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 research at the Eötvös Loránd University focuses 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 *Trypanosoma 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).

### FUCCI Cell Cycle reporter construct

Fluorescent ubiquitination-based cell cycle indicator (Fucci) is a sophisticated technology which can easily determine G1 and/or S/G2/M phases of the cell cycle. The technology analyzes living cells in a spatio-temporal manner using a dual color scheme of orange and green.

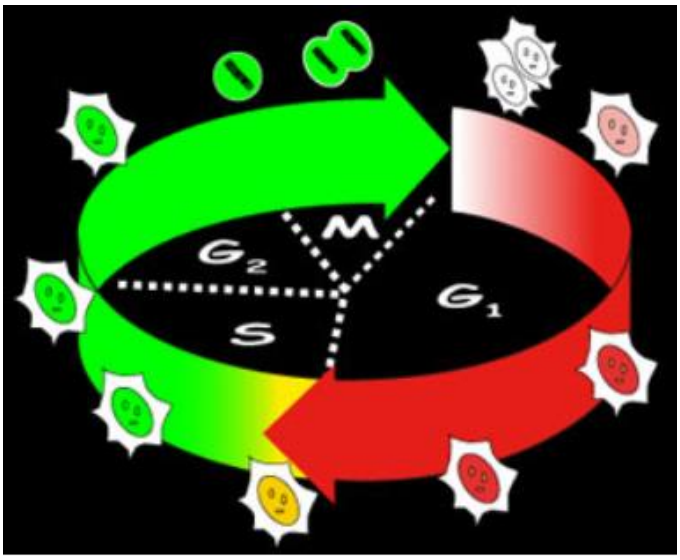


**Figure 1.** Schematic representation of Fucci cell-cycle labeling Fucci-G1 Orange labels G1 phase nuclei in orange. Fucci-S/G2/M Green labels S/G2/M phases nuclei in green.

In my current project, we aim to integrate our developed cell intensity calculation pipeline code written in python language with an existing software written in C# language. The pipeline includes several levels as shown in Figure 2. To calculate the intensity value for each cell, we utilized a generalist algorithm called Cellpose [2] for cell segmentation. So, we could segment cells and generate mask images and outlines for each cell in the unlabeled phase-contrast images. The next part of this stage is to crop the image of cells in all 3 channels as well as the mask image. To do so, we need to achieve coordinates of each cell to be able to draw a rectangle around them.

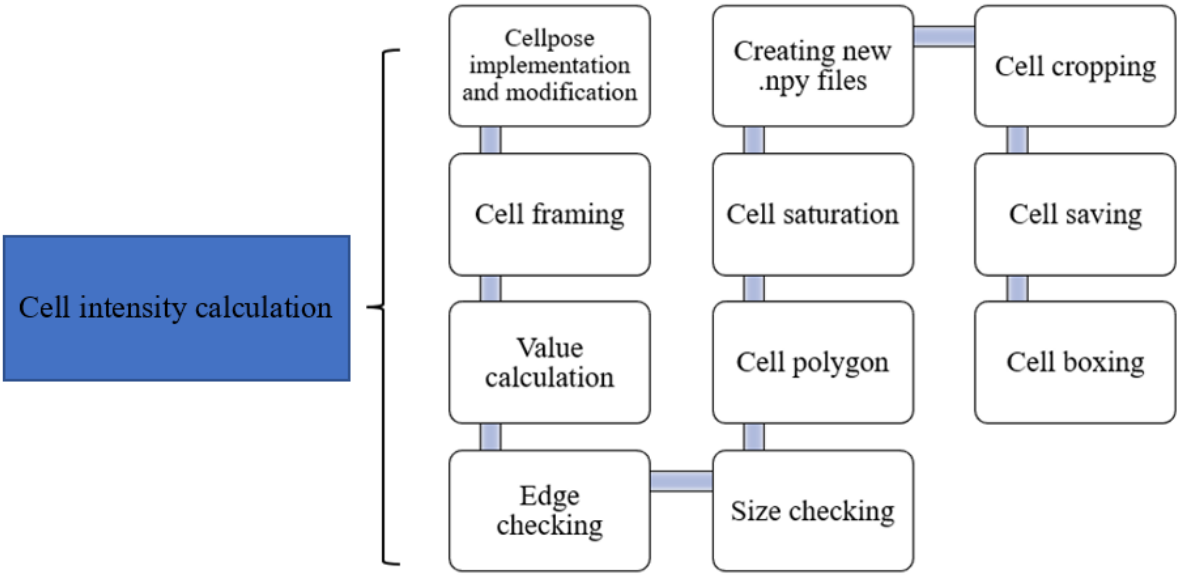


**Figure 2.** Cell intensity calculation pipeline.

After finishing my complex exam, I will start writing the manuscript with this tittle: Automated pipeline for cell segmentation of phase contrast images and subsequent calculation of fluorescent intensity

Authors of the paper are: Hamid Cheraghi, Kinga Dóra Kovács Inna Székács, Robert Horvath, Bálint Szabó

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. HeLa cells are the most widely used human cell line in biological research [3].

In another project that I am working on, we have a dataset of fluorescent and phase contrast HeLa cell images. My dataset contains 324 Phase contrast images as well as the corresponding Green Fluorescent Protein (GFP) and the Red Fluorescent Protein images to the Phase contrast images. Images were acquired on an automated microscope by me. The second step is implementing a deep Convolution Neural Network (CNN) for cell cycle estimation (prediction). In this step, we use a

pre-trained CNN networks named Xception model. The overall structure of this model is as follows:

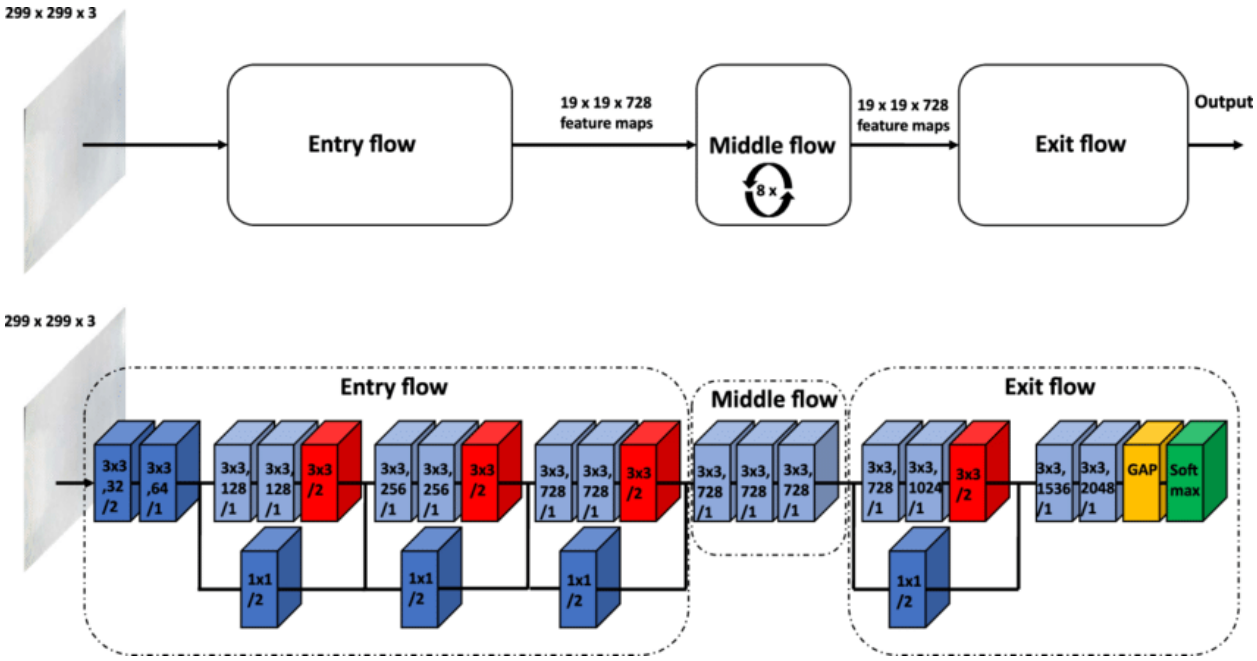


**Figure 3.** Xception CNN architecture before finetuning the network. We will feed this network using the provided images from previous steps and generated labels to predict the intensity value for each cell.

During this semester, I joined a laboratory at the Institute of Molecular Medicine (IMM) in Lisbon, Portugal for one month. Below, I shall explain a summary of main activities during my secondment at IMM [4].

*Trypanosoma brucei* is a unicellular and extracellular parasite which provokes sleeping sickness in humans and nagana in cattle; both of these diseases are endemic in Africa [5]. This parasite present two different stages in the mammalian host: a replicative form called “slender”, and a non-replicative form called “stumpy”. These two different stages can be cultured in the laboratory, being essential for the research of the diseases. Figure 4. demonstrates the cellular heterogeneity in the *T.brucei* life cycle.

As I mentioned before, this particular parasite presents some morphological differences (as the name suggested, “stumpies” are smaller and fatter, while the “slenders” are longer and more active) but it is very difficult to differentiate them by eye in the lab, since their small size (an average of $15\mu m$). To classify them, some markers are needed to detect if they are in one stage or other. This can be very difficult to get since you need to modify genetically the parasites, and this can be very challenging sometimes. A tool to let us know the number of slenders and stumpies in a population could be very useful for the *T.brucei* field. Therefore, developing an algorithm based on Artificial Intelligence (AI) is one of our future goals.

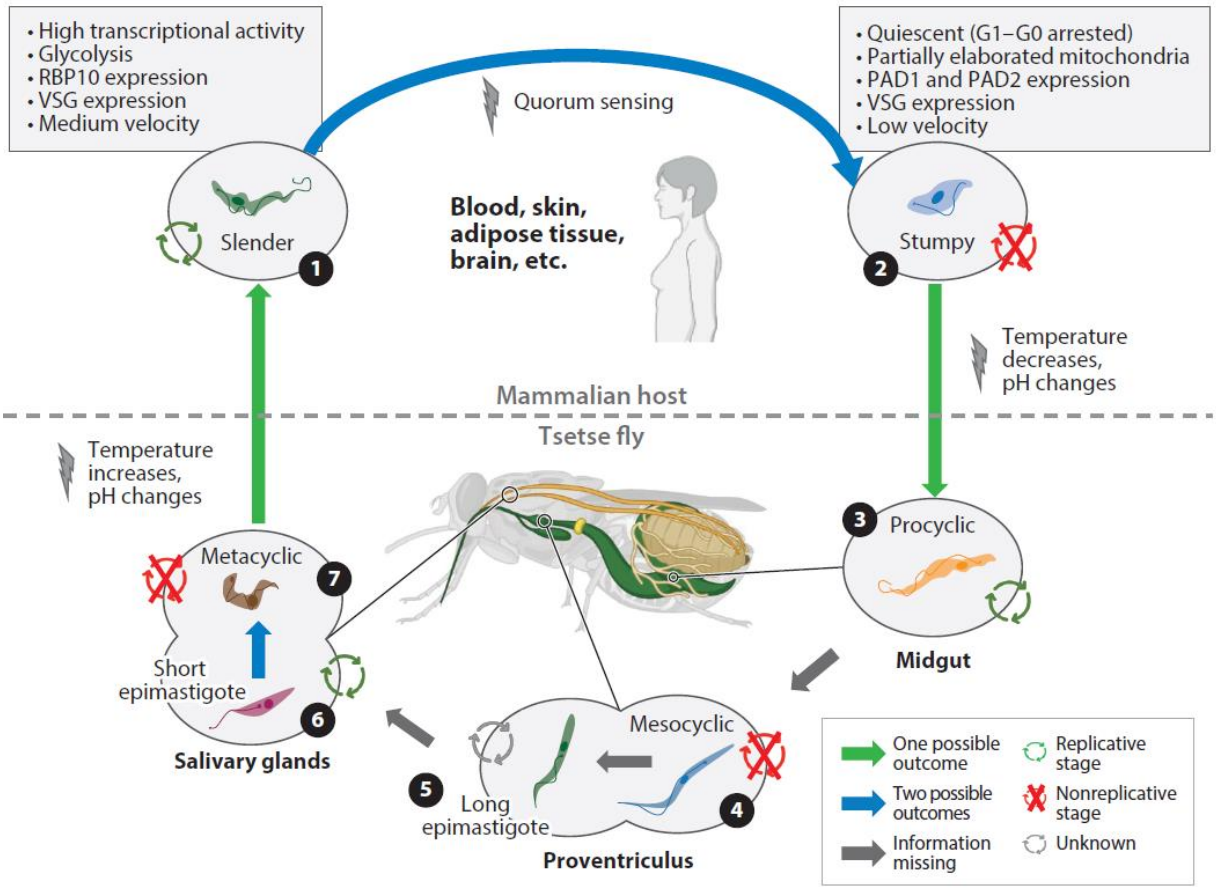


**Figure 4.** Cellular heterogeneity in the Trypanosoma brucei life cycle. The life cycle comprises at least seven stages in the mammalian host and the tsetse fly. (❶) Slender forms are replicative forms that ensure chronicity of the infection. These cells use quorum sensing to induce differentiation to the stumpy form (95). (❷) Stumpy forms are arrested in G1–G0 and preadapted to the tsetse fly environment. All stumpy forms appear to differentiate to procyclic forms (❸) in the midgut of the fly. Here, parasites differentiate to the (❹) mesocyclic forms, which are cell-cycle arrested and have a long body.Mesocyclic forms migrate to the proventriculus, where they eventually differentiate into the (❺) long epimastigote form. This is one of the most motile forms (67). It undergoes asymmetrical division to the (❻) short epimastigote, which is a replicative form responsible for establishing an infection in the salivary glands (109). The short epimastigote is attached to the epithelium of the salivary gland, and it is a replicative stage with two possible outcomes: It remains a short epimastigote or differentiates to a (❼) metacyclic form, the last stage in the fly and the infective one for mammalian hosts (53). This form is nonreplicative and is preadapted to the mammalian environment (108). When the metacyclic form enters in the blood, it encounters a higher temperature, which induces differentiation to the (❶) slender form.

To achieve this, I joined Luisa Figueiredo's lab at IMM, which is a well-known lab specialize in these parasites. The first goal of my secondment was to study and learn about *T.brucei* cells. Likewise, working with these parasites which becomes two important tasks was my major goals. These two tasks are:

First: Parasite classification in culture (in vitro), second: Parasite classification in blood (in vivo).

For this work we needed to get good pictures from stumpies and slenders populations and develop a network which will be able to recognize and classify them just using their morphological characteristics (area, length, width). The first aim for me was to keep the parasite in culture, to get familiar with their behavior and recognize what is the normal morphology and motility they have in vitro. After this, I learnt how to prepare them for imaging without changing their morphology: how to centrifuge, fix and wash them. I also learnt how to prepare a slide with mounting oil to conserve the parasites in perfect conditions for them to imagining.

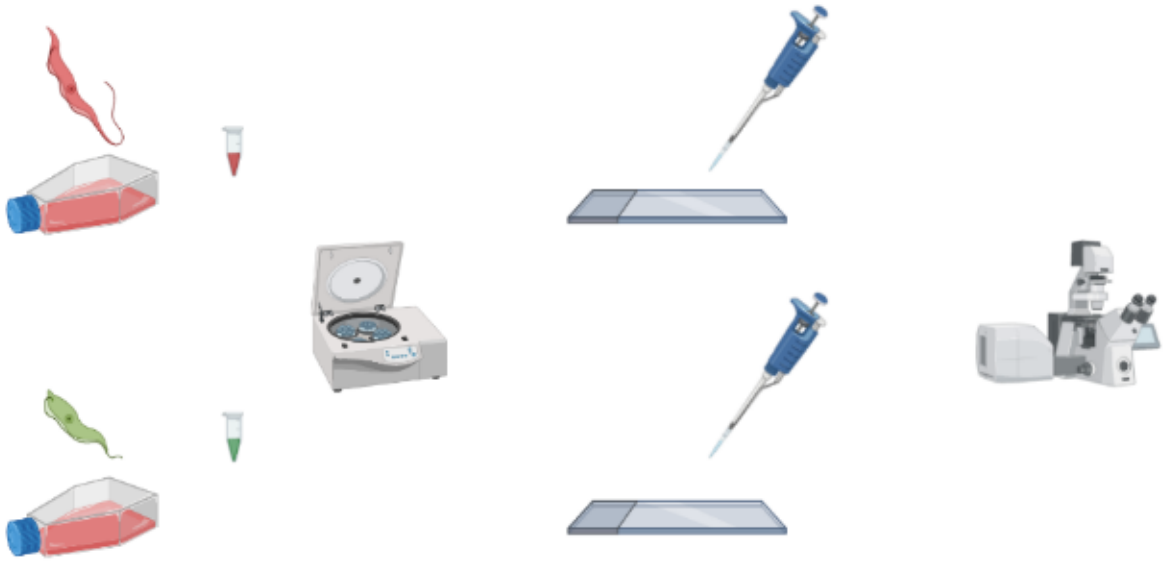

**Figure 5.** Preparation process for *T.brucei* parasites to be scanned and provide the best conditions for the parasites to be ready for capturing the images.

To differentiate what cells are slenders and what cells are stumpies, I worked with a cell line which express a Green Fluorescent Protein (GFP) when they become stumpy. Working in the Luisa Figueiredo's lab helped me to produce two different populations which almost the 90% of the cells are stumpies or slenders, for then the classification of the network can be easier. Also, Hoechst as used to die the DNA from the parasites, being like that more easier for the network to discard what are cells what are background. Figure 5. shows the process of preparing parasites to be ready for taking the images.

**Study activities**

In this semester, I continued my studies by taking courses entitled "Sensory biophysics II" (Neptun code: FIZ/3/045E). The obtained grade was: 'Excellent' for this course. Also, I completed the course with the title "Theoretical evolutionary biology" (Neptun code: FIZ/3/005E).

During the semester, I participated in weekly lectures organized by the Ludwig Maximilian University (LMU) of Munich in the Cell2Cell ITN: https://cell2cell.eu [6]

Chromatin Biology by Dr. Claudia Keller Valsecchi, Institute of Molecular Biology Mainz (IMB), Germany.

The second lecture with the title "Gene regulatory mechanisms governing Plasmodium development and stage transitions" was taught by Prof. Dr. Shruthi Vembar

The third lecture with the title "Regulation of Toxoplasma developmental switches through transcriptional rewiring" by Prof. Dr. Shruthi Vembar.

**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.

[4]. https://imm.medicina.ulisboa.pt/

[5]. Luzak, Vanessa, et al. "Cell-to-Cell Heterogeneity in Trypanosomes." Annual review of microbiology 75 (2021): 107-128.

[6]. https://cell2cell.eu/