UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

SELF-TUNED OBJECT TRACKING ALGORITHM FOR LIVE-CELL BRIGHT FIELD TIME-LAPSE MICROSCOPY

Tesis sometida a la consideración de la Comisión del Programa de Posgrado en Ingeniería Eléctrica para optar al grado y título de Maestría Académica en Ingeniería Eléctrica

SERGIO ARTURO MOYA VALERÍN

Ciudad Universitaria Rodrigo Facio, Costa Rica

2022

Inscription

To Yulieth, for your unwavering love and support—this journey would have been much harder without you.
Thanks.

Acknowledgments

To Yulieth, for your unwavering love and support. This journey would have been much harder without you. Your presence has been a constant source of strength, and I'm endlessly grateful for the countless ways you've helped me along the way.

To Professor Francisco Siles Canales, whose guidance has shaped both my academic and professional paths. More than a tutor, you've been a friend, offering advice that has made a profound impact on my life. Thank you for your unwavering support throughout and master's studies, as well as in those, critical decisions in my career.

To the PRIS-Lab, a place that feels like home and continually inspires me. Every challenge faced there has been a learning opportunity, and the willingness of everyone to assist makes the experience truly special. A heartfelt thank you to all who have contributed to my growth. Special thanks to Daryl Ross and Ignacio Gomez for their efforts in developing the annotation tool that will benefit many.

Finally, I extend my gratitude to the Postgraduate Program in Electrical Engineering at the Postgraduate Studies System, UCR, for their support throughout my Master's journey.

Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Ingeniería Eléctrica de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Ingeniería Eléctrica.

FEDERICO ALBERTO RUIZ UGALDE (FIRMA)
PERSONA FISICA, CPF-01-1059-0203.
Fecha declarada: 18/11/2024 01:18:44 PM

Dr. rer. nat. Federico Ruíz Ugalde
Representante del Decano
Sistema de Estudios de Posgrado

Digitally signed by: FRANCISCO DE PAULA SILES CANALES (FIRMA)
Date and time: 11/15/2024 4:28:21 PM

Dr. rer. nat. Francisco Siles Canales
Director de Tesis

STEVE QUIROS BARRANTES (FIRMA) Date: 2024.11.15 16:33:24 -06'00'

Dr. rer. nat. Steve Quirós Barrantes
Asesor

MSc. Marco Villalta Fallas
Asesor

Dr. Diego Dumani Jarquín
Representante del Director
Programa de Posgrado en Ingeniería Eléctrica

SERGIO ARTURO MOYA VALERIN (FIRMA) Digitally signed by SERGIO ARTURO MOYA VALERIN (FIRMA) Date: 2024.11.18 13:52:40 -06'00'

Sergio Arturo Moya Valerín
Candidato

Table of Contents

Resumen

This thesis presents a novel self-tuned object tracking algorithm specifically designed for live-cell bright-field time-lapse microscopy. The primary motivation for this project stems from the challenge of analyzing cell behavior over time using bright-field microscopy images, which are typically more challenging to process compared to fluorescence microscopy due to lower contrast and noise. The Damage Proliferation Phenomenon (DPP), a significant contributor to multiresistant and aggressive cancer phenotypes, plays a central role in this investigation.

The proposed tracking algorithm leverages machine learning and pattern recognition techniques to address this challenge. By utilizing Bayesian optimization and gradient descent methods, the algorithm achieves robust performance in tracking cells across multiple time-series images, offering a solution to limitations posed by traditional manual observation and segmentation methods. This tool can significantly aid in understanding cancer cell behavior, improving diagnosis, and informing treatment strategies.

The study includes a comprehensive review of the current state of the art, comparisons of various segmentation algorithms, and tests the developed algorithm using real-world datasets from live-cell microscopy. Results show improvements in tracking accuracy, particularly in detecting DPP-related behaviors, and provide insights into potential future enhancements.

Abstract

Esta tesis presenta un novedoso algoritmo de seguimiento de objetos autotunado, diseñado específicamente para la microscopía de campo claro de células vivas con intervalos de tiempo. La principal motivación para este proyecto proviene del desafío de analizar el comportamiento celular a lo largo del tiempo utilizando imágenes de microscopía de campo claro, que suelen ser más difíciles de procesar en comparación con la microscopía de fluorescencia debido al menor contraste y al ruido. El Fenómeno de Proliferación de Daños (DPP), un importante contribuyente a los fenotipos de cáncer multirresistentes y agresivos, desempeña un papel central en esta investigación.

El algoritmo de seguimiento propuesto utiliza técnicas de aprendizaje automático y reconocimiento de patrones para abordar este desafío. Mediante el uso de optimización bayesiana y métodos de descenso de gradiente, el algoritmo logra un rendimiento robusto en el seguimiento de células a lo largo de múltiples imágenes de series temporales, ofreciendo una solución a las limitaciones impuestas por los métodos tradicionales de observación y segmentación manual. Esta herramienta puede ayudar significativamente a comprender el comportamiento de las células cancerosas, mejorando el diagnóstico y proporcionando estrategias informadas de tratamiento.

El estudio incluye una revisión exhaustiva del estado del arte, comparaciones de varios algoritmos de segmentación, y prueba el algoritmo desarrollado utilizando conjuntos de datos del mundo real provenientes de microscopía de células vivas. Los resultados muestran mejoras en la precisión del seguimiento, particularmente en la detección de comportamientos relacionados con el DPP, y proporcionan información sobre posibles mejoras futuras.

List of Tables

List of Figures

Acronyms

AP	Apoptosis
DNA	Deoxyribonucleic Acid
DPP	Damage Proliferation Phenomenon
PR	Pattern Recognition
RDD	Repaired DNA Damage
RNA	Ribonucleic Acid
GD	Gradient Descend
SGD	Stochastic Gradient Descend
BO	Bayesian Optimization
MPG	Multipartite Graph / K-partite Graph

Glossary

Algorithm	A specific procedure for solving a well-defined computational problem.
Apoptosis	A form of cell death in which a programmed sequence of events leads to the elimination of cells without releasing harmful substances into the surrounding area.
Brightfield	One of the simplest optical microscopy. In bright-field microscopy, illumination light is transmitted through the sample and the contrast is generated by the absorption of light in dense areas of the specimen.
Chemosensitivity	The susceptibility of tumor cells to the cell-killing effects of anticancer drugs.
Dataset	A collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer.
Fluorescense	Is based on the principle that fluorescent materials emit visible light when they are irradiated with ultraviolet rays or with violet-blue visible rays.
Genotoxic	Substances that may bind directly to DNA or act indirectly leading to DNA damage by affecting enzymes involved in DNA replication, thereby causing mutations which may or may not lead to cancer or birth defects.
Microscopy	The technical field of using microscopes to view objects and areas of objects that cannot be seen with the naked eye.
Multiresistance	Resistance of cancer cells to multiple chemotherapy drugs.
Phase Contrast	Optical microscopy technique that converts phase shifts in light passing through a transparent specimen to brightness changes in the image.
Phenotype	The set of observable characteristics of an individual resulting from the interaction of its genotype with the environment.
Phototoxicity	A condition in which the skin or eyes become very sensitive to sunlight or other forms of light.

Time Series	A series of values of a quantity obtained at successive times, often with equal intervals between them.
Time Steps	Moving from one image to the next on a time series.

Chapter 1

Introduction

Metastatic tumors are the second commonest cause of WHO-client fatalities after cardiovascular diseases and, after injuries, are also one of the leading causes of death in most regions of the world. They are the most frequent neoplasia among men and women on a global scale, being responsible for -8888 new cases worldwide in 2018. Lung, breast, and gastrointestinal neoplasia are the most frequent diagnosed cancers globally. The International Agency for Research on Cancer also included that the malignancy is a serious public health issue as there were a total of 14 million new cases and 8.2 million deaths due to cancer in 2012. Lung, breast, and colorectal cancer are the most prevalent forms of cancer around the world (INEC, 2018).

Detection of the disease in advanced stages is a key factor in the management of its course since it is very much connected with the outcome and the treatments that are going to be used. However, to just find the presence of the cancer without applying any adequate therapeutic procedure that will, in a certain degree, improve the life of the patient is not enough. Furthermore, it is also very important to establish whether the cancer form is special and what are the other factors that relate to it so that we can provide targeted and personalized care. Negligence in such things may result in the development of drug resistance and the patient's condition could get less favorable. The DPP is a concept that has been referred in connection with the phenomenon of aggressive and multi-drug resistant tumors. The capability to detect and comprehend such subaskets and their specifications is a big advance in the long-term race against the malignancy Torre et al. 2015 Organization, 2022.

So, after you get the cancer report, one of the most important steps will be making the perfect treatment decision. However, the treatment plan you will follow will have the choice of the modalities - whether it will be chemotherapy, radiotherapy, or they may include surgery. The specific treatment plan may involve a single modality or a combination of approaches, which will depend on the characteristics of the cancer and the patient. Chemotherapy, which is one of the most commonly used treatment options, involves the administration of DNA-damaging drugs known as genotoxic. These drugs stop the divide and consequentially affect the cell proliferation. There are various ways to give chemotherapy,

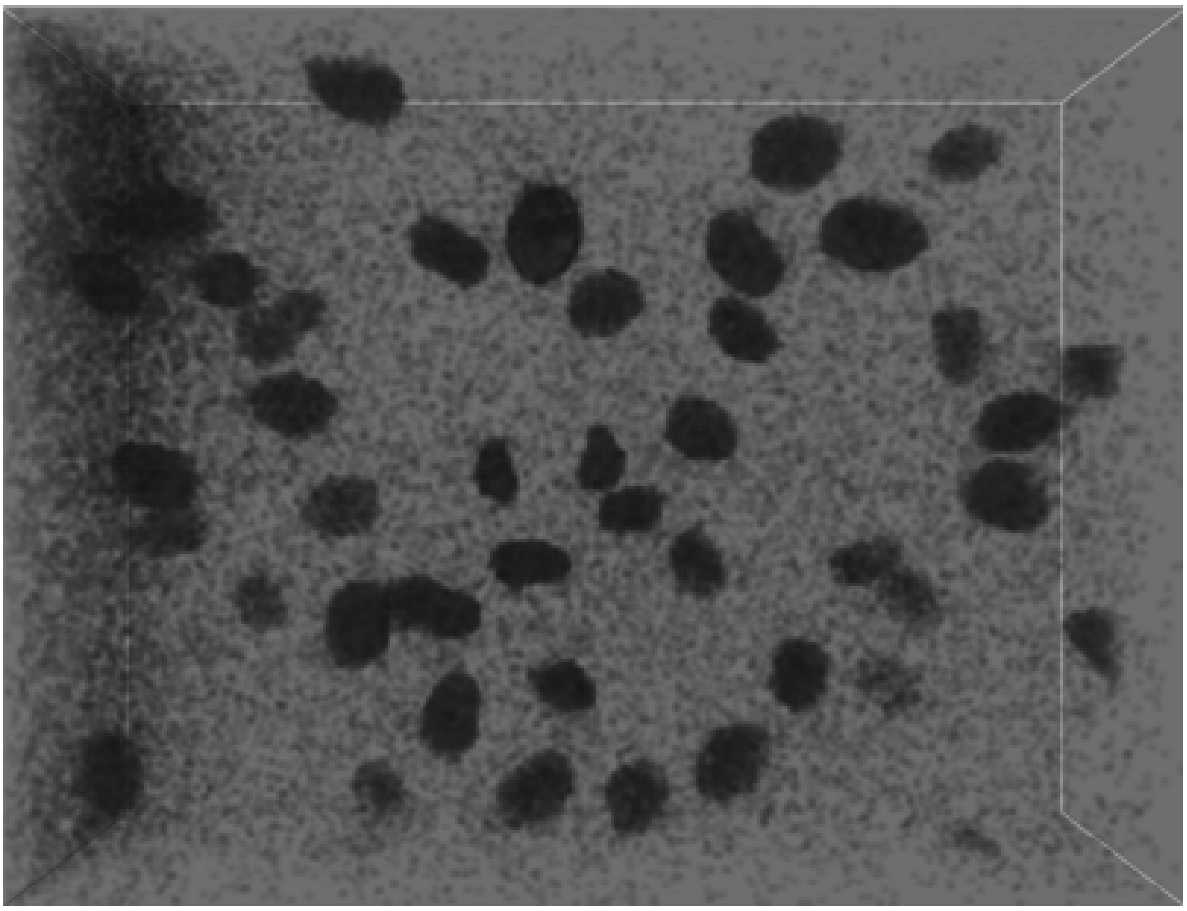
including injection, oral medication, infusion, and topical application. The choice of method will depend on the specific circumstances of the patient and the cancer (Institute, 2017).

Take for example, the application of human vision through microscopes. It has been the major method used to study cellular behavior during the last decades, but quite few studies have incorporated machine vision with cancer cells. Even though classical visual analysis can expose numerous biological processes, there will always be a human factor-influencing making decisions and perceptive judgment decisions-biases. People can have different points of view and even the same expert can get different results. On the contrary, a machine that is labeling only those characteristics that are relevant will have a much more uniform output, and consequently, medicine will become simpler, yet, of course, if the machine has a wrong diagnosis, it will lead to chronic misdiagnosis.

The essence of machine learning and pattern recognition techniques is their generalizability, meaning the way they can apply to as many situations as possible (Collins, 2003). As a consequence of this, a large number of algorithms are created to be variable in which they can be used not only in original but also in different scenarios. Nevertheless, it should be emphasized that these algorithms may indicate trends in states that are not pertinent to the initial problem being solved, especially when using complex deep learning models. These errors in the algorithm's outputs may lead to even unexpected but useful results that can help with later upgrades or pitch the idea for further investigation.

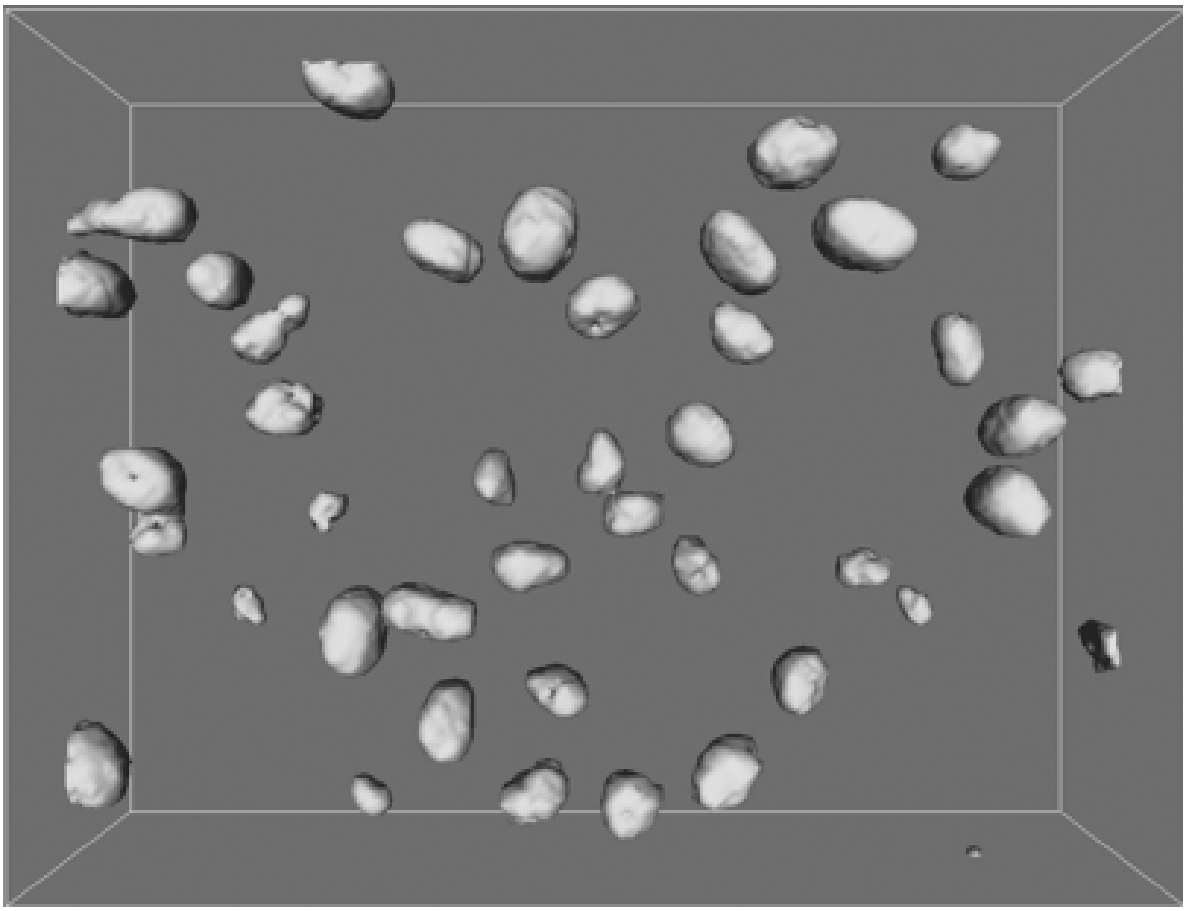
Computational biology includes many different foci, and the well-structured use of these computational methods is a must in the research conducted by natural scientists that want to go deeper in biological science. The most relevant computational methods are the development and the implementation of analytical and theoretical methods, mathematical modeling, and computation. As seen in the results presented in Figures 1.1 and 1.2, these techniques can be employed to analyze and extract relevant information from images. The complication of the given biological problems often calls for computational power as a tool to achieving good results.

Figure 1.1: Stained nuclei of proliferating MDCK cells.



It is easily observable that the image taken from the microscope have a lot of noise, and it is hard to observe the nuclei characteristics Source: Dufour et al. 2005

Figure 1.2: Segmented stained nuclei of proliferating MDCK cells.



Source: Dufour et al. 2005 After process the image with image processing, pattern recognition and machine learning techniques, the cell nuclei are easily observable, and some characteristics were enhanced.

In contemporary scientific research, it is common to utilize advanced technology that generates large quantities of data that cannot be effectively analyzed through manual methods. One example of this is the use of high-end microscopes, which can produce millions of images from plates and other sensors, resulting in an overwhelming amount of data. Without the use of computational methods, the sheer volume of data produced by these sensors can make it difficult to extract meaningful information, potentially leading to the loss of valuable insights that could be used to advance scientific investigations.

While traditional computer vision techniques have been successful in addressing a wide range of biological problems, it is essential to stress the fact that these methods may not always yield the expected result. The biological systems' intricate nature can give rise to the diverse results produced by these techniques. In situations like these, it is necessary for computer scientists to work hand in hand with natural sciences experts, using a mix of the present algorithms and designing new ones, in

addition to changing the existing code, to attain the aim.

Bright-field microscopy is a typical and reliable method for visualizing cells and identifying potential diseases, that is, it helps both in diagnosis and medication of diverse conditions. However, the manipulation of images acquired through this technique and the extraction of relevant information remains an ongoing area of exploration. Up until now, there are still no established methods for cell lineage information extraction from bright-field microscopy images, which implicitly delays advancements in the treatment of cancer. Scientists have mainly concentrated on clarifying the underlying biological mechanisms rather than using computational methods for analyzing the images. In contrast, when utilizing the knowledge from the past explorations, it can be possible to create algorithms to be used in this image thus lessening the workload of the health professionals and, in turn, the misdiagnosis.

Research investigations conducted by Universidad de Costa Rica specifically through the *Red de investigación traslacional en cáncer y biocomputación* ITCB founded by the *Pattern Recognition and Intelligent Systems* (PRIS-Lab) and the *Laboratorio de Quimiosensibilidad Tumoral* (LQT), among others, in association with Universitätsmedizin der Johannes Gutenberg-Universität Main, have disclosed that a fraction of cancer cells subjected to genotoxic chemotherapy can still endure conditions that should usually cause cell death by apoptosis (Hayashi y Karlseder, 2013). This Damage Proliferation Phenomenon (DPP) not only enables these cells to multiply but also to pass on the genetic instability that results from cell damage. Along with that, DPP promotes the increase in aneuploidy (which is a state of having an abnormal number of chromosomes) (Santaguida y Amon, 2015) that, in some cases, can result in the change of healthy cells to malignant ones thus producing more aggressive phenotypes or even leading to the emergence of more resistant varieties of cancer (Chen et al. 2015).

On a global scale, the Damage Proliferation Phenomenon (DPP) has become one of the hot topics in the scientific community owing to its relevance for unraveled cellular regulatory networks. Researches performed there not only aim to discover ways through which cells decide on cycle arrest or apoptosis (Choi et al. 2012) but also, to start the genotoxic environments perturbations, that bring about the DNA damage responses, which subsequently affect the transduction cascades therefore indicating the chemosensitivity or radiosensitivity markers during the stress of the DNA replication. Studies like the one by the Vice-Rectory of the University of Costa Rica (UCR) named *Analysis of the role of the cellular response to DNA damage in the induction of cell death generations after chemotherapy treatments (Análisis del papel de la respuesta celular al daño en el ADN en la inducción de muerte celular generaciones posteriores a tratamientos quimioterapéuticos)* provide quite good insights on

the DPP's posttransational reaction where it plays a crucial role in early decision-making of cellular functions like cell cycle arrest, and apoptosis, and repairing the DNA (Quirós et al. 2014).

The prior investigations have brought the necessity to understand the DPP (Damage Proliferation Phenomenon) and chemotherapy link. They have mainly dealt with investigating the mechanisms responsible for the aggressive phenotypes and the multiresistant cancer varieties. This work has managed to expose the association between these two themes while employing a mixture of methods like protein analysis and classical laboratory & statistical analysis. While the results created by these queries have proven to be of eminent value still they pave the way for further exploration in the future.

It is paramount to conduct standardized and repeatable experiments in which the treatment of each sample is identical, in order to be able to sufficiently characterize and quantitatively assess the Damage Proliferation Phenomenon (DPP) occurrence. The additional presence of other common cellular phenomena, i.e., cell cycle arrest, repaired DNA, and apoptosis, can introduce noise and cause research in this area to be impeded. Therefore, it becomes imperative to isolate every phenotype in a perfect analogy to a control experiment to see the effect of each one on the cells independently. This task becomes even more complex, even for the great specialists, if we take into account the limitations of fluorescent visualization, which is not an option because of the phototoxicity of the fluorescence stain. Thus, in most cases, the researchers would be left with bright field microscopy, which can sometimes lead to difficulties in human vision and therefore could produce less accurate results. Moreover, keeping track of a single cell over the course of time is quite essential to correctly identify the event of DPP. However, even if an individual instance of observation is required, but to do it in a plate with hundreds or even thousands of cells which makes it nearly impossible, even the expert human observers (Mora, 2018) often cannot achieve this.

1.1. Justification

The Damage Proliferation Phenomenon (DPP) is a unique manifestation by which specific cells are capable to replicate although there is a huge damage inflicted on their DNA. This unique mechanism can be responsible for the appearance of the multi-resistant and aggressive cancer types even after the complete course of chemotherapy or radiation therapy is administered to the patient. At this time, the qualified healthcare providers suffer from the real ordeal in detecting the existence of DPP. The integration of state-of-the-art technology, such as biocomputing, machine learning, and pattern recognition algorithms, can assist healthcare professionals in the segregation and study of DPP. This

would allow further investigations into how DPP operates among cellular regulation networks and subsequently result in the formulation of innovative approaches for the identification and curtailment of the genetically unstable cell's proliferation, thus the pursuit of multiresistant and aggressive cancer phenotypes would also be carried on.

To meet this challenge, the creation of an innovative technology that combines advanced computational methods with biocomputing, machine learning, and pattern recognition algorithms is extremely important. This technology would support medical staff in their analysis of samples, facilitating the discovery of DPP and observing its interactions with cellular regulation networks. Besides, this tech can offer a dependable and effective way of the cell's tracking through time, which is the premise for the identification of DPP. By allowing the recognition of DPP, this technology may reduce the prevalence of genetically unstable cells and thus, deal with the origin of multi-resistant and aggressive cancer phenotypes.

On top of that, the employment of machine learning and pattern recognition tools can digit the boundaries of human observations through decreasing the bias and variability arising from the engagement of multiple people in the sample analysis. Through the implementation of machine learning algorithms, the outcomes can be more constant and dependable thus the cases of misdiagnosis can be diminished while the accuracy of the diagnosis can be increased.

To sum up, the innovation of a technology that applies modern computation methods for cell tracking on bright-field microscopy is certainly fundamental for DPP detection and thus, for the progress of cancer treatment. It can be an effective and dependable way of isolating DPP and thus, gathering knowledge about its interactions with cellular regulation networks. This technology will considerably affect the field of cancer research and therapy, as it will be responsible for the reduction of the proportion of genetically unstable cells and the tackling of the emergence of multiform and aggressive cancer phenotypes.

1.2. State of the Art

The tremendous amount of data generated by various devices globally has fostered a growing necessity to utilize this data to acquire additional information on the world around us, especially in the biological realm. One of the medical data types that are crucial in the understanding and treatment of different diseases is microscopy image data. Nevertheless, the efficient analysis of this data is a challenge that needs the application of diverse statistical, mathematical, and probabilistic tools. Therefore, computational science has turned out to be a becomingly significant resource in our quest to decode biological systems through data processing and comprehension.

In the area of cancer research, there are several models of hypotheses concerning the considerations of tumor size and treatment resistance. Yet, the dynamics of tumor growth and its changes are substantial, because fast-growing tumors are often the most sensitive to chemotherapy. Furthermore, there are models that suggest the sigmoid growth pattern of tumors, where at lower loads, the rate of increase is more, only after some time a plateau state with lower growth rate is reached. The diversity of cancer cells is the main reason for the problem of resisting treatment. Techniques like endogenous barcoding are introducing the potential of monitoring intratumoral heterogeneity, thus giving a better view of cancer. Also, there is a Can mapping the effects of a tumor on both the patient and the tumor actually result in wider antitumor responses? (Vasan et al. 2019) change in this view.

The presence of intratumoral heterogeneity, which manifests as genomic disarray within tumor cells, is a primary reason for the development of resistance to cancer therapy. The high cost of single-cell sequencing techniques has precluded their use as related devices that permit isolating and characterizing individual cells in a mixed population. Apart from genomic disorder, some cancer types get rid of mutation-generating pathways by activation of parallel signaling pathways or through changes in their location of the cells. Thus, cancer treatment approaches have to identify intratumoral heterogeneity as a priority for successful elimination of the disease. (Dagogo-Jack y Shaw, 2018)

The application of Bayesian filtering methods to track nanometer-scale targets in cellular environments has been established to be a reliable, robust and efficient method even in realistic motion conditions. This is especially beneficial in a medical decision process since it provides the consistency of the solution. The multiple hypothesis tracking algorithm is found to outperform its predecessors, with the Jaccard similarity values ranging between -0.05 to 0.5 or even higher. Moreover, these methods have been assessed through the use of **fluorescence microscopy images**, and hence their effectiveness in a practical setting is well supported. (Chenouard et al. 2013)

The heuristic methodologies of tracking cells are typically economically efficient, but the majority of those are greedy. Those often search for a global optimal option by averaging several local optimal solutions. The albumin free single tracking in time-lapse undergoes the whole of that timespan: the particles that are detected during the imaging sequence. The algorithm first links the detected particles in two successive frames. Then links the track segments generated in the first step to simultaneously close gaps and capture particle merge and split events. So despite the assigning being greedy at the beginning of the process, the further track segment assignment is temporally globally optimized. These issues have been so doable from the computational point of view that it is not very likely to be strongly biased on the local optimal (Jaqaman et al. 2008). All things considered, this type of a tracker enhances the performance of tracking while maintaining the performance on the time-steps and has been displaying over **fluorescence microscopy images**.

Heuristic approaches to tracking cells in microscopy images are usually applied due to their computational effectiveness. Nevertheless, these operations are generally greedy, thus they may not offer globally optimal conditions. To resolve this issue, robust single-particle tracking has been implemented tracking cells in alive-cell time-lapse sequences. The methods typically utilize a set of discovered particles measured over specific time intervals, thereby linking between consecutive frames and accounting for merge and split events. This method's computational viability is demonstrated because its result is not significantly biased toward local optimal solutions (Jaqaman et al. 2008). Besides, this method has been previously proven to efficiently track cells in **fluorescence microscopy images** while providing high performance.

Inferring parameters and initial conditions from measurements of single cells using cytometry and fluorescence microscopy is a perplexing challenge. One of the proposed strategies is to apply a maximum likelihood estimator (MLE) accompanied by a single-cell model. The MLE technique utilizes statistical tools to determine the parameters and initial conditions which provide the best fit to the observed data. The single-model, on the other hand, provides a mathematical representation of the population of cells dynamics in the form of a set of ordinary differential equations like:

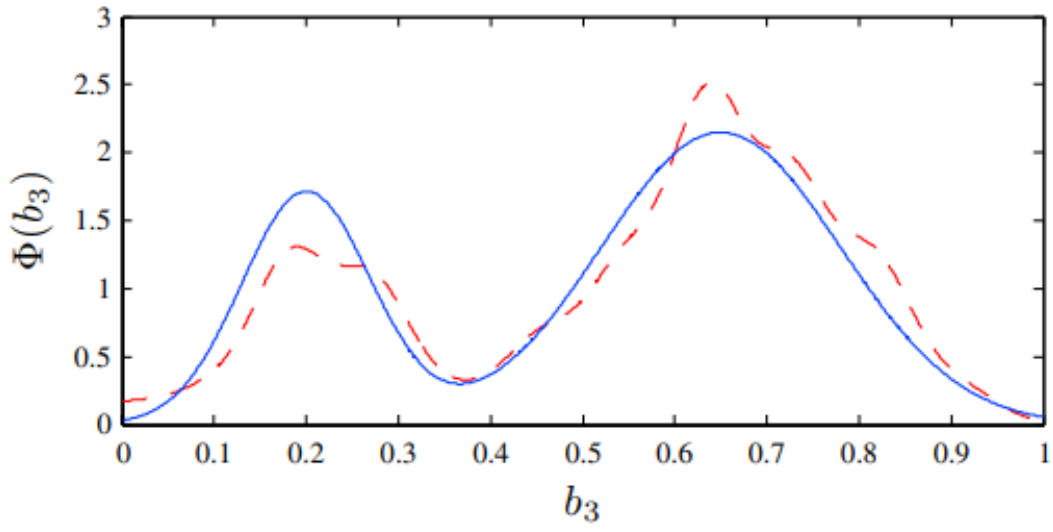
$$x^{(i)} = F(x^{(i)}, p^{(i)}), x^{(i)}(0) = x_0(p^{(i)}), \tag{1.1}$$

$$y^{(i)} = H(x^{(i)}, p^{(i)}), i \in 1, ..., N \tag{1.2}$$

F being a continuous vector, H is the maximum likelihood function.

This approach maximizes the likelihood to assign the parameters and initial conditions of single-cell measurements acquired from cytometry and fluorescence microscopy. The state and output distribution of the population are inferred by means of a single-cell model, which leads to the efficient estimation. The modeling of biochemical reactions networks in a population of N cells can be presented in the form of a system of differential equations. This technique is capable of handling even the most realistic noise scenarios in the models based on the approvals in Hasenauer et al. 2010. The antecedent paper lacks direct scores to show this but it is obvious that normal deviation between the actual population and the observed is slight, rarely higher than a value of 0.5. The above-mentioned technique is notably superior to the other models because it can handle a larger number of cell measurements, unlike those which for necessity have to approximate the measured population response through a density estimator. As a consequence, the previous approach lost some precision.

Figure 1.3: Real parameter distribution vs estimated parameter distribution.



Actual (—) vs. estimated (– –) parameter distribution.

Source: Hasenauer et al. 2010

The first step in the development of cancer therapy is to know the growth rate of a given tumor. For example, by knowing the increase in the size of the tumor, doctors can identify the right chemotherapy drugs, change the doses, or find out if the case needs an operation. Nevertheless, some models can be constructed, but because of their expensive computational overhead, the real-world application often

becomes impossible. On the other hand, statistical methods of busting the problems such as maximum likelihood estimation have a track record of proving their effectiveness, as well as achieving good performance with respect to their low computational costs when applied to **fluorescence microscopy images** (Patmanidis et al. 2019). The form of a model that describes growth dynamics can be expressed in making use of the following mathematical formulation:

$$x_{k+1} = f(_k, \theta_1, \theta_2) + w_k$$

$$y_k = x_k + v_k$$

with:

$$f(_k, \theta_1, \theta_1) = \theta_2 e^{\left(ln\left(\frac{x_k}{\theta_2}\right) e^{\left(-\frac{1}{\theta_1} t_{k+1}\right)} \right)} \tag{1.3}$$

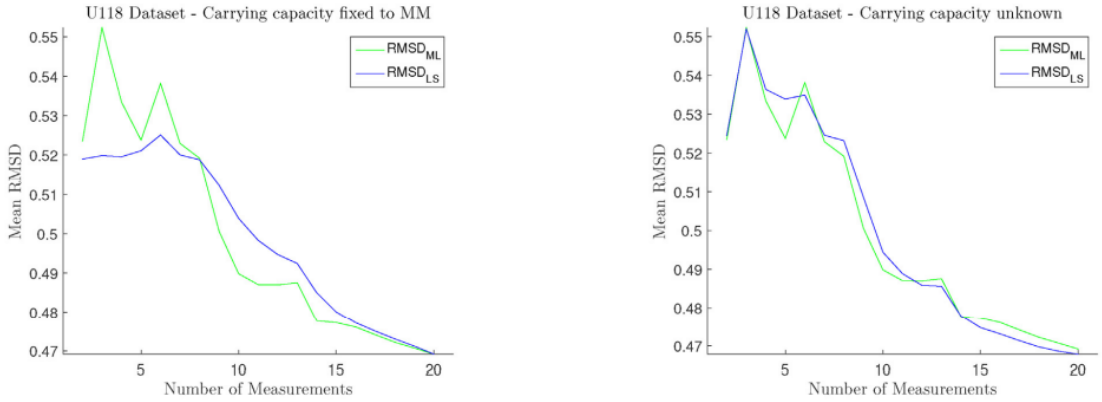
where

θ_1 (days) is related to the cells' proliferative ability

θ_2 (mm3) is the carrying capacity and t_{k+1} (days) is the time between k and k+1

In this case, the authors do not provide specific numerical values for the root mean square deviation (RMSD) between the predicted and observed growth dynamics. However, it is apparent that the maximum likelihood estimator (MLE) model shows a slightly better performance compared to the other model. Despite this, the simplicity of the MLE method is an important advantage, as it results in improved interpretability and lower computational cost. This is particularly valuable in the context of analyzing fluorescence microscopy images, where computational efficiency and interpretability are critical considerations. (Patmanidis et al. 2019)

Figure 1.4: Real parameter distribution vs estimated parameter distribution.



RMSD comparison between MLE and LS Source: Patmanidis et al. 2019

The acquisition of microscopy images via fluorescence techniques is a vital aspect of various natural science fields, particularly in the realm of biological research. However, many image processing techniques are inherently stochastic in nature and require extensive annotation of various events. In certain scenarios, obtaining large amounts of labeled data is infeasible. In such cases, the use of automatic methods becomes necessary. The proposed method, referred to as MAMLE, is specifically designed for fluorescence microscopy images and comprises a combination of seven steps:

- i) Image denoting
- ii) Foreground and background segmentation
- iii) Multi-scale morphological edge detection
- iv) Threshold decomposition and initial segmentation
- v) Shape learning from the initial segmentation
- vi) Likelihood optimization-based splitting
- vii) Maximum likelihood-based merging

The proposed method combines several techniques, including cell shape learning and segmentation, to improve the accuracy of microscopy image processing. This approach, known as MAMLE, utilizes a maximum likelihood function to update the cell shape and can be run iteratively until the shape reaches a stable state. This method was specifically developed for the analysis of fluorescence microscopy images

and aims to address the challenges of processing large amounts of data and the stochastic nature of the underlying processes (Chowdhury et al. 2013). Additionally, the iterative nature of this method ensures that the cell shape is optimized for the specific dataset it was trained on.

The Active Surfaces With and Without Edges method and its generalization for 3-D images is a technique for image segmentation that seeks to partition the image by minimizing the piece-wise constant Mumford–Shah function. While it has demonstrated relatively satisfactory performance in some restricted scenarios, such as separating the foreground from the background, it has been found to be less effective in handling cell clusters. The error rates reported in literature range from 0.038 to 0.12 (Dufour et al. 2005). In order to address these challenges, the Coupling Multiple Active Surfaces variation of the Active Meshes method was proposed. This approach utilizes multiple level set functions to capture more information about the image, thereby enabling improved handling of clustered cells.

The Active Surfaces With and Without Edges method and its generalization for 3-D images seeks to partition the image by minimizing the piece-wise constant Mumford–Shah function for image segmentation. It is showing relatively proper performance on some restricted scenarios, there are some challenges to address like the background, and it is not working well with cell clusters. The error rates goes from 0.038 to 0.12 (Dufour et al. 2005). On Active Meshes, the Coupling Multiple Active Surfaces variation came to overcome some of these issues, primarily the clustered cells issues, in this method, by using more level set functions to capture more information about the image.

The use of 3D active triangular mesh in image segmentation, as a generalization of the Mumford-Shah model, has demonstrated superior performance in comparison to other methods. The error rates, as reported in the literature (Dufour et al. 2011), range between 0.038 and 0.086, indicating high segmentation quality. Furthermore, the utilization of this method is computationally efficient, making it a suitable framework for processing large datasets, especially in the context of fluorescence microscopy images. The application of active meshes and active segmentation algorithms has been shown to be effective in various aspects, including segmentation quality and computational efficiency.

In order to improve cell segmentation and tracking in fluorescence microscopy images, an advanced level-set based algorithm was developed and compared to the traditional mean-shift algorithm. This algorithm presents several enhancements, including improved performance, enhanced separation of initially touching cells, and the ability to handle cell division events. Additionally, the algorithm includes a stopping criterion, a new internal energy of the level-set function, and detection of cells entering the observation frame. According to the results presented in (Dzyubachyk et al. 2010), the

proposed algorithm demonstrates superior performance compared to other existing methods such as DUF05, CellProfiler, and DCellIQ in several scenarios. However, in some cases, the algorithm's performance may decrease with the use of more time steps. Overall, this algorithm presents promising results for fluorescence microscopy image processing.

The clustering algorithm, K-Means, has been employed in the past for tracking cells in fluorescence microscopy images. However, as demonstrated in a comparative study involving 9 related methods, the performance of this approach is inadequate. Specifically, the concordance between the algorithm's output and human identification of pixels associated with cells is poor, in comparison to traditional methods such as OTSU and Maximum Entropy. In this context, the evaluation of the algorithm is primarily qualitative, rather than quantitative, as the output of simply labeled pixels may not be sufficient for identifying cells (Dima et al. 2011). This highlights the importance of utilizing algorithms that are capable of detecting objects, rather than just labeling pixels based on their color, when working with fluorescence microscopy images.

Table 1.1: Segmentation Algorithms comparison.

Table 2. Euclidean distances from ideal segmentation for different segmentation methods and imaging conditions

METHOD	CONDITION 1	CONDITION 2	CONDITION 3	CONDITION 4	CONDITION 5	ALL CONDITIONS
k2	0.664	0.651	0.470	0.599	0.581	0.593
Ot	0.663	0.638	0.355	0.598	0.560	0.563
mx	0.702	0.666	0.130	0.697	0.601	0.559
k3	0.422	0.387	0.255	0.311	0.290	0.333
k4	0.310	0.286	0.174	0.247	**0.216**	0.247
k5	0.237	0.223	0.130	0.221	0.217	0.206
cn	0.256	0.242	0.209	0.312	0.242	0.252
glo	**0.141**	**0.122**	**0.099**	**0.215**	0.283	**0.172**
iso	0.885	0.651	0.422	0.718	0.680	0.671
2nd Man. Seg.	–	–	0.107	–	–	–

Euclidean segmentation distances as determined from the bivariate similarity index plot are reported as a measure of segmentation accuracy. The smallest values (bold) represent the highest accuracy with respect to the reference data. See text or Supporting Information Table 1a for algorithm abbreviations. The last row gives the corresponding result from the second non-reference manual segmentation, which was performed on image condition 3 only.

Source: Dima et al. 2011

Tracking algorithms usually consider using a sequential two-step method for detection and tracking of the moving objects (Jaqaman et al. 2008) (Gustavsson et al. 2022). A major weakness of this arrangement is that the errors during the detection step may get separated and become almost infeasible to rectify in the subsequent tracking stages. In contrast, Conservation Tracking solves this problem by merging the segmentation and tracking stages through which the information is shared and propagated to improve the overall performance (Schiegg et al. 2013). Though more computationally intensive, this method has been proven to outperform classical techniques in fluorescence microscopy contexts.

The non-parametric, statistical algorithm known as mean shift has been utilized in the tracking of scale-space objects, with demonstrated efficacy. Specifically, this algorithm has been utilized in the tracking of blobs based on appearance (Debeir et al. 2005) and even in the tracking of cells under Phase Contrast imaging (Collins, 2003). This approach is primarily based on color, that it's based on color but modules can be added for the texture, contour and others properties (Collins et al. 2003)furthermore, the due its factor for object tracking it. Additionally, this algorithm has been demonstrated to be effective in microscopic imaging scenarios (Collins et al. 2003), with mean error per pixel typically ranging from 10 to 20 when utilizing **phase contrast microscopy images**.

Filters such as Kalman Filter and Particle Filter are considered reliable in Visual Object Tracking; however, they exhibit certain flaws. The Maximum Likelihood Finite Impulse Response (ML-FIR) filter was suggested as a resolution to these uncertainties related to the parameters and the noisy data Min Pak et al. 2016. Nevertheless, it is important to understand that the ML-FIR filter has not been utilized in cell tracking yet. The sift consists of two vectors $s_k = [x\dot{x}y\dot{y}]^T$, indicating the spatial position of a moving object on a plane, while the measurement vector, is defined as $z_k = [xy]^T$. It is grounded on the premise that the measurements and process noises (v, w) are of zero mean while, the measurement models are expressed as:

$$s_{k+1} = As_k + w_k$$

$$z_k = Cs_k + v_k$$

with:

$$A = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.4}$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

where T is the sampling time.

Bipartite graphs are a type of graph that can be used with traditional machine learning algorithms,

and K-partite graphs can also be implemented. The application of these methods to the tracking process can be very helpful as they can deal efficiently with the non-linear relations among the objects that are being processed, thus benefiting from the machine learning capabilities while not losing the pattern recognition strengths of the system. (Na et al. 2020) This is an example of their incorporation of non-linear relationships that object-focused processing can handle better. The State-of-the-art bipartite graphs used in the core engine of Maximum Likelihood, Classification, Clustering, and so on have demonstrated very good results when compared to the traditional methods. In terms of the specifics, the evaluations were performed on three different datasets, and the resulting improvements ranged from 0.05 to 0.5 on the relative error metric. However, it should be noted that cell tracking has yet to incorporate this method.

It is an undisputed fact that tracking multiple objects in video sequences is a very complicated problem, as the number of objects can happen to vary and segmentation errors can exist. Notwithstanding these issues, using the objects and k-partite graph as the core solution has been demonstrated as a powerful approach in overcoming these problems. One of the weaknesses of this method is the hardware that it consumes, but the results obtained from the specific applications are first-rate even with this shortcoming. For example, in Zhong et al. 2014, the application of k-partite graphs was for object tracking which was successfully improved in various different situations. Even though this method has not been tried much for cell tracking, there is a likelihood it could solve some issues in that field.

Multipartite graphs are a form of graph that have gained popularity as a tool in object tracking as indicated by the study of parallelizing a multipartite graph matching algorithm for tracking multiple football players - Siles et al., 2014. It is an approach having flexibility, easy to modify for specific applications, and, it can also be parallelize relatively easy. A bipartite graph on its own may fulfill the role of only one object tracking, whereas a multipartite graph can do the same by presenting more information from the sequent. While this method is not the standard practice in object tracking, it has produced remarkable outcomes with recalls and precision above 0.90 for passes, dribbles, and interceptions around 0.70 for shots. Due to the high optimization potential such as reducing the time by using parallel computing (Villalta et al., 2018), this method may be pursued on the other applications in the future even if it is not used currently in relation to cell tracking.

One of the areas where the tracking algorithm can improve is that it does not offer a statistical evaluation of which features are important for tracking. The current way of determining features

importance is through manual configuration or by the premise that all features are equally important. This is a great shortcoming apparently, given that other algorithms like tracking that use statistical determination to decide the features.

The K-partite graph is a good way to implement multimedia retrieval, which has a further advantage in reconstructing the error content. In doing so, it can be a trapdoor to resolving problems with the tracking items, especially in the presence of gaps between frames where these items go missing. Although not stating clearly, the method has not been applied directly to the cell tracking penguins, it can be inferred that the principles of the method can be used to tackle the task. (Gao et al. 2012). An interesting thing is that as an application of the k-partite the graph in the field of cell tracking this technique can be imagined, however, it has not been used so far.

.

Similarity-based methods

Thresholding, traditionally a prominent image-processing technique, encompasses a full spectrum, from bare threshold values to more sophisticated decision-making. The idea is to utilize multi-channel images, like those recorded on three or more channels, and collect information on different variables per channel to carry out individual processing and treatment. Every channel can have its own image processing technique, and then the results can be fused with weight or heuristics (Hong et al. 2013). This method is especially useful for brightfield images that are not in grayscale as the multichannel RGB images help keep information/separate details through different channels compare. The application of this approach leads to the use of different thresholding algorithms to extract useful information from the images. It is a big plus for the process of image tracking as this method makes the pipeline more flexible and versatile.

A comprehensive assessment of cell morphology is the cornerstone of finding out pathological or cancerous cells, since it is a reliable and non-chemical method. The shape of cells is a widely known factor that pathologists use to diagnose those conditions in the first place, especially in **bright field microscopy images** Kallen y Hornick, 2021. One of the ways to evaluate cell morphology is to extract properties such as Hu's Moment Invariants that tell about the means, variances, and density distributions like mass $m(x; y)$, probability density $p(x; y)$, geometric shapes. Other methods could be elliptic and circular fitting, as well as the distance from the nearest point to the boundary of the cell for cell morphology analysis. The methods that are used for statistical evaluation are

feature extraction, principal component analysis, distributed stochastic neighborhood embedding, and unsupervised classification that are well-known to determine the phenotype of each cell (Bhaskar et al. 2019).

The use of a three-step process through image analysis, namely, background correction using machine learning, thresholding based on maximally stable external regions, and finally cell splitting by means of watershed and fragment merging, has proven to be a successful method for processing **bright field microscopy images**, with a match ratio of around 0.9 (Buggenthin et al. 2013). The employment of an image correction step before the thresholding and object identification process has been shown to lead to an enormous improvement in the cell segmentation, which stresses again the need to carefully consider each step in the image analysis pipeline.

The employment of Principal Component Analysis (PCA) on **bright field microscopy images** for the purpose of correcting illumination is essential in understanding cell behavior and programmed cell death. This is a vital factor that goes a long way in outlining the effects of a number of diseases. Moreover, the adoption of simple techniques like OTSU thresholding subsequently dilation, erosion, and contrast enhancement has played a great role in preprocessing the image and making the segmentation process smoother by avoiding issues related to overlapping cells. Besides, adding an extra cleaning stage post-segmentation can also help in removing unwanted objects as pointed out in (Chiang et al. 2018).

The use of axial information as a feature to classify pixels in a bright field microscopy image has been shown to be effective for several cell types, including E. coli, S. cerevisiae, and epithelial cells, even in mixtures. Principal component analysis (PCA) was performed for image enhancement and, after that, Support Vector Machines (SVMs) were employed for classification. Axial information, as a standalone feature for cell labeling, outweighs even the morphological differences of cells in the study, thus, it is a good candidate for additional research in cell classification (Lugagne et al. 2018).

The use of multi-focal distances, also known as z-stacks, in bright field microscopy images has been demonstrated to be an effective method for constructing a two-dimensional projection of the cell plate with enhanced contrast. This is achieved by measuring the intensity of each stack, and then utilizing variance and mean to create a projection. Additionally, the median deviation from the median of the intensities of all the z-slices for each pixel in the image is used. This hybrid technique allows for the overcoming of certain limitations associated with traditional bright field microscopy (Selinummi et al. 2009) (Lee et al. 2021).

One alternative approach that has been successfully applied in tracking yeast cells is using a classifier to detect the edge probability directly, followed by binarization of the probability and the Hough transform to determine potential cell center candidates. This has been shown to work well across a variety of cell types and could potentially be adapted to tumors through consideration of individual cell factors in **bright field microscopy images**. In this context, the association of the factors could diminish the separation of consequential noise and plant information in a cell (Zhang et al. 2014).

An algorithm that effectively calculates the total area occupied by cells in a **bright field microscopy image** is crucial for the effectiveness of segmentation. This can be done through edge detection and the application of morphological operations, with prior illustrated research (Čepa, 2018). The segmentation process has proved to be a tough and arduous task because of the interference caused by noise, variations in the intensity, and the texture of the cell regions. The concerned step has been the edge detection stage, in which techniques, such as Canny or Sobel, are used in conjunction with morphological operations. They are put through a dilation or erosion process, which enhances the overall accuracy of the cell boundary detection. Thus, this step is core in segmentation as it achieves the clear distinction of cells and the background from each other and identifying each individual in a population.

Dealing with bright field microscopy images has its own drawbacks, as the cells often exhibit low contrast with the background which makes it difficult to segment them accurately. Moreover, the picture manipulation issue is further fostered by the high level of brightness and cell density. Despite these drawbacks, the techniques such as edge detection and morphological operations have been reported effective for the detection of the total area covered by cells in bright field images (Čepa, 2018). Nonetheless, it should be noted that while these methods might work, they are often not as dependable as those in fluorescence microscopy, which if not utilized with caution can yield misleading results. Hence, fluorescence microscopy should predominantly serve as a reference and not the only method for the cell analysis.

The area of bright-field microscopy image processing has been the center of interest for the scientific public, able to elaborate on the non-invasive study of cells, and their behavior in their environment. Thus, allowing one to conduct experiments in pathology, where having knowledge of the kinds of cell behaviors and the malignancies is the first step in their diagnosis and therapy. Still, with bright field microscopy images, the work is even more challenging where low contrast, high levels of brightness, and the presence of many cells frequently complicate the segmenting problem.

The scientific community is continuously searching for good methods and instruments that could be used for bright field microsimage processing.

On a side note, using other techniques like fluorescence microscopy can sometimes create problems such as phototoxicity for the cells under investigation, albeit they are the most efficient. This in turn emphasizes the necessity of developing suitable methods for analyzing bright field microscopy images without requiring the use of fluorescence staining at all.

Several tools have been introduced recently which are aimed at enhancing bright field microscopy image processing. One of the proposed solutions involves edge detection and morphological operations, thresholding, and the incorporation of different focal distances or z-Stacks. Nevertheless, these methods are still in the development stage and they have not been fully optimized for the use in bright field microscopy imaging processing up to this point.

Moreover, the application of machine learning and computer vision techniques like PCA, SVM, and Distributed Stochastic Neighborhood Embedding that are mentioned in the literature point out the possibility of these ideas as a vehicle for the positive development of bright field microscopy image processing. Over and above that, these techniques can be imposed to yield informative morphological features as results and make the cells' decision based on their phenotype.

To sum up, the bright field microscopy imaging processing is a territory that researchers are actively working in, even if it is so difficult that it poses some challenges to deal with and even if it offers room for self-improvement in technology. The cells' surface having less contrast and being overly bright makes it rather difficult to compare images accurately and thus, segment and analyze cells. Nevertheless, due to processed imaging technology, deep learning capabilities, and computer vision, there is now greater potential for, actually, the facilitation of cell analysis through better performance in bright field microscopy images. The area is valuable because it can lead to the progress of diseases like the cancer by enhancing the diagnostics and therapies and thus influencing the health sector. Bright field microscopy, as compared to fluorescence microscopy, might also have a much smaller impact on the cells due to phototoxicity, thus, promoting the accuracy and reliability of the results. To wrap it up, this field of research is not only effective but also essential for the worldwide community, since the potential it has to health care advancements is great and thus it can impact our lives positively.

1.3. Tracking Algorithm

Definition 1. A graph (V,E) with n vertices is referred to as multipartite if there exists a partition of the vertex set, denoted as $P_1, ..., P_p$, such that no edges exist between vertices within the same partition set. (Wieser, 2013):

$$\begin{aligned} &\forall p \in P_i, q \in P_j \\ &\text{with:} \\ &(p,q) \in E \Rightarrow i \neq j \end{aligned} \tag{1.5}$$

The employed graph-based model to illustrate the dynamics of biological systems refers to the frame i as a set of nodes that connect to nodes of the adjacent frame ($i-1$ and/or $i+1$) There is a node in the network that represents each cell and the characteristics of the cell associated with the node include:

- area
- position
- intensity
- circularity
- perimeter
- radius

Cited in (Frome, 2007), the main goal of the procedure is to connect the cells of a temporal series based on their resemblance to each other with the use of a distance function, for example, through the application of a distance function, such as:

- Euclidean Distance
- Manhattan Distance
- Minkowski Distance
- Hamming Distance

For the purpose of carrying out the matching between two node sets that are part of two neighboring images, multiple alternatives can be used. Just one of the often employed methods would be the Hungarian algorithm, as stated in (Date y Nagi, 2016). Besides, it might be possible to take advantage of the N-fold Cartesian product, where the objects are turned to folds, as introduced in (Ezurike, 2016).

$$A1 \times A2 \times ... \times An = \{(a_1, a_2, ..., a_n) \mid a_1 \in A_i \forall 1 \geq i \geq n\}. \tag{1.6}$$

Which time complexity is:

$$O(N^M) \tag{1.7}$$

assuming all folds are the same (which is true for this particular application)

The main challenge associated with utilizing an N-fold Cartesian product for matching between two sets of nodes belonging to consecutive images is the computational complexity that arises from numerous possible pairwise comparisons. To address this issue, several strategies have been proposed, such as pruning objects whose distance exceeds a predefined threshold or imposing constraints on certain features prior to measuring the distance. These approaches have been shown to effectively decrease the number of pairwise comparisons and mitigate computational complexity to:

$$O(N^{M/P}) \; with: \; M/P < P \tag{1.8}$$

The critical point has always been the computational complexity which comes into action due to the availability of multiple pairwise comparisons for matching nodes from two consecutive images, thereby making it an N-fold Cartesian product problem. Proposed solutions to these challenges are manifold, for example, the exclusion of the objects that are further than a certain threshold distance or the requirement of specific features to be indicated before measuring the distance. These mechanisms have been demonstrated to significantly cut down the number of pairwise comparisons and, thus, computational complexity has been reduced to:

Figure 1.5: Hungarian Algorithm Cost Matrix.

Agent	Task							
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
A_1	300	250	180	320	270	190	220	260
A_2	290	310	190	180	210	200	300	190
A_3	280	290	300	190	190	220	230	260
A_4	290	300	190	240	250	190	180	210
A_5	210	200	180	170	160	140	160	180

Each cell represent the cost of a worker to perform a task. Source: Wang et al. 2021

A majority of the existing literature that uses the Hungarian Algorithm for object tracking do not address the complicated nature of the objects being tracked, as it tends to distribute all features of the given object equally missing the presence of unique attributes and dimensions of the objects themselves (Liberti et al. 2014).

$$d\left(p,q\right)=\sqrt{\sum_{i=1}^{n}\left(q_i-p_i\right)^2} \tag{1.9}$$

Source: R&D, 2005

An alternative approach to the current literature is proposed, which takes into consideration the fact that the importance of the features of an object may vary. Instead of using traditional distance metrics, the use of weighted distances is proposed. This approach recognizes that the hidden dimension of the problem, that is, the object itself, has been ignored in previous research (Liberti et al. 2014). It is worth noting that using a weighted distance with a constant vector in which all elements are set to 1 is equivalent to using traditional distance metrics (Liberti et al. 2014).

$$d\left(p,q,w\right)=\sqrt{\sum_{i=1}^{n}w_i\left(q_i-p_i\right)^2} \tag{1.10}$$

To sum up, weighted distance usage in object tracking is an adaptive method as it allows one

to assign different levels of significance of the object's features instead of treating all features equally. On the contrary, the biggest hurdle in this approach is to find the suitable weight values that are both statistically robust and that would boost the performance level. In-depth study is required for evaluating and developing methods for the adaptive statistical adjustment of these weight values, consequently, leading to the superior performance in object tracking.

The goal of our research is to pursue two alternative manners for the regulation of the weights in the formula for the weighted distances in object tracking. The first path would entail the optimization of the weights from the internal viewpoint by treating them as variables and employing the use of the gradient descent-based approach. The second approach entails external optimization of the weights treating them as hyperparameters. The contrast between these two will have ramifications on the way the weights are envisaged and how they are tuned to yield better results.

Gradient descent is an effective optimizer that is very popular in diverse machine learning and deep learning tasks. It is a recurrent algorithm that comes up in the optimal solution by changing in the direction of the steepest descent, a cost function. The algorithm begins with an initial set of parameters, and at each iteration, the algorithm updates the parameters towards the negative gradient of the cost function. The algorithm is iterated until the cost function attains minimum or the stopping condition is satisfied. Various techniques can be used to enhance gradient descent such as momentum, learning rate schedules, and adaptive learning rates. Overall, gradient descent is a widely popular and powerful optimization algorithm that has been addressed with considerable success in vast numbers of the machine learning and deep learning field of studies (Baldi, 1995).

Gradient descent has three major types, and the main distinction among them is the amount of computation they need for calculating the gradient of the target function. On the aspect of the performance, they are a tradeoff between accuracy and update time, i.e., the speed of convergence (Ruder, 2016).

Gradient descent, a classic algorithm, determines the cost function gradient by parameter θ over the whole training dataset. However, this mechanism can be somewhat slow as it requires looping through every data point for a single update making it not a proper fit for big data sets that are out of memory (Dogo et al. 2018).

$$\theta = \theta - n \cdot \nabla_\theta J(\theta) \tag{1.11}$$

Stochastic gradient descent (SGD) is a variant of gradient descent that computes the gradient of

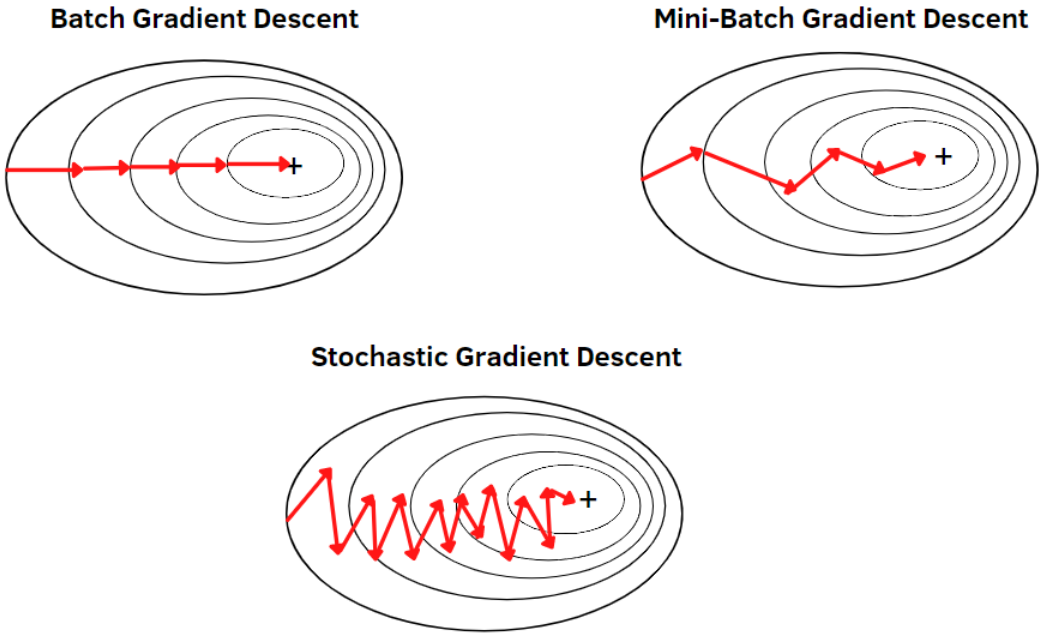
the cost function with respect to the parameters θ for each individual training example, rather than calculating it over the entire dataset as in the classic gradient descent. This approach can lead to faster convergence because it allows more frequent updates of the parameters. On the contrary, the absolute variance of the updates can give causes to oscillations in the objective function thus leading to suboptimal results (Jin et al. 2021).

$$\theta = \theta - n \cdot \nabla_\theta J(\theta; x^{(i)}; y^{(i)}) \tag{1.12}$$

Minibatch gradient descent uses a small number of training examples to compute the gradient of the cost function with respect to the parameters θ. This technique offers the benefits of the speed of stochastic gradient descent but also the reduced variance from the gradation of classic gradient descent. This method has been shown to converge more quickly than the classic gradient descent and produce more accurate results compared to stochastic gradient descent (Qian y Klabjan, 2020).

$$\theta = \theta - n \cdot \nabla_\theta J(\theta; x^{(i:i+n)}; y^{(i:i+n)}) \tag{1.13}$$

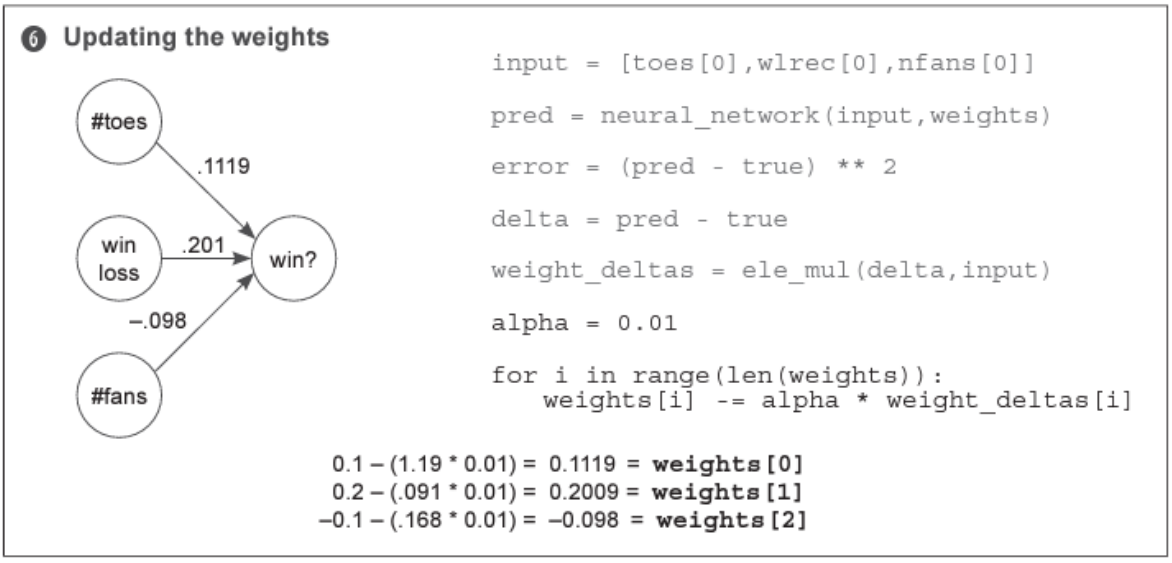
Figure 1.6: Gradient Descent Algorithm Convergence.



The illustrations presented depict the manner in which each algorithm minimizes the error and approaches convergence. It is worth noting that while batch gradient descent appears to be the most straightforward, it is also the slowest. On the other hand, stochastic gradient descent exhibits a high degree of variance due to its per-object update method. Conversely, mini-batch gradient descent represents a compromise between the two aforementioned methods, balancing speed and accuracy.
Source: Wagh, 2022

Given that graphs can be represented mathematically as matrices (as demonstrated in Islam et al. 2022), and neural networks are commonly represented as a series of matrices or tensors with various operations (as outlined in Do et al. 2017), it is logical to attempt to adapt gradient descent techniques commonly used in deep learning to this novel approach of graph-based optimization.

Figure 1.7: Gradient Descent Algorithm explanation.



This figure illustrates the effect of independent weighting on individual features, and how these weights are adjusted for each iteration of the gradient descent algorithm. Source: Trask, 2019

The idea of independently assessing and subsequently adjusting each feature by means of gradient descent algorithms is a very attractive way in image processing. You can benefit from this approach in different ways, for example, by using classic gradient descent, stochastic gradient descent, or mini-batch gradient descent, etc. On the contrary, it should be pointed out that the number of supporting sources and the literature regarding some variations seems to be not so extensive and are to be kept in mind when practicing this method. In addition to that, this method may even be tailored to different strategies, thus allowing for more versatile applications.

Algorithm 1: Gradient Descent Algorithm.

```
def neural_network(input, weights):
  out = 0
  for i in range(len(input)):
    out += (input[i] * weights[i])
  return out

def ele_mul(scalar, vector):
  out = [0,0,0]
  for i in range(len(out)):
    out[i] = vector[i] * scalar
  return out

toes =  [8.5, 9.5, 9.9, 9.0]
wlrec = [0.65, 0.8, 0.8, 0.9]
nfans = [1.2, 1.3, 0.5, 1.0]

win_or_lose_binary = [1, 1, 0, 1]
true = win_or_lose_binary[0]

alpha = 0.01
weights = [0.1, 0.2, -.1]
input = [toes[0],wlrec[0],nfans[0]]
```

(continued)

```
for iter in range(3):

  pred = neural_network(input,weights)

  error = (pred - true) ** 2
  delta = pred - true

  weight_deltas=ele_mul(delta,input)

  print("Iteration:" + str(iter+1))
  print("Pred:" + str(pred))
  print("Error:" + str(error))
  print("Delta:" + str(delta))
  print("Weights:" + str(weights))
  print("Weight_Deltas:")
  print(str(weight_deltas))
  print(
  )

  for i in range(len(weights)):
    weights[i]-=alpha*weight_deltas[i]
```

1

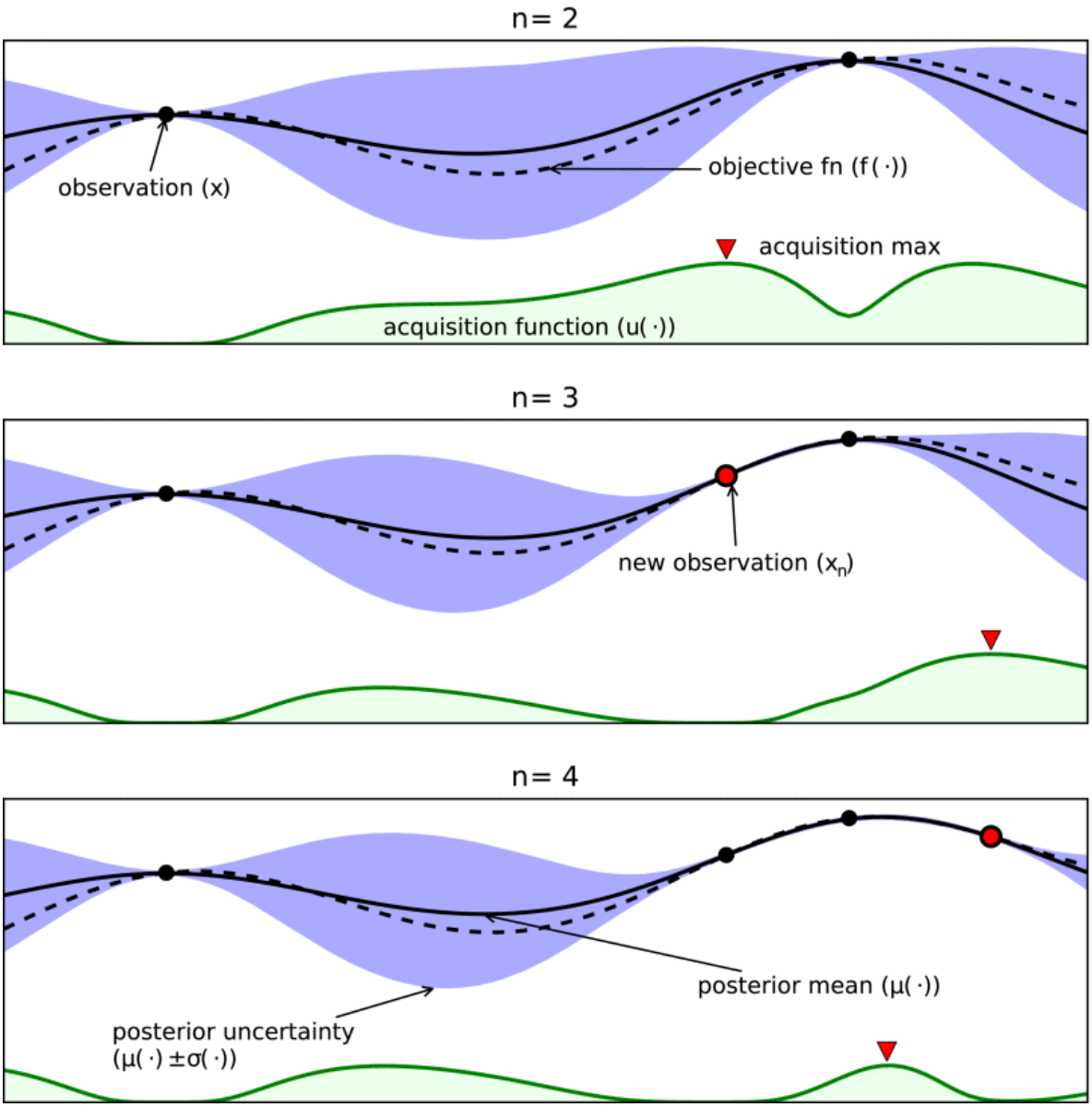
2 the pseudocode for a gradient descent algorithm, which utilizes an approximation of the derivative to calculate the error between the original value and the approximate pseudo-derivative. It is important to note that while this approximation may not be exact, the multiplication factor can be omitted as it linearly increases the weights. The utilization of gradient adjustments may mitigate this issue. Source: Trask, 2019

Instead of using the weights as parameters in the main algorithm, they can be optimized externally by treating them as hyperparameters only. Hyperparameter optimization has been a constant theme in the discipline but has only recently attracted more interest in the field of deep learning. The underlying idea stays the same, which is to pinpoint the essential hyperparameters, familiarize oneself with their hyperparameter space, gauge the performance of the model under different combinations of hyperparameters, reveal the best-operating combination, and finally, present the optimal set of hyperparameters. (Yu y Zhu, 2020)

Bayesian optimization is a method that has been predominantly utilized in the area of global optimization, yet it has been demonstrated to be more effective than any other optimization algorithm that has been established to date on many difficult optimization problems (Jones, 2001). This approach is distinct from others since it creates a probabilistic model of the objective function, f(x), which is then used to make decisions about the next point to evaluate based on this model. The method makes use of all the information obtained from previous evaluations of f(x) to decide where to sample next by considering both the accumulated data and the model's uncertainty. This allows Bayesian optimization

to implement the process of optimization efficiently in a data-lean manner (Snoek et al. 2012).

Figure 1.8: Gradient Descent Algorithm Convergence.



The image depicts a visual representation of Bayesian hyperparameter tuning with three observations. It illustrates the process of how the optimization algorithm learns by considering the parameters in a probabilistic space, without the need for a derivative of the cost function. It can be observed that the algorithm updates its understanding of the parameter space based on new observations, and adjusts its predictions accordingly. The use of a probabilistic model in this approach allows for the incorporation of uncertainty, enabling the algorithm to make more informed decisions and optimize the parameters more effectively. Source: Wagh, 2022

1.4. Problem

Formal definition

The insufficiency of information obtained from a single bright field microscopy image poses a significant challenge in the context of cancer research. Cell tracking is paramount to getting a thorough knowledge of the origin and the continuing process of cancer through cell lineage. This is particularly relevant in the case of the Damage Proliferation Phenomenon (DPP), a phenomenon in which certain cells are able to proliferate despite significant DNA damage, which can result in the emergence of multiresistant and aggressive cancer phenotypes. The difficulty in observing the DPP in a single image highlights the need for a comprehensive analysis of the cell behavior over time.

The target of this project is to invent a cell tracking tool based on bright field microscopy images. This instrument will help researchers to create cell lineages and explore the behavior of cells in a temporal perspective. The cell tracking capability will lead to a more precise understanding of DPP and its relationships with regulation networks. In addition, it will assist the detection and reversal of DPP, by shifting it back to cell apoptosis or cell cycle arrest. Accordingly, the restriction of genetically unstable cell proliferation if accomplished, can directly prevent the emergence of multiresistant and aggressive cancer phenotypes.

Though the primary aim of this project is building a tool for cell tracking, it must be stated that the question of classifying cell phenotypes, which is associated with the cell lineage, is not part of this project. So, even if the tool will not address classification of the various cell phenotypes directly, it will nevertheless serve as a resource for researchers to do further investigation.

1.5. Hypothesis

It is proposed the development of a tracker R, which, for a given set of images $I = \{I_1, I_2, \ldots, I_n\}$ and a set of cells $C = \{c_1, c_2, \ldots, c_m\}$, would be able to assign a feature vector $v(c_i)$ to each cell c_i. The tracker would be capable of identifying and following the individual cells over time by analyzing the specific features of each cell and employing a machine learning model to capture the patterns within these features. The ambition is to achieve the maximum distance between the cells in vector space and the minimum concatenation error between two consecutive images, thus permitting tracking of cells in sequences of images with at least 85 % accuracy.

1.6. Objectives

Main Objective

The objective of this project is to create an efficient cell tracking algorithm that will be able to give precise identification and tracking of single cells throughout a range of bright field microscopy images, by integrating the factors of the cell life span such as apoptosis, mitosis, and the phase of the cell cycle.

Specific Objectives

1. To meticulously analyze the available literature from the scientific and professional fields.
2. To build the validation dataset that will be used for evaluation in the future experiments.
3. To formulate the algorithm which will be able to follow the cells throughout the bright field image series.
4. To verify the experiment with the dataset that has been generated prior.
5. To share the knowledge obtained with the scientific community.

Dataset

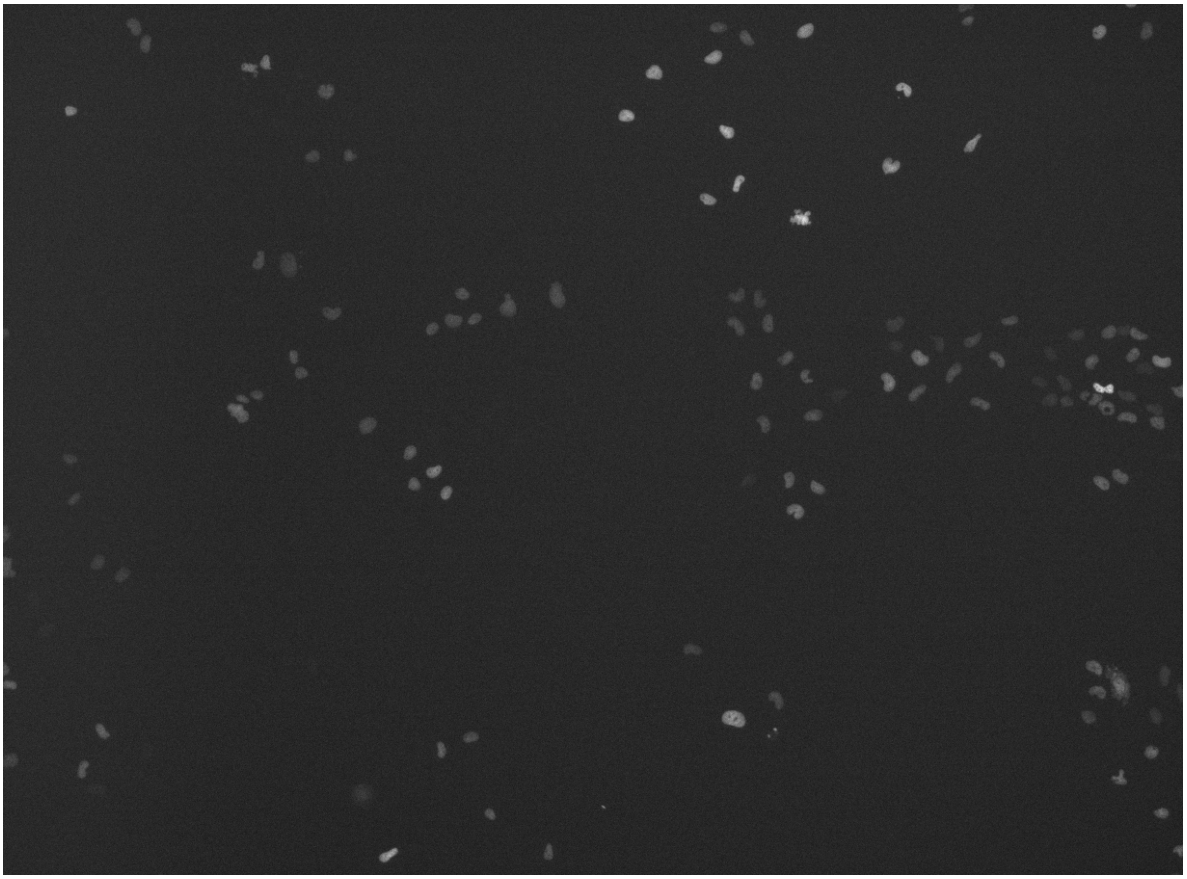
The material used for this work is an image series or time series, which is a dataset that has been made from images taken from identical sources over a time period. This dataset consists of two types of images that are fluorescent microscopy and bright-field microscopy, which were obtained by a Citation 3 microscope with 4x magnification. The images were taken at hourly intervals, with every capture consisting of one fluorescent image followed by a bright-field image, thus creating a pseudo-equivalence between the two. The reason for this is to use the advantages of fluorescence microscopy for checking the movement of cells in the period of time, while the bright-field images only act as the primary dataset for analysis.

A critical point is that with the specific set of cellular events, as in the case of cellular contraction, the fluorescent images capture them before the bright-field images. This is because of the ability of the fluorochrome to mark only the cell nuclei, allowing the exact hour of the cell division to be denoted in the fluorescence image, while it is not in the bright-field images by the time. However, the fluorescence images were not utilized for direct analysis in this work; they were the reference for the mapping and

assisted in the identification of important events such as division to provide accurate ground truth labeling for the bright-field images.

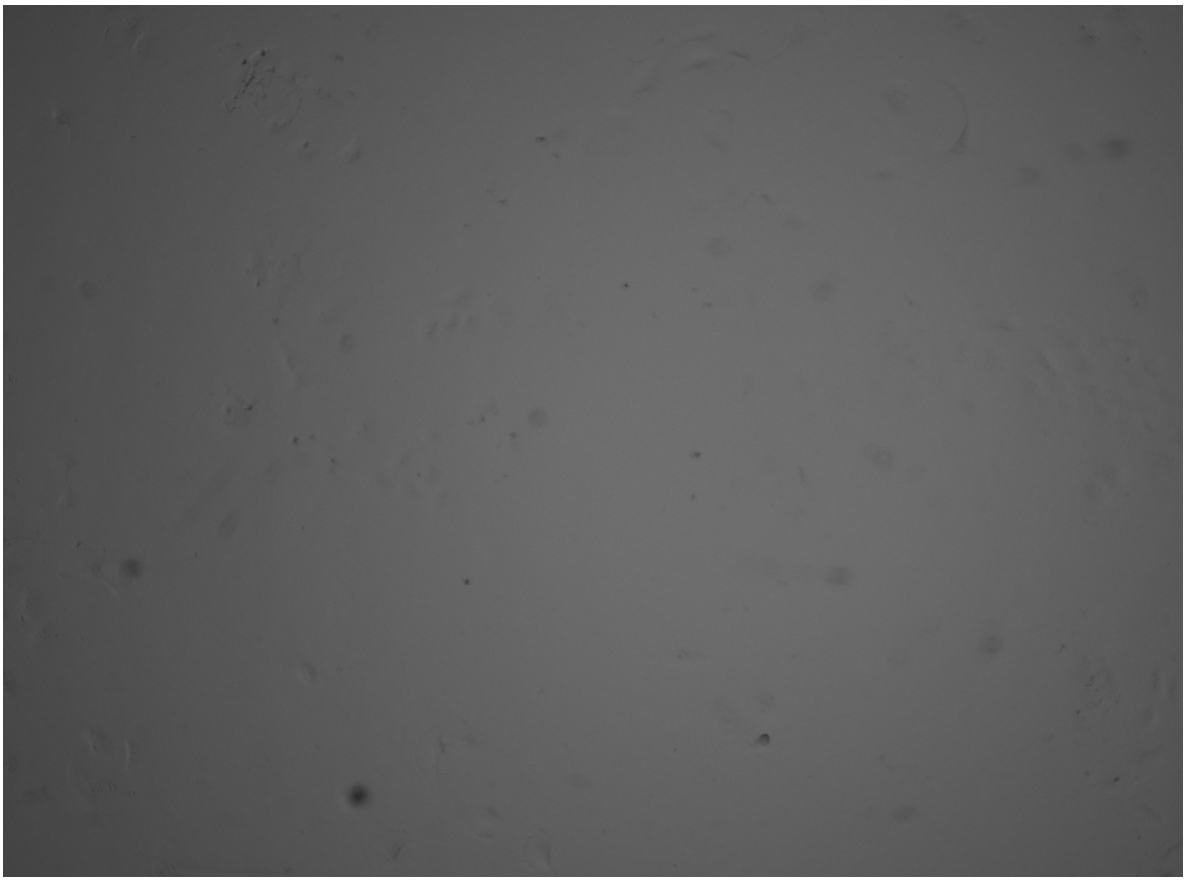
The main objective of this venture is to study the behavioral changes in cells on bright-field microscopy images. The study uses the fluorescence images as a reference for the annotation of the images to ensure the accuracy and reliability of the bright-field tracking algorithm even though certain features may not be as observable in the bright-field microscopy technique alone.

Figure 1.9: Fluorescence microscopy.



Images from fluorescence microscopy have better contrast, and it is easier for machines and the human eye to identify each element and its attributes. This sample is from cancer cells with DPP present.

Figure 1.10: Bright Field.



Images from Bright Field Microscopy have lower contrast than fluorescence, and it is harder for machines and for the human eye to identify each element and its attributes. Also, fluorochromes are genotoxic, so they damage the cell which bias the tests on those. This sample is from cancer cells with DPP present.

Currently, the investigation has 5 datasets, they are comprised of more than 50 images from both fluorescence and bright fields each, and for most of the images. There is one fluorescence image that corresponds to a bright field one, which allows us to have more than 250 images for each kind to complete the task.

1.7. Contributions

The successful creation of the self-tuned object-tracking algorithm for bright-field time-lapse microscopy images was the principal aim of this project, which attributed it to the overcoming of the challenges like contrasting low and noisy images. This project enhances the ways of cell tracking by the adoption of machine learning and optimization technological techniques. Key contributions are

described as follows:

Launch of the new cell tracking algorithm: A machine-learning-based algorithm was proposed for bright-field microscopy and it was the first of its kind. With that, it was possible to track cells in time series data with better accuracy. Bayesian optimization usage: A Bayesian optimization technique was applied to tune the parameters of the algorithm, thus increasing the performance of the tracking in different experimental conditions. Gradient descent techniques integration: The state-of-the-art gradient descent methods were put into application in order to adjust feature weights, so that precision increases and computational load reduces. Real-joke using microscopy data: The algorithm was tried, tested, and validated with several microscopy datasets, thereby proving its practicality and effectiveness in biological research contexts. Tracking without human bias: The algorithm minimized the manual observation errors through the automation of the cell tracking process, thus providing more consistent and reliable results.

Chapter 2

Methodology

2.1. Literature Review

To successfully implement the project, comprehensive analysis and critical research of the existing literature were performed. The section of the problem provides the detailed description of the area of investigation, in brief, this term has to do with tracking cells on bright field microscopy for cancer cells.

The literature review was focused on addressing the following:

- The top performant tracking algorithms used in the State of the Art literature for similar problems.
- The main image processing techniques for bright field microscopy images to be able to extract valuable information from these images.

To ensure a full understanding of the research question, multiple databases were used for the research. This was done in order to consider various alternative methods, perspectives, techniques, etc. The noticeable databases include:

- IEEE Xplore Digital Library
- ACM Digital Library
- SpringerLink
- Elsevier Home
- biorxiv
- arxiv
- medxiv

- AccessEngineering
- ACS Publications
- Alpha Editorial Cloud
- American Institute of Physics (AIP)
- Annual Reviews
- CAB
- EBSCOhost Computers & Applied Science
- EBSCOhost eBook Subscription Academic
- IOPscience
- Knovel
- ProQuest One Academic
- ScienceDirect
- Taylor & Francis Online
- Web Of Science
- Wiley Online Library

The content in those databases has considerable amounts of the actual literature. The content is precise and very well accepted in the academy.

The research was executed using the following keywords (or combinations thereof):

- Microscopy
- Chemosensitivity
- Bright field
- Phenotype
- Cancer

- Cell
- Cell textures
- Object tracking
- Time series
- Pattern recognition
- Image processing
- Machine learning

You can combine any of these generic categories for finding the respective literature. Moreover, it is also not rejected to employ the bibliography of each already approved reference.

The first exclusion criteria utilized as the criteria to discard literature citations were:

- Non-English or Spanish language articles.
- The article does not present pertinent data for the study at hand.

Subsequently, the exclusion criteria for the literature concerned the abstract, followed by the introduction and conclusions, if a piece of literature passes those filters it was read entirely.

The bibliographic research outcomes were included as part of the State of the Art section.

2.2. Validation sets

To produce a substantial quantity of ground truth annotations, we utilized manual methods that were embedded into bright field images while making use of fluorescence images as visual input.

- Must the fluorescent cells in the microscope be segmented manually, automatically, or in a mixed way?
- Which segmentation algorithm (or the combination of the algorithms) should be used?
- How to handle mitosis, apoptosis, and other biological phenomena?

After the validation set creation, there are some things left to be decided, such as:

- How many time steps should have been used?

- How much does the difference between fluorescence and bright field segmentation affect the results?
- How to measure and compensate the difference between bright field and fluorescence segmentations?

2.3. Software infrastructure

Cell tracking is based on the creation of several software modules. These may require different programming languages which even may run on different software architectures, which is why it is necessary to define:

- The programming languages and frameworks.
- The software architecture.
- The pipeline.

The software will be able to run experiments on numerous occurrences as it consists of gathering the images, segmenting them, tracking and scoring the entire experiment, and returning the results. In this pipeline, we could clearly define as:

- Gather the images
- Segment the images
- Track the cells
- Score the experiments

2.4. Pattern Recognition Algorithm

A Pattern Recognition Algorithm was built as a cell-tracking on the bright field images across time based on the mentioned software infrastructure earlier.

The algorithm has to include the following sections:

- Data processing
- Features extraction
- Tracking

2.5. Validation

For the purpose of tracking experiment evaluation, we utilized the validation set that was previously detailed. Additionally, we executed extra validations to ensure that the outcomes were unbiased. At least the following metrics must be evaluated:

- Accuracy
- Loss

2.6. Results Dissemination

In order to disseminate the results of this research, at least the following activities must be carried out:

- Create at least one scientific article in English with the most relevant results of the project.
- All the outcomes from the investigation were uploaded to PRIS-Lab servers (cloud.prislab.org).
- Ideally, lectures were being made at conferences or university classes.

2.7. Data

Multiple datasets are available for analysis, wherein the majority of these datasets are available in the fluorescence microscopy-based and bright field-based formats. Furthermore, to demonstrate the concept of the algorithm, a Dataset generator was developed, the dataset generated synthetic cell data of which certain biology phenomena were modeled.

The phenomena that were taken into account to create these synthetic cells are:

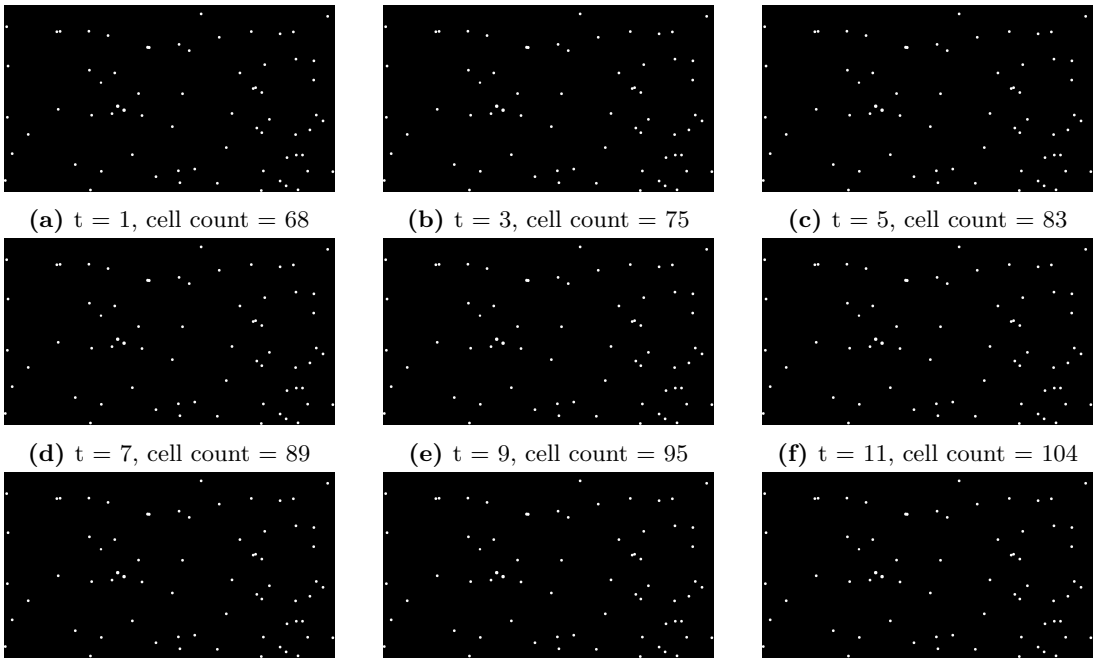
- Cell area.
- Cell growth through time.
- Cell perimeter.
- Cell perimeter changes through time.
- Cell circularity.
- Cell shape.

- Cell shape changes through time.
- Cell color intensity.
- Cell color intensity changes through time.
- Cell M phase.

Also, it takes into account some other pieces of data that are needed to create the data model, such as:

- Cell position (uniformly random generated).
- Cell microscopy image size (fixed).

As the data generator is exclusively designed for modeling purposes, the generated outcomes were not considered as authentic, a normal distribution was assumed for all modeled biological phenomena. We only require the abovementioned features; hence, the generation of images is not necessary, but it was also made for debugging reasons. It is significant to keep in mind that the images are just representing the cells as points with their location, not more.



(a) t = 1, cell count = 68 **(b)** t = 3, cell count = 75 **(c)** t = 5, cell count = 83

(d) t = 7, cell count = 89 **(e)** t = 9, cell count = 95 **(f)** t = 11, cell count = 104

(g) t = 13, cell count = 115 **(h)** t = 15, cell count = 127 **(i)** t = 17, cell count = 135

Figure 2.1: Started with 68 cells. (a) The number of cells increased to 75 (c) Some cell movement is observable (e). The cell count in (h) is almost twice as (a)

Algorithm 2: Data generation algorithm

Result: Annotated K-Partite Graph

```
mpg ← [ ];
for epoch in epochs do
    graph ← [ ];
    for cell in cells do
        x, y, radius, perimeter, area, intensity = modifyRandomCell(cell);
        if area >MAXAREA then
            area1, area2 = divideRandomCell(random);
            x, y, radius, perimeter, area, intensity = modifyRandomCell(cell);
            graph.append(x, y, radius, perimeter, area1, intensity, cell);
            x, y, radius, perimeter, area, intensity = modifyRandomCell(cell);
            graph.append(x, y, radius, perimeter, area2, intensity, cell);
        end
        else
            graph.append(x, y, radius, perimeter, area, intensity, cell);
        end
    end
    mpg.append(graph)
end
```

The parameters used to create the previous images (Figure 2.1) were:

- INITIAL POPULATION = 64
- EPOCHS = 50
- MAX FRAME SIZE X = 1920
- MAX FRAME SIZE Y = 1080
- CELL AREA = 228
- CELL AREA MU = 16
- CELL GROWTH = 1

- CELL GROWTH MU = 0.5
- CELL MOVEMENT = 0
- CELL MOVEMENT MU = 4
- MPHASE MOVEMENT = 0
- MPHASE MOVEMENT MU = 16
- CELL INITIAL CIRCULARITY = 0.7
- CELL INITIAL CIRCULARITY MU = 0.1
- CELL DEFORMATION = 0.1
- CELL DEFORMATION MU = 0.1
- CELL INITIAL INTENSITY = 128
- CELL INITIAL INTENSITY MU = 16
- CELL COLOR CHANGE=16
- CELL COLOR CHANGE MU=4
- MAX AREA = 256

The Multipartite Graph Tracker (MPGT)

The Multipartite Graph Tracker (MPGT) is a compound algorithm, based on the tracking objects in a sequence of images across multiple frames, that is particularly useful in the context of the tracking cell behavior in bright field microscopy images. This tracker uses a multipartite graph structure where every layer of the graph shows a different frame in the sequence and the nodes represent the segmented objects in those frames.

Formalization of MPGT

Graph Construction: Let $G = (V, E)$ be a multipartite graph where $V = \bigcup_{t=0}^{T} V_t$ and $E = \bigcup_{t=0}^{T-1} E_t$. Here, T is the total number of frames.

- Each $V_t = \{v_{t,i}\}$ represents the set of nodes (objects) detected in frame t, where $v_{t,i}$ denotes the i-th object in frame t.
- Each $E_t = \{(v_{t,i}, v_{t+1,j})\}$ represents the set of edges between objects in consecutive frames t and $t+1$, indicating potential matches.

Feature Extraction and Similarity Calculation: For each pair of objects $(v_{t,i}, v_{t+1,j})$ in consecutive frames, a set of features is extracted. Let $f_{t,i}$ and $f_{t+1,j}$ be the feature vectors for objects $v_{t,i}$ and $v_{t+1,j}$, respectively.

- The similarity score $S(v_{t,i}, v_{t+1,j})$ is calculated using a weighted sum of feature differences:

$$S(v_{t,i}, v_{t+1,j}) = \sum_{k=1}^{K} w_k \cdot |f_{t,i,k} - f_{t+1,j,k}|$$

  where K is the number of features, w_k is the weight for the k-th feature, and $f_{t,i,k}$ is the k-th feature of object $v_{t,i}$.

Optimal Matching with the Hungarian Algorithm: The goal is to find the optimal matching between objects in consecutive frames that minimizes the total similarity score. This is achieved using the Hungarian algorithm.

- Let C be the cost matrix where $C_{i,j} = S(v_{t,i}, v_{t+1,j})$. The Hungarian algorithm finds the matching $M \subseteq E_t$ that minimizes the sum of costs:

$$\text{mín} \left(\sum_{(v_{t,i}, v_{t+1,j}) \in M} C_{i,j} \right)$$

Weight Optimization with Genetic Algorithms: The weights $w = \{w_1, w_2, \ldots, w_K\}$ are optimized using a genetic algorithm to improve the accuracy of object matching.

- The genetic algorithm iteratively updates the weights to maximize the overall match rate. The fitness function for the genetic algorithm is defined as:

$$\text{Fitness}(w) = \frac{\text{Number of Correct Matches}}{\text{Total Number of Matches}}$$

Parameter Limits with Bayesian Optimization: To further refine the tracking process, Bayesian optimization is used to find the optimal parameter limits for the tracking algorithm. This involves setting upper and lower bounds $L = \{L_1, L_2, \ldots, L_K\}$ and $U = \{U_1, U_2, \ldots, U_K\}$ for each feature.

- The objective of Bayesian optimization is to find L and U that maximize the match rate while ensuring the tracking process remains within realistic limits. The optimization problem can be formulated as:

$$\text{máx}(\text{Match Rate}(L, U))$$

subject to:

$$L_k \leq f_{t,i,k} \leq U_k \quad \text{for all } k \in \{1, \ldots, K\}$$

Tracking Consistency and Error Correction: The MPGT has built-in checks to ensure that it keeps a consistent track over time. In the case of an object that disappears and then for some reason reappears in the later frames, the algorithm is able to reconnect the correspondence that is, it can correct the potential tracking errors.

- By using the historical data of tracked objects and applying this to make decisions about re-establishing matches, the object will be recognized and matched.

Handling Complex Scenarios: The tracker is also fine-tuned to deal with complex situations like cell division cases. When an object separates into two or even more new objects, the algorithm detects the event and keeps track of the new objects as planned.

Performance Optimization: Algorithm decreases the computational complexity and introduces efficient data structures and parallel processing to deal with large datasets and high-throughput image sequences with little or no performance degradation.

By applying these strategies, including weight optimization and parameter limits, the MPGT stands as a formidable instrument for researchers to study dynamic biological processes, delivering accurate and reliable tracking of objects in time-lapse image sequences.

Multipartite Graph Tracker (MPGT)

Crafted for cutting-edge research, the Multipartite Graph Tracker (MPGT) is a state-of-the-art algorithm designed to track objects across multiple frames in a sequence of images, which is notably

beneficial for the analysis of cell behavior in bright field microscopy images. This tracker follows a multipartite graph structure, where each layer of the graph is a different frame in the sequence, and the nodes are the segmented objects in those frames.

Formalization of MPGT

Graph Construction: Let $G = (V, E)$ be a multipartite graph where $V = \bigcup_{t=0}^{T} V_t$ and $E = \bigcup_{t=0}^{T-1} E_t$. Here, T is the total number of frames.

- Each $V_t = \{v_{t,i}\}$ represents the set of nodes (objects) detected in frame t, where $v_{t,i}$ denotes the i-th object in frame t.
- Each $E_t = \{(v_{t,i}, v_{t+1,j})\}$ represents the set of edges between objects in consecutive frames t and $t+1$, indicating potential matches.

Feature Extraction and Similarity Calculation: For each pair of objects $(v_{t,i}, v_{t+1,j})$ in consecutive frames, a set of features is extracted. Let $f_{t,i}$ and $f_{t+1,j}$ be the feature vectors for objects $v_{t,i}$ and $v_{t+1,j}$, respectively.

- The similarity score $S(v_{t,i}, v_{t+1,j})$ is calculated using a weighted sum of feature differences:

$$S(v_{t,i}, v_{t+1,j}) = \sum_{k=1}^{K} w_k \cdot |f_{t,i,k} - f_{t+1,j,k}|$$

  where K is the number of features, w_k is the weight for the k-th feature, and $f_{t,i,k}$ is the k-th feature of object $v_{t,i}$.

Optimal Matching with the Hungarian Algorithm: The goal is to find the optimal matching between objects in consecutive frames that minimizes the total similarity score. This is achieved using the Hungarian algorithm.

- Let C be the cost matrix where $C_{i,j} = S(v_{t,i}, v_{t+1,j})$. The Hungarian algorithm finds the matching $M \subseteq E_t$ that minimizes the sum of costs:

$$\min_{M} \sum_{(v_{t,i}, v_{t+1,j}) \in M} C_{i,j}$$

Weight Optimization with Genetic Algorithms: The weights $w = \{w_1, w_2, \ldots, w_K\}$ are optimized using a genetic algorithm to improve the accuracy of object matching.

- The genetic algorithm iteratively updates the weights to maximize the overall match rate. The fitness function for the genetic algorithm is defined as:

$$\text{Fitness}(w) = \frac{\text{Number of Correct Matches}}{\text{Total Number of Matches}}$$

Parameter Limits with Bayesian Optimization: To further refine the tracking process, Bayesian optimization is used to find the optimal parameter limits for the tracking algorithm. This involves setting upper and lower bounds $L = \{L_1, L_2, \ldots, L_K\}$ and $U = \{U_1, U_2, \ldots, U_K\}$ for each feature.

- The objective of Bayesian optimization is to find L and U that maximize the match rate while ensuring the tracking process remains within realistic limits. The optimization problem can be formulated as:

$$\max_{L,U} \text{Match Rate}(L, U)$$

  subject to:

$$L_k \leq f_{t,i,k} \leq U_k \quad \text{for all } k \in \{1, \ldots, K\}$$

Tracking Consistency and Error Correction: The main functional unit of the MPGT is the set of rules to keep tracking consistency over time. If after some contradictory observations the algorithm restores the links, it is possible to correct potential tracking errors.

- The key to this is to keep a historical record of tracked objects and use these data to make confident decisions on re-establishing matches.

Handling Complex Scenarios: The tracker is customized to cope with complicated scenarios such as cell division events. Once an object partitions into several other objects, the algorithm detects this situation and tracks the newly appeared objects accordingly.

Performance Optimization: To deal with the computational complexity, the algorithm features the efficient data structures and a parallel processing routine, which consequently allows it to receive a vast amount of information from high-throughput imaging without significant performance degradation.

Furthermore, the use of different techniques such as the optimization of weights and limits of parameters makes the MPGT a very powerful tool for researchers to study departures in biological systems through time, providing object tracking with maximum fidelity and reliability.

Genetic Algorithms Overview

Simulating the process of natural evolution, genetic algorithms operate. They initially have a population of the individuals that are possible solutions and then, through the process of evolution, the individuals create better solutions. The hallmark stages of a genetic algorithm are:

Initialization: This is the stage where individuals, which are randomly generated, make up a population. An individual is the representation of possible solutions as a chromosome (a list of parameters).

Selection: Fitness scores are used to choose the best individuals out of the population. These are metrics that evaluate the performance of the individual in implicating the question under study. Among the methods used for selection are tournament selection, roulette wheel selection, and rank-based selection.

Crossover (Recombination): The selected individuals (parents) form pairs to exchange the segments of their chromosomes and create offspring. Crossover, which is the process of exchanging parts of chromosomes, is an introduction to the genetic variability and helps in a further exploration of the search space. Some of the most commonly used crossover techniques include single-point, multi-point, and uniform crossover.

Mutation: The mutation acts on offspring whereby changes, which are random and new, are applied to the parent chromosomes. The migration of genetic diversity is a very early stage of stopping a premature convergence to suboptimal solutions. Mutation can be as simple as flipping a bit or adding noise to genetic sequences based on probability distributions.

Evaluation: The performance level of new chromosomes (offspring) is assessed by the achievement of tasks.

Replacement: The new individuals form the next generation by replacing some or all of the old population. This repeats for a predefined number of generations or until the solution is acceptable.

Genetic Algorithms in This Research

In this work, genetic algorithms are being used to find the best weight set for tracking parameters in bright field microscopy images. The steps involved are:

Data Preparation: The annotation tool has the capacity of segmenting cancer cells and tracking them through various images. The data structure that will contain the information includes features such as radius, centroid position, intensity, and shape descriptors, which are collected into a DataFrame.

Initialization of MultiPartiteGraph: With the help of a MultiPartiteGraph, a dataset is initialized with the number of unique epochs (time points). Each epoch is attached to a graph that comprises the nodes that show the segmented cells and their features.

Node and Graph Reconstruction: Nodes are reconstructed from the DataFrame, and the parent-child relationships are established based on the tracking data. This one structure represents the both lineages and migration of cells.

Evaluation Function: The evaluation function quantifies the fitness of each individual based on the precision of the tracking matches. The parameters targeted here, radius, centroid position, intensity, and shape descriptors, are optimized as a means to achieve the desired tracking performance.

Genetic Algorithm Execution: The genetic algorithm walks through the main steps of selection, crossover, mutation, and evaluation. Individuals that get higher fitness scores are the ones that usually will be represented in the next generation, consequentially guiding the population to better solutions.

Optimization and Results: The optimum parameters are selected after multiple generations. With the improved parameters, the cell tracking in bright field microscopy images has achieved a higher level of accuracy and reliability.

Greedy Approach with Exponential and Binary Search for Parameter Limits

Greedy methods may be viewed as a more effective complement to genetic algorithms in optimization as we experienced while tracking parameters. Instead, the hyperparameter space was difficult to deal with; the main challenge it posed was its complexity. To tackle this, a quickly alternative method was employed for rough tuning of the parameter limits. The algorithm used in this case is a

combination of the binary and the exponential search that was used to determine the optimal bounds for every parameter.

Exponential Search

The exponential search was the most rapid initial choice to find a parameter limit range where it probably lies. The starting value was very small and was used in a small step size of approach, but with each iteration, we increased the limit until the unmatch rate, which measures performance, was over a fixed number. Thus we could narrow the range of the parameter space down to more suitable values for the next step of the adjustment.

Algorithm 3: Exponential Search

Result: Optimized Parameter Limit

```
Initialize the parameter limit with a small value;
while unmatch_rate ≤ threshold do
    limit = limit × exponential_factor;
    evaluate(limit);
end
```

Binay Searh

As soon as a potential range was traced with an exponential search, binary search was used to set the limits with more exactness. The binary search approach was applied to the division of the span into halves and to the evaluation of the performance at the halfway point done repeatedly. This loop procedure went on till the difference between the upper and lower bounds became minor, thus guaranteeing a more exact determination of the optimal limit.

Algorithm 4: Binary Search

Result: Optimized Parameter Limit Range

```
Define the upper and lower bounds of the parameter limit range identified by the exponential search;
while (upper_bound - lower_bound) > epsilon do
    midpoint = (upper_bound + lower_bound) / 2;
    if evaluate(midpoint) < threshold then
        lower_bound = midpoint;
    end
    else
        upper_bound = midpoint;
    end
end
```

Evaluation Criteria

The bound evaluation of each parameter was derived from the unmatch rate that measures the untracked elements. By keeping the match rate high, the major objective was to minimize the unmatch rate. The parameters analyzed were radius, position (x and y coordinates), centroid position, intensity measures, and shape descriptors like eccentricity, perimeter, and circularity.

Challenges and Observations

Even though there was a methodical way, several issues arose:

High Feature Correlation: A number of features were identified as being highly correlated, and thus, it was hard to find the proper settings for individual parameters. This mutual dependence often resulted in suboptimal values where changing one parameter unfavorably affected others.

Unmatch Rate Variability: The unmatch rate significantly fluctuated, and it was often the case that many combinations caused a low match rate. This variance made it hard to single out the stable and reliable parameter limits.

Complex Parameter Interactions: Complex interplays hindered the optimization work of different parameters a lot. The modification of one parameter compensatively dictated another one, which resulted in a dynamic and unstable optimization landscape.

Conclusion of the Greedy Approach

The exponential and binary search methods availed of a structured way to explore the field of parameters. However, the strong feature correlation and the ensuing match rate variability made this technique less effective. Even after observing the rigorous attempts to adjust the parameters, the majority of the combinations could not provide satisfactory results, which threw light on the complexity and the interrelationship of the tracking parameters in bright field microscopy images.

Bayesian Optimization for Parameter Tuning

After struggling with genetic algorithms and greedy approaches to tuning parameter tracking, we opted for Bayesian Optimization (BO) as a fine-tuning limits strategy. The section describes Bayesian Optimization's principles, benefits, and the reason for choosing this approach for the research.

Bayesian Optimization Overview

Bayesian Optimization is an efficient tool to maximize black-box functions which are costly to evaluate. This is particularly significant where the function has no known analytical shape, and each evaluation entails a cost of time or other resources. The central point of Bayesian Optimization is to create a probabilistic model of the objective function and then employ it for decisions around where to evaluate the function next.

How Bayesian Optimization Works

1. **Surrogate Model:**
    - Causative Bayesian Optimization is a paradigm that relies on a Gaussian Process (GP) as a surrogate model, which is a very good method of estimating the objective operation. The proxy model allows you to calculate the value of the function, which is assigned to a probability that represents the uncertainty.
    - The Gaussian Process is characterized by its mean and covariance functions, which makes it possible to represent the uncertainty in the predictions with it. The Mean function shows

how much value is expected of the objective function, on the other hand, the Covariance function expresses the correlations between the points in the search space.

$$f(x) \sim \mathrm{GP}(m(x), k(x, x'))$$

where:

- $m(x)$ is the mean function.
- $k(x, x')$ is the covariance function.

2. **Acquisition Function:**

   - As you may know, the primary goal of the acquisition function is to determine the next best location for sampling. In doing so, it is necessary to weigh the benefits of both exploration (for example, sampling in areas with the highest model uncertainty) and exploitation (that is, sampling at places with the highest model predictions). The most frequently used acquisition functions for Bayesian optimization are Expected Improvement (EI), Probability of Improvement (PI), and Upper Confidence Bound (UCB)". .
   - The present work was conducted using the Expected Improvement (EI) acquisition function. EI is a criterion that assesses the mean and uncertainty of the surrogate model's predictions, to measure the expected value of improvement of the best observation.

$$\mathrm{EI}(x) = \mathbb{E}[\mathrm{máx}(0, f(x) - f(x_{\mathrm{best}}))]$$

3. **Iterative Process:**

   - Bayesian Optimization is a mechanism that works through an iterative process. In the course of each iteration, the surrogate model is refreshed with additional data points (parameter evaluations), and the acquisition function is optimized in order to select the next point to evaluate.
   - First, the chosen point is assessed by the real objective function, and the statue; is incorporated into the dataset. This cycle continues until a predefined stopping criterion is reached, which might be either a maximum number of iterations or a full convergence to an optimal solution.

$$x_{\text{next}} = \arg\max_{x} \text{EI}(x)$$

Advantages of Bayesian Optimization

Bayesian Optimization offers several advantages that make it suitable for optimizing tracking parameters in this research:

- **Efficiency:** Bayesian Optimization is efficient in terms of the number of function evaluations required. By using the surrogate model to guide the search, it can find optimal solutions with fewer evaluations compared to other optimization methods.
- **Handling Expensive Evaluations:** Because it minimizes the number of evaluations needed, BO is well-suited for problems where each evaluation is time-consuming or costly.
- **Probabilistic Framework:** The use of a probabilistic model allows BO to quantify uncertainty in the predictions, leading to more informed and effective exploration of the search space.
- **Adaptability:** BO can handle a wide range of optimization problems, including those with noisy or non-differentiable objective functions.

Application of Bayesian Optimization in This Research

In this study, Bayesian Optimization was employed to find the optimal limits for the tracking parameters. The process involved the following steps:

1. **Initialization:**
    - The parameter bounds were defined based on initial experiments and domain knowledge. These bounds included limits for various features such as radius, position, centroid coordinates, intensity measures, and shape descriptors.
2. **Surrogate Model and Acquisition Function:**
    - A Gaussian Process was used as the surrogate model to approximate the objective function. The Expected Improvement (EI) acquisition function guided the selection of new points for evaluation.
3. **Optimization Process:**

- Bayesian Optimization iteratively sampled the parameter space, updating the surrogate model with each new evaluation. The optimization aimed to maximize the match rate and minimize the unmatch rate of the tracking algorithm.
- The process involved an initial set of random evaluations (*init_points*) followed by a specified number of iterations (*n_iter*) to refine the parameter limits.

Algorithm 5: Bayesian Optimization

Result: Optimized Parameter Limits

for *i in 1 to (init_points + n_iter)* **do**

  $x_{\text{next}} = \arg\max_x \text{EI}(x)$;

  $y_{\text{next}} = f(x_{\text{next}})$;

  update_GP(x_{next}, y_{next});

end

Results of Bayesian Optimization

The use of Bayesian Optimization brought about considerable advancement in the tracking performance. The suggested limits led to achieving a higher match rate and lower unmatch rate in comparison to previous techniques. This was a clear illustration of the Bayesian optimization's ability to manage the complicated and tightly interdependent mode of the tracking parameters' behavior.

To sum up, Bayesian optimization was adopted primarily due to its efficient parameter space exploration capacity and its successful use in the optimization of expensive and complex objective functions. The good results achieved in this study prove its capacity to enhance tracking parameters in bright field microscopy images considerably. This is the essential method in which the overall approach to the enhancement of the accuracy and the reliability of cell tracking in cancer research is made.

Limitations

The test results highlighted a number of issues that affected the proper performance of the tracking algorithm. These include problems associated with the precision of the segmentation, the manual annotation, the field of vision, afterimage displacement, inconsistent recall strategy, the effectiveness of the genetic algorithm, and the validation of the segmentation algorithm for tracking applications.

1. **Segmentation Accuracy:**

- In most cases, the segmentation algorithm was not able to identify all cells in one frame, thus there were cell types that were present in the frame but were not detected by the software despite being visible to the human eye.
- Mainly in instances where the cell was partially hidden or the contrast was weak, the algorithm frequently detected only parts of the cells instead of the whole cell body.
- The tracking algorithm's perception of cell motion and had been misplaced by the inaccurate lineage information and erroneous feature computations due to the errors in segmentation.

2. **Manual Annotation:**

   - The dependence on manual assessment to confirm and rectify the segmentation and tracking results was a backbreaking task and it took a great deal of time.
   - The activity of manual annotation is dealt with human errors, which can potentially bring bias or inaccuracy to the results. As a consequence, tracking data can be inconsistent.

3. **Limited Field of View:**

   - The photos were taken with a central view of the Petri dish as a focus, which made it possible to see only a small area and not the entire sample space.
   - Cells at the boundaries of the frame might disappear or appear by the way of light, which, in turn, would make the tracking algorithm handle the continuity across the frames more difficult.

4. **Abrupt Image Changes:**

   - The bright field microscopy films experienced the rapid change between frames which was characteristic of abrupt or direct changes i.e. imaging process or biological variations.
   - The tracking continuity was disrupted by these sudden changes, which resulted in mismatches or loss of tracking data.

5. **Unmatched Recall Strategy:**

   - The unmatched recall strategy filtered objects with attribute values beyond a certain specific limit in order to improve tracking precision, but it also resulted in a higher untracked rate.
   - This strategy underscored the dilemma between precision and comprehensiveness, since pertinent links between cells in multiple frames could be veiled.

6. **Performance of the Genetic Algorithm:**

   - The hyperparameter space was so complex that the genetic algorithm ran into difficulties, which was primarily attributed to the high correlations that occurred between the features and the corresponding difficulty in isolating the individual parameter effects.
   - The algorithm frequently came to local optima, causing unacceptably low to the tracking problem.

7. **Validation of Segmentation for Tracking:**

   - Although the segmentation algorithm was validated for general detection, it was not proven for tracking performance specifically, thereby leading to possible errors in the tracking process.
   - The tracking applications not being strictly validated may have been the source of the inconsistencies, especially in situations where cells were partially detected or not detected at all.

Chapter 3

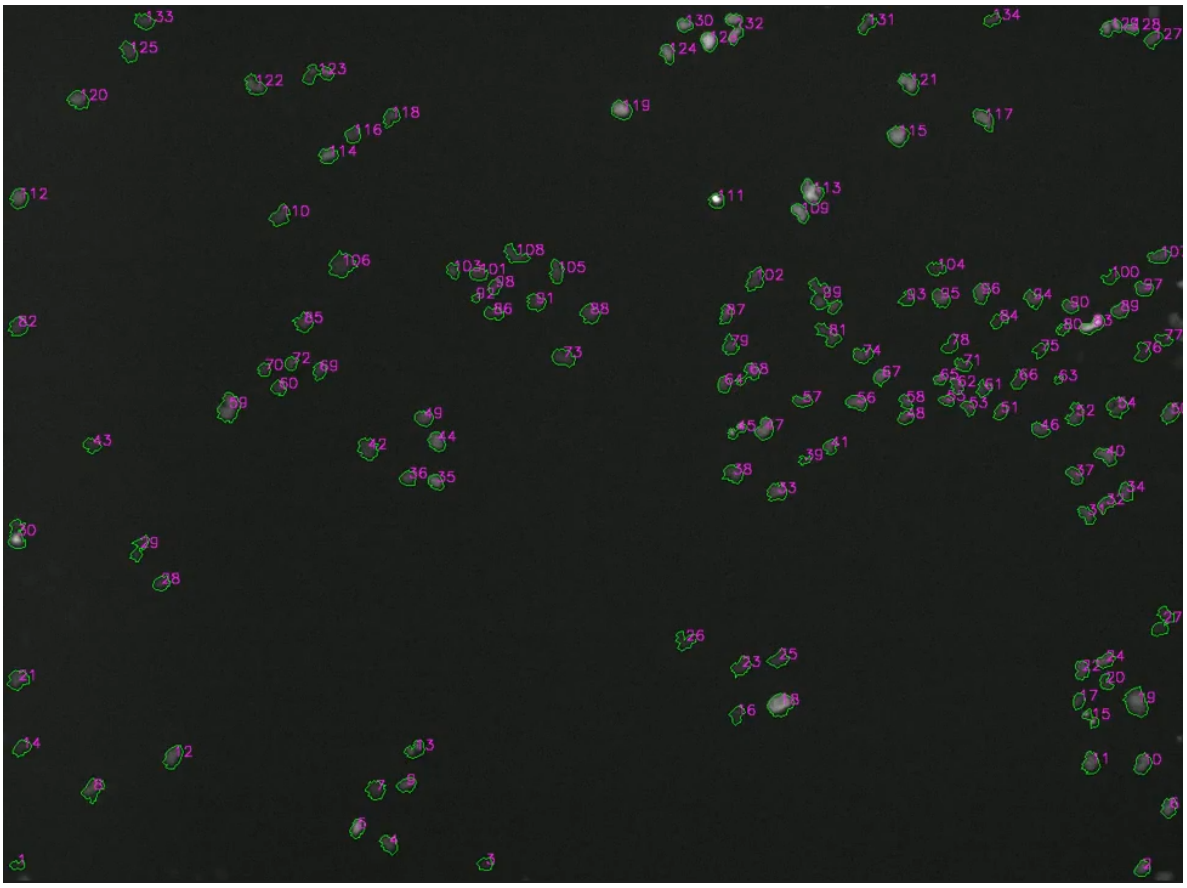
Results and Discussion

Synthetic Data Results

The segmentation of fluorescence images has been done with a very high accuracy that has been obtained. The accuracy is over 90 % for images that are characterized by low to medium cell density. Nonetheless, it should be known that there is room for improvement, especially concerning the detection of cells situated on the edges of the image. Such shortfalls could possibly be solved by resorting to the traditional computer vision techniques. In contrast, in assessing the object tracking capabilities, the outcome is marked by the less favorable side, with less than 20 % accuracy being reported for the case of images with low to medium cell insertion levels by applying these algorithms:

1. BOOSTING Tracker
2. MIL Tracker
3. KCF Tracker
4. CSRT Tracker
5. MedianFlow Tracker
6. TLD Tracker
7. MOSSE Tracker
8. GOTURN Tracker

Figure 3.1: Fluorescence — Segmented.



This is a segmented version of a fluorescence microscopy image, in which the majority of cells have been identified and labeled.

Results

The performance of the suggested graph-matching solution will be assessed by comparing it to other methods available. This comparison will utilize the similarity index given by the cost matrix and ankle surface grounding truth deviations. The performance will be gauged through the to the ground matching discrepancy of a distance score, that is, the distance between two nodes - (tn, 0i) and (tn+1, 0i‘) - matched by them. The loss of the features used for the matching will be the score calculated based on this distance. The same way of measuring is applied consistently with the methodology found in other literature like R&D, 2005 and de Leeuw y Pruzansky, 1978.

$$score = \sum_{i=1}^{n} d\left(p_{(n,i)}, q_{(n+1,i')}, w\right) - d\left(p'_{(n,i)}, q'_{(n+1,i')}, w\right) \tag{3.1}$$

In which a perfect match, in which 2 equal objects means a $score = 0$

The algorithm named Irregular Mini Batch Gradient Descent which has been suggested for conducting the experiments to be carried out mainly focuses on the purpose of optimizing the main tracking performance. The operation of this algorithm is achieved by the application of a specific technique called the Mini Batch Gradient Descent, which is a different of the conventional algorithmic learning. In comparison to the conventional technique that substitutes elements from different epochs to obtain a specified batch size, this algorithm ships all elements of the current epoch to the gradient update. At this point, the error is split by the current epoch's number of elements instead of the batch size, which is often used in traditional Mini Batch Gradient Descent. This series of experiments will provide evidence of its proposed algorithm as well as the improvement of tracking performance.
The goals were reached with an experiment that was implemented for 100 iterations at a learning rate of 0.001 using the previously described algorithm.

Figure 3.2: Graph Track Matches Through time (More is better).

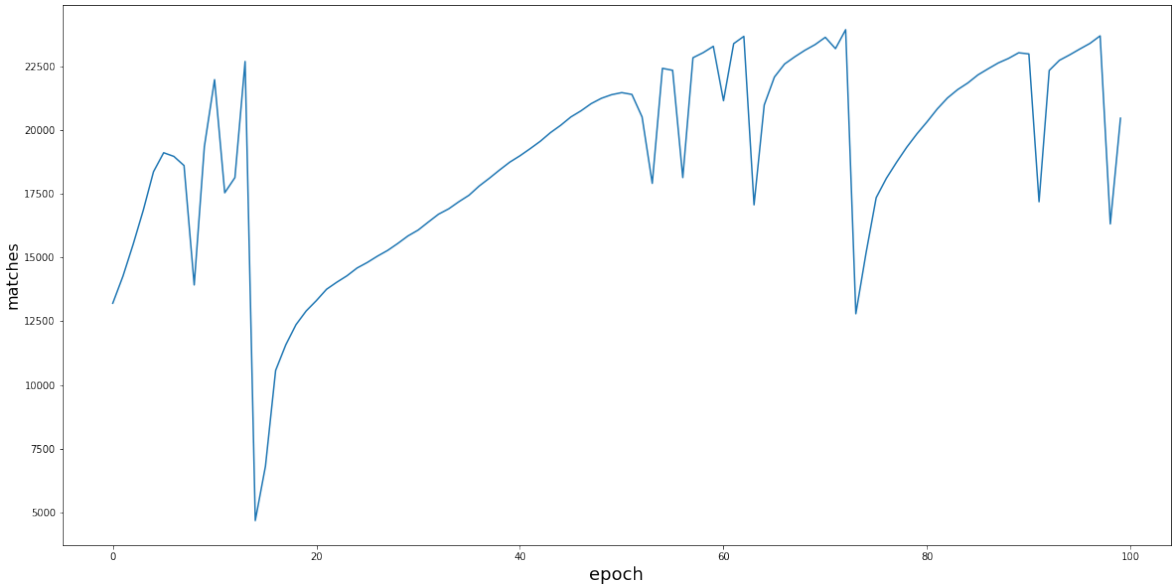
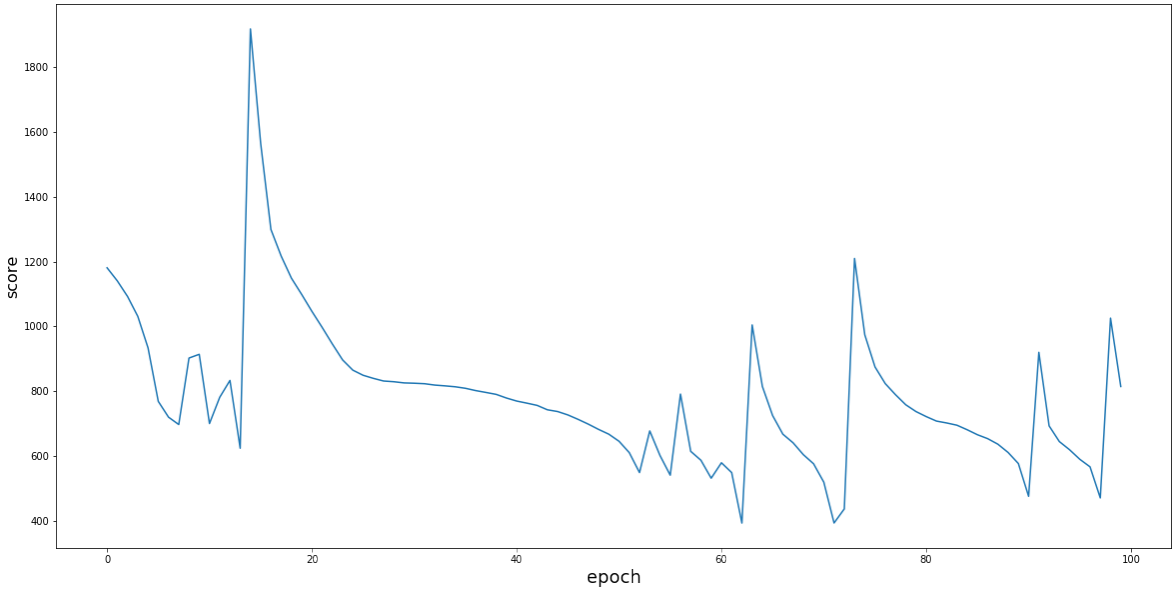
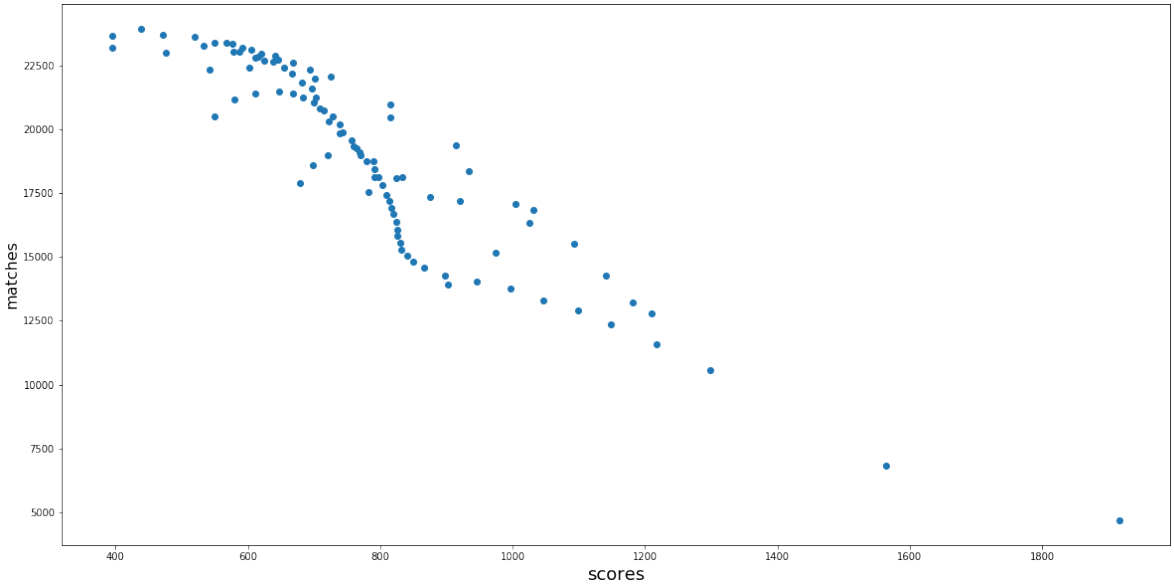


Figure 3.3: Graph Track Score Through time (Less is better).



Apart from this, it should be observed that the Irregular Mini Batch Gradient Descent algorithm that is being proposed has this duality of computational efficiency and precision as it can, in a matter of little time, arrive at a solution that is satisfactory, while also exhibiting a high-level performance in terms of matching accuracy. In addition, it can be inferred that this is the only algorithm that can cater to the specific data requirements as it has the capacity to define the significance of each targeting variable based on its importance to tracking instead of the traditional methods of defining all as equal. All in all, the findings highlight the capability and promise of the proposed algorithm to solve the difficult issue of object tracking in bright-field microscopy images.

Figure 3.4: Graph Track Score Through time.



Further, I would like to stress that the direct consequence of the optimization of the algorithm is the performance of the object-tracking task. As the score turns lower, the number of matches rises, reflecting that the algorithm is successful in identifying and linking the different objects exposed at different time points. Also, the note of the linkage of the convergence observed in the results with the typical behavior of gradient descent optimization algorithms makes it truly remarkable. It is important to notice that the correlation between the score and the number of matches, together with the convergence observed in the results, distinctly points out the legitimacy of the proposed Irregular Mini Batch Gradient Descent algorithm as a robust solution for object tracking in bright-field microscopy images. The outcomes demonstrate the algorithm's potential of effectively learning the needed features for object tracking and applying them for better performance.

3.1. Discussion

Optimized Tracking Performance: Evaluating the Impact of Frame Selection, Untracked Recall, and Bayesian Optimization

The assessment of our tracking algorithm included two dominant strategies: frame selection and Untracked recall. Both methods were crucial in overcoming the difficulties brought by the abrupt changes in the images and increasing the match rate of the tracking algorithm.

Difficulties with Image Data

During the work, it became clear that the images had sudden changes, which posed a major concern for image segmentation and tracking. These abrupt alterations were the result of the very nature of bright field microscopy images, where light and contrast variations, and cell morphology could be very noticeable. To reduce these problems and enhance the algorithm performance, we implemented frame selection and untracked recall as two primary strategies.

Frame Selection

Frame selection was applied to find and exclude specific frames that negatively impacted the match rate. The approach was to use a Laplacian filter on each image and then do the pixel-by-pixel differences between consecutive frames. The Laplacian filter accentuates the areas where the intensity change is rapid, taking these as edges in the images.

By plotting the Laplacian difference between images, we found frames that caused the largest variability. The L-method removing these frames from the analysis led to a significant increase in the match rate. The selection process showed that the algorithm's performance could be improved by a large extent through the removal of a very small number of frames that are problematic.

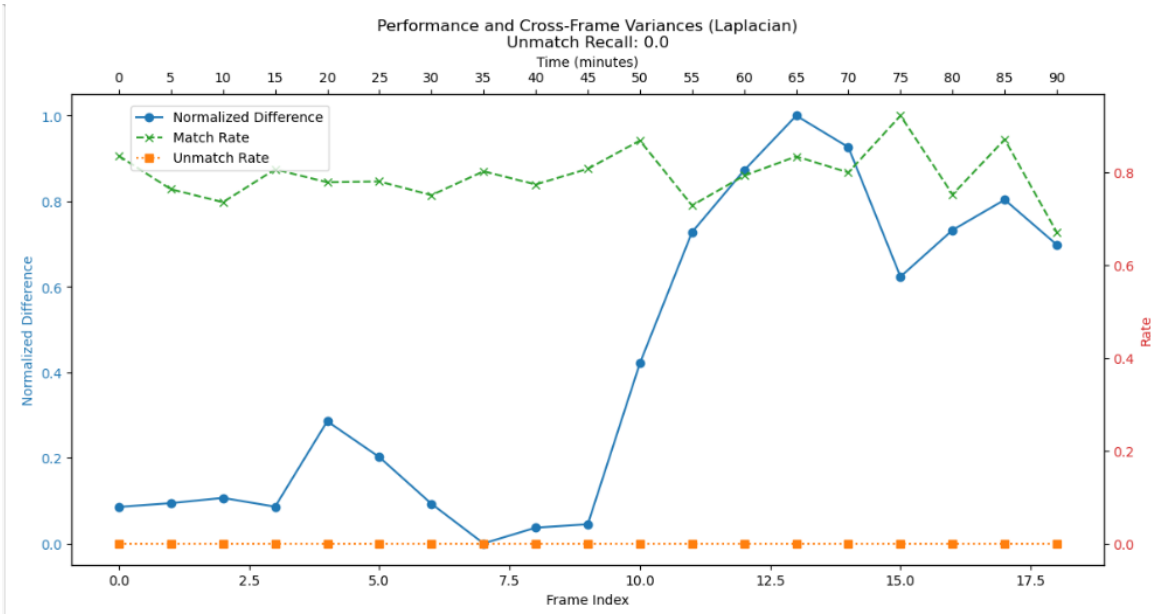


Figure 3.5: Plot showing the Laplacian differences and the impact on match rate.

Plot Analysis of Frame Selection:

Frame Selection Results: The match rates and untracked object rates with and without frame selection are as follows:

- **Without Frame Selection:**
  - Match Rates: [0.792, 0.837, 0.856, 0.891, 0.961, 0.966]
  - Untracked Object Rates: [0.0, 0.223, 0.264, 0.368, 0.450, 0.500]
- **With Frame Selection:**
  - Match Rates: [0.810, 0.856, 0.874, 0.906, 0.962, 0.971]
  - Untracked Object Rates: [0.0, 0.222, 0.248, 0.350, 0.421, 0.475]

The enhancement in match rates in the case of the frame selection method is a clear affirmation of the growth of this approach. The exclusion of those frames that are highly variable makes the algorithm considerably more concentrated on the comparison of the stable ones, and this ultimately results in the unparalleled tracking accuracy. On the whole, the principle is functional, but the picture is a quite reduced increase in the effectiveness.

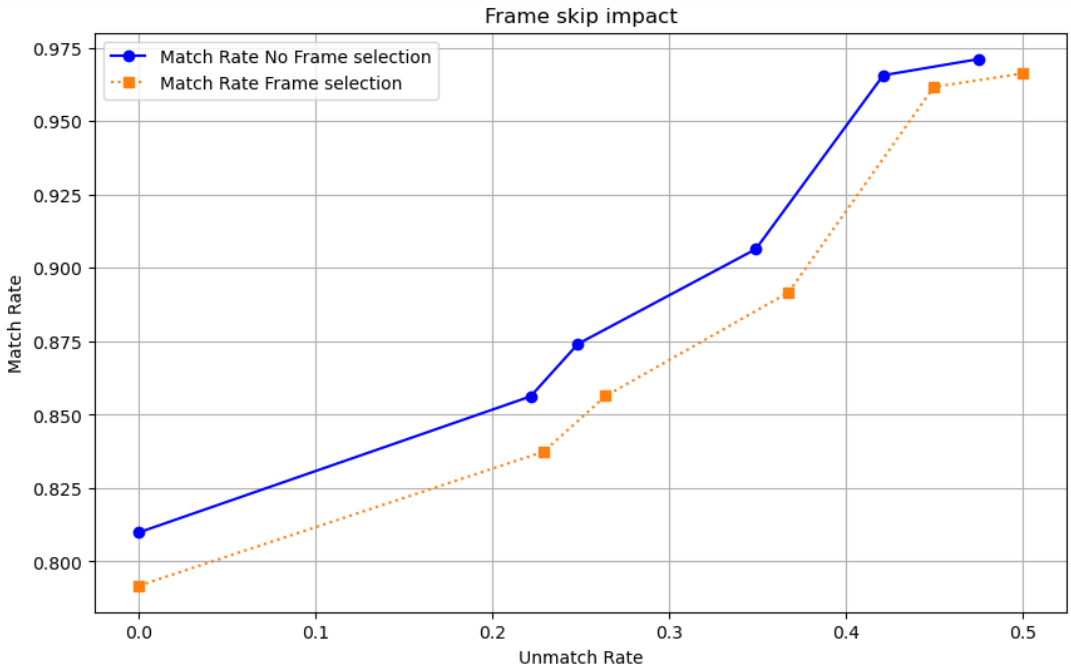


Figure 3.6: Plot showing the impact of frame selection.

Plot Performance Analysis of Frame Selection:

Untracked Recall

An additional strategy for upgrading the tracking protocol and reinforcing the assertion of our algorithm is through the entitlement of untracked recall. The method is cashing in on the imposition of certain objects from the matching process through the defined limits of various attributes. To be specific, untracked recall enables us to sift out objects that are essentially different such as position, intensity, or shape. To illustrate, when the spatial contrast between two objects across successive frames surpasses the allowable limit, the two items will get flagged as untracked and thus, may be classified as; such, subsequently, they will be excluded from the matching endeavor.

The employment of untracked recall stands for a number of salient tasks. First of all, it plays the critical role of abolishing edge cases where the chances of making accurate matches are very low mostly due to the significant deviations of object attributes. Such deviations may result from compulsive movement, abrupt alterations in object appearance, or inconsistencies in segmentation. The elimination of these troublesome issues allows the tracking algorithm to operate exclusively on more stable and trustworthy matches, which, in the end, can lead to overall improvement of the tracking process.

The dominant realization that has been made through the utilization of untracked recall is the relation between the match rate and the untracked object rate. The untracked increase of specific unmatched objects results in the valorization of the fidelity of the resided pairs. This median is significantly consequential in situations where accuracy is paramount over the sum of assorted connected objects. By means of manipulating the drawn limits during the untracked recall, we would manage to find an optimal state of equilibrium between the two impacts which result in a stronger tracking unit.

The use of this technique also carries the impact of the extended research taking root, particularly the ones that target finding, pinpointing, and analyzing various cell phenotypes. The principle of untracked recall, being quite robust, this methodology can be tailored and developed to fit certain phenotypic behaviors, thus it is a versatile tool for the scientists to monitor the cellular responses to ots of microbial and other stimuli. For example, within the frame of cancer directing, pinpointing cells that develop a DNA Damage Proliferation (DPP) tendency is a key task. DPP, which is a process wherein cells continue to multiply even though there is DNA damage present, is a factor often observed at the very early stages of the tumor's development and agression.

DPP cells can be more easily identified in the future iterations of this untracked recall technique that has been customized. This can be accomplished by tightening the unmistakable and recent for the attributes that bothe phenomena should be exposed to imply DNA damage and genetic mutations

in cells. With this logical filtering, which, apart from the usability, includes the cutting of irrelevant data will be actually the people standing most of the time next to; the ones behaving in a way only they would know that they are doing DPP.

In addition to that, flexibility is untracked recall's trait that is easily adaptable to other phenotypes. The example for that is the pench of using it for the identification of cells that are carrying out undergoing apoptosis, which is a genetically programmed cell death process, through the exclusion of cells that are not showing the morphological changes characteristic for this particular fate. In like manner, in the case of DNA Damage Repair, untracked recall could be utilized to segregate the cells that actively take part in the repair mechanisms from those that do not meet the specific criteria for this process.

The untracked recall method can be further improved to directly target DPP, apoptosis, cell cycle arrest, and DNA damage repair. Major objectives that the projects will accomplish include the sorting of cellular behavior, precise cellular abilities, and the exploration of corrective factors, which will be achieved through the advancements in the mechanism of tracking through imaging.

Untracked Recall Implementation:

- **Attribute Limits:**

  Limits were set on various attributes such as position, intensity, and shape descriptors. Objects that exceeded these limits were excluded from the matching process.

- **Performance Evaluation:**

  The performance of the tracking algorithm was evaluated based on the match rate, mismatch rate, and untracked object rate. We could control the balance between matches and untracked objects by adjusting the limits.

- **Results:**

  Implementing untracked object recall resulted in a higher match rate, indicating more accurate tracking. However, this came at the cost of a higher Untracked rate, which was acceptable given the overall improvement in performance.

  Match and Untracked Object Rates with Untracked Recall:

  Example results with untracked recall:

- Match Rate: 0.9064
- Untracked Rate: 0.3490

The untracked recall strategy turned out to be not merely a good, but indeed a very effective and powerful method for the subsequent increase in the correctness of the tracking algorithm. Through the use of a filtering method that included less reliable matches, the tracking algorithm could even achieve a better overall match rate despite the fact that the number of untracked objects increased.

Bayesian Optimization for Parameter Tuning

In addition to the tracking performance, the Bayesian Optimization (BO) was utilized to readjust the parameter's limits thus achieving extreme condition untracked recall values. BO is the most effective way of optimizing black-box functions, which means that it is the most advantageous and best way to achieve the optimization for those functions, where the evaluation of a function is expensive.

Bayesian Optimization Overview: Bayesian Optimization is a method that is based on using a surrogate model, which is normally a Gaussian Process (GP), in order to make the objective function more precise. The GP is able to furnish a probabilistic assessment of the function's value at any given point of space search which can be used by the optimization process to guide.

The acquisition function in BO, such as Expected Improvement (EI), combines exploration and exploitation by selecting points that are likely to yield the most improvement over the current best observation. The optimization process runs continue until the process converges to the optimum solution.

Advantages of Bayesian Optimization:

- Efficiency: BO requires fewer function evaluations compared to other optimization methods.
- Handling Expensive Evaluations: Ideal for problems where each evaluation is costly.
- Probabilistic Framework: Quantifies uncertainty in predictions, leading to more informed decisions.
- Adaptability: Can handle noisy or non-differentiable objective functions.

Application and Results of Bayesian Optimization: Bayesian Optimization was the method applied in this research in order to determine the optimal parameters to be used for tracking. As a result of the optimized limits, the tracking quality was improved significantly; the matching rates were high, and the undetected object rates were in the acceptance range.

Example Bayesian Optimization Results:

- Optimized Limits:

Parameter	Value
radius	0.94877092
x	0.93998146
y	0.50893338
centroid_x	0.39160934
centroid_y	0.15923940
position	0.09371845
average_intensity	0.97286628
intensity_variance	0.39441120
eccentricity	1.00000000
perimeter	0.75954542
area	0.41004958
circularity	0.76173495
major_axis	0.99563719
minor_axis	0.93485105
angle	0.86856704
hull_area	0.98669717
solidity	0.93870039
equivalent_diameter	0.98357111
extent	0.90054477
aspect_ratio	0.68700012
convexity	0.89149526

Table 3.1: Optimized Parameters from Bayesian Optimization

- Match Rate: 0.856
- untracked object Rate: 0.222

Plot Analysis of Bayesian Optimization:

Detailed Results Comparison

Comparison of Methods: Laplacian filter was the most satisfying to visualize the edge detection filters (Laplacian, Prewitt, Scharr, Roberts, LoG) used that exhibited mostly the same information at

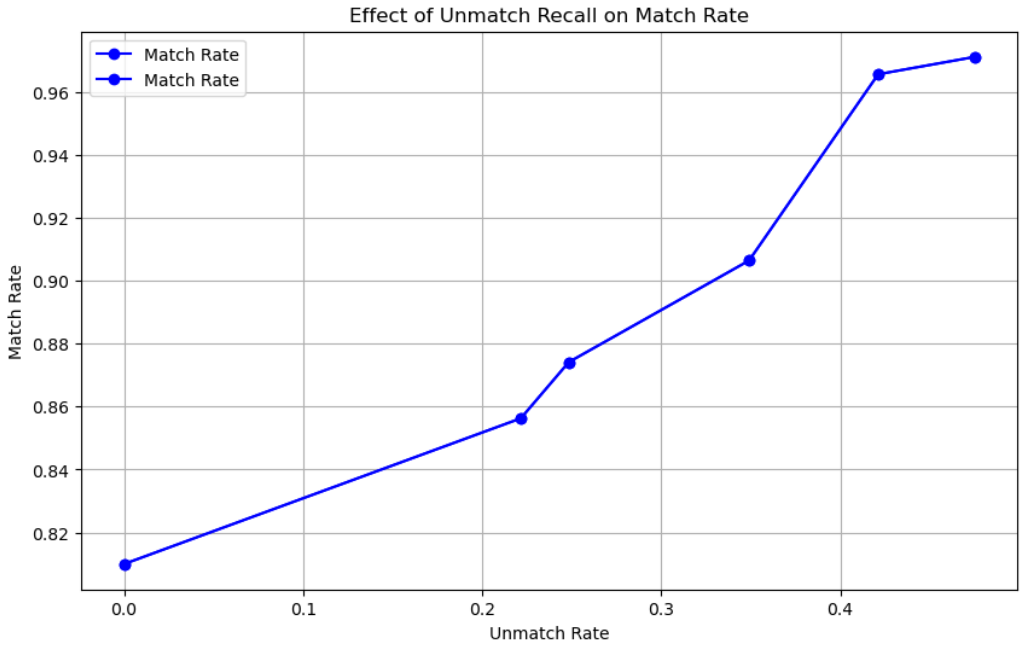


Figure 3.7: Plot showing Bayesian Optimization results, including match rate and untracked object rate.

the time. The tables below illustrate the match rates and untracked object rates that were reached through various methods and configurations:

Match Rate	Untracked Object Rate
0.792	0.0
0.837	0.223
0.856	0.264
0.891	0.368
0.961	0.450
0.966	0.500

Table 3.2: Match and Untracked Object Rates with Frame Selection

Match Rate	Untracked Object Rate
0.810	0.0
0.856	0.222
0.874	0.248
0.906	0.350
0.962	0.421
0.971	0.475

Table 3.3: Match and Untracked Object Rates without Frame Selection

The data in these tables provide a clear picture of the merit of frame selection in raising match success rates and of the various edge detection techniques. Apart from that, they emphasize the fact

that higher match precision comes at the cost of more untracked object rates.

General Performance Evaluation

The data reveal that the integration of frame selection and Untracked recall with Bayesian Optimization leads to a considerable increase in the efficiency of the tracking algorithm. The frame selection step is successful in decreasing the variability of the dataset to a certain extent, while untracked recall enhances the matching process by the exclusion of unreliable matches. The tuning of the parameters is carried out through Bayesian Optimization at which the exact match rate and untracked object rate are achieved.

Key Observations:

- **Frame Selection:**
  - Removing problematic frames with high Laplacian differences led to substantial improvements in match rates.
  - The Laplacian filter was particularly effective in identifying frames with abrupt changes.
- **Untracked Recall:**
  - Setting attribute-specific limits helped filter out less reliable matches, improving overall match accuracy.
  - This approach demonstrated that increasing the number of untracked objects could lead to a higher match rate.
- **Bayesian Optimization:**
  - Bayesian Optimization efficiently navigated the parameter space, resulting in optimal limits for tracking parameters.
  - The probabilistic framework of BO provided a robust approach to handling the complexity and interdependence of tracking parameters.

Limitations

The pairing of frame selection, untracked recall, and Bayesian Optimization has been the most successful method for improving the tracking algorithm; however, it has some limitations that need

to be fixed. The challenges encountered in tagging data and segregation accuracy are central to the problems mentioned above.

Data Tagging Challenges

Data tagging for this project is a highly labor-intensive task. The complexity and density of the microscopy images make it an especially arduous task; for instance, annotating just 20 images can take several days. The meticulous effort involved in accurately identifying and labeling each cell within the pictures is the main reason behind the prolonged duration. The complex shapes of the cells and the overlapping characteristic of the cells are the primary obstacles that contribute to the difficulty of the process, making it both time-consuming and prone to human errors.

The substantial time that goes into tagging the data becomes a limiting factor to the scalability of the dataset. The potential to improve the tracking algorithm in a larger dataset is there, but the current tagging process is the most significant contributing factor that limits this. Automation of the process can reduce the burden to some extent but it would require the technology to analyze images to develop significantly before achieving the same level of accuracy as manual tagging does.

Challenges with Current Segmentation Technology

The current segmentation technology that is in use in this research encounters several major challenges that affect the tracking algorithm's accuracy and reliability. One of the principal concerns is the varying detection rates of the different cells from different frames. Frequently, there are cells that are visible in a frame that are not detected in other frames, having been easily pointed out by someone. This inconsistency, on the other hand, creates huge interruptions in the tracking operation that the algorithm depends on the continuous identification of cells for their movement to be properly tracked.

Also, the segmentation algorithm often captures cell parts that are missing crucial features for accurate tracking which could result in confusion. The fragmentary readings result in irregular computations of the shape, volume, and other measures, which may lead to misalignment and decline in the overall efficiency of the tracking algorithm. Yet even with these setbacks, the segmentation project has made impressive strides; however, it has not been originally validated for tracking purposes. The validation criteria were not adjusted regarding the continuity and precision required for effective tracking, thus more tailored segmentation techniques were needed for tracking applications.

Merging of Multiple Independent Contours

The merger of multiple independent contours between two frames is one noticeable segmentation issue. This merger interrupts the tracking continuity and confuses the algorithm, as it is not sure if these contours are the same object or if the merging process has happened.

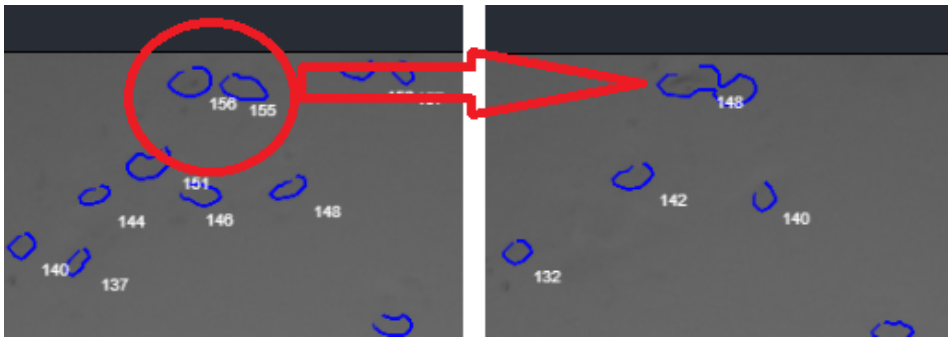


Figure 3.8: The image shows multiple independent contours in one frame that have merged into a single contour in the subsequent frame, creating confusion in the tracking process.

Fragmentation of a Single Object

Another critical issue is the fragmentation of a single object into multiple independent contours within the same frame. This fragmentation causes the algorithm to interpret the fragments as separate objects, leading to inaccurate tracking data. The fragmentation of objects distorts the calculated features, such as size and shape, further complicating the tracking process.

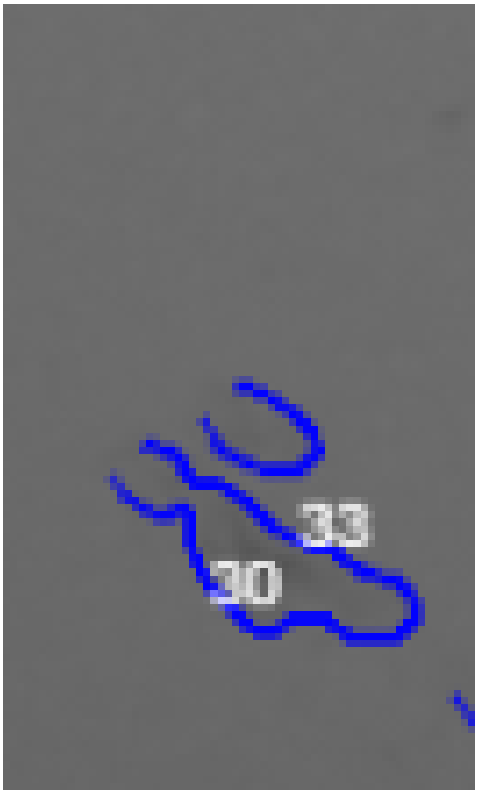


Figure 3.9: The image depicts a single object that has been fragmented into multiple independent contours within the same frame, resulting in tracking errors.

Failure to Detect Visible Objects

Moreover, it is worth mentioning that the segmentation algorithm sometimes has difficulty locating objects that are clearly visible under normal lighting conditions. Oracle Disrupts the continuation of tracking, for an example, the object being not depicted in tracked cells.

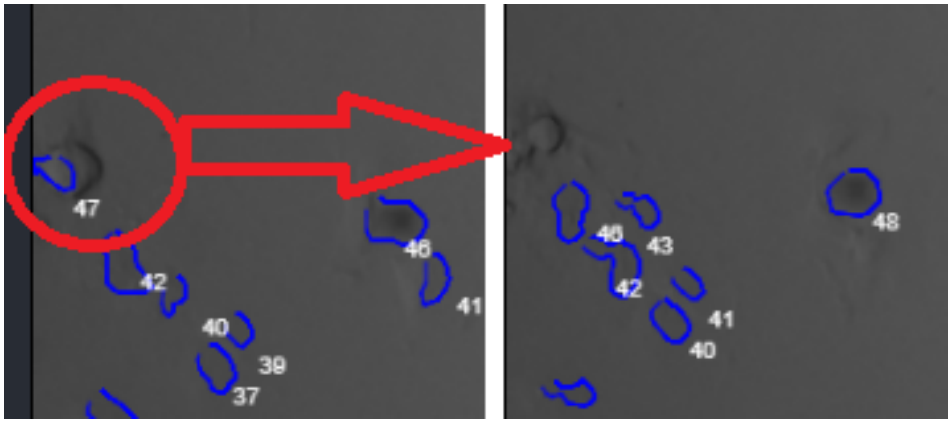


Figure 3.10: The image highlights an object that is prominently visible but was not detected by the segmentation algorithm, leading to gaps in the tracking sequence.

New Object Appearing at the Edge of the Image

An additional complexity comes into play when a new object is seen at the periphery of the image. The lack of segmentation for these objects may mislead the tracking algorithm, as it is now facing an object that is not being traced without having any past information.

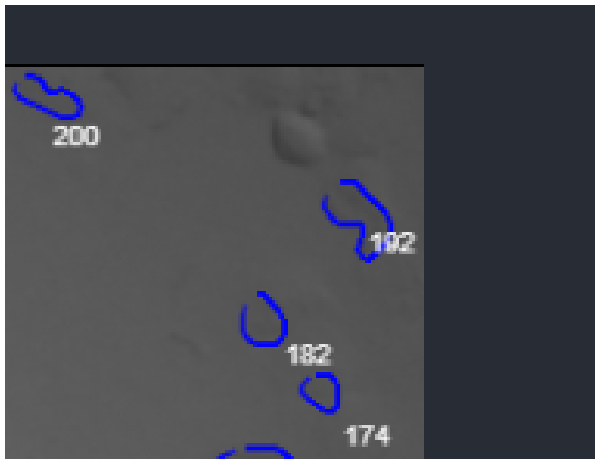


Figure 3.11: The image shows a new object entering from the edge of the frame that was not segmented, creating a sudden, untracked object in the dataset.

Partial Capture of Cells

Another issue is the partial capture of cells, where only a portion of the cell is detected by the segmentation algorithm. This problem leads to inaccuracies in the calculated features and subsequent tracking errors.

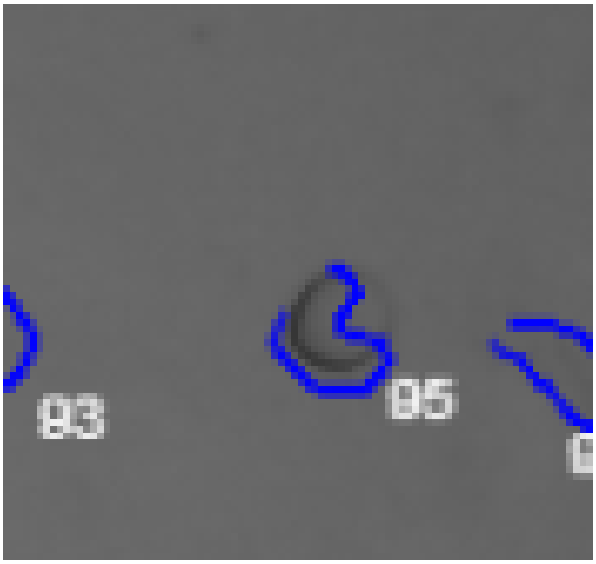


Figure 3.12: The image shows a cell that has been only partially captured by the segmentation algorithm, leading to incomplete feature calculation and tracking errors.

Partial Capture with Additional Irrelevant Areas

Segmentation algorithm not only discovers portions of the cell but it can also misclassify the irrelevant areas into the contour being drawn. These foreign areas may mislead the tracking algorithm, thus resulting in additional errors.

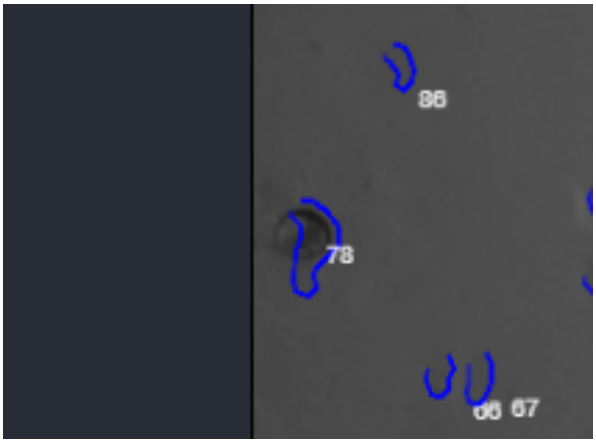


Figure 3.13: The image shows a cell that has been partially segmented along with additional irrelevant areas, which can mislead the tracking algorithm and reduce accuracy.

Demonstrating Tracking Performance

Three images displaying the identical cell being tracked throughout 18 frames are presented here to illustrate the effectiveness of our tracking algorithm. Each of these images consists of two views: the top shows the cell's positional variation through the frames, and the bottom marks the cell with the drawn contour making it more apparent to visualize the tracking accuracy.

The whole movie, the cell is continually and precisely tracked. The contours drawn evidently feature the cell in each frame, which is an indicator of the algorithm's capability of tracking exactly the same way through position and shape alteration of the cell.

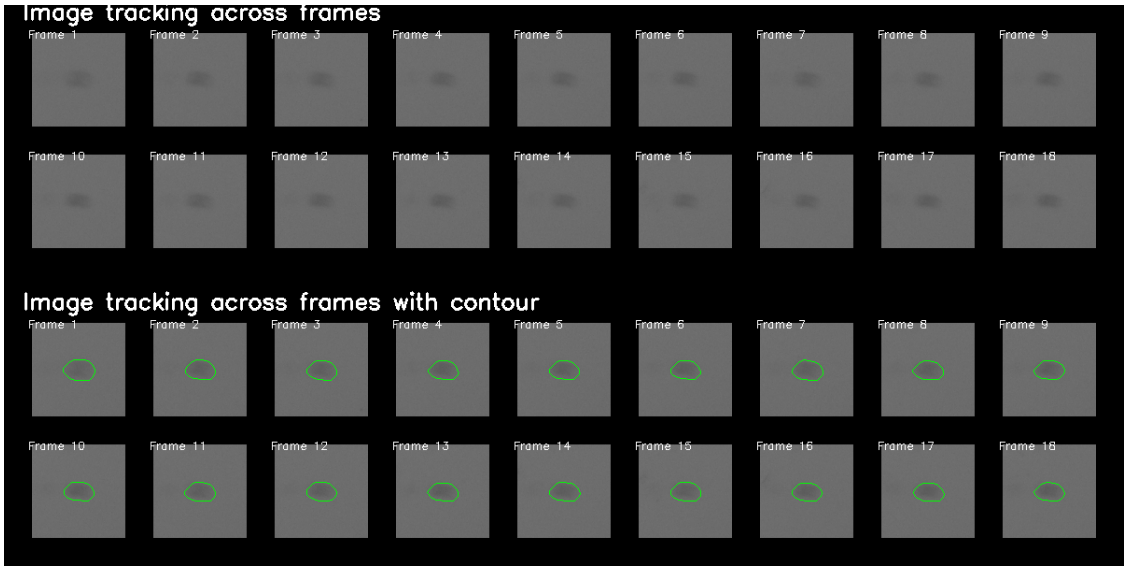


Figure 3.14: The cell is being tracked across 18 frames, with a contour highlighting the cell's boundary.

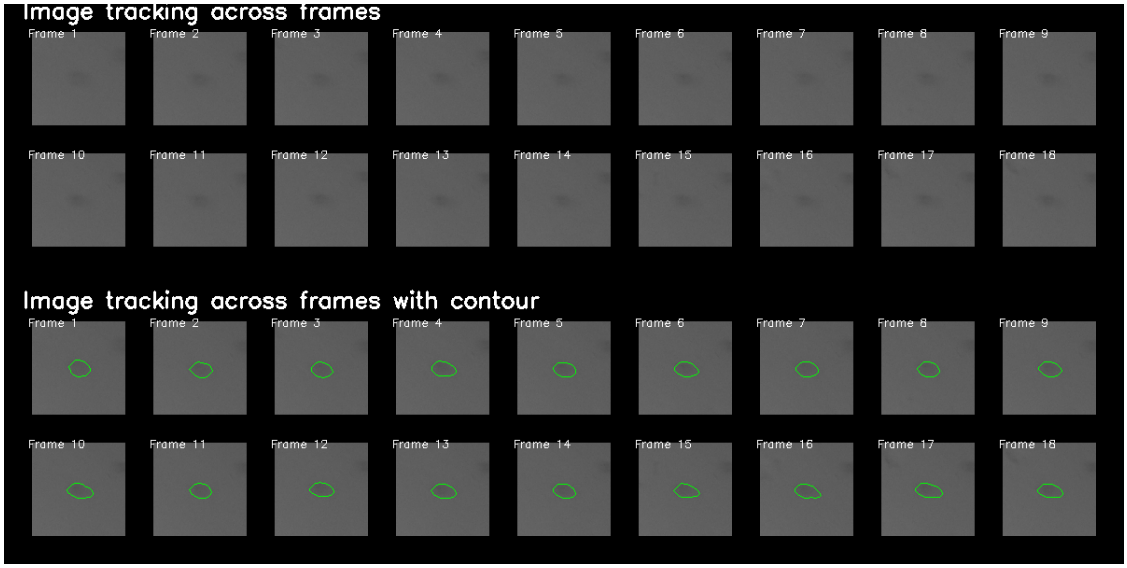


Figure 3.15: The cell is being tracked across 18 frames, with a contour highlighting the cell's boundary.

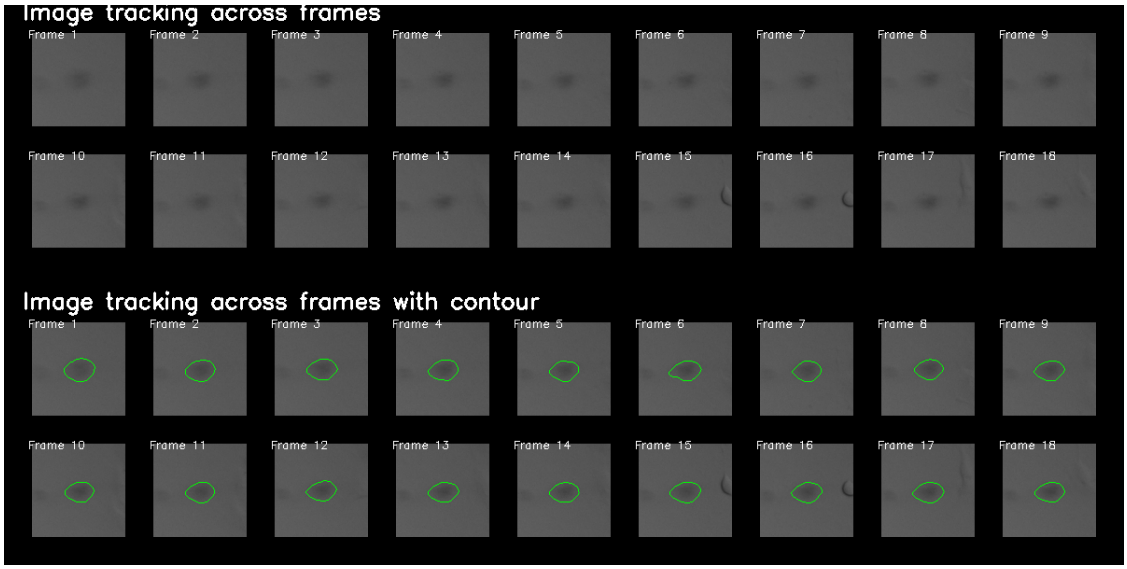


Figure 3.16: The cell is being tracked across 18 frames, with a contour highlighting the cell's boundary.

Chapter 4

Conclusions, Recommendations, and Future Work

4.1. Conclusions

Effectiveness of Frame Selection

The application of frame selection significantly improved the tracking performance. By identifying and excluding frames with abrupt changes through the use of the Laplacian filter, we were able to focus on more stable frames. This led to a substantial increase in match rates and demonstrated the importance of selective preprocessing in handling complex datasets.

Untracked Recall Enhancement

Introducing Untracked recall as a strategy to filter out objects exceeding specific attribute limits proved effective in refining the tracking process. While this increased the untracked object rate, it enhanced the overall match accuracy. This balance between precision and recall is crucial in ensuring reliable tracking in bright field microscopy images.

Bayesian Optimization for Parameter Tuning

Bayesian Optimization was successful in optimizing the parameter limits for the tracking algorithm. By efficiently navigating the complex parameter space, it identified optimal limits that significantly improved the match rate while maintaining an acceptable untracked object rate. This method proved to be a robust solution for handling the interdependent nature of tracking parameters.

Challenges in Data Segmentation

The research highlighted significant challenges in the segmentation of bright field microscopy images. The existing segmentation technology often failed to detect all cells or captured only fragments, complicating the tracking process. This limitation underscores the need for improved segmentation algorithms tailored for this type of imagery.

Overall Improvement in Tracking Accuracy

The combined use of frame selection, untracked recall, and Bayesian Optimization led to a comprehensive improvement in the tracking algorithm's performance. These strategies collectively addressed the inherent challenges of the dataset, contributing to more accurate and reliable tracking of cancer cells.

Implications for Cancer Research

The advancements achieved in this study have significant implications for cancer research. Improved tracking accuracy allows for better analysis of cell behavior, providing valuable insights into the dynamics of cancer cell movement and interaction. This can contribute to a deeper understanding of cancer progression and the development of targeted therapies.

4.2. Recommendations and Future Work

Enhancement of Segmentation Algorithms

To significantly improve the quality of tracking, there is a pressing need to develop and integrate more advanced segmentation algorithms. Current segmentation methods often fail to detect all cells or capture only fragments, leading to inaccuracies in tracking. By leveraging machine learning techniques, particularly convolutional neural networks (CNNs), we can achieve more precise and reliable segmentation of cells in bright-field microscopy images. Improved segmentation will directly enhance the tracking algorithm's performance by providing accurate input data.

Implementation of K-Partite Graphs

Transitioning from bipartite to k-partite graph structures can offer more robust and flexible tracking capabilities. K-partite graphs allow for the inclusion of multiple frames in the matching process, providing a more holistic view of cell movement and interaction over time. However, implementing k-partite graphs requires significant computational resources and optimizations, including parallel processing and GPU acceleration. Future work should focus on developing efficient algorithms and leveraging high-performance computing to handle the increased complexity.

Optimization Techniques

Further research should explore advanced optimization techniques to fine-tune the parameters of the tracking algorithm. While Bayesian Optimization has proven effective, other methods such as evolutionary algorithms, reinforcement learning, or hybrid approaches could be investigated to achieve even better performance. These techniques can help in navigating the complex parameter space more efficiently and finding optimal solutions that enhance tracking accuracy.

Integration of Real-Time Processing

Developing capabilities for real-time processing of bright field microscopy images would be highly beneficial, especially for live cell tracking applications. This requires optimizing both the segmentation and tracking algorithms to process data swiftly and accurately. Implementing real-time processing will enable researchers to monitor and analyze cell behavior as it happens, providing immediate insights and potentially informing experimental adjustments on the fly.

Cross-Validation and Benchmarking

To ensure the robustness and generalizability of the tracking algorithm, extensive cross-validation and benchmarking against standard datasets are recommended. This will help in identifying any biases or limitations of the algorithm and provide a clear comparison with existing methods. Establishing standardized evaluation metrics and protocols will also facilitate more consistent and reliable assessments of tracking performance.

Leveraging Fluorescence Microscopy for Model Training

Although fluorescence microscopy cannot be used directly due to the genotoxicity of fluorochromes, it can still play a role in enhancing bright field microscopy. By creating models based on paired fluorescence and bright field images taken simultaneously, we can train a model to enhance bright field images. This model would be trained to recognize and compensate for the features identified in fluorescence images, potentially leading to improved segmentation and tracking in bright field microscopy without direct use of fluorescence during live cell tracking.

High Magnification Imaging

Utilizing higher magnification imaging can help capture finer details of the cells, reducing the likelihood of missing cells at the borders of images. This approach, however, will result in larger image datasets and slower processing times. Therefore, it will be necessary to develop optimization strategies to manage the increased data volume and ensure efficient processing.

Comprehensive Petri Plate Imaging

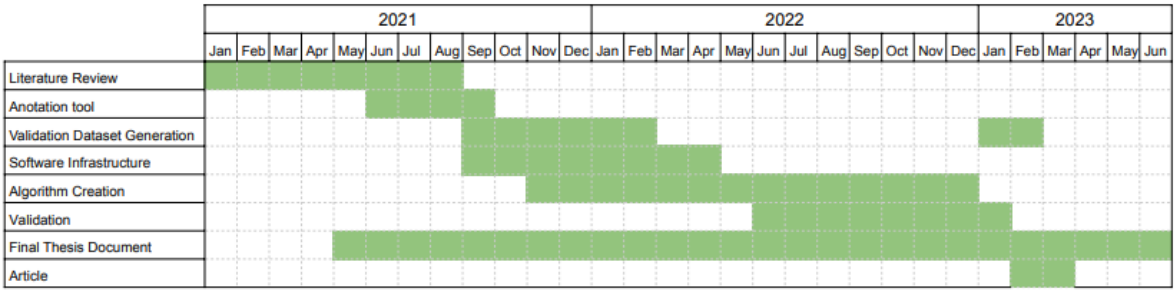
Currently, images capture only the center of the Petri plate, leading to cells disappearing at the borders. Expanding the imaging field to encompass the entire Petri plate will prevent cells from being lost at the edges, ensuring more consistent and comprehensive tracking. While this approach will increase the number of cells to be tracked and thus the complexity of the dataset, it will provide a more complete picture of cell behavior. Optimizations in data processing and algorithm efficiency will be essential to handle the increased workload without compromising performance.

Chapter 5

Schedule

The following figure shows the schedule to complete the objectives proposed as part of this thesis.

Figure 5.1: Schedule of thesis milestone and objectives



Chapter 6

References

Baldi, P. (1995). Gradient descent learning algorithm overview: a general dynamical systems perspective. *IEEE Transactions on Neural Networks*, *6*(1), 182-195. https://doi.org/10.1109/72.363438

Bhaskar, D., Lee, D., Knútsdóttir, H., Tan, C., Zhang, M., Dean, P., Roskelley, C., & Edelstein-Keshet, L. (2019). A methodology for morphological feature extraction and unsupervised cell classification. *bioRxiv*. https://doi.org/10.1101/623793

Buggenthin, F., Marr, C., Schwarzfischer, M., Hoppe, P. S., Hilsenbeck, O., Schroeder, T., & Theis, F. J. (2013). An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC Bioinformatics*, *14*(1), 297. https://doi.org/10.1186/1471-2105-14-297

Čepa, M. (2018). Segmentation of Total Cell Area in Brightfield Microscopy Images. *Methods and Protocols*, *1*(4). https://doi.org/10.3390/mps1040043

Chen, G., Mulla, W. A., Kucharavy, A., Tsai, H.-J., Rubinstein, B., Conkright, J., McCroskey, S., Bradford, W. D., Weems, L., Haug, J. S., Seidel, C. W., Berman, J., & Li, R. (2015). Targeting the Adaptability of Heterogeneous Aneuploids. *Cell*, *160*(4), 771-784. https://doi.org/10.1016/j.cell.2015.01.026
doi: 10.1016/j.cell.2015.01.026.

Chenouard, N., Bloch, I., & Olivo-Marin, J.-C. (2013). Multiple Hypothesis Tracking for Cluttered Biological Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(11), 2736-3750. https://doi.org/10.1109/TPAMI.2013.97

Chiang, P.-J., Wu, S.-M., Tseng, M.-J., & Huang, P.-J. (2018). Automated Bright Field Segmentation of Cells and Vacuoles Using Image Processing Technique. *Cytometry Part A*, *93*(10), 1004-1018. https://doi.org/10.1002/cyto.a.23595

Choi, M., Shi, J., Jung, S. H., Chen, X., & Cho, K.-H. (2012). Attractor Landscape Analysis Reveals Feedback Loops in the p53 Network That Control the Cellular Response to DNA Damage. *Science Signaling*, *5*(251), ra83-ra83. https://doi.org/10.1126/scisignal.2003363

Chowdhury, S., Kandhavelu, M., Yli-Harja, O., & Ribeiro, A. S. (2013). Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE). *BMC Bioinformatics*, *14*(10), S8. https://doi.org/10.1186/1471-2105-14-S10-S8

Collins, R. (2003). Mean-shift blob tracking through scale space. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, *2*, II-234. https://doi.org/10.1109/CVPR.2003.1211475

Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, *15*(2), 81-94. https://doi.org/10.1038/nrclinonc.2017.166

Date, K., & Nagi, R. (2016). GPU-accelerated Hungarian algorithms for the Linear Assignment Problem. *Parallel Computing*, *57*, 52-72. https://doi.org/10.1016/j.parco.2016.05.012

de Leeuw, J., & Pruzansky, S. (1978). A new computational method to fit the weighted euclidean distance model. *Psychometrika*, *43*, 479-490. https://doi.org/10.1007/BF02293809

Debeir, O., Van Ham, P., Kiss, R., & Decaestecker, C. (2005). Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE Transactions on Medical Imaging*, *24*(6), 697-711. https://doi.org/10.1109/TMI.2005.846851

Dima, A. A., Elliott, J. T., Filliben, J. J., Halter, M., Peskin, A., Bernal, J., Kociolek, M., Brady, M. C., Tang, H. C., & Plant, A. L. (2011). Comparison of segmentation algorithms for fluorescence microscopy images of cells. *Cytometry Part A*, *79A*(7), 545-559. https://doi.org/10.1002/cyto.a.21079

Do, K., Tran, T., & Venkatesh, S. (2017). Learning Deep Matrix Representations. https://doi.org/10.48550/ARXIV.1703.01454

Dogo, E., Afolabi, O., Nwulu, N., Twala, B., & Aigbavboa, C. (2018). A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks. https://doi.org/10.1109/CTEMS.2018.8769211

Dufour, A., Shinin, V., Tajbakhsh, S., Guillen-Aghion, N., Olivo-Marin, J.-C., & Zimmer, C. (2005). Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Transactions on Image Processing*, *14*(9), 1396-1410. https://doi.org/10.1109/TIP.2005.852790

Dufour, A., Thibeaux, R., Labruyere, E., Guillen, N., & Olivo-Marin, J.-C. (2011). 3-D Active Meshes: Fast Discrete Deformable Models for Cell Tracking in 3-D Time-Lapse Microscopy. *IEEE*

Transactions on Image Processing, *20*(7), 1925-1937. https://doi.org/10.1109/TIP.2010.2099125

Dzyubachyk, O., van Cappellen, W. A., Essers, J., Niessen, W. J., & Meijering, E. (2010). Advanced Level-Set-Based Cell Tracking in Time-Lapse Fluorescence Microscopy. *IEEE Transactions on Medical Imaging*, *29*(3), 852-867. https://doi.org/10.1109/TMI.2009.2038693

Ezurike, F. O. (2016, junio). *Cartesian product in set theory* [Bachelor]. Dominican Institute.

Frome, A. L. (2007, octubre). *Learning distance functions for exemplar-based object recognition* [doctoral]. University of California.

Gao, Y., Wang, M., Ji, R., Zha, Z., & Shen, J. (2012). k-Partite graph reinforcement and its application in multimedia information retrieval [Intelligent Knowledge-Based Models and Methodologies for Complex Information Systems]. *Information Sciences*, *194*, 224-239. https://doi.org/10.1016/j.ins.2012.01.003

Gustavsson, A.-K., Ghosh, R. P., Petrov, P. N., Liphardt, J. T., & Moerner, W. E. (2022). Fast and parallel nanoscale three-dimensional tracking of heterogeneous mammalian chromatin dynamics. *Mol Biol Cell*, *33*(6), ar47.

Hasenauer, J., Waldherr, S., Radde, N., Doszczak, M., Scheurich, P., & Allgöwer, F. (2010). A maximum likelihood estimator for parameter distributions in heterogeneous cell populations. *Procedia CS*, *1*, 1655-1663. https://doi.org/10.1016/j.procs.2010.04.185

Hayashi, M. T., & Karlseder, J. (2013). DNA damage associated with mitosis and cytokinesis failure. *Oncogene*, *32*(39), 4593-4601. http://dx.doi.org/10.1038/onc.2012.615

Hong, D., Lee, G., Jung, N. C., & Jeon, M. (2013). Fast automated yeast cell counting algorithm using bright-field and fluorescence microscopic images. *Biological Procedures Online*, *15*(1), 13. https://doi.org/10.1186/1480-9222-15-13

INEC. (2018). *Estadísticas Vitales 2018* (inf. téc.). INEC.

Institute, N. C. (2017). NCI Dictionary of Cancer Terms: Chemosensitivity.

Islam, M., Faruk, O., Kar, S., & Islam, M. (2022). Matrix Representation of Graph Theory with Different Operations. *IOSR Journal of Mathematics*, *18*, 8-27. https://doi.org/10.9790/5728-1801010827

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., & Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*, *5*(8), 695-702. https://doi.org/10.1038/nmeth.1237

Jin, K., Latz, J., Liu, C., & Schönlieb, C.-B. (2021). A Continuous-time Stochastic Gradient Descent Method for Continuous Data. https://doi.org/10.48550/arXiv.2112.03754

Jones, D. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *J. of Global Optimization*, *21*, 345-383. https://doi.org/10.1023/A:1012771025575

Kallen, M. E., & Hornick, J. L. (2021). The 2020 WHO Classification: What's New in Soft Tissue Tumor Pathology? *The American Journal of Surgical Pathology*, *45*(1). https://journals.lww.com/ajsp/Fulltext/2021/01000/The_2020_WHO_Classification___What_s_New_in_Soft.1.aspx

Lee, G., Oh, J.-W., Her, N.-G., & Jeong, W.-K. (2021). DeepHCS++: Bright-field to fluorescence microscopy image conversion using multi-task learning with adversarial losses for label-free high-content screening. *Medical Image Analysis*, *70*, 101995. https://doi.org/https://doi.org/10.1016/j.media.2021.101995

Liberti, L., Lavor, C., Maculan, N., & Mucherino, A. (2014). Euclidean Distance Geometry and Applications. *SIAM Review*, *56*(1), 3-69. https://doi.org/10.1137/120875909

Lugagne, J.-B., Jain, S., Ivanovitch, P., Ben Meriem, Z., Vulin, C., Fracassi, C., Batt, G., & Hersen, P. (2018). Identification of individual cells from z-stacks of bright-field microscopy images. *Scientific Reports*, *8*(1), 11455. https://doi.org/10.1038/s41598-018-29647-5

Min Pak, J., Ki Ahn, C., Hak Mo, Y., Taeg Lim, M., & Kyou Song, M. (2016). Maximum likelihood FIR filter for visual object tracking. *Neurocomputing*, *216*, 543-553. https://doi.org/10.1016/j.neucom.2016.07.047

Mora, A. (2018). *Cell phenotype classification using M-phase features in live-cell bright field time-lapse microscopy* [master]. Universidad de Costa Rica.

Na, S., Luo, Y., Yang, Z., Wang, Z., & Kolar, M. (2020, julio). Semiparametric Nonlinear Bipartite Graph Representation Learning with Provable Guarantees. En H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (pp. 7141-7152, Vol. 119). PMLR. https://proceedings.mlr.press/v119/na20a.html

Organization, W. H. (2022). Cancer.

Patmanidis, S., Chignola, R., Charalampidis, A. C., & Papavassilopoulos, G. P. (2019). A comparison between Nonlinear Least Squares and Maximum Likelihood estimation for the prediction of tumor growth on experimental data of human and rat origin. *Biomedical Signal Processing and Control*, *54*, 101639. https://doi.org/10.1016/j.bspc.2019.101639

Qian, X., & Klabjan, D. (2020). The Impact of the Mini-batch Size on the Variance of Gradients in Stochastic Gradient Descent. https://doi.org/10.48550/arXiv.2004.13146

Quirós, S., Siles, F., & Mora, R. (2014). *Análisis del papel de la respuesta celular al daño en el ADN en la inducción de muerte celular generaciones posteriores a tratamientos quimioterapéuticos* (inf. téc. N.º B4653). Universidad de Costa Rica.

R&D, A. P. (2005). *Euclidean Distance raw, normalized, and double-scaled coefficients* (inf. téc.). Advanced Projects R&D.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. [cite arxiv:1609.04747Comment: Added derivations of AdaMax and Nadam]. https://doi.org/10.48550/arXiv.1609.04747

Santaguida, S., & Amon, A. (2015). Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nat Rev Mol Cell Biol*, *16*(8), 473-485. http://dx.doi.org/10.1038/nrm4025

Schiegg, M., Hanslovsky, P., Kausler, B. X., Hufnagel, L., & Hamprecht, F. A. (2013). Conservation Tracking. *2013 IEEE International Conference on Computer Vision*, 2928-2935. https://doi.org/10.1109/ICCV.2013.364

Selinummi, J., Ruusuvuori, P., Podolsky, I., Ozinsky, A., Gold, E., Yli-Harja, O., Aderem, A., & Shmulevich, I. (2009). Bright Field Microscopy as an Alternative to Whole Cell Fluorescence in Automated Analysis of Macrophage Images. *PLOS ONE*, *4*(10), 1-9. https://doi.org/10.1371/journal.pone.0007497

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. En F. Pereira, C. Burges, L. Bottou & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin*, *65*(2), 87-108.

Trask, A. (2019). *Grokking Deep Learning* (1st). Manning Publications Co.

Vasan, N., Baselga, J., & Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature*, *575*(7782), 299-309. https://doi.org/10.1038/s41586-019-1730-1

Wagh, A. (2022). Gradient Descent and its Types. https://www.analyticsvidhya.com/blog/2022/07/gradient-descent-and-its-types/

Wang, L., He, Z., Liu, C., & Chen, Q. (2021). Graph based twin cost matrices for unbalanced assignment problem with improved ant colony algorithm. *Results in Applied Mathematics*, *12*, 100207. https://doi.org/10.1016/j.rinam.2021.100207

Wieser, C. (2013, abril). *Ranking on Multipartite Graphs* [doctoral]. Universitat München.

Yu, T., & Zhu, H. (2020). Hyper-Parameter Optimization: A Review of Algorithms and Applications. https://doi.org/10.48550/arXiv.2003.05689

Zhang, C., Huber, F., Knop, M., & Hamprecht, F. A. (2014). Yeast cell detection and segmentation in bright field microscopy. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 1267-1270. https://doi.org/10.1109/ISBI.2014.6868107

Zhong, J., Tan, J., Li, Y., Gu, L., & Chen, G. (2014). Multi-Targets Tracking Based On Bipartite Graph Matching. *Cybernetics and Information Technologies*, *14*(5), 78-87. https://doi.org/doi: 10.2478/cait-2014-0045

Chapter 7

Appendix

You can find additional images, code, video, and others at: https://cloud.prislab.org/s/zCdnSfaiyzXKrGW