Semester report

by **Hamid Cheraghi** (hamidchr@student.elte.hu)

PhD Program: Biological physics, Statistical physics, and Quantum physics

Supervisor: Bálint Szabó

Ph.D. Thesis title: Image-based single cell isolation with deep learning

Introduction

My current research at the Eötvös Loránd University has focused on single-cell isolation of yeast and unicellular parasites. In this project, one of the main objectives is to develop hardware and software tools to isolate *T brucei* cells for subsequent DNA/RNA sequencing. Further goal of the project is to establish different strategies to isolate cells by an automated micropipette from agar plates and, if possible, from microfluidic devices with the aim to maintain single-cell lineages separated for further downstream analysis (e.g. next generation sequencing).

Automatic and reliable characterization of cells in cell cultures is key to several applications such as cancer research and drug discovery. Given the recent advances in light microscopy and the need for accurate and high-throughput analysis of cells, automated algorithms have been developed for segmenting and analyzing the cells in microscopy images.

Artificial Intelligence (AI) has been used to develop and advance numerous fields and industries including healthcare. Nowadays, many biological applications require the segmentation of the image of single-cells in microscopic images. There are several deep learning-based algorithms for 2D and 3D cell segmentation in microscopy cell images.

Cell Cycle

Cell cycle is the series of events that takes place in a cell as it grows and divides. A cell spends most of its time in what is called interphase, and during this time it grows, replicates its chromosomes, and prepares for cell division. The cell then leaves interphase, undergoes mitosis, and completes its division. The resulting cells, known as daughter cells, each enter their own interphase and begin a new round of the cell cycle [1]. Figure.1 shows the illustration of cell cycle.

FUCCI Cell Cycle

The FUCCI (Fluorescent Ubiquitination-based Cell Cycle Indicator) system was developed by Sakaue-Sawano et al. as a tool to visualize the dynamics of cell cycle progression. HeLa-Fucci cells periodically express two fluorescent protein tagged molecules namely Geminin fused with monomeric Azami green (mAG-hGem) and Cdt1 fused with monomeric Kusabira Orange (mKO2-hCdt1) the expression levels of which oscillate periodically during the cycle: mKO2-hCdt1 levels are high during G1, while mAG-hGem levels are high during the S/G2/M phases.

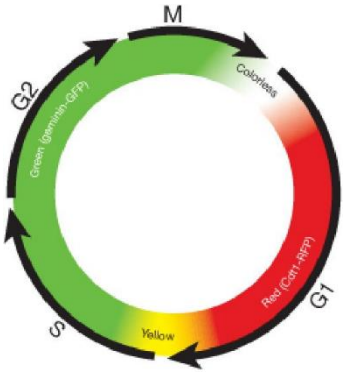


Figure.1 Two-color FUCCI cell cycle.

In this project, I am working on a dataset of Hela cell images. HeLa cells have since become the most widely used human cell line in biological research [2]. The dataset contains 221 Phase contrast images as well as the corresponding Green Fluorescent Protein (GFP) and the Red Fluorescent Protein images to the Phase contrast images.

Description of my current research

As I mentioned, utilizing AI algorithms is a major objective in our single cell isolation project. Thus, the project is divided into several stages and, the first step is to generate a ground truth dataset using AI. To do so, we utilized a generalist algorithm called **Cellpose** [3] for cell segmentation. So, we could segment cells and generate mask images and outlines for each cell in the unlabeled phase-contrast images. Figure 2 demonstrates an example of segmentation of the phase contrast images.

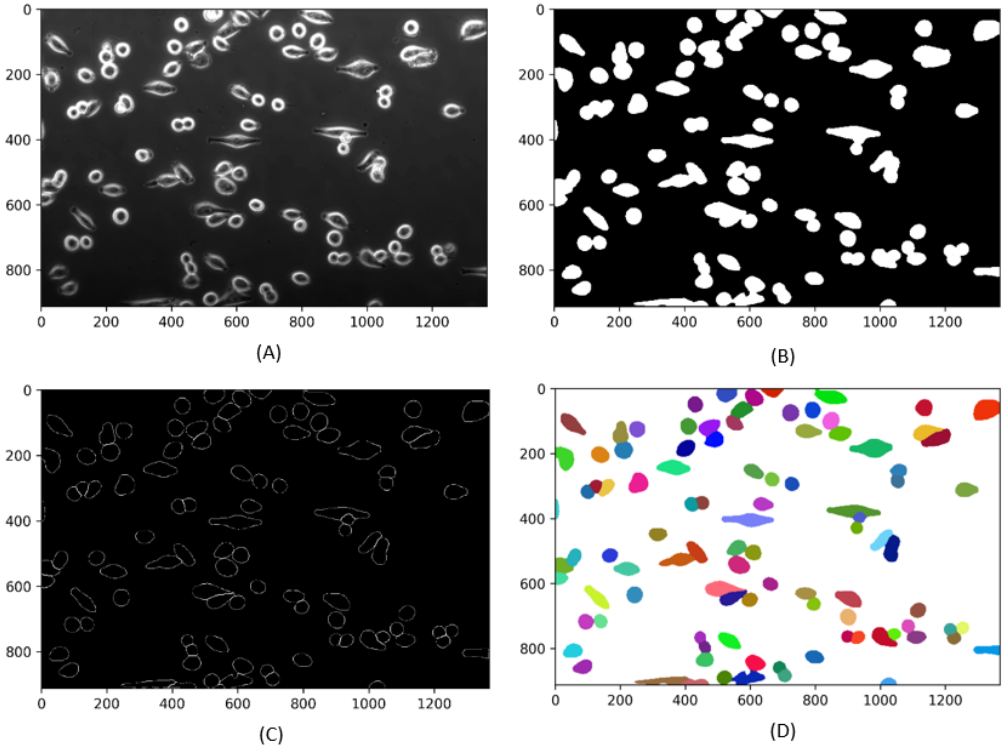


Figure.2 Generated mask images and outlines for a given input image using Cellpose. (A) an input phase image, (B) a generated gray mask, (C) outlines of each cell in a phase image, (D) a generated RGB mask that helps to see border of every cell.

When we are able to obtain the mask and outline for each and every cell in a given phase contrast image, we segment the fluorescent red and green images using the mask determined in the phase contrast images. The next part of this stage is to crop the image of cells in all 3 channels. To do so, we need to achieve coordinates of each cell to be able to draw a rectangle around them. Also, we can annotate each rectangle and put the number of cells next to the rectangle. Figure 3 shows an example of the mask, the red, and the green images along with rectangles and the annotations.

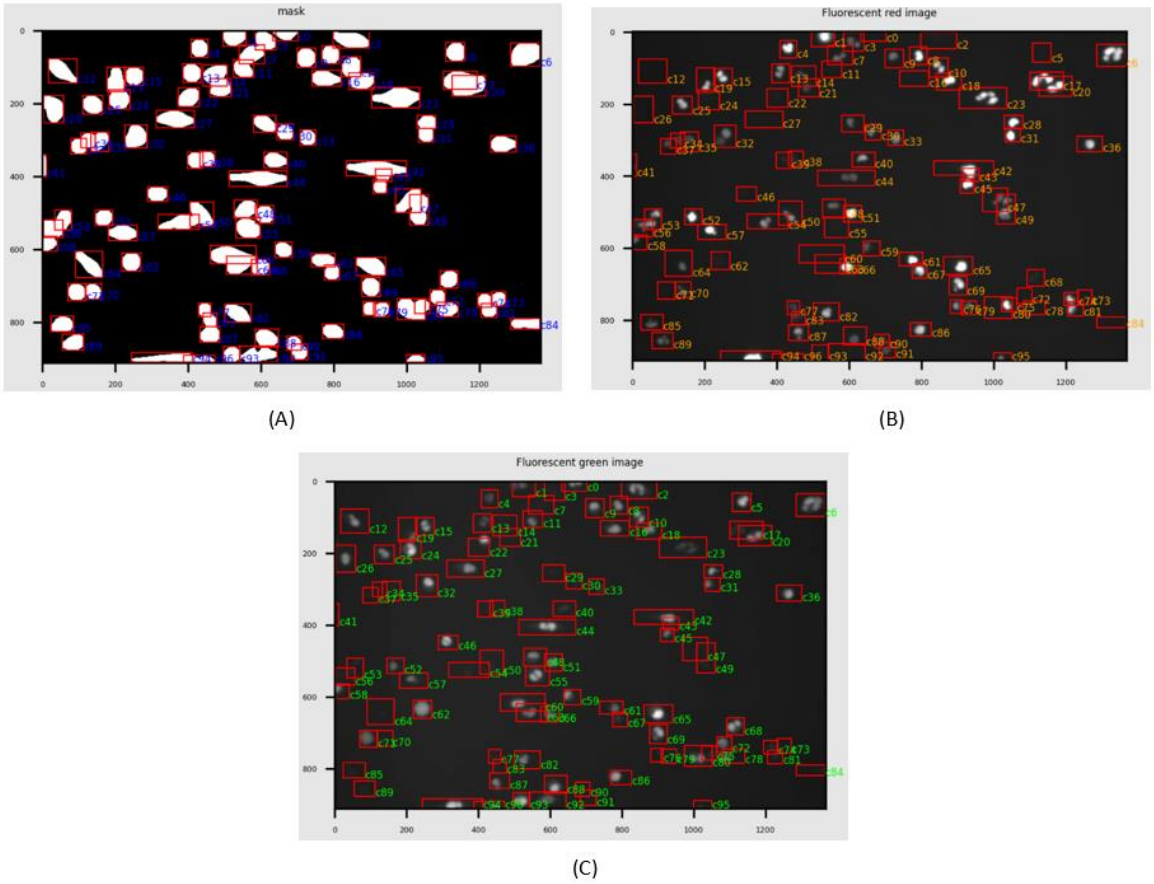
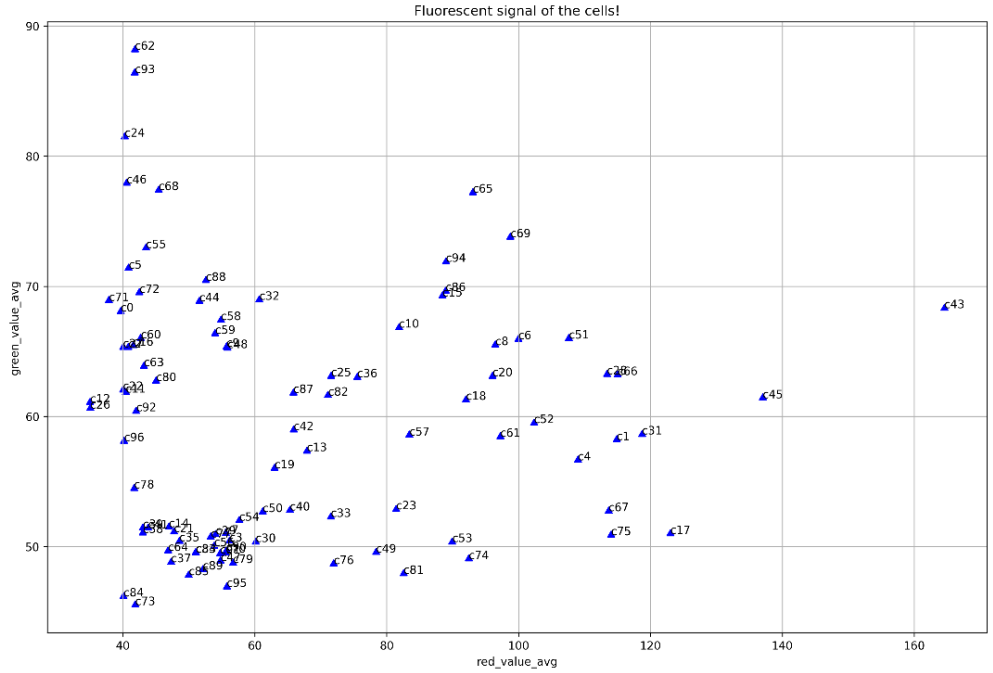
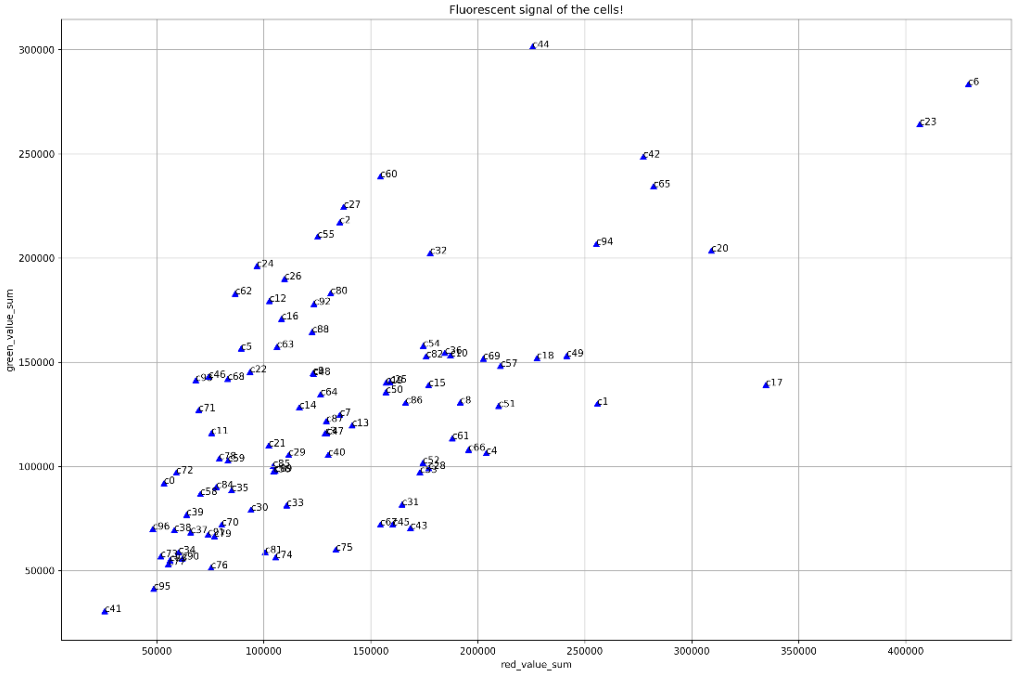


Figure.3 An example of the mask, the red, and the green images along with rectangles and the annotations. (A) a generated gray mask, (B) Fluorescent red image, (C) Fluorescent green image.

The next step of this part is to calculate the sum and the spatially averaged fluorescent intensity of each cell in both red and green channels. By this, we can see the fluorescent signal of each cell. It should be mention that the following plots show the average value and sum value for only one image with almost **100** cells as an example. I have similar results for **30,000** cells in a dataset containing **221** images.





Study activities

In this semester, I continue my studies by taking courses entitled “Deep learning and machine learning in natural sciences” (Neptun code: FIZ/3/089). My grade was: ‘Excellent’. Also, I completed the course with the title “Data exploration and visualization” (Neptun code: FIZ/3/085). The result of this course hasn’t been announced yet.

Membrane Biology was another course that I took. This course was held in a block within two days. It was supposed to be in English when I registered for it but, before the beginning of the course, I received an email from the professor that said this semester, it is in the Hungarian language.

During the semester, I participated in weekly lectures organized by the Ludwig Maximilian University (LMU) of Munich in the Cell2Cell ITN: https://cell2cell.eu

In these ten lectures professors from different universities including University of Tokyo, University of California San Francisco (UCSF), LMU gave presentations of their current research and state of the art in single-cell science, single-cell heterochromatin spreading sensor (HSS), and single-cell analysis of transcriptome (RNA-sequencing).

I participated in a Biostatistics workshop offered by Cell2Cell ITN: https://cell2cell.eu.

This workshop organized by the institute of Epigenetiks (Epigenetiks Genetik Biyoinformatik Yazılım), Istanbul, Turkey.

The Biostatistics Course Sylabus:

- Day1: Descriptive Statistics
- Day2: Experiment design (Sampling Theoren, Normal Distribution, Confidence Intervals, Sample size determination, Propensity Matching)
- Day3: Hypothesis Testing (Parametric and Non Parametric tests, Paired Tests)
- Day4: Linear Regression, Logistic Regression
- Day5: Anova and its different Applications
- Day6: Principal Component Analysis and its applications in experimental design (Batch Effects, mislabelling, etc.)
- 2hours Lectures+2 Hours Practicals

References

[1]. National Human Genome Research Institute: https://www.genome.gov/genetics-glossary/Cell-Cycle

[2]. British Society for Immunology: https://www.immunology.org/hela-cells-1951

[3]. Stringer, Carsen, et al. "Cellpose: a generalist algorithm for cellular segmentation." Nature Methods 18.1 (2021): 100-106.