

Masking Approach for Homogeneous Image Plane: In case of Bright Field Microscopy

Surajit Das, *Info Chemistry, ITMO*, Pavel Zun, *Infochemistry, ITMO*

Abstract— Bright-field microscopy is a crucial technique in biomedical imaging. Segmenting live cell culture images in bright- field microscopy presents challenges due to low contrast, uneven lighting, and noise. The following study introduces a novel and innovative unsupervised masking technique that segments bright- field microscopy pictures without the necessity of pre-trained models or operator annotations. Our approach utilizes statistical background modeling, spatial statistics (Moran's I, adjusted variogram, Cumulative Squared Shift of Nodal Intensity), fuzzy inference, and shape analysis to differentiate foreground from background. In contrast to region-growing approaches, our methodology does not depend on distance metrics but rather establishes thresholds adaptively. Comparative assessments of our algorithm with Cellpose and StarDist on challenging unstained cell culture images demonstrate that our method yields visually coherent segmentation. The proposed algorithm has been presented through a graphical user interface (GUI) and we offer an expert-driven evaluation to ascertain segmentation quality and as well as the standard evaluation metrics including Accuracy, Recall, Precision, IoU and F1 score. This paper advances efficient, training-free segmentation methods, with prospective developments in instance segmentation. The code is available at <https://github.com/SurajitDaz/ITMO-info-chemistry/tree/main/LiveCell>

Index Terms— Computer vision, image analysis, live cell, quantitative microscopy, unsupervised segmentation.

I. INTRODUCTION

THE use of bright-field microscopy is widely accepted in the field of observing stained or naturally pigmented specimens, routine analysis in clinical, educational, and research settings, examining fixed cells, tissues and microorganisms, quality control in manufacturing processes, etc. [1]. Nowadays, the study of biological samples and their analysis is greatly accompanied by machine learning and data science [2]. On the other hand, analysing unstained images remains a challenge. However, solving this challenge will allow for much faster analysis of live cell images without killing, staining and fixing them, and hence also allows for continuous imaging of the cells. In this article, a novel methodology is introduced for

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment.

Surajit Das, ITMO University, Russia, 191002, St. Petersburg, Lomonosova Str., 9 (e-mail:author@itmo.ru). He is employed in infochemistry lab.

Dr. Pavel Zun, ITMO University, Russia, 191002, St. Petersburg, Lomonosova Str., 9 (e-mail:author@itmo.ru). He is employed in infochemistry lab

processing bright-field microscopy live unstained cell culture images. The proposed model is found to be robust under some presumptions listed below.

- The environment involving optical conditions and imaging system of Bright Field microscopy is set up to achieve nearly homogeneous image plane (defined in subsequent section).
- The specimen or samples are recommended to be placed sparsely within the Field of View (FOV) for a good result. However, this is not a bottleneck condition.
- One of two image sets (either background or specimens) must have optical heterogeneity caused by its inherent texture.

We recapitulate some crucial information pertaining to bright field microscopy as they are highly related to the context and principle followed in this research. A typical bright-field microscope consists of a light source, a condenser lens, a stage, objective lenses, an eyepiece (or ocular lens), coarse and fine focus knobs and a diaphragm/iris. Light passes through the condenser, illuminating the specimen from below. The transmitted light, refracted, absorbed, and/or scattered by the sample, enters the objective lens. The image is magnified and projected to the eyepiece or a digital camera (an imaging sensor, e.g. CMOS) for observation or documentation. Contrast between the specimen and the background is due to the differences in light absorption or scattering [3]. The sample/ specimen types used in bright-field microscopy are often thin sections of tissues, cell cultures, bacteria, algae, protozoa, and small invertebrates.

The experiences gained during regular lab work show that analysing bright-field microscopy images using machine learning (ML) or deep learning (DL) can be challenging due to the following issues: I) Low contrast and noise, which grow for unstained or thin specimens that transmit most of the light, cause them to blend into the background. Low contrast makes it difficult for models to differentiate between sample features and background noise. II) Although theoretically the field of view (FOV) of the imaging system is illuminated with constant illumination, in practice, it is found that uneven illumination across the field of view can lead to brightness variations which affect the ML/DL model's interpretation and reduce the accuracy. III) Optical aberrations and artifacts caused by dust, air bubbles, glare, or imperfections in the glass slide create confusing and irrelevant

patterns. IV) Biological specimens often contain overlapping cells or structures that complicate object segmentation. This makes it hard for models to distinguish individual features. V) Labeling objects (or manual masking) in microscopy data demands accurate annotation, requires expert knowledge, and is both error-prone and time-consuming. As an outcome, limited availability of annotated datasets hampers model training. VI) Data imbalance is a common challenge which occurs in many cases while we are interested in sparse occurrences of certain biological events (e.g., abnormal cells). This can lead to biased models that perform poorly on rare but critical features. VII) Lack of standardization makes a ML/DL model, which is trained on one dataset, often fail to generalize to new datasets or imaging conditions. Different microscopes, lighting conditions, and sample preparations yield variable image characteristics even for the same staining [4]. VIII) Bright-field microscopy lacks 3D information due to its reliance on transmitted light (poor depth information), and hence the flattened images can obscure features such as depth and thickness variations.

Normally, traditional masking techniques are categorized as spatial masking [5], spectral filtering [6] and computational/algorithmic masking [7]. Spatial masking involves physically blocking or modifying portions of the image plane to control the light distribution and reduce unwanted artifacts. Spectral filtering uses filters to selectively pass or block specific wavelengths of light, enhancing contrast and reducing background noise. Computational or algorithmic masking involves using algorithms to process the image data and remove unwanted artifacts or enhance specific features. Masking techniques can enhance contrast, reduce artifacts, improve image quality, segment images, etc.

In this study, we introduce a novel algorithmic masking technique, which demands no training cost to the dataset in order to mask the background followed by the foreground. The objective of this masking, in ML/DL terms, lies with semantic segmentation. The technique involves analysis with classical & spatial statistics, fuzzy logic and geometrical shape identification methods; additionally, it is accomplished by traditional techniques used in computer vision.

It is important to mention that this approach is completely different from the approach called region growing [8] which is facilitated with seed point selection followed by comparison with neighbouring pixels to cluster based on the similarities. The similarities are normally determined by applying some distance metrics (e.g Euclidian distance). The selected seed is considered as the reference point. Unlike region growing, this approach uses no distance metric; rather, we sample some pixels from background for statistical analysis and to set a threshold for background masking. This process does not signify that the pixels having values above the defined threshold will belong to the complementary set. Whereas, in the region-growing method, if the distance of a neighbouring pixel is far away from the seed point and if it does not belong to the cluster of the seed point, it automatically belongs to

the complementary set. In this case, the threshold of the background ensures certain cases where pixels belong to the background set but does not conclude anything about the uncertainties existing above the threshold. The uncertainty is handled by the intelligent fuzzy system and subsequently by statistical methods.

II. BACKGROUND

This section describes some important criteria and technical terminologies used in the subsequent sections.

A. Introducing Image Plane

Image Plane: A theoretical and physical 2D space, which is a collection of pixels that are arranged in an array and have intensity values I , is regarded as an image plane in this study. The values of I are related to transmittance or reflectance function T , illumination factor governed by the physical plane of sensor L and optical distortions/aberrations or sensor non-uniformity Δ . Let the projection matrix P , which maps the 3D object space (X, Y, Z) to 2D image coordinates (x', y') in many-to-one mapping, be written as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

- P is the 3×4 projection matrix. The matrix describes the perspective or orthographic projection.
- $(X, Y, Z, 1)^T$ are the homogeneous coordinates of a point in 3D space.
- $(x', y', 1)^T$ are the 2D homogeneous coordinates of the corresponding point on the image plane.

The generalized expression of intensity is:

$$\begin{aligned} I(x', y') &= T(X, Y, Z) \cdot L(x', y') + \Delta(x', y') \\ \implies I(x', y') &= T(P^{-1}(x', y', 1)^T) \cdot L(x', y') + \Delta(x', y') \end{aligned} \quad (2)$$

B. Notion of Homogeneous Image Plane & Homogeneity

1) Homogeneous Image Plane: According to the hypothesis considered, in the context of microscopy, a homogeneous image plane is defined as an image plane where the intensity and quality of the image remain uniform for the entire surface. This uniformity implies that any observed variation in the image arises solely due to the object or sample being imaged, not due to inconsistencies in the optical system, illumination, or imaging medium. For an image plane to be truly homogeneous:

- Illumination must be uniform (evenly distributed): Light intensity must be evenly distributed across the field of view. Any variations in intensity on the image plane arise solely from the object's properties, not from inconsistent lighting or transmission/reflection factors.

- Optical components (properties) must be ideal: The system (lenses and other optical elements) must not introduce aberrations, distortions or inconsistencies that could alter the object's true appearance.
- Imaging medium must be consistent: The data should be uni-modal and the sensor (or photographic film) must have a uniform response to light across its surface.

In this context, if an image plane is regarded as "homogeneous," it doesn't mean the image itself is uniform (e.g., a flat color), rather, it refers to the system properties being uniform across the plane and hence the intensity function remains invariant ($I(x', y') = I_0$) of space (not the co-ordinate of image plane) and time for that particular image plane. Then the equation (2) can be rewritten as:

$$\begin{aligned} I(x', y') &= T(X, Y, Z) \cdot L_0 \\ \implies I(x', y') &= T(P^{-1}(x', y', 1)^T) \cdot L_0 \end{aligned} \quad (3)$$

[where $L(x', y') = L_0$ and $\Delta(x', y') = 0 \quad \forall (x', y')$]

2) Homogeneity of Region of Interest (ROI) in Image: A homogeneous image plane does not ensure homogeneity of a region of interest (ROI) in an image. A homogeneous image plane only ensures a unimodal dataset with an even (constant) illumination factor, L as mentioned in the above equation (3). The intensity variations of a homogeneous image plane are influenced only by the object's inherent properties and uniform illumination. The homogeneity of ROI depends on the homogeneous image plane and the inherent properties of the objects which are imaged. Traditionally, the homogeneity of the ROI in an image is measured with the help of gray level co-occurrence matrix (GLCM) [9] and formulated as:

$$H_d^\theta = \sum_{i=1}^n \sum_{j=1}^n \frac{p_d^\theta(i, j)}{1 + |i - j|} \quad (4)$$

where, n is the total number of gray levels in the neighbourhood selected, $p_d^\theta(i, j)$ denotes the probability of a pixel pair having intensities i and j respectively at a certain d distance and angle θ . Therefore, homogeneity is computed for various senses of direction, and it is computationally expensive [9].

In order to avoid the computational complexity, we measure the spatial standard deviation with respect to localized mean (SSDLM) for a neighbourhood. Unlike the homogeneity measures in the GLCM, this technique has limitations in capturing the homogeneity along every sense of direction, but it is proved below that the SSDLM inversely varies with the metric homogeneity (if it exists) along any sense of direction.

Let us consider a 3×3 neighbourhood around any pixel intensity $\omega_{(i,j)}$. In this case, i and j represent the cartesian coordinate of that pixel and ω is a function that maps to the intensity of that pixel.

$$\begin{pmatrix} \omega_{(i-1,j-1)} & \omega_{(i-1,j)} & \omega_{(i-1,j+1)} \\ \omega_{(i,j-1)} & \omega_{(i,j)} & \omega_{(i,j+1)} \\ \omega_{(i+1,j-1)} & \omega_{(i+1,j)} & \omega_{(i+1,j+1)} \end{pmatrix}$$

Therefore, SSDLM for the neighbourhood of $\omega_{(i,j)}$ will be represented as $SSDLM_{(i,j)} = \sqrt{\sum_i \sum_j \frac{(\omega_{(i,j)} - \bar{\omega})^2}{N}}$ [$\bar{\omega}$ is the mean value of ω].

From the context of GLCM matrix, we know that the homogeneity directly varies with the contrast [9]. Hence,

$$\sum_{i=1}^n \sum_{j=1}^n \frac{p(i, j)}{1 + |i - j|} \propto \frac{1}{\sum_{i=1}^n \sum_{j=1}^n (i - j)^2 \cdot p(i, j)}$$

$$Again, \sum_{i=1}^n \sum_{j=1}^n (i - j)^2 \cdot p(i, j) \propto \sqrt{\sum_i \sum_j \frac{(\omega_{(i,j)} - \bar{\omega})^2}{N}}$$

[according to the rules of variation]

Therefore, $SSDLM_{(i,j)}$ is inversely proportional to homogeneity.

C. Other Spatial Statistical Metric Used:

1) Moran's I: Moran's I is a measure of spatial autocorrelation, which is defined as:

$$I = \frac{n \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j w_{ij} \cdot \sum_i (x_i - \bar{x})^2}$$

n : Number of spatial units,

x_i, x_j : Values at locations i and j. [where, i and j are the labels of two locations]

\bar{x} : Mean of all values.

w_{ij} : Spatial weight between locations i and j (e.g., inverse distance).

Interpretation:

$I > 0$: Positive spatial autocorrelation (similar values cluster together).

$I < 0$: Negative spatial autocorrelation (dissimilar values are neighbour).

$I = 0$: No spatial autocorrelation (random spatial pattern).

2) Cumulative Squared Shift of Nodal Intensity (CSSNI):

The **Cumulative Squared Shift of Nodal Intensity (CSSNI)** quantifies local intensity variations in greyscale images by computing the sum of squared intensity differences (SSID) between each pixel and its 8-connected neighbour. The mathematical formulation of this metric is given below:

Let $\omega_{(i,j)}$ denote the intensity value of a pixel at the location (i, j) in a patch size of $(M \times N)$ in a greyscale image. The CSSNI measure is defined as:

$$CSSNI = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N \sum_{(m,n) \in \mathcal{N}_{(i,j)}} (\omega_{(i,j)} - \omega_{(m,n)})^2 \quad (5)$$

where $\mathcal{N}_{(i,j)}$ represents the set of 8-connected neighbouring pixels. The division by 2 prevents double-counting since each pixel pair contributes to the sum twice.

3) Adjusted Variogram: Adjusted variogram is a variogram-like measure for spatial intensity variation. The spatial variation of the intensities of the pixels within an image patch is commonly analyzed using the semivariogram [10], which quantifies the relationship between the differences in the intensity of the pixels and their spatial separation. A conventional empirical semivariogram $\gamma(h)$ is defined as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (x_i - x_j)^2 \quad (6)$$

Where h is the spatial lag, x_i and x_j are pixel intensities at locations i and j (which are two different labels of the location), and $N(h)$ is the number of pixel pairs separated by h . Standard variograms rely on binning pixel pairs based on discrete lags to estimate spatial dependence.

In this study, we introduce an adjusted variogram that provides a global estimate of intensity variation without explicit lag binning. The function computes the average squared intensity differences between all pixel pairs, normalised by their Euclidean distance in a specified neighbourhood. It can be formulated as:

$$\gamma(h) = \frac{1}{2} \cdot \mathbb{E} \left[\frac{(z(x) - z(x+h))^2}{d(x, x+h)} \right]$$

Where:

- $z(x)$ is the pixel intensity at position x ,
- $z(x+h)$ is the pixel intensity at position $x+h$,
- $d(x, x+h)$ is the Euclidean distance between positions x and $x+h$,
- h is the spatial distance between pixels.

This method differs from conventional variograms in three key aspects:

- 1) **No explicit lag binning:** Instead of computing semivariance for specific lag distances, this approach aggregates all pairwise intensity differences.
- 2) **Global averaging:** The function provides a single scalar estimate of spatial intensity variation rather than a curve over multiple lags.
- 3) **1D transformation:** The patch is flattened into a sequence, which simplifies computation but does not fully preserve 2D spatial relationships.

The adjusted measure does not replace a traditional variogram; it serves as a computationally efficient alternative for capturing global intensity variation, here used with a goal of texture analysis and spatial feature extraction. Its overall advantages are i) fast and lightweight, ii) single-scalar output and iii) good for local analysis.

D. Computational Complexity Analysis

In this section, we analyze the computational complexity of the proposed methods to determine their efficiency in processing microscopic images.

1) Time Complexity of Functions: We analyze the time complexity of each function used in the algorithm. Here we denote the total number of pixels in the image as $N = n \times m$ and the patch size as N_p (typically 5x5, 7x7 or 9x9).

- **Calculating Moran's I for a neighbourhood:** The dominant terms arise from the distance matrix computation, weight matrix construction and dilation. Hence, the overall time complexity of the function is:

$$\mathcal{O}(N^2) = \mathcal{O}((n \times m)^2) \quad (7)$$

- **Calculating Adjusted Variogram-like Measure:** This function computes pairwise distances and squared differences in $\mathcal{O}(N_p^2)$, leading to an overall complexity of $\mathcal{O}(N^2)$, where N_p is the number of pixels in the patch.
- **Calculating Cumulative Squared Shift of Nodal Intensity (CSSNI):** This iterates over all pixels and their 8-connected neighbour, resulting in a time complexity of $\mathcal{O}(N)$.
- **Fuzzy Module:** This function performs basic arithmetic and comparisons, leading to a constant time complexity of $\mathcal{O}(1)$.

2) Overall Time Complexity: The most computationally expensive functions are Calculating variogram, Calculating Moran's I, which runs in $\mathcal{O}(N_p^2)$ per pixel ($p \leq 11$). Since these functions are not applied to every pixel in the image (used only where fuzzy cannot identify the pixels for the foreground or background set), the overall complexity of the algorithm will be much less than:

$$\mathcal{O}(N \cdot N_p^2) \quad (8)$$

Given that N_p is relatively small, the algorithm remains efficient for large images.

This analysis confirms that the proposed method is computationally feasible for high-resolution microscopic images while maintaining accuracy in segmentation and classification. However, Moran's I can be accelerated using quadtree-based spatial indexing, Variogram calculations can be optimized with multi-resolution approximations. Also, parallel processing (Graphical Processing Unit (GPU) acceleration) can significantly reduce runtime for large images.

III. RELATED WORK:

In this section, a comprehensive overview is captured based on the contemporary research that introduces unsupervised learning based on several algorithmic & computational themes. Some notable research is included to depict the general evolutions of unsupervised learning for image segmentation problems. Unlike the images derived from the macro imaging system (which indicates the standard imaging system) or the images derived from the remote sensing system, the microscopic images have very high spatial resolution. In some microscopy modalities, specific contrast and brightness (phase contrast, fluorescence) are also important. The imagery often adopts special types of noises and other challenges which are already discussed in the introduction, and these noises or challenges are not present in the images obtained from

standard or remote sensing systems. However, this literature review is not only restricted within the scope of unsupervised learning in the field of microscopy images. It examines the overall recent advancements and accompanying gaps in unsupervised image segmentation techniques, as the dedicated important methodologies developed for unsupervised learning in segmentation problems are limited.

A. Unsupervised Learning for Image Segmentation Based on CNN:

Among the several studies on unsupervised learning for image segmentation, one notable method is the convolutional neural network (CNN)-based algorithm for unsupervised image segmentation [11]. This approach optimizes feature extraction and clustering functions jointly, predicting cluster labels through differentiable functions. A spatial continuity loss enhances segmentation quality, while batch normalization normalizes response maps. PASCAL VOC 2012 and BSD500 are the used datasets for the experiment. However, the paper exhibits some research gaps as it does not address real-time segmentation performance, especially in the case of low-contrast noisy dataset which can be comparable to the scenario of microscopy images. Additionally, the paper has not argued scientifically about the diversity of input data and about the scalability of the proposed method.

B. Bayesian Statistical Model

An old study proposes the Bayesian unsupervised satellite image segmentation method based on stochastic estimation maximization (SEM) algorithm over global methods like MAP or MPM [12]. This study assesses spectral and spatial context contributions to image parameters. However, a serious research gap is discovered as the dependence on initialization affects solution reliability. Also, in certain cases, the limited exploration of spatial context contributions is observed. In case of low signal-to-noise ratio scenarios, the model needs to be developed for a good result and it should be validated for various image parameters. Another study, Image Segmentation with Adaptive Spatial Regularisation (ASR), introduced a Bayesian computation methodology accompanied by Potts-Markov random fields (MRFs) [13]. The method marginalized regularization parameters and considered small-variance asymptotic analysis. Despite achieving comparable results to supervised approaches, it lacked exploration of alternative regularization techniques and scalability assessments. Also, Independent pixel consideration neglects neighbouring pixel influence.

C. Soft Computing Based Approaches:

A research involving Evolutionary Algorithm-Based Fuzzy Clustering (EABFC) introduces an unsupervised fuzzy clustering approach for image segmentation, combining an evolutionary algorithm (EA) with fuzzy clustering to leverage both local and non-local spatial information [14]. The method employs a multi-objective evolutionary sampling strategy to optimize pixel selection while preserving image details, followed by label correction using entropy and spatial constraints. However,

the study has two key limitations: (1) it lacks experimental validation on diverse datasets, raising concerns about generalizability, and (2) it does not support user-defined parameters, limiting customization for different segmentation tasks. A novel approach combining fuzzy logic with Markov random field (MRF) has been proposed for image segmentation [15]. This method develops an adaptive fuzzy inference system and utilises spatial constraints effectively. The approach is notable for not requiring training data and implementing a new clique potential MRF function. Fuzzy logic has been widely used for unsupervised segmentation. Fuzzy Random Fields and Unsupervised Image Segmentation proposed a fuzzy statistical model incorporating Gibbs sampling and stochastic estimation maximization (SEM) methods [16]. The approach demonstrated robust segmentation by integrating fuzzy components into traditional statistical models. However, the study did not explore real-world applications extensively or compare with advanced segmentation techniques. Another notable study, Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation, applied adaptive iterative conditional estimation (ICE) to improve segmentation efficiency [17]. The model generalized statistical fuzzy segmentation and adapted it to contextual settings using SEM, ICE, and Expectation-Maximization (EM) algorithms. Future work aims to integrate the approach with existing segmentation techniques.

D. Unsupervised Domain Adaptation for Microscopy Images:

Panoptic Domain Adaptive Mask (PDAM), based on Domain Adaptive Mask R-CNN (DAM), offers a novel segmentation strategy but struggles with domain shift due to contextual discrepancies [18]. This approach works with R-CNN and uses cycleGAN with an auxiliary objects inpainting mechanism. The former is responsible for synthesising images alike to the target, while the later one is responsible for reinforcing the image construction. The method encounters a problem regarding domain shift due to contextual information discrepancies and exhibits inadequate adaptation in the feature level for large domain gaps. Another approach puts forward an encoder-decoder-based multi-task learning model to cluster pixels according to foreground, background and cell boundaries as unsupervised domain adaptation. This method requires further improvements in domain-regularising cost functions and performance metrics, as the Dice metric is insensitive to clustered cell isolation. Also, the article does not discuss fixing the issue of ad-hoc parameter estimation [19].

E. Hyperspectral Image Segmentation:

In other approach, Hyperspectral microscopy image segmentation combines both unsupervised deep learning (UHRED) for denoising and supervised deep learning (SHRED) for enhancement followed by K-means clustering and mean squared error for loss calculation [20]. The method uses the Adam optimizer for determining the parameters of the model. However, challenges pertaining to overlapping species classification and automation of saturated pixel identification remain unanswered.

F. Classical and Hybrid Approaches:

The traditional technique along with machine learning has also come into the picture to accomplish the unsupervised learning, competitive learning, fuzzy c-means clustering, and Gibbs random fields that consider spatial connectivity constraints that are used in order to improve tissue component segmentation through an iterative conditional modes (ICM) algorithm adaptation. The prominent lacks of this approach lie in limited accuracy produced, subjectivity in setting thresholds and the need for ICM algorithm adaptation [21]. Another research proposed a classical approach involving edge detection and morphological processes, called MPS-Based Image Segmentation for Bright-field Microscopy [22]. The method has been implemented in open-source software Fiji. Histogram equalisation, edge detection by the Canny edge detector and filling holes by using a maximum filter are the key steps to segment the total cell area by creating a binary image. Though it works with various cell types, it has a serious lack of handling the scenarios where the cell border is obscured. Also, the report about huge testing and comparison outcomes across diverse scenarios is absent. It requires better parameter standardisation across microscopy setups.

IV. METHODOLOGY

The schematic diagram for end-to-end workflow is attached below (Figure 2). All the steps of the workflow are described in subsequent sections.

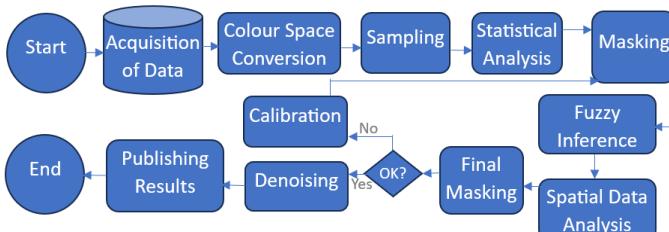


Fig. 1. End-to-end workflow

A. Acquisition of Data:

The primary experiments are carried out on a dataset consisting of 10 images of live cells with the resolution of 1920×1440 each. The dataset includes low-contrast cell culture photographs of myoblast cells. The cells are irregular in shape and without any specific colour, as they are unstained. Additionally, the cell boundaries are obscured (not prominently visible) in many instances due to the movements of the cell at the time of photography. In order to make the dataset statistically significant, the images have undergone several transformations involving modified resolutions, rotational transformations and contrast variations for the purpose of demonstrating the robustness of the model. For each image, the augmentation specification includes four different rotational transformations, which include vertical flip, horizontal flip, and 45-degree clockwise and anticlockwise rotation, which yield 50 instances (including the originals). These 50 instances are further transformed (randomly) with different contrast

and/or different resolution (1200 x 1660). So, the dataset has altogether 100 instances.

B. Colour Space Conversion:

The default colour space conversion algorithm for RGB to gray implemented in many libraries (e.g., OpenCV) employs a perceptual luminance model based on human vision sensitivity. This approach utilises a weighted method, where the grayscale image is computed as:

$$\text{Gray} = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (9)$$

In contrast, this study adopts an **average-based grayscale conversion**, formulated as:

$$\text{Gray} = \frac{R + G + B}{3} \quad (10)$$

This method assigns **equal weights** to all channel intensities, ensuring unbiased treatment across RGB components. Consequently, it provides a **more sensitive response** to the mathematical scenarios dealing with second-order exponents of intensity variations, as any channel is neither penalised nor rewarded by assigning any uneven weight. Hence, the uniform weighting scheme ensures impartial consideration of intensity values from all spectral bands, enhancing consistency in subsequent analyses.

C. Sampling in $I_{\text{background}}$:

The model includes manual random sampling. The user performs manual random sampling of the background pixels. In this experiment, 32 observations from the set $I_{background}$ (background) are picked up randomly by clicking on the pixels. For each click, the system picks a neighbourhood $n \times n$ (which is an adjustable hyperparameter, 5×5 in this case).



Fig. 2. Graphical User Interface (GUI) for Sampling Background Pixels with neighbourhood 5×5

D. Statistical Analysis of Data

For each neighbourhood, its $SSDLM_\omega$ is calculated. Finally, the mean and the standard deviation of the samples are calculated. The calculation $SSDLM_\omega$ used in this experiment is mentioned below.

- **Sample Observations:** [1.05, 0.98, 0.64, 2.71, 0.48, 0.44, 0.79, 0.55, 0.70, 5.76, 0.42, 0.52, 1.10, 0.44, 1.80, 1.18, 1.09, 1.37, 2.04, 0.42, 0.58, 0.49, 0.87, 0.46, 0.75, 0.70, 0.39, 0.27, 0.81, 0.63, 0.67, 0.81]
- **Range:** Range=5.76–0.27=5.49
- **Mean of $SSDLM_\omega$:** Mean= 0.9320
- **SD of $SSDLM_\omega$:** SD=1.066

Therefore, by default, the algorithm calculates $L_B = (\text{mean} + 3 \times \text{SD}) \implies 0.93 + 3 \times 1.06 \approx 4$ to define the lower bound of masking; however, this threshold can be further calibrated using the slider, named "Masking Threshold" in the GUI shown in Fig. 4.

E. Masking

In this step, the $SSDLM_\omega$ of the patches of all the pixels are checked; for a pixel and its patch, if its $SSDLM_\omega$ is found to be lower than the threshold value, the pixel is turned into black by replacing the pixel value with zero.

F. Fuzzy Inference System

The sub-steps of intensity transformation using fuzzy logic have been described in the following section.

1) **Fuzzy Membership Function Definitions:** Three fuzzy membership functions are defined to categorize pixel intensity. The parameters used here are α, β, a, b and c are related as following: $\alpha = \beta = (110), b = (110), c = 140$ and $a = 80$ (Fig. 3).

a) *Half-Trapezoidal Decreasing Membership Function:* $u_d(x)$

$$u_d(I_{\text{eff}}) = \begin{cases} 1, & I_{\text{eff}} \leq a \\ \max\left(\frac{\alpha - I_{\text{eff}}}{\alpha - a}, 0\right), & \text{otherwise} \end{cases} \quad (11)$$

b) *Triangular Membership Function:* $u_g(x)$

$$u_g(I_{\text{eff}}) = \max\left(\min\left(\frac{I_{\text{eff}} - a}{b - a}, \frac{c - I_{\text{eff}}}{c - b}\right), 0\right) \quad (12)$$

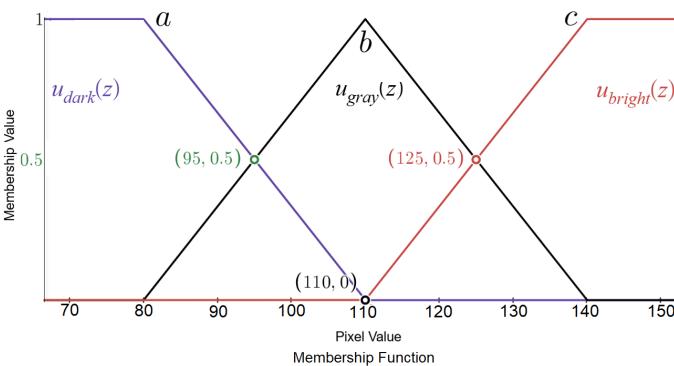


Fig. 3. Membership Function, μ_{dark} , μ_{gray} and μ_{bright} denote black, gray and white regions respectively.

c) *Half-Trapezoidal Increasing Membership Function:* $u_b(x)$

$$u_b(I_{\text{eff}}) = \begin{cases} 1, & I_{\text{eff}} \geq c \\ \max\left(\frac{I_{\text{eff}} - \beta}{c - \beta}, 0\right), & \text{otherwise} \end{cases} \quad (13)$$

2) Calibration of the Parameters of Membership Function:

The parameters a, α, b, β, c which are set to some default values ($\alpha = \beta = (110), b = (110), c = (140)$ and $a = (80)$), do not allow direct calibration through the external calibration window demonstrated in Fig.4. However, "Shift Gray" slider allows shifting the gray value which is here b and is set to 110 by default. The "Span Gray" slider allows changing the slopes of the slanting lines which are intersected at a point, b , leaving the coordinates of the point b unchanged. Once the slopes of two slanting lines are changed, the x -coordinates of the points (which have same x -coordinates as a c points respectively) of the intersections of the two slanting lines and X -axis are also changed. Hence, a , and c are also adjusted automatically. The operation is applied to the values of x -coordinates as the intensity value is an independent variable.

3) **Fuzzy System:** For each pixel , the fuzzy system considers the intensity value (grayscale) and using membership functions, it calculates the membership values of the pixel for dark, gray and white sets (μ_{dark} , μ_{gray} and μ_{bright}). Next the aggregations of three outputs of three membership functions are composed in the following formula:

$$I = \frac{v_d \cdot u_d(\text{input_px}) + v_g \cdot u_g(\text{input_px}) + v_b \cdot u_b(\text{input_px})}{u_d(\text{input_px}) + u_g(\text{input_px}) + u_b(\text{input_px})} \quad (14)$$

where, $v_d = 0, v_g = 127, v_b = 255, u_d = \mu_{\text{dark}}, u_g = \mu_{\text{gray}}, u_b = \mu_{\text{bright}}$. Finally, based on the rule defined, the pixel with an intensity value less than 80 is classified as black, the pixel with an intensity value greater than 140 is classified as white, and the pixels with any other values are called ambiguous entities and are sent to the module named 'Spatial Data Analysis'.

G. Spatial Data Analysis

The 'Adjusted Variogram' and the 'Cumulative Squared Shift of Nodal Intensity' (CSSNI) are computed and normalised by $SSDLM_\omega$ when a pixel is taken into consideration for spatial data analysis based on its neighbourhood (5×5 in this case). Here, the fuzzy system classifies the pixel based on some rules (intensity < 80 is black and intensity > 140 is white), which are described above. The pixels categorised as ambiguous entities undergo spatial data analysis. Computation of Moran's I ensures that the pixels belonging to the neighbourhood with randomness or dissimilarity are not classified as white; however, this rule can be customised by the user through the calibration window. Adjusted variogram and CSSNI are independent features; the evidence is attached in the supplementary pages.

H. Final Masking:

The final masking module checks the intensity value of the pixel in the green channel and the order of normalised adjusted variograms in the neighbourhood (5×5 in this case) of that pixel in the R, G and B channels. As the green colour contributes more to brightness for a pixel in the visual system, it should be less than both the red and blue channels if the

pixel is originally dark in nature. Also, as the images are pinkish, the R channel is dominating, and hence the normalised variogram of the neighbourhood in the R channel is higher than the two other channels. In a more sensitive case (where the adjusted variograms of the comparable neighbourhood are close to each other or to the thresholds set by the calibration window), it repeats the same process (shaking the order of R, G, B as above), considering the normalised CSSNI of three neighbourhoods from three channels.

I. Calibration & Repeat if Required:

Once the output is generated in the step called "Final Masking", calibrations may be done with several iterations of the above workflow for the desired result. Figure 5 demonstrates two different example outputs generated by two calibrations for visual impact.

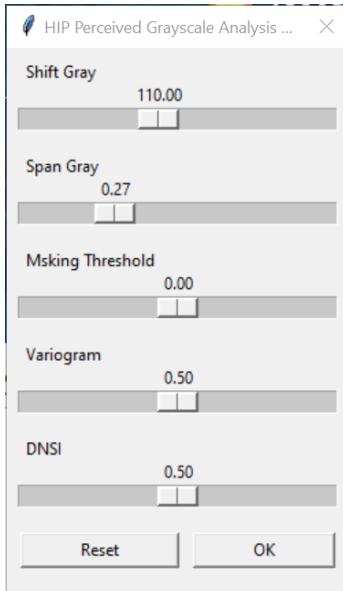


Fig. 4. Calibration Window

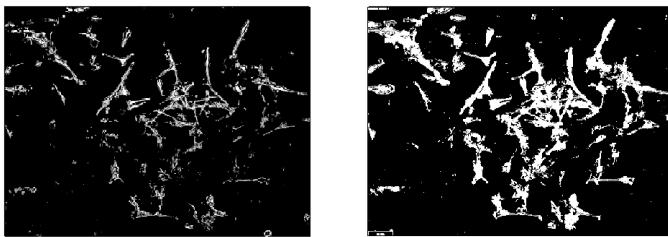


Fig. 5. Two different outputs generated by two calibrations

J. Removing Noise :

The protein blobs in an image are removed by setting up the value of the isoperimetric quotient (named as circularity in the interface) 1 as an upper and 0.31 as a lower bound, accompanied by a range of area with a lower bound 121 and upper bound 1800. Hence, the blobs which do not belong to the circularity range defined and do not lie between the range of

area are filtered out. Next, only area elimination is performed to remove the objects lying under the specific area threshold (121). Finally, the median filter is used to remove the salt-pepper noise.

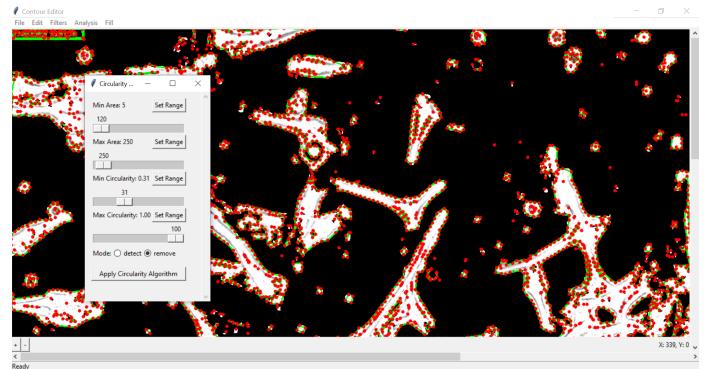


Fig. 6. Graphical User Interface for Denoising Image

K. Model Evaluation:

1) Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

2) Precision, Recall & F1 Score: Precision indicates the number of predicted positives which are actually correct, and Recall projects the number of positives which are detected. F1 computes the harmonic mean of Precision and Recall, balancing false positives and false negatives.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

3) Jaccard Index (Intersection over Union, IoU): Given a predicted binary mask $\hat{Y} \subseteq \Omega$ and the corresponding ground truth mask $Y \subseteq \Omega$, where Ω denotes the image domain, the IoU is defined as:

$$\text{IoU}(Y, \hat{Y}) = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} \quad (19)$$

where $|\cdot|$ denotes the cardinality (i.e., number of pixels) of the set. The numerator represents the number of true positive pixels, and the denominator represents the union of true positives, false positives, and false negatives. The IoU ranges from 0 to 1, with higher values indicating greater overlap between prediction and ground truth.

L. Algorithm

This section includes two algorithms based on which the unsupervised masking works:

Algorithm 1: Unsupervised Masking

Input: RGB image I , patch size $n \times n$

Output: Segmented mask S

- 01: Convert I to grayscale: $I_{\text{gray}} = (R + G + B)/3$
- 02: Sample background pixels \rightarrow compute SSDLM (Eq. 4)
- 03: Calculate threshold $L_B = \mu_{\text{SSDLM}} + 3\sigma_{\text{SSDLM}}$
- 04: for each pixel $p \in I_{\text{gray}}$ do
- 05: Classify p via fuzzy rules (Eqs. 11–13)
- 06: if p is ambiguous then
- 07: Compute CSSNI, Moran's I , adjusted variogram
- 08: Reclassify p using spatial statistics (Eqs. 5–6)
- 09: Normalize adjusted variogram and CSSNI.
- 10: If Moran's $I << 0$, no white classification.
- 11: Additionally, check intensity order in RGB:
- 12: If green $<$ red and blue, and red variogram $>$ green and blue variograms, retain pixel as foreground.
- 13: If variograms are close, resolve with CSSNI order across channels.
- 14: end if
- 10: end for
- 15: if reclassification using spatial statistics is correct, then
- 16: Apply postprocessing (Denoising Algorithm) $\rightarrow S$
- 17: else: Calibration followed by the steps 03 - steps 15

Algorithm 2: Postprocessing and Denoising

Input: Segmented mask M

Output: Cleaned mask M_{clean}

- 01: For each object $C_i \in M$, compute Circularity:
- 02: $\text{Circularity}(C_i) = \frac{4\pi \cdot \text{Area}(C_i)}{(\text{Perimeter}(C_i))^2}$
- 03: If $\text{Circularity}(C_i) \notin [0.31, 1.0]$, remove C_i from M
- 04: For each remaining $C_i \in M$, check area:
- 05: If $\text{Area}(C_i) \notin [A_{\min}, A_{\max}]$ or $\text{Area}(C_i) < A_{\text{thresh}}$, remove C_i from M
- 06: Apply median filter with kernel size $k \times k$:
- 07: $M_{\text{clean}} = \text{median_filter}(M, k)$
- 08: Return M_{clean}

V. RESULT & DISCUSSION

The model runs on 100 images out of which 10 are lab images and the rest are augmented images for evaluating the performance of the model for the job of semantic segmentation.

The model yields the following results, which are represented visually in Figure 7 in this article along with the average scores obtained through the several evaluation metrics. The corresponding results against 10 lab images are attached in the supplementary pages, both in tabular format and visual format.

The average precision achieved was 0.52 %. This indicates that false positives can still be reduced. It is caused by overestimating the borders. Recall (0.66%) indicates that the model is capturing a good amount of the actual foreground. The average

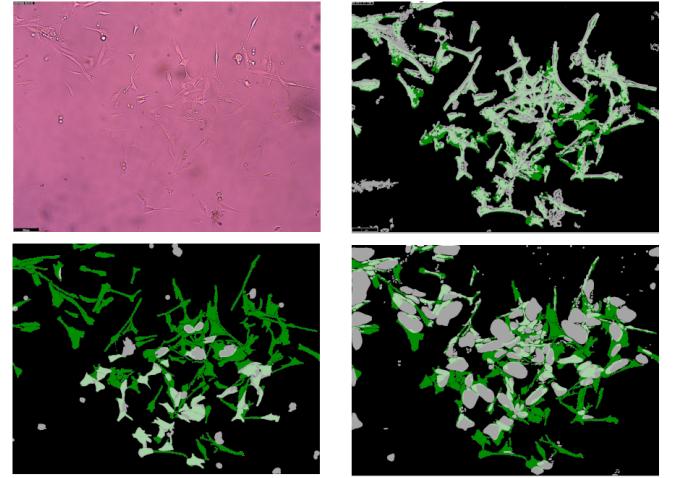


Fig. 7. The top-left image is the original image. The images placed at top-right, bottom-left, and bottom-right demonstrate the segmentation outputs of our model, cellpose, stardist respectively for the input image placed at top-left. This set of images is considered a representative set. Additional results for the full dataset, including comparisons with Cellpose and StarDist, are provided in Supplementary Figures S1–S10 (Supplementary Material). The green pixels are the true positives which are not detected by a particular model. The bright white pixels are true positives which are detected by a model too. The off-white pixels are false positives detected by a model

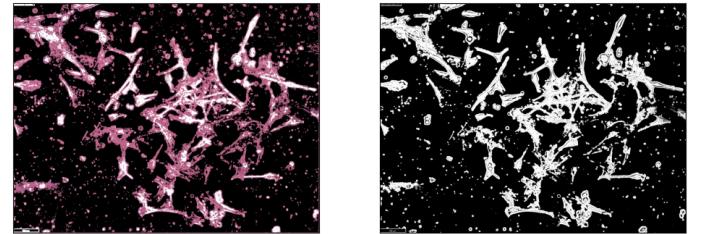


Fig. 8. Image after masking based on lower bound and upper bound where pink color denotes the uncertainty regions (left), Segmented image with Noise (Right)

accuracy is lying around 86%. The average Jaccard score & F1 score are found to be 0.43 and 0.6 (approx.), respectively. In the table provided in supplementary pages, the performance data obtained from other models are also mentioned.

TABLE I
MODEL PERFORMANCE COMPARISON (AVERAGES)

Model	IoU	Accuracy	Precision	Recall	F1 Score
Our Model	0.431	0.871	0.531	0.726	0.601
Stardist	0.087	0.672	0.130	0.267	0.172
Cellpose	0.130	0.865	0.358	0.164	0.205

VI. CONCLUSION

We present an unsupervised masking methodology for bright-field microscopy image segmentation, utilising spatial statistics and fuzzy inference. The methodology is devoid of training, resilient to changes in light, and versatile across various microscopy configurations. Expert-led validation and conventional evaluation metrics are used for model validation. Future directions encompass several areas involving better

precision, recall and instance segmentation for overlapping cellular communities. Besides this, "Integration of active learning for semi-supervised enhancement", "Application to alternative microscopy modalities (e.g., phase contrast, fluorescence)", etc. are the potential directions to extend the current work. Our research facilitates practical, efficient, and automated segmentation in biomedical imaging. next page.

ACKNOWLEDGMENT

The authors thank the Infochemistry Scientific Center at ITMO University for providing research support and computational resources. The authors also acknowledge the feedback from experts during the visual evaluation phase, which helped strengthen the segmentation performance assessment. The work was completed within the research topic No. 640103, "Development of methods for automated processing and analysis of optical and atomic force microscopy images using machine learning techniques".

REFERENCES

- [1] 3. In: Illuminators, Filters, and the Isolation of Specific Wavelengths. John Wiley Sons, Ltd; 2012. p. 35-51. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118382905.ch3>.
- [2] Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. *Nature Methods*. 2019 Dec;16(12):1233-46. Available from: <https://doi.org/10.1038/s41592-019-0403-1>.
- [3] Murphy D. *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. New York. 2001.
- [4] Hoque MZ, Keskinarkaus A, Nyberg P, Seppänen T. Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. *Information Fusion*. 2024;102:101997. Available from: <https://www.sciencedirect.com/science/article/pii/S1566253523003135>.
- [5] Netravali AN, Prasada B. Adaptive quantization of picture signals using spatial masking. *Proceedings of the IEEE*. 1977;65(4):536-48.
- [6] Albanwan H, Qin R. A novel spectrum enhancement technique for multi-temporal, multi-spectral data using spatial-temporal filtering. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2018;142:51-63. Available from: <https://www.sciencedirect.com/science/article/pii/S0924271618301539>.
- [7] Netravali AN, Prasada B. Adaptive quantization of picture signals using spatial masking. *Proceedings of the IEEE*. 1977;65(4):536-48.
- [8] Tang J. A color image segmentation algorithm based on region growing. In: 2010 2nd International Conference on Computer Engineering and Technology. vol. 6; 2010. p. V6-634V6637.
- [9] Singh S, Srivastava D, Agarwal S. GLCM and its application in pattern recognition. In: 2017 5th International Symposium on Computational and Business Intelligence (ISCBI); 2017. p. 20-5.
- [10] Oliver MA. In: Fischer MM, Getis A, editors. *The Variogram and Kriging*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 319-52. Available from: https://doi.org/10.1007/978-3-642-03647-7_17.
- [11] Kim W, Kanezaki A, Tanaka M. Unsupervised Learning of Image Segmentation Based on Differentiable Feature Clustering. *IEEE Transactions on Image Processing*. 2020;29:8055-68.
- [12] Masson P, Pieczynski W. SEM algorithm and unsupervised statistical segmentation of satellite images. *IEEE Transactions on Geoscience and Remote Sensing*. 1993;31(3):618-33.
- [13] Pereyra M, McLaughlin S. Fast Unsupervised Bayesian Image Segmentation With Adaptive Spatial Regularisation. *IEEE Transactions on Image Processing*. 2017;26(6):2577-87.
- [14] Mukhopadhyay A, Maulik U, Bandyopadhyay S. Multiobjective Genetic Algorithm-Based Fuzzy Clustering of Categorical Attributes. *IEEE Transactions on Evolutionary Computation*. 2009;13(5):991-1005.
- [15] Nguyen TM, Wu QMJ. A fuzzy logic model based Markov random field for medical image segmentation. *Evolving Systems*. 2013 Sep;4(3):171-81. Available from: <https://doi.org/10.1007/s12530-012-9066-1>.
- [16] Caillol H, Hillion A, Pieczynski W. Fuzzy random fields and unsupervised image segmentation. *IEEE Trans Geosci Remote Sens*. 1993;31:801-10. Available from: <https://api.semanticscholar.org/CorpusID:2812600>.
- [17] Caillol H, Pieczynski W, Hillion A. Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*. 1997;6: 3425-40. Available from: <https://api.semanticscholar.org/CorpusID:18027402>.
- [18] Liu D, Zhang D, Song Y, Zhang F, O'Donnell L, Huang H, et al. PDAM: A Panoptic-Level Feature Alignment Framework for Unsupervised Domain Adaptive Instance Segmentation in Microscopy Images. *IEEE Transactions on Medical Imaging*. 2021 Jan;40(1):154-65.
- [19] Mukherjee S, Sarkar R, Manich M, Labruyere E, Olivo-Marin JC. Domain Adapted Multitask Learning for Segmenting Amoeboid Cells in Microscopy. *IEEE transactions on medical imaging*. 2023 Jan;42(1):42-54. Place: United States.
- [20] Abdolghader P, Ridsdale A, Grammatikopoulos T, Resch G, Légaré F, Stolow A, et al. Unsupervised hyperspectral stimulated Raman microscopy image enhancement: denoising and segmentation via one-shot deep learning. *Opt Express*. 2021 Oct;29(21):34205-19. Available from: <https://opg.optica.org/oe/abstract.cfm?URI=oe-29-21-34205>.
- [21] Gaddipati A, Vince DG, Cothren RM, Cornhill JF. Enhancement of unsupervised segmentation using Gibb's random fields for microscopy image analysis. In: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136). vol. 2; 1997. p. 586-9 vol.2.
- [22] Čepa M. Segmentation of Total Cell Area in Brightfield Microscopy Images. *Methods and Protocols*. 2018;1(4). Available from: <https://www.mdpi.com/2409-9279/1/4/43>.