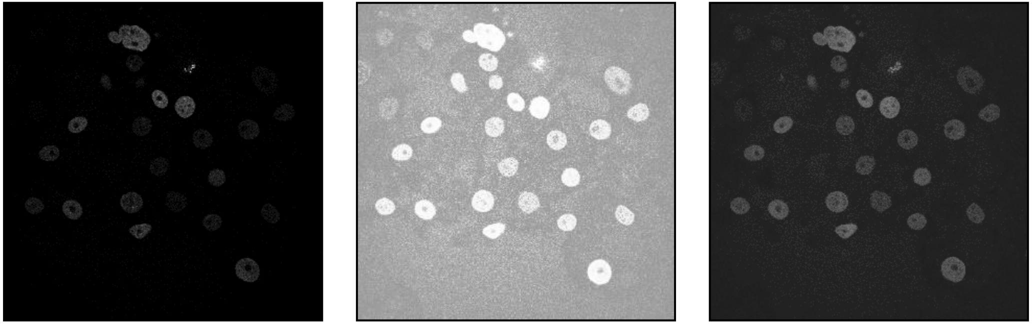

a) b) c)

Figure 3. a) Maximized contrast, b) Equalized Histogram, c) CLAHE

Each method produced a distinct dataset for comparative analysis of segmentation performance.

3.2 Training and Segmentation

The model consists of a DeepLabV3+ network with a ResNet-18 backbone [12], [13] pretrained on the ImageNet dataset [17]. The network, optimized for cross-entropy loss with a final classification layer, contains approximately 20.6 million trainable parameters and is configured for binary semantic segmentation (*cell* vs. *background*).

The binary masks were mapped from the 255 and 0 intensities to categorical labels for "*cell*" and "*background*" respectively.

To enhance generalization, the training set underwent extensive augmentation. Geometric transformations included random rotations (0°–360°), horizontal/vertical flips, scaling (0.7–1.4), shear (±10°), and translations (−50 to 50 pixels). A second variant added photometric changes like brightness/contrast adjustments (±5%) and Gaussian noise (σ=0.05) to simulate diverse acquisition conditions.

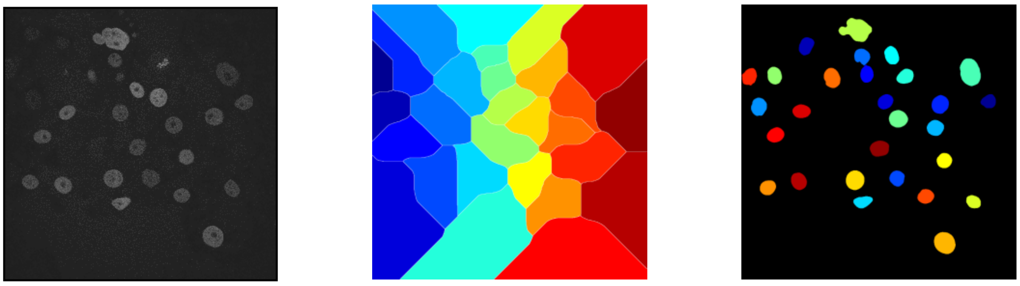
The network was trained with a batch size of 8, using the Adam optimizer, learning rate of 0.001, and 10 epochs. These parameters balanced GPU memory constraints with convergence speed. Data were shuffled each epoch, and validation metrics were monitored every five iterations. Checkpointing allowed recovery in case of interruptions. After training, the network produced binary segmentation images with white (255) representing estimated cells and black (0) representing the background.

3.3 Postprocessing

Postprocessing was applied to all segmentation results to correct local artefacts. Small regions were removed to eliminate possible noise or artefacts. Then, the binary image obtained is dilated, expanding the edges of the objects to close any discontinuities or gaps that may have appeared during segmentation. Subsequently, the internal holes within the binary objects are filled, ensuring a coherent representation of the cells as compact and continuous regions. Finally, a morphological erosion using the same structuring element as the morphological dilation is used, in order to restore object sizes while maintaining boundary improvements.

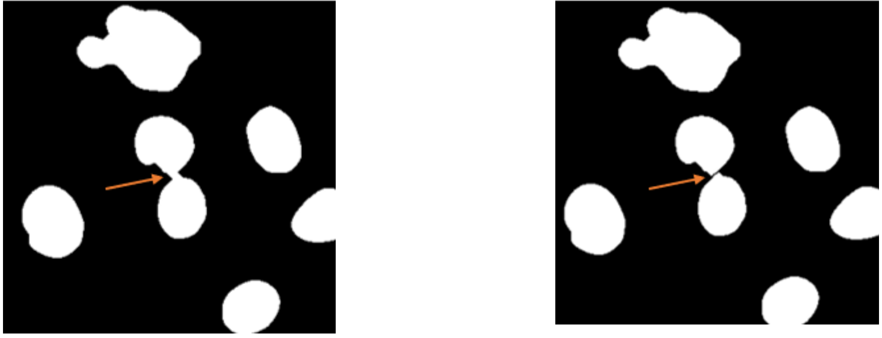
In some cases, the segmentation may produce merged regions corresponding to multiple adjacent cells (as in *Figure 5*). To separate merged neighboring cells, the *watershed transform* [11] was applied. The watershed algorithm starts from the *distance transform* [15] of the complement of the segmented image after morphological operators. The *distance transform* calculates, for each pixel, the minimum distance to the edge of the object, resulting in a three-dimensional map where the central areas of the objects appear as valleys and the edges as peaks. To avoid over-segmentation caused by minor intensity variations, less significant local minima are suppressed. Watershed lines then separate individual cells, effectively resolving overlaps.

This postprocessing pipeline produces cleaner, more consistent segmented images (as in *Figure 4*), providing a reliable foundation for the quantitative evaluation.



a) b) c)

Figure 4. a) preprocessed image b) watershed space c) segmented image (each cell with a color)



a) b)

Figure 5. a) Merged cells b) Result after applying watershed transform

4 Results and Conclusions

The segmentation performance of the DeepLabV3+ network with a ResNet-18 backbone (including pre-processing and post-processing) was evaluated by comparing the binary images obtained after segmentation with ground-truth annotations. Two aspects were analyzed:

a) **Semantic segmentation:** how well the pixels were correctly identified
b) Instance segmentation: how well the cells were identified

4.1 Semantic segmentation

In order to determine how well the pixels were correctly classified as *background* or *cell* pixels, the confusion matrix was computed from:

- True Positives (TP) = *cell* pixels classified correctly as *cell*
- True Negatives (TN) = *background* pixels classified correctly as *background*
- False Positives (FP) = *background* pixels classified as *cell*
- False Negatives (FN) = *cell* pixels classified as *background*

The F1-score (eq. 2) and Intersection of Union (IoU - eq. 3) were also derived.

$$F1\ Score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{2}$$

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

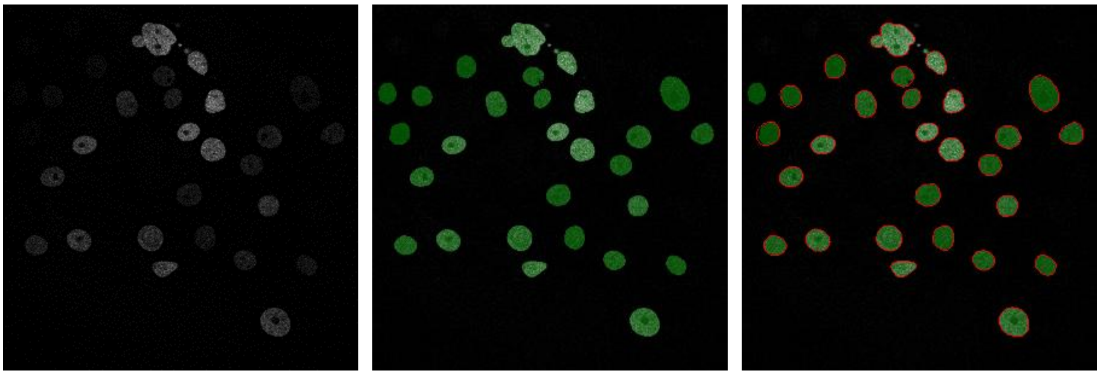
Six preprocessing strategies were tested: contrast maximization, histogram equalization, and adaptive histogram equalization (CLAHE), each of them with and without photometric changes. Among them, CLAHE without photometric changes produced the best results, as shown in *Figure 6*.

For postprocessing, several structuring element shapes and sizes were evaluated, and the diamond-shaped element of size 3 provided the most accurate results, effectively removing small artifacts while preserving cell contours.

With this configuration, the model consistently achieved an F1-score of 95.40%, indicating high segmentation accuracy and confirming the effectiveness of the DeepLabV3+ architecture for fluorescence microscopy cell images. For a complete overview of the evaluation metrics, the full results are provided in *Table 1*.

Table 1. Segmentation evaluation metrics for the validation dataset

	Structural element and size	**Diamond**					
		3 pixels		5 pixels		7 pixels	
	Metric	F1-Score	IoU	F1-Score	IoU	F1-Score	IoU
No photometric changes	Contrast maximization	94.66%	89.86%	94.66%	89.86%	94.65%	89.86%
	Histogram Equalization	92.72%	86.48%	92.75%	86.53%	92.77%	86.56%
	CLAHE	**95.40%**	**91.23%**	95.40%	91.22%	95.38%	91.20%
Photometric changes	Contrast maximization	83.31%	71.57%	83.34%	71.62%	83.38%	71.67%
	Histogram Equalization	94.38%	89.37%	94.38%	89.36%	94.37%	89.36%
	CLAHE	95.09%	90.66%	95.11%	90.69%	95.14%	90.75%
	Structural element and size	**Disk**					
		3 pixels		5 pixels		7 pixels	
	Metric	F1-Score	IoU	F1-Score	IoU	F1-Score	IoU
No photometric changes	Contrast maximization	94.65%	89.84%	94.66%	89.86%	94.66%	89.86%
	Histogram Equalization	92.72%	86.47%	92.74%	86.51%	92.77%	86.57%
	CLAHE	**95.40%**	**91.22%**	95.40%	91.22%	95.39%	91.20%
Photometric changes	Contrast maximization	83.29%	71.54%	83.34%	71.62%	83.39%	71.69%
	Histogram Equalization	94.38%	89.36%	94.38%	89.37%	94.37%	89.36%
	CLAHE	95.09%	90.65%	95.11%	90.69%	95.15%	90.77%



a) b) c)

Figure 6. Segmentation results for the best parameters
a) Preprocessing using CLAHE, b) green overlay of the detection, c) the mask contour (red) overlaid on image 6.b.

4.2 Instance segmentation

To determine how well the cells have been identified, for each frame, each cell from the segmented image must be matched to the corresponding cell from the reference image. For the reference cells, the size of each cell is additionally extracted by estimating the effective radius as half of the diagonal of the bounding box rectangle. This approximation is justified by considering the cell as an elliptical shape, and the bounding box diagonal provides a reasonable estimate of the maximum extent.

In order to establish the association between predicted and actual cells, a matrix of Euclidean distances is constructed between all the centroids of the cells in the segmented image and those in the reference image (*Figure 7*). The matrix is then filtered to retain only the values below the effective radius associated with each reference object; larger values are marked as ∞ (infinity), indicating that the prediction is considered too far to be a valid match. This restriction serves to prevent false matching between spatially distant objects.

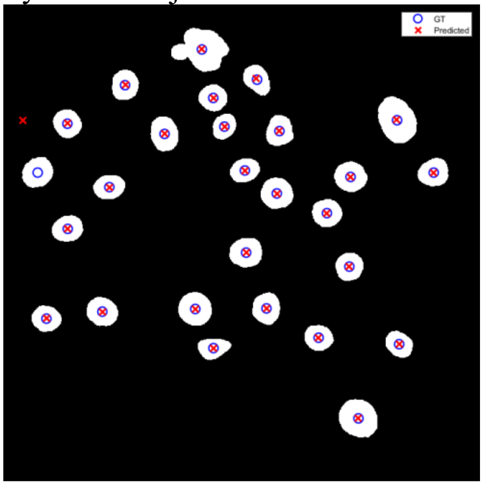


Figure 7. Image with the centroids of the segmented cells
and the centroids of the cells from the reference images

For the best configuration for semantic segmentation (diamond 3px, CLAHE, no photometric changes), the segmentation performance was further evaluated by counting the detected and missed cells, resulting in the values from *Table 2*.

Table 2. Results for instance segmentation

Metric	Subset 1 (validation)	Subset 2 (training)
TP (*cells* as *cells*)	No. of detected cells: 1982	No. of detected cells: 2282
FP (non-*cells* as *cells*)	No. of detected cells: 138	No. of detected cells: 167
FN (*cells* as non-*cells*)	No. of detected cells: 26	No. of detected cells: 5
$Precision: \frac{TP}{TP+FP} \cdot 100$	93.49%	93.18%
$Recall = \frac{TP}{TP+FN} \cdot 100$	98.70%	99.78%
The average distance between the centers of the reference cells and the centers of the associated predicted cells	1.03px	0.54px

5 Conclusions

The proposed DeepLabV3+ model with a ResNet-18 backbone achieved reliable segmentation of GOWT1 stem cells, reaching on the validation dataset a semantic segmentation F1-score of 95.40% under optimal preprocessing and postprocessing conditions and an instance segmentation precision score of 93.49% and a recall of 98.70%. These results confirm the effectiveness of context-aware CNNs for fluorescence microscopy, offering a strong balance between precision and computational efficiency. The approach proved robust to variations in illumination and noise, providing consistent, biologically meaningful segmentations suitable for downstream quantitative analyses and future extensions toward tracking or division detection.

References

[1] E. Meijering, *Cell Segmentation: 50 Years Down the Road [Life Sciences]*, IEEE Signal Process. Mag., vol. 29, no. 5, pp. 140–145, doi: 10.1109/MSP.2012.2204190, Sept. 2012.

[2] C. F. Koyuncu, R. Cetin-Atalay, and C. Gunduz-Demir, *Object-Oriented Segmentation of Cell Nuclei in Fluorescence Microscopy Images,* Cytometry Part A, vol. 93, no. 10, pp. 1019–1028, doi: 10.1002/cyto.a.23594, 2018.

[3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, *arXiv*: arXiv:1606.00915. doi: 10.48550/arXiv.1606.00915, May 12, 2017.

[4] S. Zhang, Z. Ma, G. Zhang, T. Lei, R. Zhang, and Y. Cui, *Semantic Image Segmentation with Deep Convolutional Neural Networks and Quick Shift*, Symmetry, vol. 12, no. 3, p. 427, doi: 10.3390/sym12030427, Mar. 2020.

[5] M. Maška *et al.*, *The Cell Tracking Challenge: 10 years of objective benchmarking*, Nat Methods, vol. 20, no. 7, pp. 1010–1020, doi: 10.1038/s41592-023-01879-y, 2023.

[6] D. K. Gupta, N. de Bruijn, A. Panteli, and E. Gavves, *Tracking-Assisted Segmentation of Biological Cells*, *arXiv*: arXiv:1910.08735. doi: 10.48550/arXiv.1910.08735, Oct. 19, 2019.

[7] A. R. Menezes and B. Ramsundar, *Open Source Infrastructure for Automatic Cell Segmentation*, *arXiv:* arXiv:2409.08163. doi: 10.48550/arXiv.2409.08163, Sept. 12, 2024.

[8] G. Rahmon, I. E. Toubal, D. Cornelison, and K. Palaniappan, *Marker and Motion Guided Deep Networks for Cell Segmentation and Detection Using Weakly Supervised Microscopy Data*, in 2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–12. doi: 10.1109/AIPR60534.2023.10440692, Sept. 2023.

[9] A. Arbelle and T. R. Raviv, *Microscopy Cell Segmentation Via Convolutional LSTM Networks*, in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy: IEEE, pp. 1008–1012. doi: 10.1109/ISBI.2019.8759447, Apr. 2019.

[10] M. Arvidsson, S. K. Rashed, and S. Aits, *An annotated high-content fluorescence microscopy dataset with Hoechst 33342-stained nuclei and manually labelled outlines*, Data Brief, vol. 46, p. 108769, doi: 10.1016/j.dib.2022.108769, Nov. 2022.

[11] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer Nature, 2022.

[12] Y. Wang, L. Yang, X. Liu, and P. Yan, *An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+*, Sci Rep, vol. 14, no. 1, p. 9716, doi: 10.1038/s41598-024-60375-1, Apr. 2024.

[13] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, *arXiv*: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385, Dec. 10, 2015.

[14] *Cell Tracking Challenge* – "Where your software moves cells…."
[Online]. Available: https://celltrackingchallenge.net/ Accessed: Nov. 24, 2025.

[15] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Fourth edition, Global edition. New York: Pearson, 2017.

[16] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*, Elsevier, pp. 474–485. doi: 10.1016/B978-0-12-336156-1.50061-6, 1994.

[17] *ImageNet*. Available [Online]: https://image-net.org/, Accessed: Nov. 24, 2025.