Norges miljø- og
biovitenskapelige
universitet

Master's Thesis 2021   30 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Development of a method to study cell division

Øyvind Roland Svendsen

Master of Science, Chemistry and Biotechnlogy

PREFACE

This master project was performed at the clinic for laboratory medicine, department of microbiology (MIK), Rikshospitalet, OUH. The task was completed as a part of the experimental cancer therapy research group led by Stig Ove Bøe, where Emma Lång served as the main counsellor for the thesis. The thesis marked the end of the 5-year field of study civil engineer in chemistry and biotechnology at NMBU.

First of all, I wish to say thank you to my main guidance counsellor, Emma Lång, for all the help and support with this project. All of the academic knowledge that you have shared during this project is very much appreciated and made this project possible. I would also like to thank research group member Anna Lång for sharing knowledge and helping me perform technical tasks in the laboratory. Thank you to research group leader Stig Ove Bøe, for all the help given when performing imaging experiments. I also want to thank my counsellor at NMBU, Morten Sørlie for all the follow up and motivation throughout the project.

I want to say thank you to my wife, Juliane, for motivating and supporting me throughout this project.
I also wish to thank my parents and my two sisters for all the support and care that you have given me.

TABLE OF CONTENTS

ABBREVIATIONS

AI	*Artificial intelligence*
CNN	*Convolutional neural network*
ECM	*Extracellular matrix*
DME	*Dulbecco's Modified Eagle*
FBS	*Fetal bovine serum*
FN	*False negative*
FP	*False positive*
IMDM	*Iscove's Modified Dulbecco's Medium*
IoU	*Intersection over union*
MTOC	*Microtubule organizing center*
PBS	*Phosphate buffered saline*
Penstrep	*Penicillin-Streptomycin Mixture*
PFA	*Paraformaldehyde*
QC	*Quality control*
ROI	*Region of interest*
TP	*True positive*
WT	*Wild type*
YAP	*Yes-assosiated protein*

SAMMENDRAG

Cellemigrering og celledeling har en essensiell rolle innen sårheling, utvikling og kreftdannelse. Celler kan bli observert gjennom mikroskopi, og dette kan gi innsikt i cellenes oppførsel og studere cellekulturers migrasjon og celledeling i forskjellige miljøer. Målet med dette prosjektet var å utvikle en deep learning-basert metode for å analysere celledeling og cellemigrering basert på data fra live cell imaging. For å gjøre dette ble en tidligere publisert metode for behandling av celler benyttet (Lång et al. 2018). I denne metoden blir monolag av den menneskelige keratinosytt linjen HaCaT stimulert til å gjennomgå kollektiv migrering etterfulgt av celledeling. En automatisert pipeline for analyse av celledelinger var utviklet ved å generere en deep learning modell med StarDist software.

For å utvikle en StarDist modell egnet for pipelinen, ble flere modeller designet for detektering av delende cellekjerner utviklet med forskjellig størrelse på treningsdatasettet og parametere. Bildene brukt til å generere treningsdatasettet var bilder hentet fra live cell imaging eksperimenter av HaCaT celler som hadde vokst i serum fritt medium i 72 timer etterfulgt av re-stimulering. Bildene ble tatt med et widefield mikroskop. Ved å teste segmenteringsevnen til modellene på bilder ikke tidligere sett av nettverkene ble den beste modellen, '1000 2-step', valgt til å bli brukt i pipelinen for analyse av bilder av levende celler. Denne modellen var god til å segmentere formen til celler, med en mean matched score på 0,83201 og hadde en moderat treffsikkerhet når det kom til identifisering av mitotiske cellekjerner, med en treffsikkerhet på 0,67494. Ved å bruke denne modellen og en modell laget av utviklerne av StarDist, ble 12 time-lapse bildesekvenser fra eksperimentene kvantitativt analysert med den utviklede metoden (Schmidt et al. 2018). Metoden målte gjennomsnittlig omkrets og areal, samt cellepopulasjon per bilde over tid. Metoden målte også den gjennomsnittlige populasjonen av mitotiske celler over tid. For å få en forståelse for cellenes tilstand gjennom behandlingen ble fikserte celler i forskjellige stadier av behandlingen bestående av serum sult og re-stimulering IF merket med Aurora B og YAP spesifikke antistoffer, og avbildet med et konfokalmikroskop. I sultede celler var YAP uttrykket og lokalisert i cellekjernen, og dette indikerer at sultede celler er under stress. YAP er en biomekanisk stress sensor som også er involvert i endringer i celle-celle og celle-ECM forbindelser, noe som indikerer at YAP aktivering er involvert i den økte migrasjonsraten i re-stimulerte celler. Resultatene fra de kvantitative analysene av bilde sekvensene og Aurora B bildene tyder også på at synkronisert

celledeling skjer rundt 24 timer etter re-stimulering på grunn av høy aktivering av Aurora B og økningen av mitotiske celler i dette tidsrommet.

Det kan konkluderes at den utviklede metoden kunne analysere celler med moderat treffsikkerhet. Fremtidig arbeid vil fokusere på forbedring av StarDist modellens treffsikkerhet for å forbedre den utviklede metodens kvalitet.

ABSTRACT

Cell migration and proliferation play an essential role in wound healing, development, and tumorigenesis. Observing cell behavior through microscopy can give an insight to cell behavior and study cell migration and proliferation under different conditions. The aim of the present project was to develop a deep learning-based methodology for analysis of cell division and cell migration based on live microscopy data. To achieve this, a previously published experimental approach was utilized where coordinated cell migration and subsequent proliferation is stimulated in cell monolayers of the human keratinocyte cell line HaCaT (Lång et al. 2018). An automated pipeline for cell division analysis was developed by generating a deep learning model with the StarDist software.

To create a StarDist model suitable for the pipeline, several StarDist models designed to detect dividing nuclei were developed with varying training dataset sizes and parameters. The images used to generate training dataset were images acquired from live cell imaging experiments of HaCaT cells treated with serum starvation (72 h) and re-stimulation. These images were acquired with a widefield microscope. By testing the performance of the generated models on images previously not seen by the networks, the best performing model, '1000 2-step', was chosen to be used in the pipeline for cell image analysis. This model was good at segmenting cell shape, with a mean matched score of 0,83201 and was moderately accurate at predicting mitotic nuclei, with an accuracy of 0,67494. Using this model, as well as a model developed by the StarDist developers, 12 time-lapse image sequences from the imaging experiments were quantitively analyzed with the developed method (Schmidt et al. 2018). The method measured the average nucleus perimeter, area, and population per frame over time. The method also measured the average mitotic cell population over time. Additionally, fixed HaCaT cells in different stages of the treatment of serum starvation and re-stimulation were IF stained with Aurora B and YAP specific antibodies were imaged with a confocal microscope. The cellular expression of these proteins gives an insight to the state of the cells in different stages of the treatment. It was found that YAP was expressed and localized in the nucleus of starved cells, indicating that starved cells are under stress. YAP is a biomechanical stress sensor which is also involved in changing the cell-cell and cell-ECM connections, and this suggests that YAP activation is involved in the increased rate of cell migration when cells are exposed to re-stimulation. The results from quantitative analyses of the image sequences and Aurora B imaging also suggests that synchronized cell division occurs

around 24 hours after re-stimulation due to activation of Aurora B and the increase of mitotic cells in around this time.

In conclusion, the developed method was able to analyze cell images with moderate accuracy. Improving StarDist model accuracy is therefore the best course of action to improve the quality of the developed method.

1   INTRODUCTION

Human skin cells serve an important purpose as a protective outer layer that protects our inner tissues from environmental factors such as radiation and physical damage. This outer layer is known as the epithelium. When basal skin cells differentiate to form keratinocytes, they arrange themselves in complex sub-layers, in which cells of each layer has different structures and serve different purposes. Many factors that contribute to epithelial cell migration and proliferation have not been thoroughly researched. When studying keratinocytes as well as cell cultures in general, data is often acquired from microscopic images. Traditionally, image analysis of biological samples has been performed manually. However, with the recent surge in deep learning programs being developed to perform analytical tasks, detecting, segmenting and analyzing images are gradually becoming more and more automated (Yang et al. 2020).

This introduction will present the human keratinocytes and provide information of their differentiation and adhesive abilities. Information on some of the proteins involved in keratinocyte migration and proliferation will be mentioned as well. Additionally, methods applied in a deep learning-based analysis of live cell images will be described.

1.1   The cell

All living organisms are made up of small building blocks known as cells. Cells perform tasks that are essential for all living organisms, such as metabolism and growth. As cells grow and proliferate, they go through a process called the cell cycle. Cell proliferation, or division, is known as mitosis and is a part of the cell cycle. To better understand different cell behaviors and how cells grow and proliferate, the cell cycle and the different phases of mitosis will be described below.

1.1.1   The cell cycle

The cell cycle is characterized by cell division (mitosis) and cell growth (interphase). During mitosis the cell divides to form two daughter cells and during interphase the cell grows, replicates its DNA, and prepares for mitosis. This occurs during three phases called the G_1, S, and G_2 phase,

respectively. The cells can also be in a resting state, the G_0 phase, where they are not preparing for cell division or cell growth (Pizzorno et al. 2015). This occurs after a cell has completed cell division, in early G_1, due to internal or external signals that tells the cell to exit the cell cycle and stop dividing. Some cells become nondividing cells and stay in this phase permanently (differentiate), while some cells return to the cell cycle to produce more cells at a later stage.

During interphase, the DNA of the cell is also loosely packed to allow for a high rate of transcription. As cells progress through the cell cycle from the G_1 to the G_2 phase, the size of the nucleus decreases as a result of chromatin condensation (Chu et al. 2017).

Programmed cell death can also occur during the cell cycle. This phenomenon is called apoptosis and is caused by apoptotic proteins produced by the cell as a response to internal or external stimuli (Henseleit et al. 1996 and Elmore et al. 2007). During apoptosis, the cell's DNA and organelles are broken down causing the cell to shrink and ultimately fragment. The cell's cytoskeleton collapses and any bonds to neighboring cells are consequently broken. When apoptotic cells shrink, they will appear in the microscope as small, condensed cells that are not bound to other cells and often floating around in the growth medium as can be seen in figure 1.1.1.
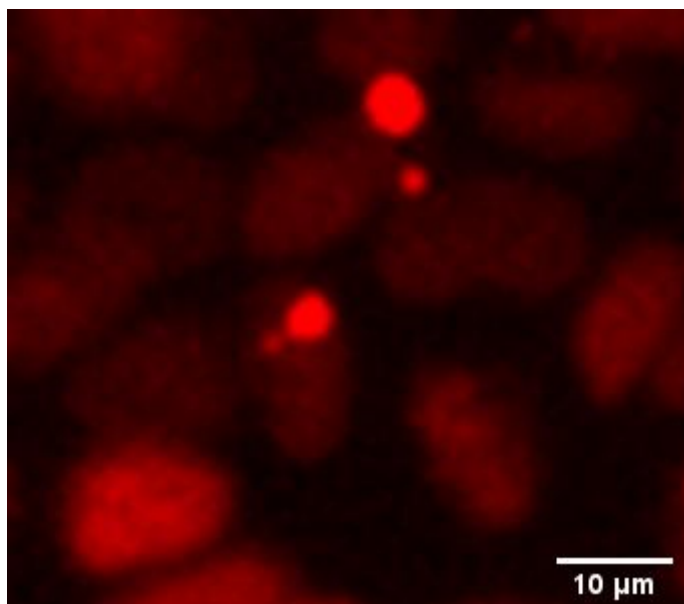

Figure 1.1.1: Image of apoptotic cells in a HaCaT cell culture. The apoptotic cells can be seen as small and condensed cells floating over the non-apoptotic HaCaT cells. The image was acquired during a live cell imaging experiment performed (05.02.21) on a Zeiss AxioObserver Z1 microscope using the 10x objective.

1.1.2   **Cell division**

Cell division, or mitosis is divided into 5 phases; prophase, metaphase, anaphase, telophase and cytokinesis. During prophase the duplicated DNA, packed as chromatin, condenses and becomes more tightly packed. In eukaryotic cells, DNA is packed in chromatin that takes up much less space than loose, unpacked DNA. Chromatin is composed of nucleosomes, where DNA is twirled around 8 proteins called histones. There are 5 main types of histone proteins; H1/H5, H2A, H2B, H3 and H4. The H2A, H2B, H3, and H4 histone proteins are involved in packing chromatin in the nucleus. Several enzymes aid in the process of chromatin condensation and decondensation. One such enzyme is the Aurora B kinase, which binds to chromatin as it condenses (Antonin et al. 2016).

As mentioned previously, the chromatin condenses during prophase, forming heterochromatin from euchromatin. The heterochromatin coils and loops around the histone proteins to form chromosomes. Meanwhile, centrosomes, composed of two centrioles orient towards two poles on each side of the cell. A cytoskeletal network of microtubules called the mitotic spindle is also formed during prophase. These spindle fibers are attached to the centrioles at the opposite poles, and their function is to separate sister chromatids at a later stage of mitosis. The centrosomes and the spindle apparatus make up a polarized tubulin-based structure. The centrosome ends are called the minus ends and attach to the microtubule organizing center, MTOC. The MTOC, made up of centrosomes, orchestrates the microtubule due to their role in orchestrating chromatid separation. The microtubule end that attaches to the centromere is referred to as the plus end.

In the prometaphase the nuclear membrane breaks down and the spindle apparatus attaches to a central part of the chromosomes called the centromere. A part of these microtubule fibers are referred to as the kinetochore spindle. The kinetochore is the part of the microtubules that is attached to the centromere. A protein in the kinetochore, CEN-C, binds to a protein packed in the centromere chromatin, CEN-A (Cheeseman et al. 2014). The centromere is made up of densely packed chromatin that not only function as the kinetochore attachment point, but also helps keep the two sister chromatids together. Subsequently, the kinetochore forms a strong interaction between the mitotic spindle and the chromosomes through the interaction between CEN-C and CEN-A.

During metaphase the mitotic spindle aligns the chromosomes at the center of the cell, in the middle of the two spindle poles. The plane that the chromosomes occupy is often referred to as the metaphase plate. At the end of metaphase, every chromosome is connected to the centrosomes at each pole through the mitotic spindle.

During anaphase, the chromosomes separate to form two sister chromatids as the microtubule connecting the centromere and the centrosomes contract and pull the chromatids towards the cell poles. Errors in chromosome segregation may occur and this leads to cell division failures. These errors usually arise due to faults in the interaction between the microtubule spindle apparatus and the chromosomes. Some examples of cell division errors are the irregular positioning of the centrosomes and errors in microtubule-kinetochore attachments leading to mis-segregation of the chromosomes (Compton et al. 2011). Furthermore, microtubule structures not associated with the separation of chromosomes elongate the plasma membrane along the axis of the cells' poles (Kiyomitsu et al. 2013). Subsequently, the poles are being pulled further and further apart. An actin and myosin-containing structure called the cleavage furrow starts to form in the equatorial cortex, or between the cell poles at early anaphase (Glotzer et al. 2004). This structure is involved in cell separation later in mitosis.

When the chromatids reach the poles, the cell enters telophase. During telophase the microtubule spindle fibers break down into independent tubulin proteins and two nuclear membranes begin to form around the chromatids at each pole. When the chromatids are encapsulated by a nuclear membrane they will begin to decondense into a more loosely packed DNA structure. As the chromatin decondenses Aurora B is released from the chromatin (Antonin et al. 2016).

Cell separation is completed by cytokinesis. Here, the membrane that makes up the dividing cell separates completely between the two cell poles to form two separate cells, the so-called daughter cells. A structure that is central to this process is the cleavage furrow mentioned above. Together, the actin and myosin proteins that makes up the cleavage furrow form contractile filament structures, which pull part of the plasma membrane towards the cell center, making the cell center narrower. During anaphase, telophase and cytokinesis the cleavage furrow contracts more and more until the membrane eventually separates to form two new daughter cells. A structure called the midbody is formed during early cytokinesis. The midbody is a complex connecting two

daughter cells at the end of cytokinesis. It consists of several proteins, including the Aurora B kinase, and is involved in detaching daughter cells during cytokinesis (Hu et al. 2012).
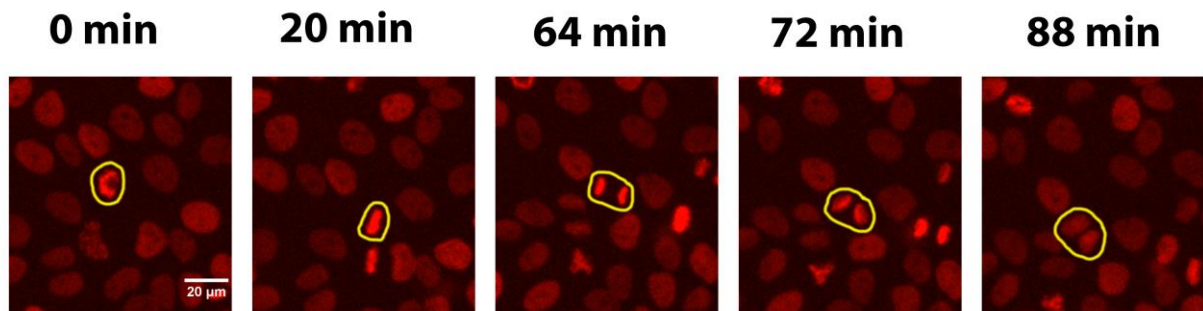

Figure 1.1.2: Images of a dividing HaCaT cell, visualized by fluorescently tagged Histone H2B proteins in the cell nucleus. The first image shows the cell in prophase, with the DNA condensed in its characteristic flower shaped form. The timeline is set to 0 to mark the beginning of cell division. The second image shows the metaphase plate formed during metaphase, 20 minutes after prophase. The third image illustrates the chromosome separation during anaphase, observed after 64 minutes of imaging. In the fourth image, at time point 72 minutes, the cell is in telophase and the DNA has begun to decondense. Finally, the fifth image shows a typical chromatin structure during cytokinesis at time point 88 minutes. The cell completes cell division in roughly 1,5 hours. The images are acquired from a live cell imaging experiment performed (19.01.21) on a Zeiss AxioObserver Z1 microscope using the 10x objective.

1.2 **Human skin**

Epithelial cells are cells specialized to form protective cell layers on the outermost surface of the human body. These cells organize themselves into a skin tissue that consists of three main cell types, keratinocytes, melanocytes and langerhans cells. Melanocytes are pigment-producing cells and langerhans cells are immune cells, which are capable of producing antimicrobial substances (Schleimer et al. 2007). The majority of human epidermal tissue consists of keratinocytes, making up 90-95% of all epithelial cells (Scieglinska et al. 2019). Keratinocytes are produced by epidermal stem cells that have differentiated to form this cell type. Furthermore, keratinocytes express several structural proteins, such as keratin, which causes the cell to become more flattened and rigid (Deo et al. 2018 and Fenner et al. 2016). Keratinocytes also express proteins like Cadherin that enables

5

keratinocytes to adhere to neighboring cells and the extracellular matrix (ECM) (Braga et al. 2016 and Mertz et al. 2013). Due to formation of strong cell-to-cell interactions, keratinocytes are mostly organized as epithelial cell sheets in the skin (Zhou et al. 2015). As a collective, keratinocytes form a protective layer that shields cells in the inner tissues of the body from harmful elements such as UV radiation, bacteria, and physical damage. As keratinocytes differentiate, they undergo a process called keratinization. During keratinization keratin proteins are produced and polymerized leading to terminal differentiation of the cell (Weedon et al. 2010). Terminal differentiation is a form of programmed cell death, where the cells are packed with keratins and their organelles are deteriorated (Eckhart et al. 2000 and Bloom et al. 1995).

As the keratinocytes differentiate, they migrate through the epidermal layers. One of these layers is known as the stratum basale. This is the bottom layer of the epidermis and its role and function will be described below.

The stratum basale consists of stem and progenitor cells with a cube like shape. The stem cells and progenitor cells are precursors to keratinocytes, and the stem cells can divide indefinitely while the progenitor cells can only divide a limited number of times. When located in the stratum basale, these cells are actively dividing. However if the cells detach from the basal membrane the progenitor cells will begin to differentiate to form keratinocytes (Yousef et al. 2017). Human skin cells only divide in the basal layer of the epidermis, and the length of its cell cycle varies depending on interaction with its environment (Gandarillas et al. 2012). However, they are known to divide approximately every 24-30 hours in-vivo and in an article by Van Erp et al. 1996 the cell cycle duration of these cells was determined to be approximately 28,2 hours. Cells in the stratum basale are involved in cell migration and formation of new skin layers after wounding.

1.2.1 The human epithelial cell line HaCaT

The cell line used for studying cell division is a human keratinocyte cell line called HaCaT (300493, CLS)(Boukamp et al. 1988). This cell line has been genetically modified to become immortal. Immortal cells are programmed to divide continuously. An advantage of using HaCaT cells to study the mechanisms and processes that occur in human skin, is that these cells have retained cellular behaviors associated with progenitor skin cells of the stratum basale. This is because they are able to migrate and proliferate, as well as interact with neighboring cells like normal skin cells, forming epithelial sheets (Boukamp et al. 1988 and Steinberg et al. 1983). In

addition, HaCaT cells have full keratinization capacity, and thus, HaCaT cells are a good tool for studying regulation of keratinization of human cells (Boukamp et al. 1988). As the cells have full epidermal differentiation capacity through keratinization, HaCaT cells are capable of terminal differentiation, however by culturing the cells in growth mediums with low calcium (Ca^{2+}) content terminal differentiation is greatly reduced (Boukamp et al. 1988).

1.3 Cell migration

Cell migration is the movement of a single cell or a group of cells, and this movement often occurs as a response to chemical or mechanical signals. These forms of stimuli can derive from the external environment or derive internally from the migrating cell itself. The table below outline which forms of stimuli activates cell migration.

Table 1.1.1: Different forms of stimuli that can activate cell migration.

Stimulus	Description
Durotaxis	Cell migration as a result of a gradient of extracellular stiffness, or stress.
Chemotaxis	Cell migration as a result of gradient of increasing or decreasing concentrations of a chemical substance.
Haptotaxis	Cell migration by means of a gradient of adhesion molecules in the ECM.
Galvanotaxis	Cell migration guided by an electric field or current

Prior to cell migration, the cells polarize. When a cell is polarized it displays some sort of asymmetry. This asymmetry can be in the form of a shape difference or organization of cellular components and this polarization process is facilitated by the cells' cytoskeleton (Vicente-Manzanares et al. 2000). During cell migration the main purpose of cell polarity is to give directionality to the cells.

Cells can either migrate as single cells, or collectively through a process called collective cell migration. Cells can perform two types of single cell migration depending on the respective cell's

morphology, meaning its size and shape, and its environment. One type of single cell migration is called amoeboid migration and it is characterized by a low degree of adhesion to its environment. The other form of single cell migration is mesenchymal migration. In mesenchymal cell migration, the migrating cell has a high degree of adhesion to its environment mediated by cadherins as mentioned in 1.2. During single cell migration, cell motility is driven by contractions and retractions of an organized actin-structure, which is a part of the cytoskeleton and is connected to extracellular adhesion complexes (Svitkina et al. 2018).

In order to understand processes in human skin like wound healing, it is important to understand molecular mechanisms that drive cell migration and proliferation. However, the mechanisms that activate collective cell migration are not well understood as opposed to mechanisms of single cell migration (Friedl et al. 2009). Below are some key characteristics of collective cell migration described.

Collective cell migration is characterized by a group of cells that interact with each other through cell-cell junctions (Friedl et al. 2004, Montell et al. 2008 and Carmona-Fontaine et al. 2008). Like with single cell migration, collectively migrating cells are polar on a single cell level. Additionally, collectively migrating cells affect each other by signaling mechanisms, resulting in a large unit of cells with the same polarization (Friedl et al. 2009).

During collective migration cells migrate as a cohesive group as a result of cell-environment interactions and cell-cell communication. Cell-environment interactions that cause migration are the interactions of proteins in the extracellular matrix (ECM) to cells. Cell-cell adhesions, or adherens junctions, are regulated by expression of adherens junction proteins (CAMs), such as cadherins and integrins. Adherens junctions are closely connected to the cytoskeleton as intracellular actin can bind to cadherin, allowing structural changes in the cytoskeleton to impact the positioning and connectivity of the cell-cell connections (Niessen et al. 2007).

Collectively migrating cells can migrate in many different group-shapes such as 3D strands, clusters or 2D sheets, as is the case for keratinocytes. 2D sheets collectively migrates over surfaces, either continuously, as is the case for gut intestinal epithelium, or as a response to signaling, as is the case for epidermal keratinocytes (Friedl et al. 2009). When epidermal cells migrate collectively, cells have been found to transit from an epithelial to a mesenchymal cell state. This

process is known as the epithelial-mesenchymal transition (EMT), which is driven in part by the downregulation of epithelial cadherins and expression of mesenchymal cadherins leading to a reduction in the strength of cellular adhesions. (Piacentino et al. 2020 and Campbell et al. 2019).

1.3.1 Live cell imaging

Microscopy is a technique commonly used to study objects that cannot be seen by the naked eye, such as cells. Live cell imaging is a method where images of living cells are acquired through a microscopic lens over a period of time. The microscope is equipped with a chamber that provides an environment suitable for the organism being imaged. This usually includes a stable temperature of 37°C and an atmosphere with 5% CO_2 (Baker et al. 2010). In live cell imaging, an image (or field of view) is also referred to as a frame. Thus, the time interval, meaning the time between imaging of a specific frame, is the same as the frame rate. The result acquired from a live cell imaging experiment consists of a collection of frames, a so called time-lapse that can be used for further analysis of changes in cells over time.

Live cell imaging can be performed on different types of microscopes, such as widefield and confocal microscopes. Widefield microscopy is a technique where the entire field of view is exposed to light from the light source by focusing the light through the objective onto the sample. In widefield microscopy many light sources can be used, but the most common light source is LED light. In confocal microscopy the light source is usually a laser. The light source is focused through an objective, as in widefield microscopy, however, in confocal microscopy the beam of light is focused to illuminate one spot of the sample. In confocal microscopy, the beam of light excites a single plane in the sample (referred to as a z-plane), the emitted light in focus pass through a pinhole and is registered by the detector, while out-of-focus light is not detected. This result in an image with higher resolution and improved signal-to-noise ratio compared to widefield microscopy.
However, since widefield imaging directly images all the cells in a frame, this technique is generally less time consuming. A simplified illustration of how light travels in widefield and confocal microscopy is shown in figure 1.1.4.
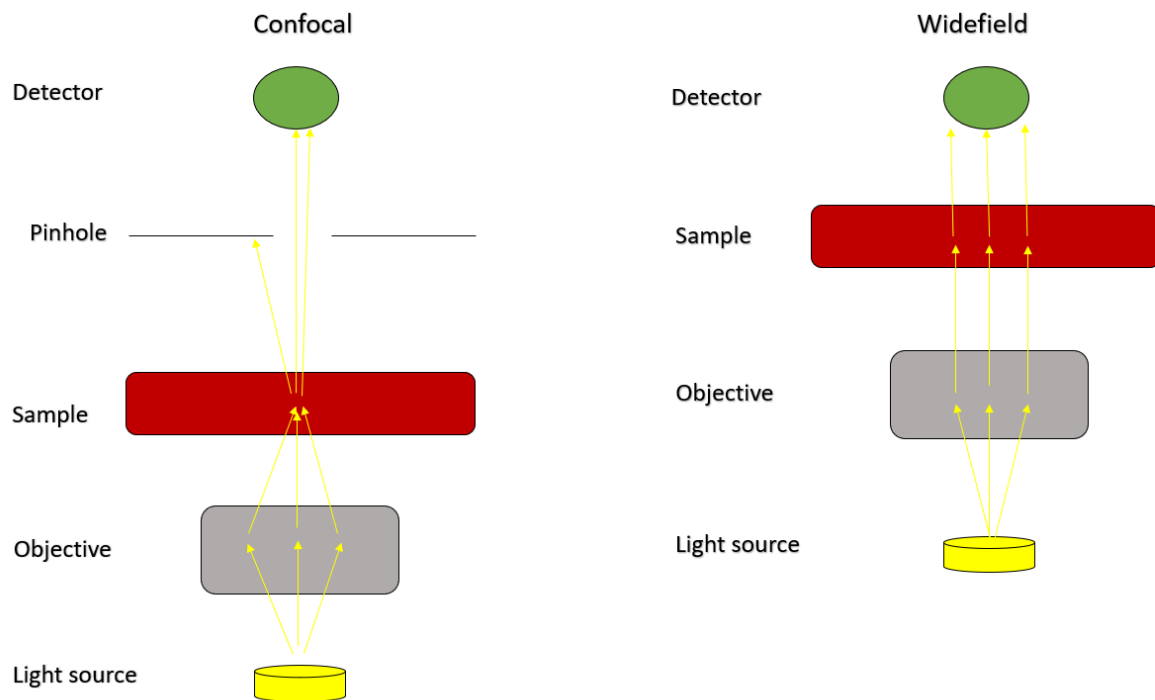
Figure 1.1.4: Illustration summarizing the main differences between widefield and confocal microscopy.

Microscopes can be either upright (objective above the sample) or inverted (objective below the sample). Inverted microscopes are most frequently used in live cell imaging, since cells adhering to the bottom of a dish or multiwell plate can be imaged directly through the glass using objectives with high numerical aperture (NA). Sterile working conditions are important when performing microscopy of live cells, as samples can become contaminated and yield unfavorable results. This is avoided when using an inverted microscope because there is no contact between the objective and the sample.

1.3.2 The established experimental system

Collective migration of keratinocytes has been studied in the local research group at OUS in order to better understand how it is coordinated with cell division, and experiments performed during the course of this master project were based on their previously published *in vitro* experimental system (Lång et al. 2018). In this experimental approach, confluent HaCaT cell sheets are exposed to three different forms of treatment, resulting in different cellular behaviors that can be observed by live cell microscopy. As the cells require serum to grow and divide, the cellular response to

serum deprivation is to exit the cell cycle and become synchronized in a G_0 resting state referred to as quiescence (Oki et al. 2014 and Lång et al. 2018).

When quiescent serum deprived cells are re-stimulated with serum the cells will resume cell growth and re-enter the cell cycle, going through the G_1, S and G_2 phases, followed by mitosis. The established system was shown to synchronize HaCaT cells and induce an increased rate of cell division compared to asynchronous cells (Lång et al. 2018). In the article written by Lång et al. 2018 cells that were serum deprived for 48 to 72 hours followed by serum re-stimulation were shown to have an increased rate of collective migration compared to cells that were not serum deprived and cells that were serum deprived for a shorter period of time. Something worth noting is that there is no chemical gradient or other external stimulus that might activate cell migration in this system. It is hypothesized that the cause of collective cell migration activation in this system is a build-up of tension across the cell sheet during serum starvation. This hypothesis is supported by previously published data, which show that the length of time cells are in quiescence regulates the collective cell migration response (Lång et al. 2018).

By performing live cell imaging with an inverted widefield microscope of re-stimulated cells that were serum deprived for 72 hours time-lapse image sequences of live cells going through synchronized migration and proliferation were acquired. Images of the cell cultures were used as datasets to train a deep learning model developed with a software called StarDist (mentioned below; section 1.6). This model was generated with the purpose of detecting and segmenting mitotic cells. In summary, the goal of this project is to perform live cell imaging of synchronized serum re-stimulated cells, analyze the images using a newly developed StarDist model that detects dividing cells and extract quantitative data using the Fiji ImageJ software (Schindelin et al. 2012).

1.4 Regulation of cell behavior

In response to external factors such as stress and access to nutrients, cells activate and repress signaling pathways. These external factors are referred to as bio signals, or stimuli, and leads to a series of chemical reactions that regulate cell functions. The intracellular molecules involved in cell signaling pathways can be smaller molecules or macromolecules, such as proteins. Cells are able to respond to stimuli that change the cell's mechanical properties or regulate cell growth and proliferation. Below, the role and properties of two proteins that are involved in the regulation of

these types of cell behavior will be discussed. These proteins are the Aurora B and the Yes-associated protein (YAP) proteins.

1.4.1   Aurora B and midbodies in dividing cells

The Aurora B protein is a kinase that is part of the chromosome passenger complex (CPC). In early mitosis, this complex is recruited to the chromosomes. The CPC re-localizes to central spindle microtubules (located between the two chromatids) at the onset of anaphase, and forms a structure called the midbody. Since the midbody is located between the separating chromatids immunofluorescent staining of Aurora B proteins can be used to detect daughter cells that originate from the same dividing cell. The midbody is involved in anaphase chromatid compaction, anaphase spindle stabilization and cytokinesis (Carmena et al. 2012). The localization of Aurora B in midbodies during mitosis is shown in figure 1.1.5 Since Aurora B activity and localization is closely linked to mitosis, imaging of immunofluorescently stained Aurora B is an effective way of observing daughter cells formed after mitosis.
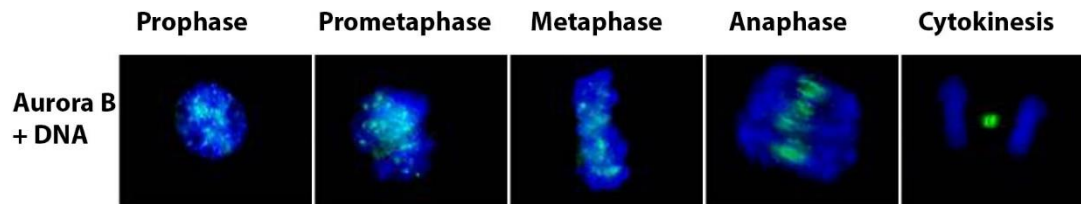
Figure 1.1.5: Localization of Aurora B(green) and DNA (blue) during different stages of mitosis. Aurora B is located in the cell's chromatin during the early stages of mitosis and is located between the daughter chromatids after chromatid-separation as a part of the midbody. Modified figure from Sun et al. 2008.

1.4.2   YAP regulation of gene transcription and stress sensing

The Yes-associated protein (YAP) is a regulator of transcription, as well as cell-cell and cell-ECM connections and is associated with the Hippo signaling pathway (Zhao et al. 2008, Kim et al. 2015). The Hippo pathway is a signaling pathway involved in regulating the levels of mitogenic lipids (Yu et al. 2012), which in turn regulates cell growth and survival (Oka et al. 2008). The Hippo pathway allows for cells to adapt to changes in the environment by inhibiting cell growth as well

as promoting cell death through apoptosis (Zhao et al. 2007, Pan et al. 2007 and Harvey et al. 2013). When the YAP protein is activated, it is transported into the nucleus where it functions as a transcriptional co-activator that promotes cell growth as well as epithelial-mesenchymal transition (EMT) (Zhao et al. 2008).

In a paper published by Zhao et al. 2007 it was found that the cell density promotes Hippo signaling, which leads to YAP inhibition. Prior to inhibition, YAP is translocated from the nucleus to the cytoplasm where it is phosphorylated (Pan et al. 2010). In summary, YAP is inhibited by the Hippo pathway and is activated by mitogenic signals (Miller et al. 2012).

In addition, YAP is mechano-sensitive, meaning that it can be activated by biomechanical stress. This activation is caused by changes in the cell's cytoskeleton structure as a response to stress or tension. The actomyosin is part of the cytoskeleton network and is composed of an actin and myosin complex that form contractile filaments (Misra et al. 2018). This network reacts to increased or decreased biomechanical tension through cadherin-containing adherens junctions.

As mentioned earlier, high cell density decreases the activation of YAP. Changes in cell density not only downregulates YAP by contact inhibition, it can also lead to activation of YAP when cell density decreases. The decreased cell density result in increased biomechanical stress in epithelial layers due to cell stretching and this stimulus causes YAP to regulate cell connections to neighboring cells as well as cell division (Misra et al. 2018 and Trepat et al. 2004).

The cell line used in this project is adherent keratinocytes that form strong intercellular connections to neighboring cells and the ECM. YAP regulates cellular connections by downregulating cadherin in cells, reducing adherences to neighboring cells and leading to EMT of cells, which is needed for collective migration (Park et al. 2019). When these cells are exposed to different treatments, such as serum starvation or serum re-stimulation, the intercellular connections will be affected as shown by the results of Lång et al. 2018. These changes may lead to an increase or decrease in YAP activity through the Hippo pathway or by changes of cytoskeletal tension.

1.5 Visualization of cellular proteins

Cell components are usually not very easy to observe in traditional light microscopy. The challenge of observing specific cell components can be addressed by performing fluorescent microscopy. In fluorescent microscopy cells are treated with a substance that emits light (when absorbing light or

another form of electromagnetic radiation) at specific wavelengths. Fluorescent labelling of proteins is performed either by adding a fluorescent protein tag to living cells or by immunofluorescence (IF) staining of fixed cells.

1.5.1 Fluorescent protein tags

A fluorescent protein tag is a protein that can chemically attach to a protein in the cell. There are three main ways by which a peptide tag binds to a protein. One way is by enzyme mediated bonding, another is bonding with small molecules such as metal ions, and the third way is through peptide-peptide interactions (Lotze et al. 2016). The protein tags contain a fluorescently active fluorophore that can be observed in microscopes equipped with a light source transmitting light in the wavelength that causes the fluorophore to emit light. An advantage of fluorescent tags is that they can be used to label proteins in live cells since they are non-toxic. One example of this is the mCherry labelled Histone H2B protein used for visualization of chromatin in the HaCaT cells used in the live cell imaging experiments performed in this project.

1.5.2 Immunofluorescence (IF) staining

Immunofluorescence can be applied to light microscopy and is a technique primarily used for microscopy of biological samples. The goal of immunofluorescence microscopy is to stain specific proteins in a cell. The fluorescent molecules used are antibodies or dyes, which form complexes with proteins in the cell and emit light in the same way as fluorescent tags. Antibodies are large proteins that bind to specific proteins or antigens.

Immunofluorescence (IF) staining of samples can be either direct or indirect. Direct IF staining includes a fluorescently tagged antibody or dye that specifically bind to the protein of interest. Indirect IF staining includes two antibodies, the primary antibody that binds specifically to the protein of interest and the secondary antibody that contains the fluorescent tag.

Immunofluorescence assays can only be performed on fixed and thus dead cells. Fixation is a treatment where the cells' organelles and molecules are immobilized and preserved in the state they were in during fixation. When cells are fixed the macromolecules in the cell are chemically bound to each other by a process referred to as cross linking. Fixed cells are also permeabilized

and blocked prior to antibody staining. When cells are permeabilized cellular membrane lipids are removed to allow large molecules like antibodies to get inside the cell. Excess antibodies introduced during staining can occupy empty space in the cell and form non-specific binding with other molecules in the cell. When cells are blocked non-specific binding of antibodies are prevented by blocking leftover spaces in the fixed cell (Ahirwar et al. 2015).

1.6 Deep Learning – manual versus automatic image analysis

Traditional computer programs can perform many tasks where an input of information is processed through a program to produce an output more effectively than the human brain. They can do things such as acquiring and storing large amounts of information in a very short time and solve advanced numerical problems effectively and precisely. However, when it comes to problems that require intelligence, meaning the ability to acquire and apply knowledge and skills, these programs fall short. Examples of tasks that require intelligence are tasks where for example perception or common-sense reasoning is required to generate satisfying outputs.

The ability of a machine to process information like a human brain is something that has been under research and development for a long time. The first artificial intelligence (AI) programs were invented in the 1950's and these programs were able to perform simple tasks which require intelligence. One example is the first successful AI program invented by Christopher Strachey that was able to play checkers against human players (Copeland et al. 2020).

Machine learning is a subcategory of AI. Machine learning programs are able to perform tasks that require intelligence, meaning they are not merely programmed to mimic human behavior, and they are also able to learn from data without being explicitly told how to do so through a complex set of rules. Machine learning can be described as a series of algorithms, called neurons, that function as a statistical model that tries to understand the structure of input and output data, provided by the researcher (Edwards et al. 2018).

Deep learning is an even more sophisticated subcategory of machine learning. Like machine learning, deep learning programs are able to learn without being explicitly programmed to do so, provided it is given a suitable dataset of inputs and outputs, a so-called ground truth. What separates deep learning from machine learning is that it is built up of a network of layers of interconnected machine learning algorithms. This network is also known as a convolutional neural

network, or CNN for short. As mentioned earlier, the neurons are structured in layers. The first layer is called the input layer, and this layer of neurons receives the initial input from the dataset. The last layer is called the output layer, and this layer generates the final output from the neural network. The layers between the input- and output layers are called the hidden layers, and the number of hidden layers in a CNN varies depending on the network. When a layer of neurons receives an input, it generates an output that is provided to another layer, and this procedure is repeated in all layers until the output layer generates a final output.

Convolutional neural network
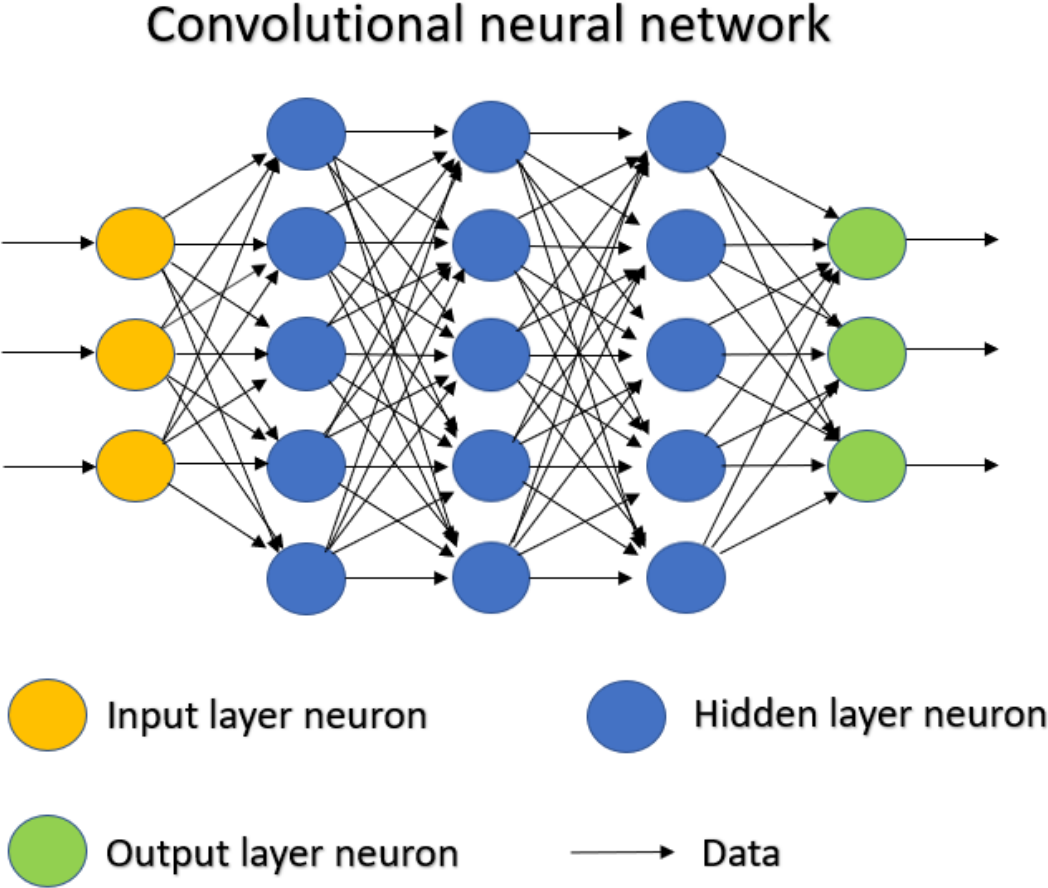
Figure 1.1.6: An illustration of the basic structure of a deep learning network. Each neuron is shown as a circle, and neurons on the same vertical axis make up a layer. The input layer (yellow) receives an input, this layer then processes the data and the input layer sends the processed data to the hidden layer (blue). This process is repeated through all layers until the data reaches the output layer (green). The output layer processes the data and provides a final output based on the data that is provided. The data exchanged between the neurons are matrix values. The figure is inspired by Xing et al. 2018.

A deep learning program analyzing images require an attribute called computer vision. When identifying objects in an image or video the program needs to be able to classify, detect and segment objects. When an object is detected, its area in the image is located, usually with a square, and when an object is segmented the object's shape is refined, locating its boundaries. An illustration of object detection by deep learning is shown in figure 1.1.7. There are two forms of segmentation, semantic segmentation classifies every pixel in a class (Shelhamer et al. 2017), while instance segmentation labels separate objects belonging to the same class as separate instances of that class (Hafiz et al. 2020).

Deep learning programs have two general methods of object detection, localization, and segmentation. One approach is the bottom-up approach. This approach first classifies every pixel into different classes, such as nucleus and background, and then detects and segment objects by grouping pixels together as objects. Another approach is the top-down approach. This approach first localizes objects with a rough shape representation, normally in the form of squares, and then segments the objects by classifying pixels to find the object shape (Borenstein et al. 2002 and Borenstein et al. 2004).
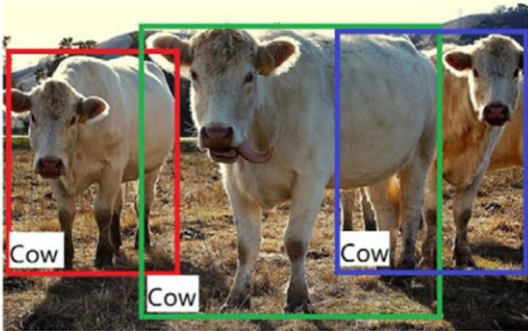
(a) Image Classification (b) Object Detection
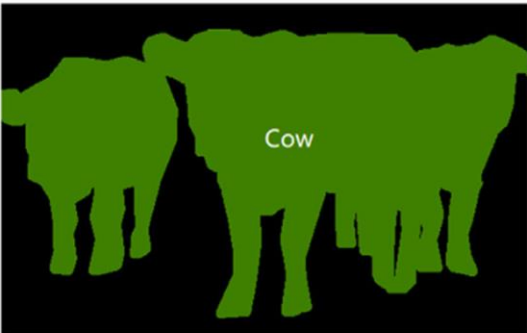
(c) Semantic Segmentation (d) Instance Segmentation

Figure 1.1.7: The hallmark steps of segmentation of multiple objects in an image. First the image is classified and then objects are detected using these classes. The objects are then segmented with semantic and instance segmentation. Adapted from Pentagonspace et al. 2020.

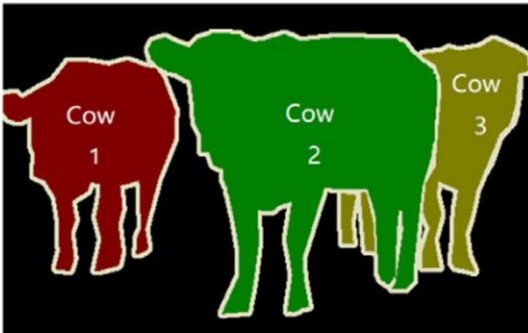In an object detection pipeline the deep learning model will generate proposals for objects from classified pixels. A proposal can be described as a proposal for a prediction. As most objects in an image consists of hundreds or thousands of pixels a large number of proposals will be generated for the same object. To prevent the same object from being registered multiple times a method called non-maximum suppression, NMS, is used (Sambasivarao et al. 2019). The first step in non-maximum suppression is to select the proposals with a confidence score over a preselected confidence, or probability, threshold for further evaluation. Other proposals with lower scores will not be considered. Each detection is given a confidence score, this score is a numeric value (0-1) describing the probability of a detected object belonging to a particular class (Koirala et al. 2019). After the proposals with the highest confidence scores have been selected, the overlapping area of the proposals is used to evaluate if the proposals are objects. If two, or multiple proposals have an intersection over union (IoU) value above a predefined IoU threshold, the proposal with the highest confidence score is selected as the object by the model. The IoU can be defined as the fraction of a detected area that overlap the area of another detection, or proposal, divided on the total area of both detections in union. This is illustrated in the figure below (figure 1.1.8).
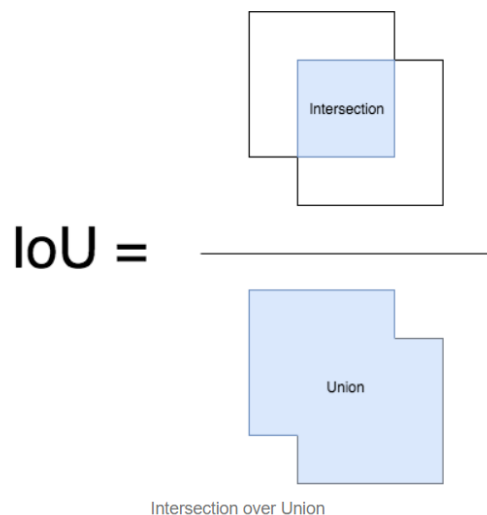

Figure 1.1.8: Intersection over union is described as the overlapping area of two objects divided by their union area. Adapted from Sambasivarao et al. 2019.

By performing non-maximum suppression multiple detections of the same object is reduced or prevented. However, this method can lead to objects in an image being undetected if the confidence

18

threshold or IoU threshold value is set too high. Due to this it is important to modify these thresholds to best fit the neural network being used and the data being analyzed.

When a CNN detects and segments objects, it makes predictions. These predictions can be correct or incorrect, and these predictions are often analyzed during and after training of a neural network. To measure the network, or model, performance its predictions are compared to a ground truth image. A ground truth image is an image where all objects are correctly classified and segmented. Correct model predictions are known as true positives, TP. Incorrect model predictions are known as false positives, FP. Objects that are not predicted by the model are known as false negatives, FN. These categorizations can be used to estimate the model's precision, accuracy and recall using the formulas shown below. Precision, accuracy, and recall show the model's ability to predict objects correctly.

$$Recall = \frac{TP}{FN+TP} \qquad Precision = \frac{TP}{FP+TP} \qquad Accuracy = \frac{TP}{FP+FN+TP}$$

These calculated values do not give any information about a model's ability to correctly segment the shape of objects. To estimate this, IoU values of model predictions and ground truths are measured. As mentioned above IoU measures the intersection area divided by the union area. This means that false positives and false negatives will be a part of the union area causing IoU to partly estimate the model's accuracy as well as object segmentation. This can be corrected by measuring the mean matched score. The term mean matched score, as defined by the developers of StarDist, measures the IoU of model predictions and ground truths, but excludes false positives and false negatives (Schmidt et al. 2018). Neural networks are often constructed with something called a loss function. The loss function is used to generate an error value and there are several different loss functions that can be applied depending on the neural network and the data being analyzed. The most common loss function is cross entropy for deep learning models performing image classification (Nichols et al. 2019 and Koech et al. 2020). The loss function requires pairs of raw data images and ground truths to estimate the loss value when using image classification models. The paired images used can be images previously seen in training, and using these images in the loss function estimates the training loss. The paired images used can also be images not seen during training, and using these images lead to the estimation of the validation loss. Comparing these values give an insight into the model's ability to segment new images compared to images used in training.

1.6.1 **StarDist**

StarDist is a supervised deep learning program that was designed by Uwe Schmidt and Martin Weigert to identify objects in images, and it is especially suited for nuclear segmentation (Schmidt et al. 2018). A supervised deep learning program is a program that is trained through training data consisting of inputs and paired correct outputs, referred to as ground truths (Wilson et al. 2019). StarDist is based on many different programming languages, as is shown in table 1.1.2, and uses the deep learning library Keras.

Table 1.1.2: The percent wise fraction of programming languages used in StarDist.

Programming language	Percentage of program
C	58,1%
C++	28,0%
Python	12,1%
Jupyter notebook	1,6%
Groovy	0,2%
Makefile	<0,0%

While many deep learning programs identify objects purely through a top-down or bottom-up approach, StarDist performs both top-down and bottom-up analysis on images (Schmidt et al. 2018). StarDist identifies objects as star-convex polygons for every pixel. To explain this further, for each pixel with an index i, j, the Euclidian distance to the nearest object edge is determined as the radial distance, $\{r_{i,j}^k\}_{k=1}^n$, along a set of n predefined radial directions with equidistant angles. This type of image analysis is that of a top-down approach. However, at the same time as the radial distances of each pixel is determined, a bottom-up image analysis is being performed and by measuring the object probability ($d_{i,j}$) of each pixel. This approach works by labelling each pixel as for example 'nucleus' and 'background'. Then the Euclidian distance to the nearest background pixel is measured for every pixel labelled a nucleus in this instance. This will result in pixels closer to the center of the object having the highest object probability values. Each radial distance measurement from a pixel labelled as an object serve as a polygon proposal, and only the objects with measurements that have a sufficiently high object probability value will be evaluated in the

model. In part, object probability has the same function as NMS for predicting star convex polygons. An illustration of object probability and radial distance is shown below (figure 1.1.9).
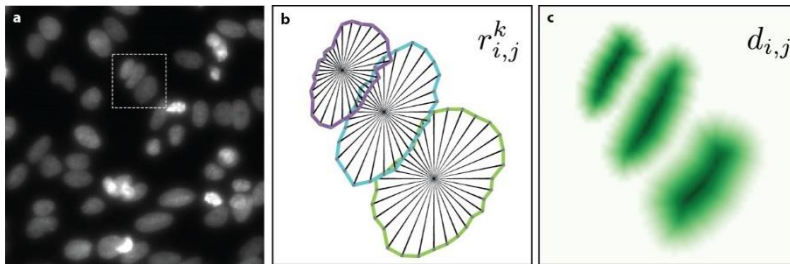

Figure 1.1.9: A projection of radial distance and object probability by StarDist. The nuclei image(a) has a box surrounding three nuclei. An illustration of the radial distance(b) and object probability(c) of these nuclei are shown. The radial distance is extracted from rays projected from the nuclei center to its edges. The object probability is illustrated by pixel intensity as the dark green pixels have a high object probability and light green pixels have a lower object probability. Modified figure from (Schmidt et al. 2018).

When generating a deep learning model for detection of dividing cells in live cell microscopy data, it is important to use software that is designed to process the data properly in order to give accurate results.

As mentioned earlier, StarDist is designed to segment nuclei images. StarDist's ability to segment nuclei images has been found to be more accurate compared to other image segmentation software like U-Net (2-class), U-Net (3 class) and Mask R-CNN, when the IoU threshold was relatively low (0,5-0,7) (Schmidt et al. 2018). However, it was outperformed by the other models when the IoU threshold was higher(0,75-0,9) (Schmidt et al. 2018). This means that StarDist is especially suitable for analysis of crowded nuclei images, due to overlapping nuclei being excluded to a lower degree. It was also found that StarDist was better equipped to segment out-of-focus images, and this is beneficial because live cell experiments that run continuously for a long period of time may eventually lose focus.

1.6.2 TrackMate plugin in Fiji ImageJ

Fiji ImageJ is an image processing software often used for biological image analysis. The ImageJ software is equipped with many functions that enable prototyping of images-processing algorithms (Schindelin et al. 2012). Examples of prototyping image-processing algorithms are adjustment of brightness to better observe nuclei in an image, changing image properties such as the pixel/μm ratio and marking, or annotating, specific regions as regions of interest (ROI). ImageJ has many

different image-processing algorithms, or plugins. Two plugins that were used for image processing together with the StarDist analysis in this project were ROI manager and TrackMate (Tinevez et al. 2017).

ROI manager is a tool that allows the user to annotate certain regions of interest in images. These annotations can be saved as a file and be further used to project a ROI map as well as be projected on the raw image itself. ROI maps that are generated from the original raw nuclei images can be used as ground truth images during model training and as mentioned previously, ground truth images are necessary to train supervised deep learning models.

When working with annotations of image objects, an object or cell is not initially recognized between frames in ImageJ. Using the TrackMate plugin an object can be tracked through multiple frames. This provides many possibilities for quantitative and qualitative analysis of live cell imaging data. Some possibilities are the analysis of cell migration patterns, as well as changes in cell size and shape over time, either on a single cell level or collectively.

2   AIM OF THE STUDY

Cell migration and proliferation play an essential role during wound healing, development, and tumorigenesis. The aim of the present project was to develop a deep learning-based methodology for analysis of cell division and cell migration based on live microscopy data. To achieve this a previously published experimental approach was employed, where coordinated collective cell migration and subsequent cell division is activated in confluent monolayers of the human keratinocyte cell line HaCaT. An automated microscopy-based pipeline was created by generating a deep learning model with the StarDist software. The secondary objective of developing the pipeline was to use the developed method to analyze the acquired nuclei images of HaCaT cells in different ways. One of these goals were to develop a StarDist model able to identify mitotic nuclei and use the developed model to quantitatively analyze mitotic cells as part of the pipeline. Another goal was to use the developed pipeline to extract information about the cell cultures, such as average nuclei shape and area, as well as the average total number of cells per frame over time. The average number of mitotic cells per frame over time and total cell population over time provides information regarding the rate of cell division in these cell cultures. Another goal of the method was to build a pipeline that was able to track cells across frames to analyze the development of cells over time to acquire information about the shape and size of nuclei prior to and after cell division as well as to analyze the relation between nuclei shape and size during cell migration.

3 MATERIALS AND METHODS

All cell lines, materials, instruments, and software used in this experiment are summarized in the appendixes A-C (table A.1, B.1 and C.1).

3.1 Cell culture work

HaCaT (300493; CLS) cells were grown in Iscove's modified Dulbecco's medium (IMDM; MedProbe) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific) and 1% PenStrep (Lonza). The cells were cultured in T75 culture flasks (VWR) at 37 °C and 5% CO2.

3.1.1 Cultivation and passaging of cells

Materials:

- Cell lines: HaCaT WT (Boukamp et al 1988) and HaCaT cells stably expressing mCherry-Histone H2B proteins (Lång et al 2012) were used.
- IMDM with 10% FBS and 1% PenStrep
- T75 culture flasks
- Phosphate buffered saline (PBS)
- Trypsin-EDTA
- 15 mL Falcon tube

Cells were repeatedly passaged throughout the project period to sustain a viable cell culture of the respective cell lines HaCaT WT (Boukamp et al 1988) and HaCaT mCherry-Histone H2B (Lång et al 2012).

1. This was done by removing the growth medium (12 mL) from the flask and rinsing the cells twice with 5 mL PBS.
2. Subsequently, 2 mL trypsin-EDTA was added to the flask and the cells were incubated with trypsin for 10-15 min at 37°C. Trypsin breaks down the intercellular bonds and the bonds between the cells and the plastic surface in the T75 flask. After approximately 10 minutes the cells should be completely detached from the flask, if the cells are still somewhat attached to the flask, the flask is lightly tapped to detach the rest of the cells.

3. When the HaCaT cells no longer adhere to the culturing flask the cells and trypsin form a cell suspension. The adhesion inhibiting trypsin is neutralized by adding 8 mL of IMDM medium to the flask and the cell suspension was thoroughly resuspended.

4. The cell suspension was transferred to a Falcon tube and then split in a 1:10 ratio, where 1 mL of the cell suspension was re-inserted in the culturing flask and 14 mL of fresh IMDM medium containing 10% FBS was added to the culturing flask.

5. The cells were then incubated at 37°C and 5% CO_2 to allow the cells to settle down in the flask forming new adhesions. For continuous growth, the cell suspension was diluted 1:10 every second day.

3.1.2 Counting of cells

Materials:

- Trypan-blue stain
- Countess 3FL Automated Cell Counter
- Countess Cell Counting Chamber Slide
- 15 mL Falcon tube
- 1,5 mL Eppendorf tube

When preparing cells for imaging, either live cell imaging or image acquisition of fixed cells, it is important to count cells and seed an appropriate number of cells for further processing. To perform a cell count, the cells in the culturing flask are exposed to trypsin and resuspended in growth medium before being transferred to a Falcon tube (as described in section 3.1.1) and counted.

1. The cell suspension is mixed by flushing the mixture up and down with a pipette. 15 µl of the cell suspension was transferred to an Eppendorf-tube and mixed with 15 µl Trypan-blue stain. Trypan-blue stain is a dye that is absorbed in the cytoplasm of dead cells and they get a dark-blue stain. In live cells the dye does not pass through the membrane and this allows for the counting of live cells by excluding all dead cells.

2. 10 µL of this mix was transferred to a Countess Cell Counting Chamber Slide and counted with the cell counter (Countess 3). The countess instrument estimates the number of live and dead cells per mL in the cell suspension.

3.1.3   Cell starvation and stimulation

Materials:

- Cell lines: HaCaT WT (Boukamp et al 1988) and HaCaT mCherry-Histone H2B (Lång et al 2012) were used.
- IMDM medium
- Fetal bovine serum (FBS)
- Penicillin-Streptomycin

As mentioned in section 1.3.2, keratinocytes that are synchronized by serum starvation respond to re-stimulation with a high degree of collective cell migration followed by a high degree of cell division compared to asynchronous cells (Lång et al. 2018). The treatments used in this project to synchronize and re-stimulate cells were as follows: Unsynchronized cell cultures, called wild type cells during the experiments, cultured in IMDM medium containing 10% FBS were synchronized by serum deprivation followed by serum re-stimulation. To increase cell growth and subsequently, the rate of cell division, IMDM growth medium containing 15% FBS were used during serum re-stimulation. HaCaT cells used for IF staining of fixed cells and live imaging were subjected to 72 hours (h) serum deprivation prior to serum re-stimulation.

1. To starve cells, the 10% FBS culturing medium was removed from the cell culture.
2. The cells were rinsed twice with serum free IMDM medium. If the cells are stored in MatTek 12-well plates, as is the case when performing live cell imaging, the volume of the serum free medium is 2 mL/well. If the cells are stored in petri dishes, as is the case for imaging of IF stained cells, the volume of serum free medium is 4 mL/dish.
3. Finally, serum free medium is applied to the cells and they were incubated at 37°C and 5% CO_2 for 72 h.
4. After cells have been starved for 72 h, they are stimulated with IMDM medium containing 15% FBS. This was done by rinsing the cell culture with serum free medium to remove dead floating cells. Then IMDM medium containing 15% FBS was added to the cell culture. The cells were incubated at 37°C and 5% CO_2 prior to imaging, either in the microscope or in an incubator.

3.2   Immunofluorescence staining (IF)

Materials:

- 12 mm glass coverslips
- HaCaT WT (Boukamp et al 1988)
- Mounting medium with DAPI
- Aurora B primary antibody
- YAP primary antibody
- Alexa Fluor 488 goat anti-mouse secondary antibody
- Alexa Fluor 594 goat anti-rabbit secondary antibody
- 0,5% BSA blocking solution (Bovine Serum Albumin fraction V dissolved in PBS)
- 4 % PFA in PBS
- 0,25% Triton X-100 in PBS
- 76x26 mm glass slides
- 6 cm petri dish
- Leica TCS SP8 STED confocal microscope

The cellular localization of YAP- and Aurora B proteins under three different cell treatments were examined through immunofluorescence (IF) staining and confocal imaging of fixed cells. The three different treatments were asynchronous cells cultured in medium containing 10% FBS, cells that were serum deprived for 72 h in serum-free IMDM medium and cells that were serum deprived for 72 h and re-stimulated for 24 h in IMDM medium containing 15% FBS as described in 3.1.3. To perform this experiment, 3 petri dishes with fixed HaCaT WT cells were used. These dishes were prepared with 10-12 12mm glass coverslips attached to the bottom of the dishes with nail polish. The reason nail polish was used instead of glue is because less chemicals affecting cell cultures dilute into the growth medium. Additionally, removing cover slips attached to the petri dish is easier when using small drops of nail polish than when using glue.

A cell suspension of HaCaT WT cells was prepared using trypsination and counting as described in section 3.1.1 and 3.1.2. The total amount of cells transferred to each dish were $0,8*10^6$ cells. The cells were counted using a Countess 3 automated cell counter as described in 3.1.2, and the volume of the cell mixture transferred was determined by the cell count. The transferred cells were allowed to settle down and attach to the coverslip overnight at 37°C and 5%CO_2.

The HaCaT WT cells were fixed after three specific treatments; asynchronously growing HaCaT cells cultured as mentioned above, cells subjected to 72 h of serum deprivation, and cells subjected

to 72 h serum deprivation and subsequent re-stimulation with serum for 24 h. The petri dish with asynchronous cells were labelled 'Wild type', the petri dish with starved cells was labelled 'starved', and the petri dish with starved and re-stimulated cells was labelled 'starved + stimulated'.

1. Cells were washed by rinsing twice with PBS on ice. It is important that PBS covers the surfaces of all the glass coverslips to avoid cells from drying. 3-4 mL PBS is required to achieve this. Then paraformaldehyde, PFA, was dissolved in PBS to form a 4% PFA solution. Cells were fixed by adding 4 mL of 4% PFA to the dishes for 10 min. Subsequently, the cells were permeabilized in 0,25% Triton X-100 solution for 4 min. Finally, the fixed cells were washed twice with PBS and 4 mL PBS were added to each petri dish. The dishes were sealed with parafilm and stored in a 4°C refrigerator. The cells are now fixed and ready to be used in IF staining.

Fixed cells were stained with protein specific antibodies against YAP and Aurora B according to the procedure described below. When staining cells, it is important to keep the surface of the coverslip wet at all times to avoid cells from drying.

2. The first step in immunofluorescence (IF) staining was blocking the cells. This was done by adding a drop of 0,5% BSA blocking solution to a sheet of parafilm, and then flipping the coverslip onto the drop of blocking solution so that the fixed cells face downwards towards the liquid. The glass coverslips were submerged for 15 min at room temperature.

3. The primary antibodies specific for YAP and Aurora B were diluted in 0,5% blocking solution (BSA) 1:100 and 1:500, respectively. After the cells were blocked as described in step 2, each coverslip was placed on top of 30 µL of the primary antibody solution. The cells were incubated with primary antibodies overnight at 4 °C.

4. Before applying the secondary antibody, the coverslips were washed twice with PBS. Washing was done by placing the HaCaT-covered side of the cover slips on drops of PBS, leaving the glass slide on PBS for 4 min for each washing step.

5. The secondary antibodies were diluted 1:400 in 0,5% blocking solution. The secondary antibodies used were Alexa Fluor 488 goat anti-mouse and Alexa Fluor 594 goat anti-rabbit. As mentioned earlier (see section 1.5.2) the secondary antibodies bind to the

primary antibodies and the fluorescent tag provides signal detection in the form of emitted fluorescent light.

6. The coverslips were submerged in the secondary antibody solution for 2 h at 37 °C.

7. After incubating with secondary antibodies, the coverslips were washed four times with PBS as in step 4.

8. Immunofluorescent staining tends to fade away over time. To reduce fading, a drop of anti-fade mounting medium was applied to a glass slide. This medium also contains DAPI, a fluorescent stain that binds to DNA. The coverslips were placed on the anti-fade mounting medium and were sealed with nail polish to keep them in place.

9. The coverslips are now ready for imaging on a fluorescent microscope. To preserve the stained coverslips for imaging at a later time point, the glass slides can be stored in the dark at 4 °C.

10. Imaging was performed on the Leica TCS SP8 STED confocal microscope equipped with a (HC PL APO CS2) 40x 1.30 NA oil immersion lens. The microscope was equipped with a continuous (470-670 nm) white light laser, a UV laser and hybrid detectors. The wavelength of the white light laser was set to 488 nm and 594 nm, for detection of the YAP and Aurora B proteins respectively. The UV laser has a wavelength of 405 nm and was used for detection of the DAPI stained nuclei.

11. Acquired images were exported from the Leica software and further processed using the Fiji ImageJ software and Adobe Photoshop.

3.3 Live cell imaging

Serum re-stimulated quiescent HaCaT mCherry-Histone H2B cells were monitored by live cell imaging (see section 1.3.1) using a widefield microscope (Zeiss AxioObserver.Z.1). By treating the cells with serum deprivation for 72 h, synchronization of cell divisions occurs after serum re-stimulation was achieved. The images acquired from live cell imaging were used to develop a StarDist model designed to segment dividing nuclei and was used to analyze model performance.

3.3.1 Preparing cells for live cell imaging experiments
Materials:
- HaCaT mCherry-Histone H2B (Lång et al 2012)
- 12-well glass-bottom MatTek plates.

- Countess 3FL Automated Cell Counter
- Countess Cell Counting Chamber Slide
- IMDM medium
- Fetal bovine serum (FBS)
- T75 culture flask
- Trypan blue stain
- 50 mL Falcon tube
- Eppendorf tube
- Zeiss AxioObserver.Z1 wide-field microscope

When performing live cell image experiments on HaCaT cells, 12-well glass-bottom MatTek plates were used. The reason for using plates with glass-bottom wells is because glass-bottom wells provide a higher image resolution than plastic-bottom wells.

1. To avoid too dense or sparse cell populations in the wells the total number of cells in the cell culture was estimated using an automated cell countess instrument, and a total of 400 000 cells were seeded in each well, using the steps described in 3.1.1 and 3.1.2. Additional IMDM 10% FBS medium was added to each well up to a total volume of 2 mL. The plate was incubated overnight at 37 °C 5 % CO_2.

2. The HaCaT cells used for live cell imaging were starved for 72 h and then re-stimulated with 15% FBS (see section 3.1.3). Some wells were used as negative controls by adding fresh serum-free IMDM medium instead of IMDM medium containing 15% FBS. Negative controls can be used to test the StarDist model designed to detect dividing cells in nondividing cell cultures, as well as evaluating the treatment's ability to stimulate synchronized cell division.

3. The MatTek plate was put in the incubation chamber of the Zeiss instrument for more than 30 min before imaging to reach equilibrium in temperature since a change in temperature will affect the focusing on the samples during image acquisition. Live imaging was started around 1-3 h after re-stimulation with serum.

3.3.2 **Image acquisition**
Materials:
- Zeiss AxioObserver.Z1 wide-field microscope

- ZEN 3.1 pro software
- HaCaT mCherry-Histone H2B (Lång et al 2012)
- 12-well glass-bottom MatTek plates containing seeded cells

For live cell experiments, HaCaT mCherry-Histone H2B cells (Lång et al 2012) seeded in MatTek plates were prepared for imaging as mentioned in 3.3.1. Live cell imaging was carried out using a Zeiss AxioObserver.Z1 wide-field microscope equipped with a 10x/0.50 NA Fluar air objective, and the Colibri 7 LED light source set to a wavelength of 590 nm, for excitation of mCherry. The Zeiss instrument was equipped with a CMOS camera (ORCA-Flash4.0; Hamamatsu) for image acquisition and the binning was set to 1. This means that each image includes the maximum number of pixels for the camera used and the pixel size was 0,65 µm. No pixels were combined resulting in the highest resolution for the 10x objective. The microscope has a built-in incubation chamber, where temperature and atmosphere can be controlled. The temperature was set to 37 °C and the CO_2 content was set to 5%. Three independent experiments were performed: 18th of January, 4th of February and 16th of March. Using the ZEN-software the sites for imaging and the time lapse settings were set for the experiments. The time interval was set to 4 minutes, meaning that an image of each site is acquired every 4 minutes resulting in a time lapse video for each site. An average of 3 sites per well was selected in each experiment performed. As mentioned earlier, cell division of HaCaT cells normally begins around 24-30 h after re-stimulation with serum. To acquire a large number of images of dividing cells, imaging was allowed to run until the rate of cell divisions decreased significantly, resulting in a total run time of around 45 hours for each imaging experiment.

3.4 Development of a deep learning model

In order to develop a deep learning model for segmentation of dividing cells, a dataset of paired images needs to be created, consisting of paired input images and ground truth images. In this instance, the input will be images acquired from live cell imaging experiments and the ground truth will be corresponding images, where mitotic cells have been marked (annotated) as regions of interest (ROIs). Creating paired ground truth images were done using the freehand selection tool and LOCI plugin, for creating ROI maps, in the ImageJ software. Processing of paired images and development of deep learning models were done through a Colab notebook for StarDist 2D which is available at the Github homepage for StarDist (Henriques Lab et al. 2020).

3.4.1 Using Fiji ImageJ to process images and create datasets

ImageJ was used to annotate mitotic nuclei as ROIs. The cells with somewhat of a coffee bean shape that showed a clear nuclear separation during anaphase were annotated as dividing cells. Conversely, dividing cells that had very loosely packed chromatin and didn't display a clear nuclear separation were not annotated, since these cells most likely had experienced mis-segregation of chromosomes during mitosis. As mentioned earlier, mis-segregation occurs due to defects in the microtubule spindle apparatus or centrosomes (Compton et al. 2011). The camera used for live cell imaging had a tendency to lose focus during image acquisition. Because of the loss of focus and resolution over time, some images of cells that seemed to have no errors during mitosis appeared rounder than in-focus nuclei, and the chromosome separation was less discernable than images of cells with perfect segregation that were in focus. To accurately annotate all dividing cells that were not subjected to chromosome mis-segregation, the duration of mitosis and the shape of the nuclei were analyzed. An example of a cell division with a successful chromosome segregation that is in-focus is shown in figure 3.1.1, an example of a cell division with a successful chromosome segregation that is out-of-focus is shown in figure 3.1.2, and an example of a cell division with chromosome mis-segregation is shown in figure 3.1.3.

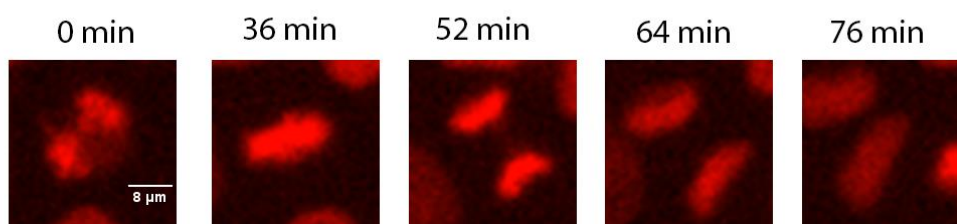| 0 min | 36 min | 52 min | 64 min | 76 min |

Figure 3.1.1: In focus images of a live cell with a functional chromosome segregation. In the first image (0 min), the cell is in prophase. In the second image (36 min), the chromatin structure clearly makes up a metaphase plate. The third image (52 min) shows the chromatin at the onset of anaphase. In the last two images (64 min and 76 min), the cell is in anaphase and telophase, respectively.
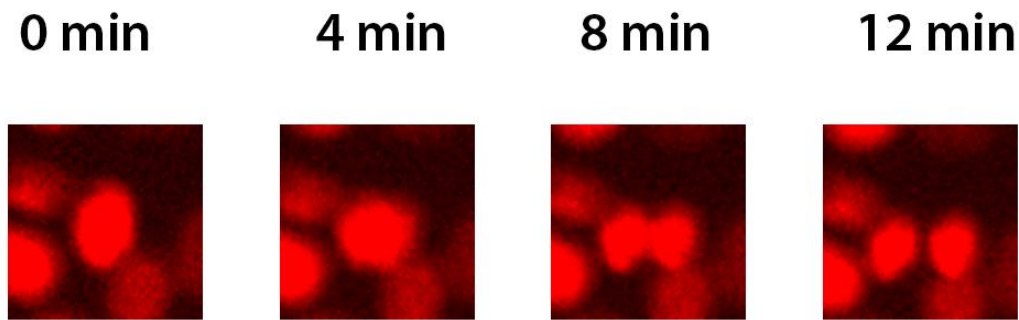
Figure 3.1.2: Images of a cell division with an atypical chromatin shape. The first and the second images (0 min and 4 min) show a cell in metaphase. In the third and fourth images (8 min and 12 min) the cell is going from anaphase to telophase. The images are out of focus, which might be causing the chromatin shape to be more circular than the typical metaphase and anaphase structures. However, the chromatin is clearly separating into two daughter nuclei and is thus considered valid for ground truth annotation.
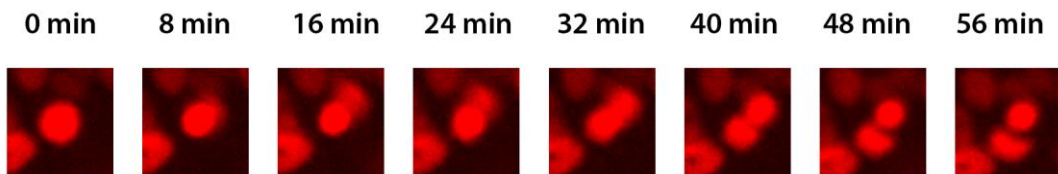

Figure 3.1.3: Image of a dividing cell with a very atypical chromatin shape. It is unclear what phase of mitosis the cell is in in the first two images (0 min and 8 min). Starting from the third image (16 min), the chromatin appears to gradually drift to eventually form two separate nuclei (48 min). The circular shape and the gradual and unclear separation of the chromatin make this cell not suitable to be annotated as ground truth for the StarDist model.

1. The time lapse images acquired with the Zeiss microscope were further processed in ImageJ. Images containing cell divisions were duplicated as single cropped images with a specific size (a width of 226 pixels and a height of 227 pixels).

2. Dividing cells were marked around the edge of the nucleus using the freehand selection tool in ImageJ. All the marked, or annotated, nuclei in each image were saved as a set of ROIs. The shape and coordinates for each ROI are now saved as a file through the ROI manager. A cell was annotated and considered a dividing cell based on the shape of the nucleus and the time of cell division.

3. The annotated nuclei in metaphase- and anaphase states, as shown in figure 1.1.1, were projected on a ROI map using the LOCI plugin. The ROI map, which is an image in TIF-format projecting the marked ROIs, was used as ground truth images combined with the original data. The result of generating ROI maps from nuclei images are paired images where mitotic cells are shown in the ROI map, as illustrated by figure 3.1.4.

4. The ROI maps paired with the original time lapse images were used as datasets for training a model in StarDist (2D). In total, 1000 paired images were generated as a dataset. The StarDist models were trained with datasets ranging from 100-1000 paired images.
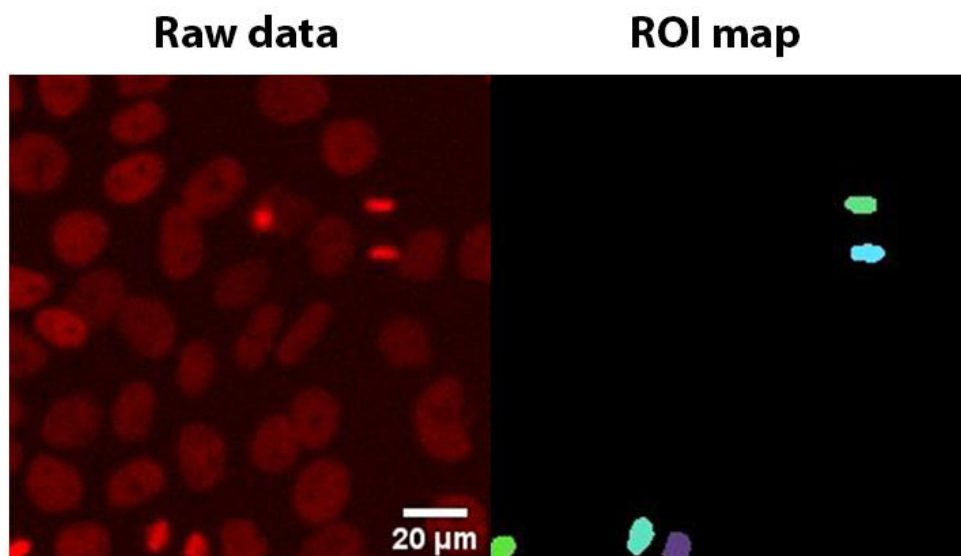

Figure 3.1.4: One of the paired images used for training the StarDist model. The image to the left is raw data acquired from live cell imaging experiments, where the cell nuclei are labelled red. The image to the right is the corresponding ground truth image with marked ROIs. The training image dataset for generating a deep learning model consist of a series of such image pairs.

3.4.2 Developing a StarDist model

Using the StarDist (2D) notebook in Colab, five StarDist models with different parameters and size of the training dataset were trained and generated. The Colab notebook used is available through the Github page of the StarDist developers (Henriques Lab et al. 2020). Training datasets used to generate the model are accessed via google drive. After training the network in StarDist, the generated model and additional data files will automatically be stored on the google drive. Subsequently, the model is downloaded from the google drive to the computer. The parameters

used in model training are chosen in conjunction with the recommended parameter values from the StarDist developers (Schmidt et al. 2018).

When developing a model, there are many parameters and settings that need to be considered. These parameters will be mentioned in the following paragraphs.

The number of epochs was set to 400. Epochs can be described as the number of rounds each image is being trained. Patch size was set to 224. The patch size is the number of pixels the training images will have during training. The patch size must be divisible by 8 and equal to or less than the pixel length and width of the paired images. Due to this parameter, it is important that all training images have the same size. The batch size was set to 4, and this is the number of patches seen by the model in each training step. The number of steps varied between the different models. This is the number of training steps per epoch. Batch size and number of steps dictate how many patches, or pixels, in an image are seen per epoch.

The percentage validation was set to 10. The percentage validation is the percentage of the paired images excluded from model training to be used in a loss function that calculates validation loss. The number of rays was set to 32. The number of rays is the n number of radial distances used when calculating radial distance, $\{r_{i,j}^k\}_{k=1}^n$ (see section 1.6.1). The grid parameter was set to 2, and this setting dictates the size of pixels needed to affect the neural network. Meaning that when training for object detection of large objects, the grid parameter should be increased, and vice versa for small objects. The initial learning rate was set to 0,0003. The initial learning rate determines how quickly the model adapts to new datasets during training. Data augmentation was not used to amplify the differences in the dataset. Data augmentation is most often used when the training dataset is small, but when the training image size is as large as 100-1000 images data augmentation is not needed.

Five models were generated for automated detection of cell divisions. One model was trained with 100 paired images, one was trained with 200 paired images, one was trained with 500 paired images and two models were trained with 1000 paired images (with the number of steps set to 1 or 2). The models were named '100', '200', '500', '1000 1-step' and '1000 2-step', respectively. The reason for creating the different models is to compare their performance based on the size of the dataset and number of steps. Based on the results, the best performing model will be used to extract quantitative data from the live cell images.

There are several different tools to measure model performance. One such tool measures the training loss vs epoch number function, and the validation loss vs epoch number function. As described earlier (1.6), training loss describes an error value for the model predictions and the ground truth for the training dataset, and validation training loss describes an error value for the model predictions compared to ground truth images for paired images not used in model training. These loss value calculations are carried out during each epoch to track training progression in the StarDist 2D notebook (Colab). The graphs plotting training loss and validation loss in relation to epoch number were automatically exported to google drive.

3.5 Using the newly developed model to analyze cell divisions

3.5.1 Analyzing time lapse images with the StarDist model

The cell division models, created in the StarDist notebook, can be applied in the StarDist plugin of ImageJ to create predictions from time lapse images (raw data). The StarDist plugin can be downloaded and installed directly into the ImageJ software. The ideal settings for running StarDist on ImageJ were to set the probability score to 0,5 and the overlap threshold to 0,3. The images are normalized, meaning that the range of pixel intensity values are set to 1-99,8%. When a StarDist model is run on an image sequence in ImageJ ROI files and ROI maps annotating the nuclei predictions are generated.

3.5.2 Analyzing model performance

One way to measure the model performance is to perform error mapping and quality metrics estimation, or quality control analysis (QC). This is done in the StarDist notebook and requires a quality control dataset of paired images previously not seen by the network. When running this step in the notebook, the created StarDist model will predict the ROIs of the nuclei images. This prediction will then be compared with the coupled ground truth images of the dataset and an analysis of the model performance will be executed. The measurements used to evaluate model performance was IoU, precision, recall, accuracy and mean matched score (see section 1.6). These values give a good indication of the models' ability to predict cell shape and to correctly predict

dividing nuclei. 30 paired images not previously used in network training were used to perform quality control analysis.

3.5.3 Quantitative analysis of cells

A method, or pipeline, to analyze live cell images was developed and used to analyze 12 image sequences. The 12 datasets, with a frame size of 354x366 pixels, or 230x237 μm, consisting of 600 frames were extracted from the live cell imaging experiment performed on the 5th of February. The image sequences analyzed started 2 h and 30 min after serum stimulation and the duration 40 hours in total.

When running a StarDist model on an image or a sequence of images in ImageJ, a set of ROIs and mask images corresponding to the model predictions will be generated. The ROIs from the model predictions were used to extract data from the original images. Subsequently, the model predictions (ROIs) are measured. ImageJ measures things such as nucleus area and x and y coordinates, and the frame of each ROI. Each measurement was sorted in columns in a measurement table. These tables were saved and further processing of the data was performed using in-house python scripts made locally by Anna Lång. The scripts retrieved these measurements and used them to perform quantitative analysis of the cells in the image sequences. One of the scripts was used to plot the average nucleus area and perimeter (appendix D.7) in each frame. Another script was used to estimate the average cell populations (appendix D.8) in each frame.

By performing these steps, a pipeline was set up where a deep learning model can analyze nuclei images and provide information about the cell cultures studied.

Additionally, ROI measurements were used in an attempt to develop a method that could track cells across frames. By using a python script locally made by Anna Lång (appendix D.9) so called tracking files were generated, where the coordinates of each cell in the raw data is marked by a white dot on a black background. By processing the tracking files in the ImageJ plugin TrackMate, the goal was to be able to use the data from tracking of single cells and further collect various cell parameters. However, this was not achieved due to time restrictions, and as a result, scripts sorting the measurements of specific cells across frames in an efficient way were not written.

Two StarDist models were used to acquire quantitative data: the *in-house* made '1000 2-step' model was used to identify dividing cells and a model designed for nuclear segmentation called 'versatile (fluorescent nuclei)' (Schmidt et al. 2018) was used to identify all cells.

4 RESULTS

4.1 YAP expression in HaCaT cells

As shown in Lång et al. 2018, serum starvation of confluent HaCaT cell sheets leads to synchronization of the cells into a common quiescent cell state. Subsequently, serum re-stimulation of the quiescent cells activates a highly coordinated and collective cell migration response followed by cell division. The culturing conditions of these cells differ from the culturing conditions of HaCaT cells that have not been deprived of serum, and subsequently these cells should express different proteins as a response to their environment.

Aurora B proteins are involved in cell proliferation as part of the midbody, and YAP is expressed as a result of biomechanical stress (Gumbiner et al. 2014 and Qi et al. 2007). Thus, IF staining of cells with Aurora B and YAP specific antibodies can be used to study cell proliferation and the biomechanical stress response of cells exposed to different culturing treatments.

HaCaT wild type cells subjected to three different forms of treatment were stained with YAP- and Aurora B specific antibodies, as described in 3.2 The cell cultures were fixed, IF stained successfully and imaged with the Leica confocal microscope. The samples were imaged with the light source emitting three different wavelengths, one giving a detector signal for YAP, one giving a signal for Aurora B, and one giving a signal for the DAPI stain used to stain DNA. Images of the fixed cell cultures can be seen in the figure below (figure 4.1.1).
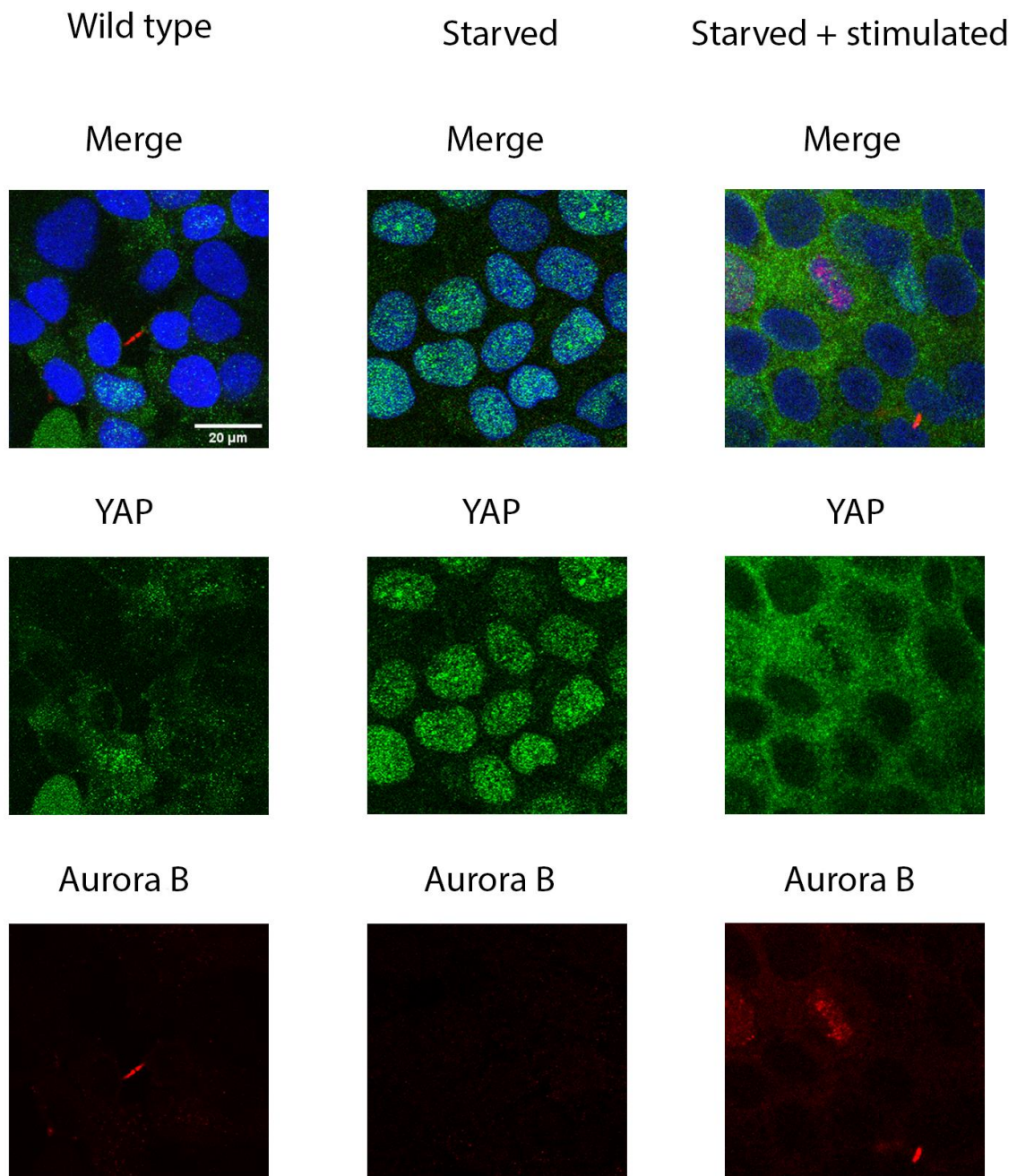
Figure 4.1.1: IF staining of HaCaT cells. The figure shows images of cells under three different forms of treatment: The wild type cells have not been starved or stimulated but were cultured in IMDM 10% FBS medium. The starved cells were starved for 72 hours before being fixed and imaged. The starved + stimulated cells were starved for 72 hours before being serum stimulated for 24hours. DAPI is shown in blue color, Aurora B is shown in red color, and YAP is shown in green color. In the top images, all three signals have been merged into one image and are therefore labelled 'merged'.

Aurora B proteins were only observed in the wild type cells and the cells that had been starved and re-stimulated with serum. The lack of Aurora B expression in the starved cells show that the cells are synchronized in the G_0 phase of the cell cycle and thus are non-dividing. The "starved + stimulated" images showed the strongest expression of Aurora B. Moreover, as the cells have been re-stimulated for 24 hours synchronized cell division was observed in this cell culture. The expression of Aurora B in the wild type cells was not as strong as the re-stimulated cells. As seen in figure 4.1.1, Aurora B is mainly observed as part of a midbody formed between two daughter cells. Conversely, Aurora B is not only detected in the midbodies of re-stimulated cells, but also gives a stronger signal of Aurora B outside the cell nucleus. This may, however, be due to background noise during microscopy. These results indicate that wild type cells are unsynchronized and that these cells have a lower rate of cell division compared to the re-stimulated 'starved + stimulated' cells.

The acquired images showed different localization of the YAP protein in wild type, starved, and starved and stimulated cells. The YAP expression was scattered in wild type cells and no clear distributional pattern was observed. In the starved- and the starved and stimulated cells the intensity of the YAP signal was stronger, and a clear subcellular localization was evident. In the starved cells, YAP was observed densely packed in the nucleus of the cell. As mentioned in the introduction, cells respond to stress by nuclear accumulation of YAP (Kim et al. 2016), indicating that deprivation increase the level the level of stress or tension in the cell monolayer. In the stimulated cells YAP proteins were observed to be located in the cytoplasm, indicating that serum re-stimulation of quiescent cell sheets results in reduced stress or tension across the cell monolayer. Serum re-stimulation also activates cell proliferation, which in turn leads to YAP deactivation through the Hippo signaling pathway.

4.2 Results from running StarDist model validation

Paired raw data and ground truth images (see section 3.4.1) were used as datasets when training the models as described in 3.4.2. One challenge that occurred when creating the datasets was that the microscope lost focus on some frames during acquisition and dividing cells with proper chromosome separation in these frames were not easily to identified. These cells were evaluated as described in 3.4.1.

The amount of data needed to create an efficient model differs greatly depending on the kind of data the model is supposed to analyze. The developers of the StarDist (2D) network have been able to produce models using only 5-10 images containing 10-20 annotated ROIs each (Schmidt et al. 2021). However, this model will probably need a much larger dataset to function adequately, since it is supposed to detect and analyze dividing cells, as well as distinguish these cells from apoptotic and nondividing interphase cells. To address this, models were made using training dataset sizes of 100-1000 paired images. In total, 5 models with differing training dataset sizes and parameters were successfully generated. Table 4.1.1 shows the models generated and their differences.

Table 4.1.1: StarDist models generated and their differences. All other parameters for model training were the same for all models.

Model name	Training dataset size (paired images)	Number of steps
100	100	1
200	200	1
500	500	1
1000 1-step	1000	1
1000 2-step	1000	2

During model training it became clear that the '1000 2-step' model could not be generated in one session. The reason for this is that this model required more memory than the other models, most likely due to a large training dataset and that the number of steps used during training was doubled (table 4.1.1). This was solved by training the model in two sessions, using a dataset of 500 different paired images during both training sessions.

4.2.1 Results from running StarDist model validation

Training loss vs epoch number, and validation loss vs epoch number, was estimated during model training and these calculations provide a numeric error value for the models throughout network training as mentioned in section 1.6. A downward trending training loss curves indicate that the model requires more training, while a flat curve indicates that the model will not benefit from more rounds of training. The training loss vs epoch number and validation loss vs epoch number curves

for all models were calculated in the StarDist notebook as described in 3.4.2. These graphs were used to monitor the training progression as well as give an indication of model performance and compare the model's ability to segment images not seen during model training (see section 1.6). The loss curves are shown in the figures below (figures 4.1.2-4.1.6).
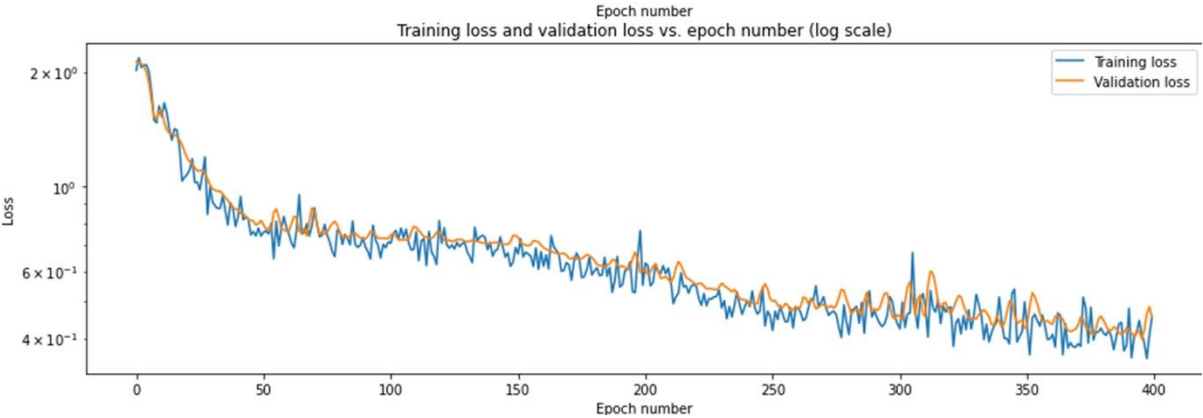

Figure 4.1.2: The training loss and validation loss as a function of epochs during training of the '100' model. The training loss and validation loss is shown on a log scale. The '100' model was trained using a dataset of 100 training images and a step size = 1 as shown in table 4.1.1.

Figure 4.1.3: The training loss and validation loss as a function of epochs during training of the '200' model. The training loss and validation loss is shown on a log scale. The '200' model was trained using a dataset of 200 training images and a step size = 1 as shown in table 4.1.1.
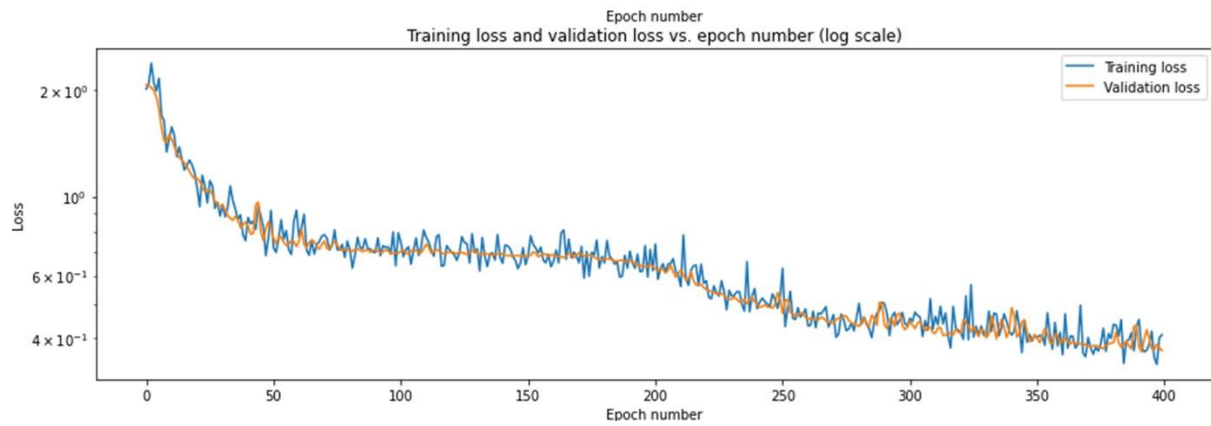
Figure 4.1.4: The training loss and validation loss as a function of epochs during training of the '500' model. The training loss and validation loss is shown on a log scale. The '500' model was trained using a dataset of 500 training images and a step size = 1 as shown in table 4.1.1.

Figure 4.1.5: The training loss and validation loss as a function of epochs during training of the '1000 1-step' model. The training loss and validation loss is shown on a log scale. The '1000 1-step' model was trained using a dataset of 1000 training images and a step size = 1 as shown in table 4.1.1.
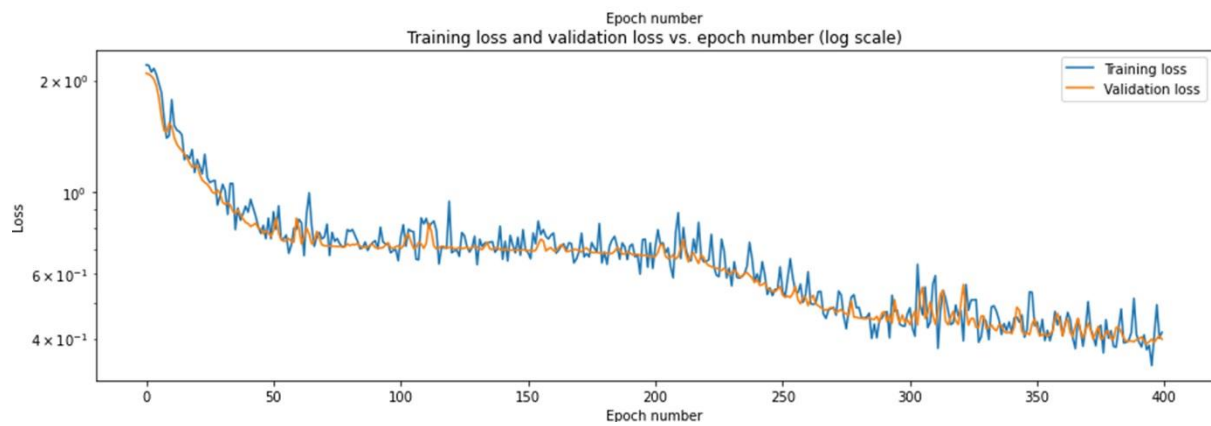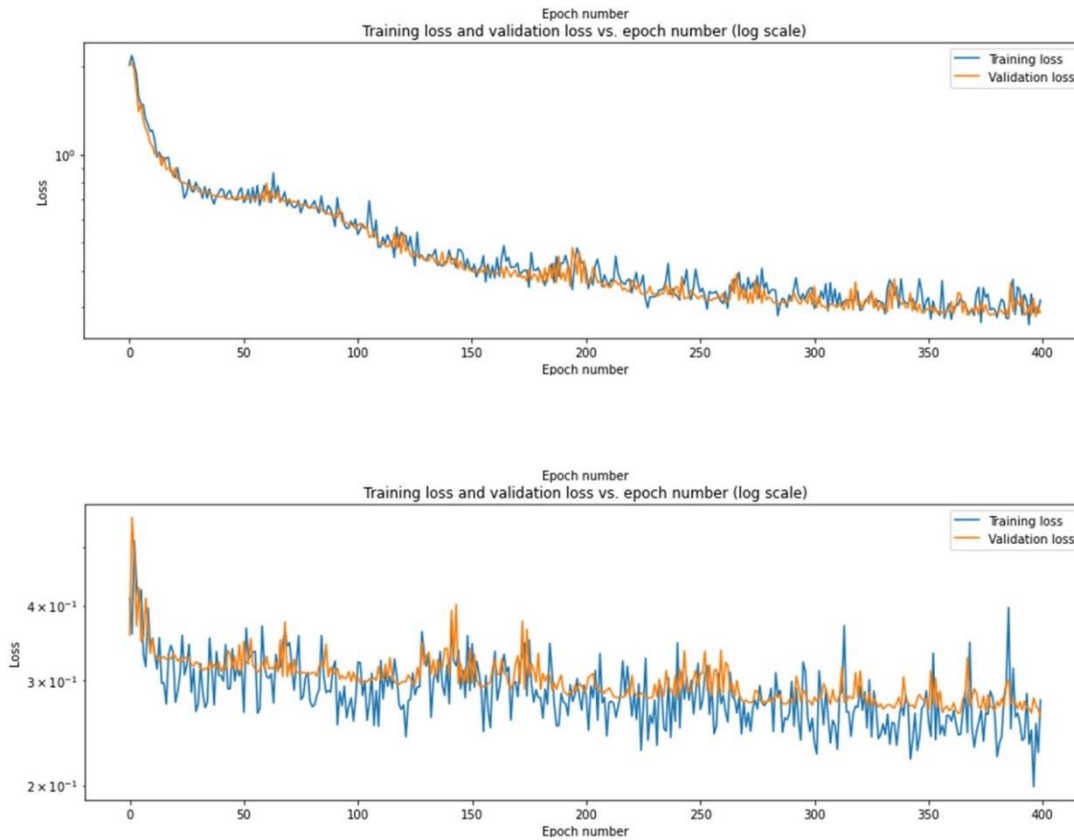
Figure 4.1.6: The training loss and validation loss as a function of epochs during training of the '1000 2-step' model. The training loss and validation loss is shown on a log scale. The '1000 2-step' model was trained using a dataset of 1000 training images and a step size = 2 as shown in table 4.1.1. This model had to be trained in 2 sequential sessions resulting in 2 separate graphs showing training and validation loss. The upper loss curve is the loss curve from the first training session and the loss curve at the bottom is from the second training session, after which the model was completed.

The graphs above show that the training loss and validation loss values are approximately the same for all models throughout training. This indicates that the model is just as capable of analyzing images not previously seen by the model as the images used during training. The '100', '200' and '500' models show a slight decrease in training loss and validation loss in the final 100-150 epochs of training. This means that a larger dataset or more epochs are needed for the loss curves to stabilize at the end of training and yield a model with more accurate segmentation of dividing nuclei. The '1000 1-step' and '1000 2-step' models show curves that seems to flatten out towards the end of the session. This is especially the case for the '1000 2-step' model, where the bottom graph in figure 4.1.6 show flattened loss curves during the second session of training. This

indicates that these models would not have an increase in their respective ability to predict and segment dividing nuclei by increasing the epoch number or size of the training dataset.

The final values of training and validation loss provides information regarding the accuracy of the model, as mentioned earlier. The final training and validation loss of '1000 2-step' was in the 0,25-0,3 range and the final training and validation loss of the other models were in the 0,4-0,5 range. The numeric value plots of training loss and validation loss can be found in figures D.1-6 in the appendix.

4.3 Analyzing the StarDist models' ability to make accurate predictions and segment nuclei

The graphs in section 4.2.1 give an indication of the models' ability to detect and segment dividing nuclei. To further estimate how well the models performed, each model was subjected to a quality control test (QC).

In this test, each model analyzed 30 nuclei images, and their predictions were compared to the ground truth of each respective image. The QC test was performed through the StarDist notebook (Colab). This QC test provides the IoU, false positives, false negatives, true positives and mean matched score for a model's prediction of all 30 nuclei images. The notebook (Colab) also provides images for comparing ground truths and predictions of raw data images. An example of this is shown in figure 4.1.7.

A   Input   Ground truth   Prediction   Ground truth and prediction. Intersection over union: 0.472

B   Input   Ground truth   Prediction   Ground truth and prediction. Intersection over union: 0.421

C   Input   Ground truth   Prediction   Ground truth and prediction. Intersection over union: 0.374

D   Input   Ground truth   Prediction   Ground truth and prediction. Intersection over union: 0.550

E   Input   Ground truth   Prediction   Ground truth and prediction. Intersection over union: 0.679
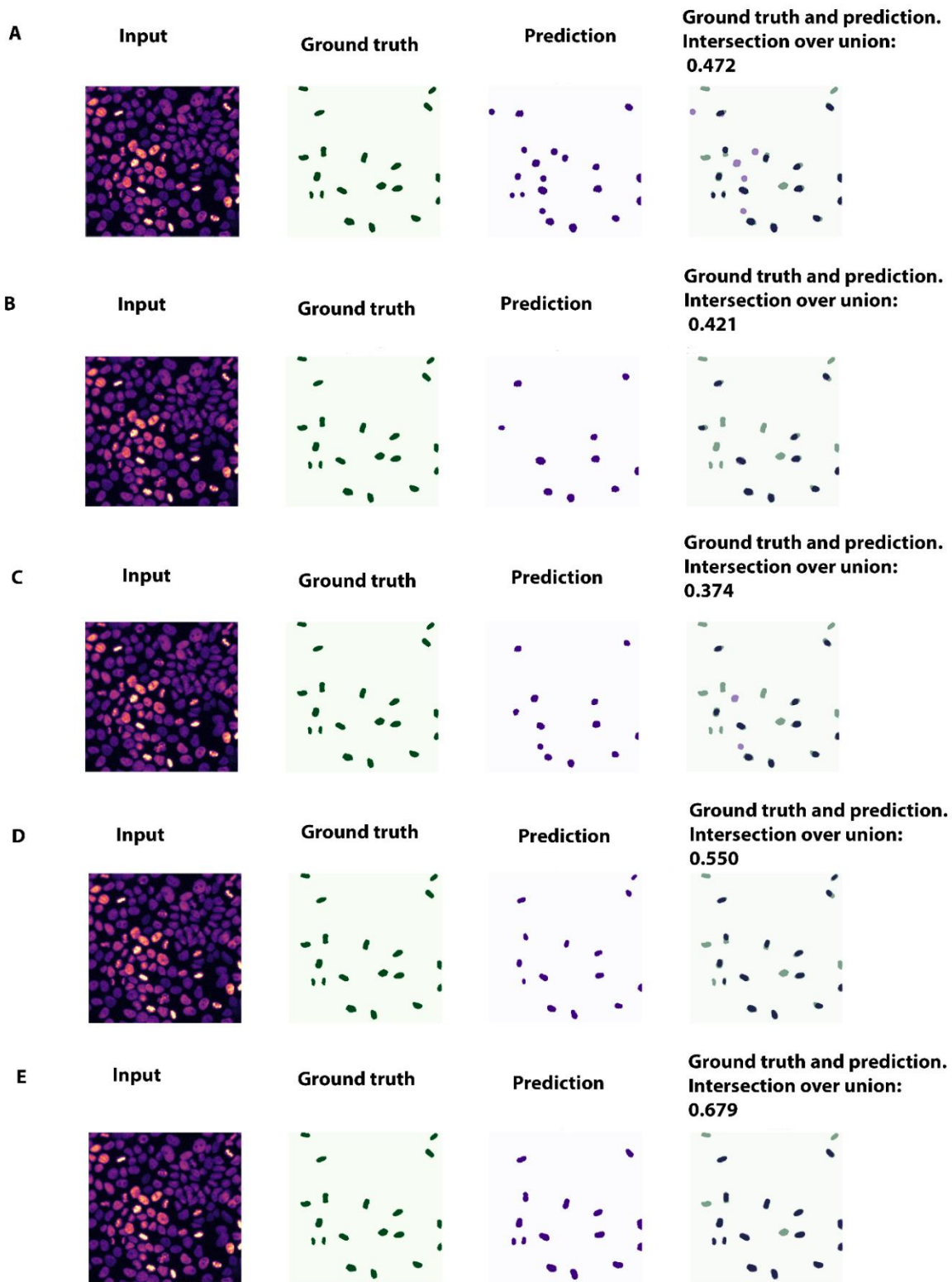
Figure 4.1.7: An illustrative comparison of ground truth and predictions of the different models for a specific nuclei image. The '100 model'(A), '200 model'(B), '500 model'(C), '1000 1-step model'(D), and '1000 2-step model'(E) are displayed vertically. For each panel, the nuclei image being analyzed is shown to the far left. This image was one of the 30 images used in the QC test. Its ground truth image is shown in the middle left and the models' prediction of the nuclei image are seen in the middle right. The images overlapping the ground truth images and the model predictions are shown in the far right.

The results of the QC test were averaged to give an overview of the models' ability to correctly predict cell shape and size and to make correct predictions of dividing cells.

Table 4.1.2: Analysis of the different models' average performance based on predictions done on 30 quality control paired images. The IoU value is the intersection over union value for predictions on the ground truth including false positives and false negatives. Mean matched score is the intersection over union of only true positive predictions. Precision calculates the rate of false positives relative to true positives and recall calculates the rate of false negatives to true positives as mentioned in section 1.6. Accuracy shows the total accuracy of the model, calculating the rate of false positives and negatives relative to true positives.

Modell	IoU	Mean matched score	Precision	Recall	Accuracy
1000. 2-step	0,58797	0,83201	0,83795	0,78704	0,67494
1000 1-step	0,47555	0,73776	0,83083	0,71258	0,60787
500	0,41533	0,68151	0,79147	0,67556	0,56243
200	0,36324	0,69793	0,8725	0,52311	0,48745
100	0,43384	0,68539	0,6999	0,79550	0,57420

As table 4.1.2 shows, the '1000 2-step' model had the highest mean matched score and accuracy of all models in the QC test. These results show that this model is most suited for detecting and segmenting dividing cells and that increasing the number of steps improves model performance. The '1000 1-step' model had the second highest accuracy and mean matched score, suggesting that increasing the size of the training dataset increases model performance. However, the fact that the '100' and '200' models had a higher mean matched score than the '500' model conflict with this result. Additionally, the test measured the accuracy of the '100' model to be 0,57420, which was higher than the accuracy of the '200' and '500' models. The conflicting results show that introducing new training data can affect model performance in different ways that are not easily quantified, as can be seen by the varying precision and recall values. The general trend, however,

is that an increased training dataset leads to a higher mean matched score and higher accuracy of the model.

4.4 Creating an automated pipeline to analyze cell cultures

Nuclei images from live cell imaging experiments, and StarDist models were acquired and generated as described in sections 3.3 and 3.4, respectively. The images and the model were used to develop a multi-step method for quantitatively analyzing re-stimulated quiescent cell cultures. By using python-based scripts locally made by Anna Lång (see D.7-9 in the appendix) coupled with measurements from the StarDist ImageJ plugin and in combination with the TrackMate plugin, several methods to collect quantitative measurements such as average nuclei area and circularity can be performed (see section 3.5.3). Because it is the chromatin of the cells that is fluorescently tagged, all measurements of cells are in reality measurements of the nuclei (see section 1.5.1). A pipeline that quantifies the cell culture qualities was created successfully, and the results are shown in sections 4.4.1 and 4.4.2.

Another goal was to develop a method that tracks single cells between multiple frames by integrating TrackMate in the pipeline. A python script for creating tracking files was developed (D.9 in appendix), however no data was extracted from these files due to time constraints.
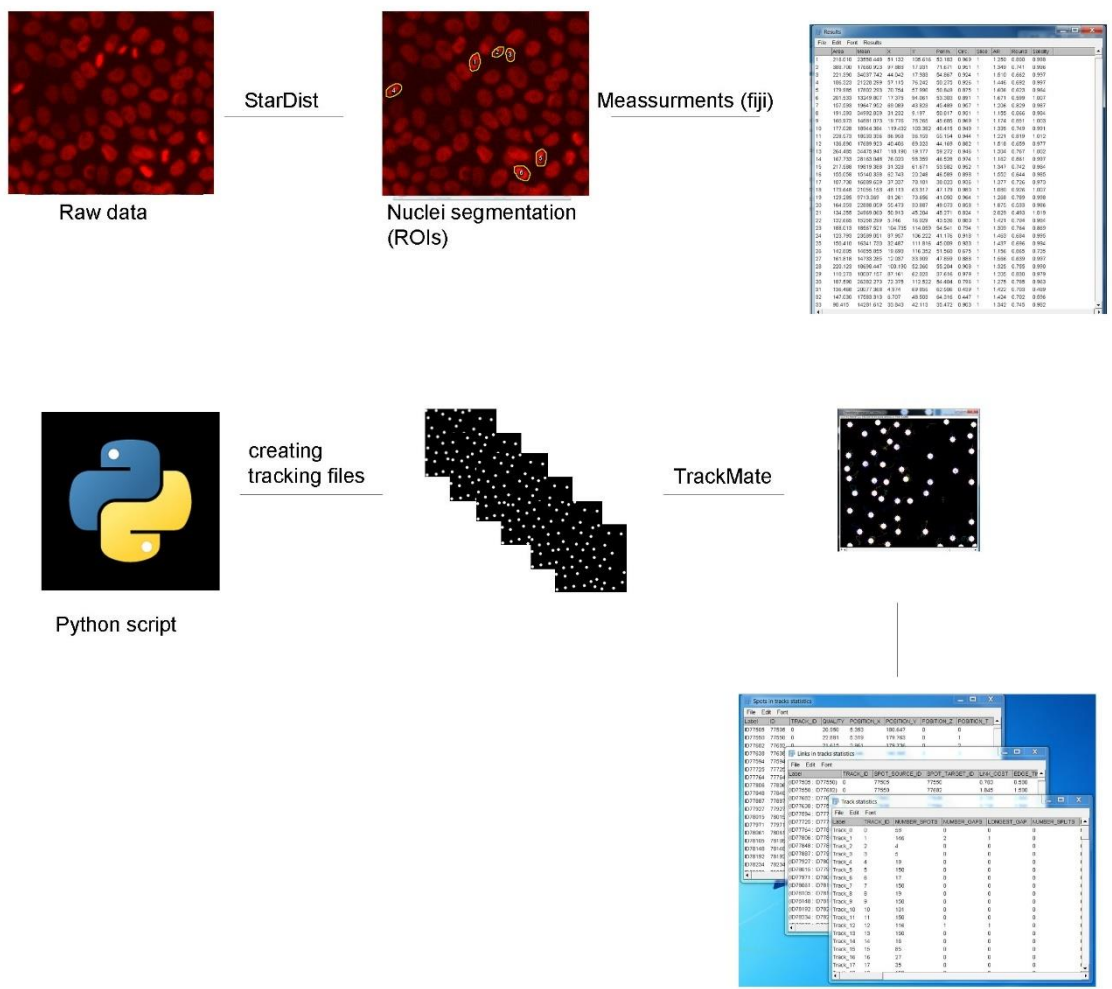
Figure 4.1.8: A visual illustration of the pipeline used in quantitative analysis of image sequences acquired from live cell imaging. The images are first classified and segmented by StarDist. Then the ROIs are measured by ImageJ and the measurements are used in a python script. Python scripts generating quantitative data was created and applied in analysis of nuclei images. A method where fluorescent nuclei were tracked was attempted to be made. In this method a python script would generate an image sequence of ROI maps, called tracking files, from ROI measurements. These tracking files would then be run through TrackMate which produce information about single cell tracks/ migration.

To summarize, the main goal for this experiment was to develop a pipeline that can be applied to any live cell imaging experiment, where a deep learning model generated via the StarDist notebook can be used to acquire quantitative data of dividing cells.

4.4.1 Collecting quantitative data based on the StarDist model

The StarDist model 'Versatile (fluorescent nuclei)' was used to identify and segment nuclei of all cells in 12 live cell imaging datasets. The image sequences consisted of 600 frames. The model predictions, in the form of ROIs, were measured, and the acquired nuclear area and cell perimeter were plotted in graphs using one of the *in-house* python scripts (D.7 in appendix). The results are shown in figure 4.1.9 and 4.1.10. The main goal of plotting nuclei area and perimeter over time was to show that quantitative measurements can be extracted from live cell imaging data using the developed pipeline.
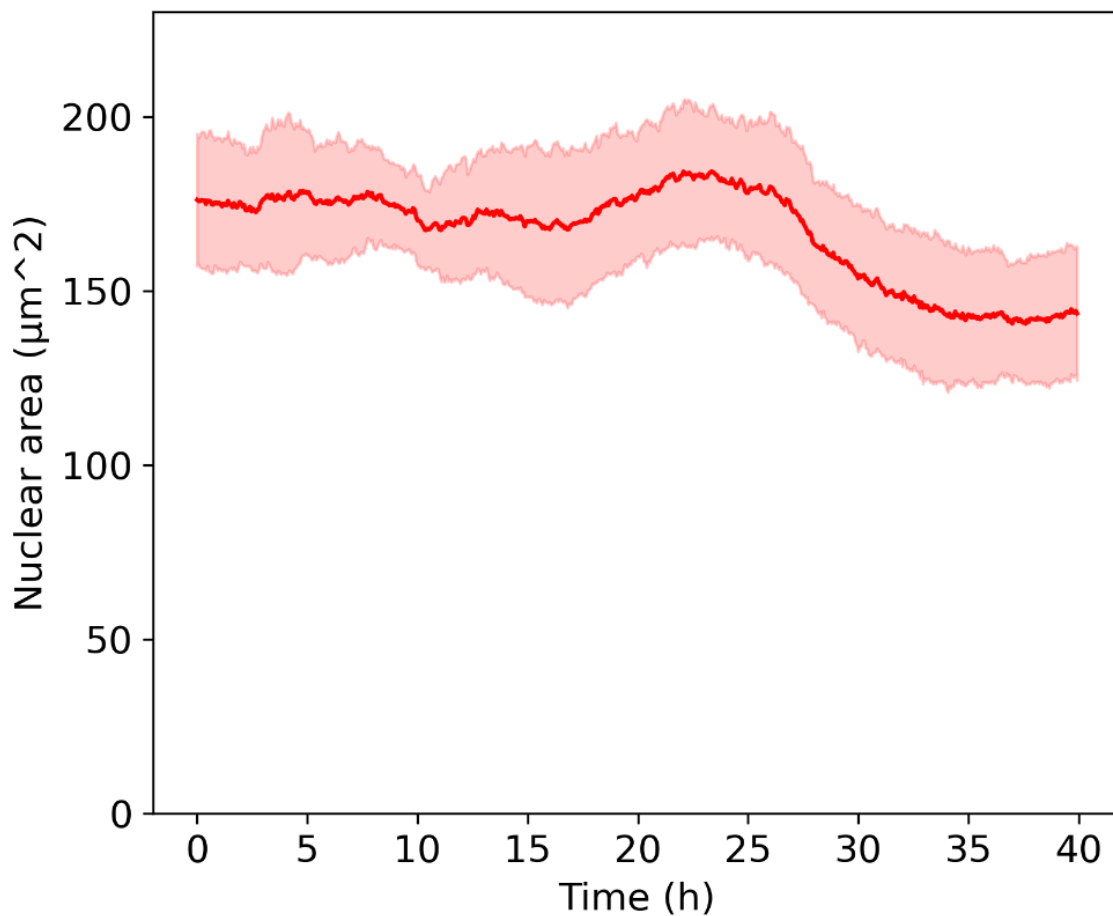

Figure 4.1.9: Mean area in μm² of all cells in live cell image sequences (n=12) as a function of time. The start time of the imaging is plotted as 0 hours, and imaging started 2,5 hours after serum stimulation. The cells were imaged for a total of 40 hours. The thick red line represents the mean area of the 12 sites, while the standard deviation is represented by the red shaded area in the graph.
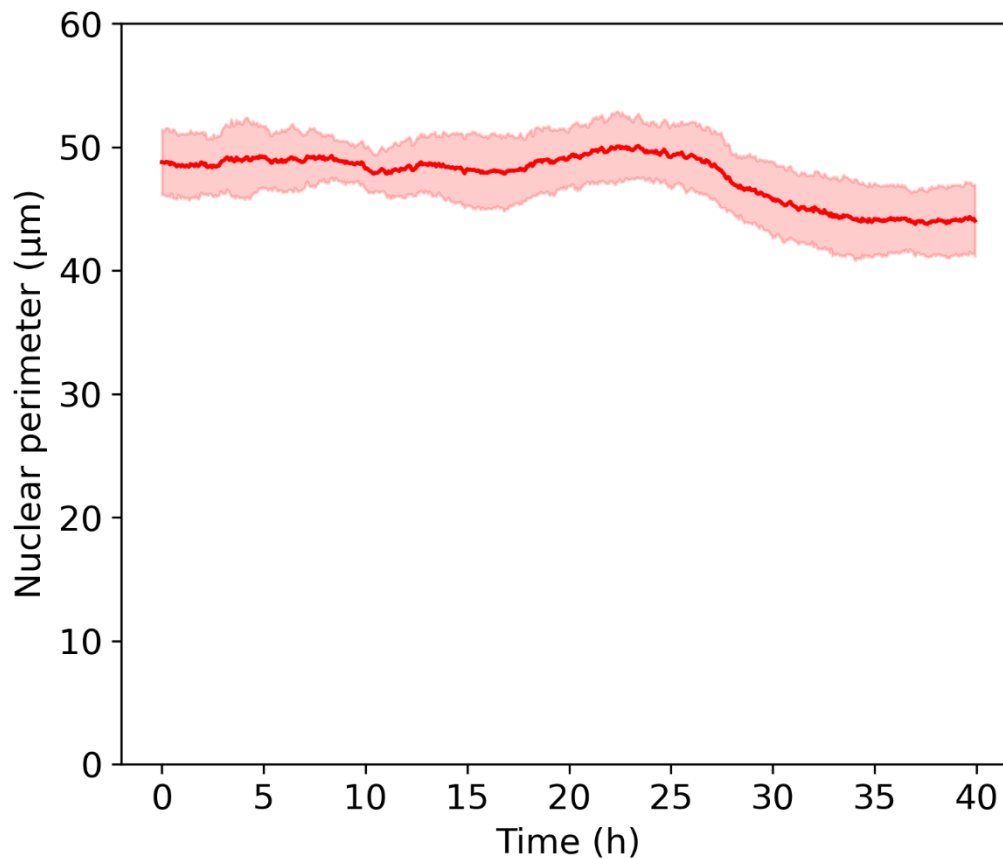
Figure 4.1.10: Mean perimeter in µm of all cells in live cell image sequences (n=12) as a function of time. As in the figure above, imaging lasted for 40 hours and began 2,5 hours after serum stimulation. The thick red line represents the mean perimeter of the 12 sites, while the standard deviation is represented by the red shaded area.

Figures 4.1.9 and 4.1.10 show a reduction in mean nuclei area and perimeter starting approximately at 25 hours of live imaging. From 25 to 30 hours into the image sequence, the mean nuclear perimeter is reduced from approximately 49 µm to 45 µm and the mean nuclear area was reduced from approximately 175 $µm^2$ to 150 $µm^2$. Since the cells were stimulated 2,5 hours before imaging the figures above indicate that the peak in cell division occurred between 27,5 to 32,5 hours after serum stimulation. There was also a decrease in nuclear perimeter and area around 10 hours into the image acquisition period, which was 12,5 hours after stimulation. This might be due to synchronized cell migration as cells in the image sequences were observed to migrate collectively around this time.

As mentioned in 1.1.1, the nuclear size changes depending on what phase of the cell cycle the cells are in. As measuring nuclear area is a measurement of the nuclear size, and the average

nuclear area changes throughout the imaging experiment, figure 4.1.9 show that the cell cultures are synchronized.

4.4.2 Cell population data from live cell imaging experiments

The average number of all cells, including both interphase and mitotic cells, and the number of mitotic cells per frame was estimated using the developed pipeline. In this pipeline the StarDist models 'versatile (fluorescent nuclei)' and '1000 2-step' were used to analyze the image sequences in ImageJ. The extracted ROI measurements on cell count were plotted using a python script written locally by Anna Lång (appendix D.8). The same 12 datasets from 4.4.1 were analyzed, consisting of 600 frames. To estimate the average population of all cells the 'versatile (fluorescent nuclei)' model was used.
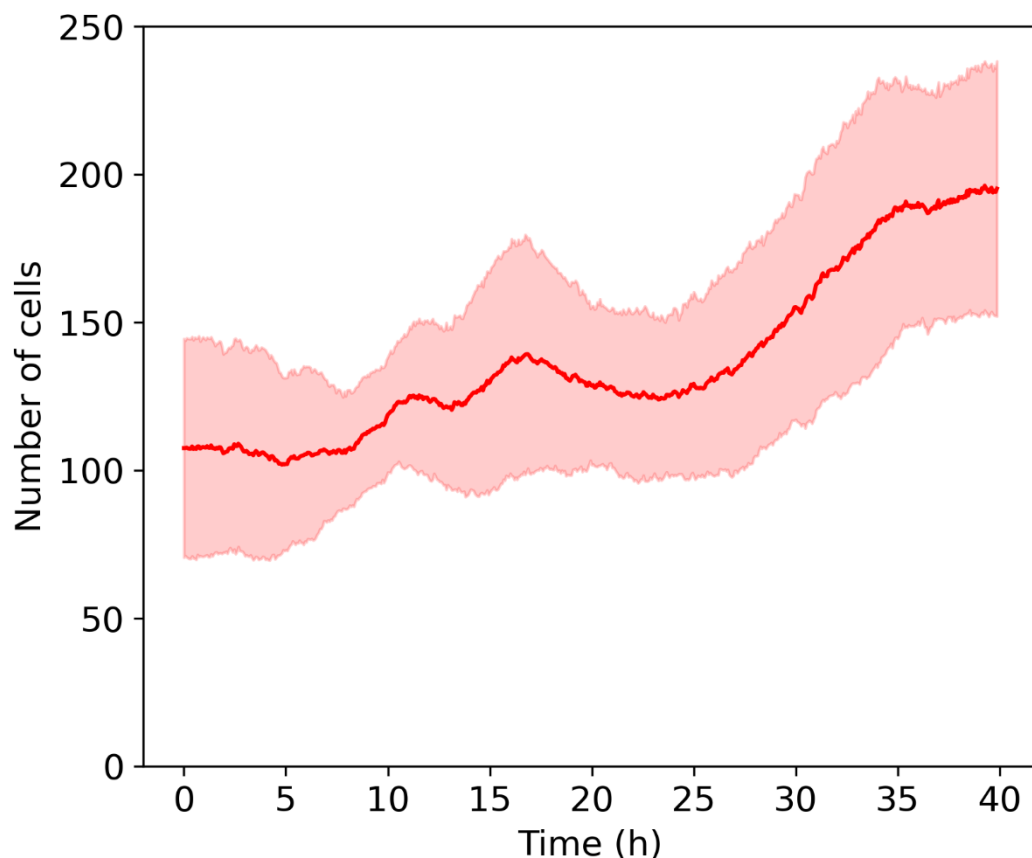

Figure 4.1.11: The average number of cells per frame in the live cell image sequences (n=12) plotted as a function of time. As in figure 4.1.9 and 4.1.10, imaging lasted for 40 hours and began 2,5 hours after serum stimulation. The

model "versatile (fluorescent nuclei)" was used to detect all nuclei in each image. The thick red line represents the average number of cells in the 12 sites, while the standard deviation is represented by the red shaded area.

As shown in figure 4.1.11, the average number of cells gradually increased, and approximately 25 hours into imaging the number of cells drastically increased from around 120 to 200 cells over a time span of 10 hours. The spike in cell population suggests that synchronized cell division occurred around 25 to 35 hours into imaging. Figure 4.1.11 also show an increase in average cell count around 10 to 15 hours into imaging. As mentioned in 4.4.1, collective migration was observed around this time, and the increased cell density can probably be explained by cells migrating into the field of view, or frame of the image. This most likely explains the small cell population peak observed around 15 hours into imaging.

The pipeline was also used to analyze the population of mitotic cells in the same 12 datasets that were analyzed by the pipeline in 4.4.1. To perform this analysis the '1000 2-step' StarDist model was used to identify mitotic nuclei, and the script for plotting the cell count was used (appendix D.8).
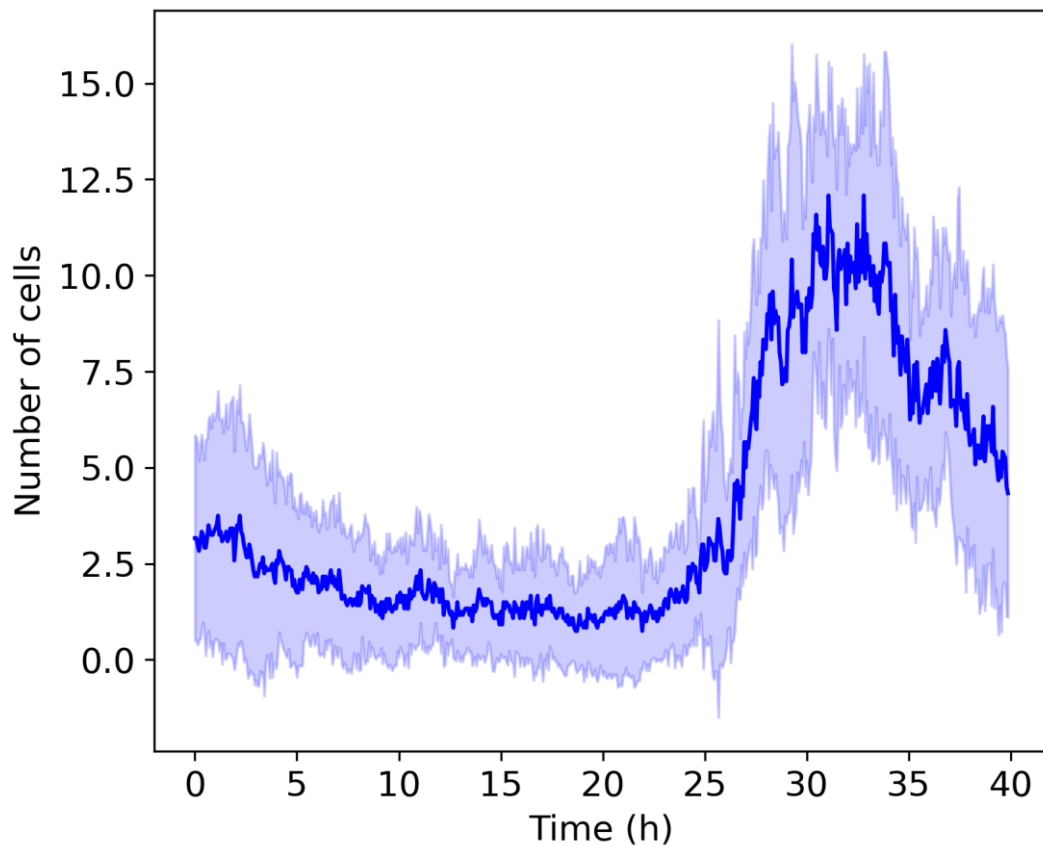
Figure 4.1.12: The average number of dividing cells in the live cell image sequences (n=12) plotted as a function of time. Imaging lasted for 40 hours and began 2,5 hours after serum stimulation. The thick blue line represents the average number of mitotic cells in the 12 sites, while the standard deviation is represented by the blue shaded area.

The figure above shows the average number of dividing cells as a function of time. The results show a large increase in the number of mitotic cells that starts around 25 h into imaging. The curve peaks with an average of 12,5 mitotic cells per frame after 30 h of imaging. From 30 to 40 hours of imaging the number of mitotic cells decrease from 12,5 dividing cells per frame to about 5,0 mitotic cells per frame. This graph shows that synchronized cell division in serum stimulated cell cultures occur between 27,5 to 37,5 hours after stimulation which correlates well with previously published data by Lång et al. 2018.

5   DISCUSSION

The goal of this project was to develop a deep learning-based methodology for analysis of cell division and cell migration using microscopy data. In the sections below, the effectiveness and performance of the StarDist models will be discussed, an evaluation of the pipeline will be given, and the response of HaCaT cells to the treatments in the established system will be discussed.

5.1   The StarDist model performances

StarDist deep learning models were generated with the goal of detecting dividing HaCaT cells in live cell imaging data. The StarDist models were trained under different conditions, with varying training dataset sizes and parameters, specifically the number of steps.

The number of paired images needed as training data to create an efficient model differs greatly depending on the kind of data the model is supposed to analyze. The developers of the StarDist (2D) network have been able to produce models using only 5-10 images containing 10-20 annotated ROIs each (Schmidt et al. 2020). However, as mentioned in section 1.6.1, the StarDist software is designed to detect round objects, and the models generated in this project are designed to segment metaphase and anaphase cells, which are less circular than interphase cells. Due to this, it was assumed that that such a model would require a much larger dataset to function adequately. Another reason to increase the training dataset was that the models were prone to detect apoptotic cells as false positives during testing in ImageJ. These cells have, as described in 1.1.1, condensed nuclei, and this might be the reason why the models occasionally falsely segmented these cells as mitotic cells.

Curves showing the training loss and validation loss during each epoch of training were generated by the StarDist notebook (figures 4.1.2-4.1.6). These curves were used to show if the model was finished with training or not. The loss curves of the '100', '200' and '500' models were still declining towards the end of training, while the '1000 1-step' and '1000 2-step' models had loss curves that flattened out towards the end of training. This indicates that the dataset size and the number of epochs used in training of the models affect their performance. Notably, the '1000 2-step' had a final training loss and validation loss that was lower than all other models (in the 0,25 to 0,3 range). As the loss function used estimates a numeric error value of model predictions, this indicates that this model is the best at predicting and segmenting the mitotic cells.

.

The StarDist model with the most accurate predictions and highest mean matched score would be included in the developed pipeline. As table 4.1.2 shows, QC analyses determined the '1000 2-step' model to be the best performing model. When comparing the mean matched scores and the accuracy measurements provided from QC analyses, models tended to increase in both aspects when the size of training dataset was increased. This was more of a general trend however, as the '100' model outperformed the '200' and '500' models in both accuracy and mean matched score. These results further prove that dataset size affect model performance. Another key factor influencing a model's ability to segment nuclei images was shown to be the number of steps used during training, as the accuracy and mean matched score of the '1000 2-step' model (0,67494 and 0,83201, respectively) was significantly higher than for the '1000 1-step' model (0,60787 and 0,73776, respectively). As mentioned in 3.4.2, batch number and number of steps are parameters that dictate how many pixels the model observes during each training epoch. This shows that both the size of the training dataset and the number of pixels observed by the model during training affect the model performance.

The training loss and validation loss curves flattened out at the end of training for the models with a training dataset larger than 500 frames. Flattening loss curves are characteristic of models that will not benefit from more epochs or increasing the size of the training dataset. A suggestion for generating a better performing model is to increase the number of steps during training. This will lead to a more memory-demanding model, however, by using a google drive with a larger storage capacity this can easily be done without drastically increasing the time it takes to create a StarDist model. Due to time restrictions, the effect of other parameters such as batch size and grid parameter, on model performance were not determined. Consequentially, creating models with different parameters and testing their performance with QC analyses might lead to development of a model that is more effective at mitotic nuclei segmentation than the '1000 2-step' model developed during this project.

StarDist is designed to segment cells. However, most cells are non-dividing and thus have a more circular shape. When StarDist makes predictions, it estimates an object's radial distance and the object probability of all pixels. This method of prediction is most suited for round objects, and thus it was uncertain if StarDist was well suited to segment dividing cells. The accuracy and mean matched score of the best performing model ('1000 2-step') show that a StarDist model can predict mitotic nuclei with moderate accuracy and segment nuclei shape fairly well. As mentioned earlier,

increasing the number of steps, and altering the other parameters used during training might improve the accuracy even more.

5.2 Evaluation of the method for analyzing cells

The deep learning-based pipeline was developed to measure the population of mitotic cells, as well as analyze nuclei metrics and the population of all cells. The deep learning was also intended to measure differences between mother and daughter cells and study the relation of nuclear shape and migration direction of cells.

The developed pipeline consisted of acquiring nuclei images by live cell imaging experiments, segmenting nuclei with StarDist, and extracting measurements from the annotations through ImageJ, as illustrated in figure 4.1.8. These measurements were then used to quantitatively analyze the cell cultures by plotting the acquired information using *in-house* written python scripts.

The quality of the nuclei images used in training and analysis might affect the performance of the pipeline. As mentioned in 3.4.1, it was challenging to determine if the dividing cells were properly separating or if mis-segregation of chromosomes occurred, as some cell divisions occurred out-of-focus. However, training the models with training dataset including out-of-focus cell divisions might also increase the robustness of the model as it learns to segment mitotic nuclei with different pixel intensities in the images.

The StarDist predictions were used to identify mitotic nuclei. As mentioned in 5.1, the best performing developed StarDist model detected mitotic nuclei with moderate accuracy and was good at predicting nuclear shape. This means that improving the performance of the StarDist model is the best way to increase the accuracy of the pipeline analysis. The model performance can either be directly improved by training the model under different conditions or possibly indirectly by only using in-focus live cell images when developing the model and analyzing cell cultures.

As demonstrated by figures 4.1.9-4.1.12, the pipeline was able to perform quantitative analysis of cells. The purpose of using the TrackMate script shown in appendix D.9, as described in section 2, was to develop a method for analyzing cells by tracking them between frames. The tracking files and ROIs provided can then be used to extract measurements of certain cellular parameters and thus analyze the cells over time. However, due to time constraints and problems connecting data from specific cells to retrieve information about specific cells and their development, this

method was not completed. Because tracking files were generated successfully, it can be concluded that there is a potential for analytical methods where single cells are being tracked and measured across image frames.

When comparing the developed method to manual analysis of live cell imaging data, a benefit of using this automated method is that it is much less time consuming, providing a way to quickly analyze dividing nuclei in large datasets.

5.3 Cell behaviors in the established system

The established experimental system used for synchronizing cells by serum starvation, and subsequent activation of collective cell migration and proliferation by serum re-stimulation was examined by studying the cellular localization of the YAP and Aurora B proteins in IF stained HaCaT cells. As shown in figure 4.1.1, YAP and Aurora B were expressed and localized differently in asynchronously growing cells, serum deprived cells and serum re-stimulated cells. Aurora B was observed in midbodies of asynchronous daughter cells and re-stimulated daughter cells which shows that cell divisions occur in these cell cultures. This is in contrast to the serum deprived, or starved, cells, in which no Aurora B containing midbodies was detected proving that cell division is not occurring when cells are subjected to starvation. The re-stimulated cells had the strongest Aurora B signal, which indicates that re-stimulated cells have the highest degree of cell division of all treatments. The localization of YAP in the nucleus of starved cells show that nutritional cell stress leads to biomechanical cell changes that increase the level of stress or tension across the monolayer. An increase in biomechanical stress leads to YAP activation and translocation of YAP into the nucleus. As mentioned in section 1.4.2, YAP is involved in cellular EMT by downregulating cadherin protein expression and subsequently, cellular adhesion complexes. Also, cells in the mesenchymal state are more prone for cell migration. However, the cells in this project remain as confluent cell sheets during serum starvation and re-stimulation. This is not consistent with cells in a mesenchymal state, which means that the cells do not go through the EMT but stay epithelial throughout the experiments. Cell migration occurs around 10 to 15 hours after serum stimulation and YAP was shown to be located in the cytoplasm of re-stimulated cells. This indicates that YAP in re-stimulated cells is no longer involved in downregulating cadherins as YAP is not located in the nucleus and thus not functionally active. This might indicate that the migrating cells will begin to form more adherens junctions and become less migratory.

Because re-stimulated cells are dividing, the Hippo pathway is activated which is what causes the YAP localization in the cytoplasm where it will eventually be phosphorylated. These findings show that the established system synchronizes the cells and creates a stress response. Observing YAP and Aurora B activation and localization at different stages of the treatment gives insight to the synchronization of cells, and the activation of collective cell migration and subsequent cell division and gives an insight to the processes involved in epithelial wound healing.

6 CONCLUSION

The findings acquired during this project show that automated models for analysis of nuclei images can be developed and are capable of extracting quantitative information about the cell cultures. The cell culture was subjected to treatment that synchronized the cells, and lead to collective migration and subsequent cell division roughly 24 hours after re-stimulation. The developed StarDist model used in the pipeline to detect dividing nuclei was evaluated with quality control and was shown to have a moderate accuracy and a good ability to predict nuclei shape. The python scripts used to analyze the StarDist predictions via ImageJ as part of the pipeline were able to extract quantitative data from the cell cultures. Future work includes improving the accuracy model, which is the key for further improvement of the pipeline. It will also include developing python scripts that are able to effectively retrieve information about specific cells and their development.

REFERENCES

Ahirwar, R, Bariar, S., Balakrishnan, A. & Nahar, P. (2015). BSA blocking in enzyme-linked immunosorbent assays is a non-mandatory step: A perspective study on mechanism of BSA blocking in common ELISA protocols. *RSC Advances*, *5*(121), 100077-100083. https://doi.org/10.1039/C5RA20750A

Antonin, W. & Neumann, H. (2016). Chromosome condensation and decondensation during mitosis. *Current Opinion in Cell Biology, 40*, 15-22. https://doi.org/10.1016/j.ceb.2016.01.013

Baker, M. (2010). Cellular imaging: Taking a long, hard look. *Nature, 466*(7310), 1137-1140. https://doi.org/10.1038/4661137a

Bloom, M. & Mourttsen, O. (1995). The evolution of membranes. In Lipowsky R., Sackmann E. (ed) *Handbook of Biological Physics* (1. Edition, pages 65-95). Elsevier.

Borenstein, E. & Ullman, S. (2002). Class-Specific, Top-Down Segmentation. In: Goos G., Hartmanis J. and Leeuwen J (ed). *Lecture Notes in Computer Science* (2351. Edition, pages 109-124). Springer.

Borenstein, E., Sharon, E. & Ullman, S. (2004, 27 June-2 July). *Combining Top-Down and Bottom-Up Segmentation* [Paper presentation]. 2004 Computer Vision and Pattern Recognition Workshop, Washington, DC, USA. https://ieeexplore.ieee.org/document/1384838

Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A. & Fusenig, N. E. (1988). Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *Journal of Cell Biology, 106*(3), 761-771. https://doi.org/10.1083/jcb.106.3.761

Braga, V. (2016). Spatial integration of E-cadherin adhesion, signalling and the epithelial cytoskeleton. *Current Opinion in Cell Biology*, *42*, 138-145. https://doi.org/10.1016/j.ceb.2016.07.006

Campbell, K., Rossi, F., Adams, J., Pitsidianaki, I., Barriga, F. M., Garcia-Gerique, L., Batlle, E., Casanova, J. & Casali, A. (2019). Collective cell migration and metastases induced by an epithelial-to-mesenchymal transition in Drosophila intestinal tumors. *Nature Communications, 10*(1), article number 2311. https://doi.org/10.1038/s41467-019-10269-y

Carmena, M., Wheelock, M., Funabiki, H. & Earnshaw, W. C. (2012). The chromosomal passenger complex (CPC): From easy rider to the godfather of mitosis. *Nature Reviews Molecular Cell Biology 13*(12), 789–803. https://doi.org/10.1038/nrm3474

Carmona-Fontaine, C., Matthews, H. K., Kuriyama, S., Moreno, M., Dunn, G. A., Parsons, M., Stern, C. D. & Mayor, R. (2008). Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature, 456*(7224), 957-961. https://doi.org/10.1038/nature07441.

Cheeseman, I. M. (2014). The kinetochore. *Cold Spring Harbor Perspectives in Biology, 6*(7), article number 015826. https://doi.org/10.1101/cshperspect.a015826

Chu, F., Haley, S. & Zidovska, A. (2017). On the origin of shape fluctuations of the cell nucleus. *National Academy of Sciences of the United States of America, 114*(39), 10338-10343. https://doi.org/10.1073/pnas.1702226114

Compton, D. (2011). Mechanisms of aneuploidy. *Current Opinion in Cell Biology, 23*(1), 109–113. https://doi.org/10.1016/j.ceb.2010.08.007

Copeland, B. (2020, August 11). Artificial intelligence. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Deo, P. & Deshmukh, R. (2018). Pathophysiology of keratinization. *Journal of Oral and Maxillofacial Pathology, 22*(1), 86-91. https://doi.org/10.4103/jomfp.JOMFP_195_16

Eckhart, L., Declercq, W., Ban, J., Rendl, M., Lengauer, B., Mayer, C., Lippens, S., Vandenabeele, P. & Tschachler, E. (2000). Terminal differentiation of human keratinocytes and stratum corneum formation is associated with caspase-14 activation. *Journal of Investigative Dermatology. 115*(6), 1148-1151. https://doi.org/10.1046/j.1523-1747.2000.00205.x

Edwards, G. (2018, 18. Nov*.). Machine Learning, An Introduction*. Towards data science. https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0

Elmore, S. (2007). Apoptosis: A Review of Programmed Cell Death. *Toxicologic Pathology, 35*(4), 495-516. https://doi.org/10.1080/01926230701320337

Fenner, J. & Clark, R. (2016). Anatomy, Physiology, Histology, and Immunohistochemistry of Human Skin. In: Albanna, M. Z. & Holmes, J. H. *Skin Tissue Engineering and Regenerative Medicine* (1. Edition, pages 1-17). Elsevier.

Friedl, P. & Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology, 10*(7), 445–457. https://doi.org/10.1038/nrm2720

Friedl, P., Hegerfeldt, Y. & Tusch, M. (2004). Collective cell migration in morphogenesis and cancer. *The International Journal of Developmental Biology, 48*(5-6), 441-449. https://doi.org/10.1387/ijdb.041821pf

Gandarillas, A. (2012). The mysterious human epidermal cell cycle, or an oncogene-induced differentiation checkpoint. *Cell cycle, 11*(24), 4507–4516. https://doi.org/10.4161/cc.22529

Glotzer, M. (2004). Cleavage furrow positioning. *Journal of Cell Biology, 164*(3), 347-351. https://doi.org/10.1083/jcb.200310112

Gumbiner, B. M. & Kim, NG. (2014). The Hippo-YAP signaling pathway and contact inhibition of growth. *Journal of Cell Science, 127*(4),709-717. https://doi.org/10.1242/jcs.140103

Hafiz, A. & Bhat, G. (2020). A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval, 9*(3), 171-189. https://doi.org/10.1007/s13735-020-00195-x

Harvey, K., Zhang, X. & Thomas, D. (2013). The Hippo pathway and human cancer. *Nature Reviews Cancer, 13*(4), 246–257. https://doi.org/10.1038/nrc3458

Henriques Lab. (2020, 8. Des). *StarDist*. Github. https://github.com/HenriquesLab/ZeroCostDL4Mic/wiki/Stardist

Henseleit, U., Rosenbach, T. & Kolde, G. (1996). Induction of apoptosis in human HaCaT keratinocytes. *Archives of Dermatological Research, 288*(11), 676-683. https://doi.org/10.1007/BF02505277

Hu, CK., Coughlin, M. & Mitchison, T. J. (2012). Midbody assembly and its regulation during cytokinesis. *Molecular Biology of the Cell, 23*(6), 1024-1034. https://doi.org/10.1091/mbc.E11-08-0721

Kim, M., Kim, T., Johnson, R. L. & Lim, DS. (2015). Transcriptional Co-repressor Function of the Hippo Pathway Transducers YAP and TAZ. *Cell reports, 11*(2), 270-282. https://doi.org/10.1016/j.celrep.2015.03.015

Kim, M. H., Kim, J., Hong, H., Lee, SH., Lee, JK., Jung, E. & Kim, J. (2016). Actin remodeling confers BRAF inhibitor resistance to melanoma cells through YAP / TAZ activation. *The EMBO Journal, 35*(5), 462-478. https://doi.org/10.15252/embj.201592081

Kiyomitsu, T. & Cheeseman, I. M. (2013). Cortical dynein and asymmetric membrane elongation coordinately position the spindle in anaphase. *Cell, 154*(2), 391-402. https://doi.org/10.1016/j.cell.2013.06.010

Koirala, A., Walsh, K. B., Wang, Z. & McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture, 162*, 219-234. https://doi.org/10.1016/j.compag.2019.04.017

Lotze, J., Reinhardt, U., Seitz, O. & Beck-Sickinger, A. G. (2016). Peptide-tags for site-specific protein labelling: In vitro and in vivo. *Molecular BioSystems, 12*(6), 1731-1745. https://doi.org/10.1039/c6mb00023a

Lång, E., Grudic, A., Pankiv, S., Bruserud, O., Simonsen, A., Bjerkvig, R., Bjørås, M., Bøe, S. O. (2012) The arsenic-based cure of acute promyelocytic leukemia promotes cytoplasmic sequestration of PML and PML/RARA through inhibition of PML body recycling. *Blood, 120*(4), 847-57.https://doi.org/10.1182/blood-2011-10-388496

Lång, E., Połeć, A., Lång, A., Valk, M., Blicher, P., Rowe, A. D., Tønseth, K. A., Jackson, C. J., Utheim, T. P., Janssen, L. M. C., Eriksson, J. & Bøe S. O. (2018). Coordinated collective migration and asymmetric cell division in confluent human keratinocytes without wounding. *Nature Communications, 9*(1), article number 3665. https://doi.org/10.1038/s41467-018-05578-7

Mertz, A., Che, Y., Banerjee, S., Goldstein, J. M., Rosowski, K. A., Revilla, S. F., Niessen, C. M., Marchetti, M. C., Dufresne, E. R. & Horsley, V. (2013). Cadherin-based intercellular adhesions organize epithelial cell-matrix traction forces. *Proceedings of the National Academy of Sciences of the United States of America, 110*(3), 842-847. https://doi.org/10.1073/pnas.1217279110

Miller, E., Yang, J., Deran, M., Wu, C., Su, A., Bonamy, G., Liu, J., Peters, E. & Wu, X. (2012). Identification of serum-derived sphingosine-1-phosphate as a small molecule regulator of YAP. *Chemistry and Biology, 19*(8), 955-962. https://doi.org/10.1016/j.chembiol.2012.07.005

Misra, J. & Irvine, K. (2018). The hippo signaling network and its biological functions. *Annual Review of Genetics, 52*, 65-87. https://doi.org/10.1146/annurev-genet-120417-031621

Montell, D. J. (2008). Morphogenetic cell movements: Diversity from modular mechanical properties. *Science, 322*(5907), 1502-1505. https://doi.org/10.1126/science.1164073

Nichols, J. A., Herbert, C. H. & Baker, M. (2019). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews, 11*(1), 111-118. https://doi.org/10.1007/s12551-018-0449-9

Niessen, C. M. (2007). Tight junctions/adherens junctions: Basic structure and function. *Journal of Investigative Dermatology, 127*(11), 2525-2532. https://doi.org/10.1038/sj.jid.5700865

Koech, K. E. (2020). *Cross-Entropy Loss Funcion*. Towards data science. https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e#_=_

Oka, T., Mazack, V. & Sudol, M. (2008). Mst2 and Lats kinases regulate apoptotic function of Yes kinase-associated protein (YAP). *Journal of Biological Chemistry, 283*(41), 27534-27546. https://doi.org/10.1074/jbc.M804380200

Oki, T., Nishimura, K., Jiro, K., Togami, K., Maehara, A., Izawa, K., Sakaue-Sawano, A., Niida, A., Miyano, S., Aburatani, H., Kiyonari, H., Miyawaki, A. & Kitamura, T. (2014). A novel cell-cycle-indicator, mVenus-p27K−, identifies quiescent cells and visualizes G0–G1 transition. *Scientific Reports, 4*, article number 4012. https://doi.org/10.1038/srep04012

Pan, D. (2010). The hippo signaling pathway in development and cancer. *Developmental Cell*, *19*(4), 491-505. https://doi.org/10.1016/j.devcel.2010.09.011

Pan, D. (2007). Hippo signaling in organ size control. *Genes and Development, 21*(8), 886-897. https://doi.org/10.1101/gad.1536007

Park, J. S., Kim, D. K., Shah, S. R., Kim, H. N., Kshitiz, Kim, P., Quiñones-Hinojosa, A. & Levchenko, A. (2019). Switch-like enhancement of epithelial-mesenchymal transition by YAP through feedback regulation of WT1 and Rho-family GTPases. *Nature communtications, 10*(1), article number 2797. https://doi.org/10.1038/s41467-019-10729-5

Pentagonspace. (2020, 7. May). *History of Deep Learning on object detection*. Medium. https://medium.com/@pentagonspace4747/history-of-deep-learning-on-object-detection-7416c35f8a72

Piacentino, M. L., Li, Y. & Bronner, M. E. (2020). Epithelial-to-mesenchymal transition and different migration strategies as viewed from the neural crest. *Current Opinion in Cell Biology, 66,* 43-50. https://doi.org/10.1016/j.ceb.2020.05.001

Pizzorno, J., Murray, M. & Joiner-Bey, H. (2015). *The Clinician's Handbook of Natural Medicine* (3. Edition, 158-179). Elsevier.

Qi, G., Ogawa, I., Kudo, Y., Miyauchi, M., Siriwardena, B. S. M. S., Shimamoto, F., Tatsuka, M. & Takata, T. (2007). Aurora-B expression and its correlation with cell proliferation and metastasis in oral cancer. *Virchows Archiv, 450*(3), 297-302. https://doi.org/10.1007/s00428-006-0360-9

Sambasivarao, K. (2019). *Non-maximum Suppression (NMS)*. Towards data science. https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c

Shelhamer, E., Long, J. & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 640-651. https://doi.org/10.1109/TPAMI.2016.2572683

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S. & Schmid B. (2012). Fiji: An open-source platform for

biological-image analysis. *Nature methods, 9*(7), 676-682.
https://doi.org/10.1038/nmeth.2019

Schleimer, R. P., Kato, A., Kern, R., Kuperman, D. & Avila, P. C. (2007). Epithelium: At the interface of innate and adaptive immune responses. *The Journal of Allergy and Clinical Immunology, 120*(6), 1279-1284. https://doi.org/10.1016/j.jaci.2007.08.046

Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. (2018 8. Jun.). *Cell Detection with Star convex Polygons* [paper presentation]. In Frangi, A.F., Schnabel, J.A., Davatzikos, C., López Alberola, C. & Fichtinger G. (ed). Medical Image Computing and Computer Assisted Intervention 2018 - MICCAI 21st International Conference, Granada, Spain. https://doi.org/10.1007/978-3-030-00934-2_30

Schmidt U, Weigert M, Broaddus C and Myers G. (2020*). Labelling/annotation: How many images or nucleus (object) instances do I have to annotate for good results?*. StarDist. https://stardist.net/docs/faq.html#how-many-images-or-nucleus-object-instances-do-i-have-to-annotate-for-good-results

Scieglinska, D., Krawczyk, Z., Sojka, D. R. & Gogler-Pigłowska, A. (2019). Heat shock proteins in the physiology and pathophysiology of epidermal keratinocytes. *Cell Stress Chaperones, 24*(6), 1027-1044. https://doi.org/10.1007/s12192-019-01044-5

Steinberg, ML. & Defendi, V. (1983). Transformation and immortalization of human keratinocytes by SV40. *Journal of Investigative Dermatology, 81*(1 Suppl),131–136. https://doi.org/10.1111/1523-1747.ep12540905

Sun, L., Gao, J., Dong, X., Liu, M., Li, D., Shi, X., Dong, J. T., Lu, X., Liu, C. & Zhou J. (2008). EB1 promotes Aurora-B kinase activity through blocking its inactivation by protein phosphatase 2A. *Proceedings of the National Academy of Sciences of the United States of America, 105*(20), 7153-7158. https://doi.org/10.1073/pnas.0710018105

Svitkina, T. (2018). The actin cytoskeleton and actin-based motility. *Cold Spring Harbor Perspectives in Biology, 10*(1), article number 018267. https://doi.org/10.1101/cshperspect.a018267

Tinevez, J. Y., Perry, N., Schindelin, J., Hoopes, G. M., Reynolds, G. D., Laplantine, E., Bednarek, S. Y., Shorte, S. L. & Eliceiri, K. W. (2017). TrackMate: An open and extensible platform for single-particle tracking. *Methods, 115*, 80-90. https://doi.org/10.1016/j.ymeth.2016.09.016.

Trepat, X., Grabulosa, M., Puig, F., Maksym, G.N., Navajas, D. & Farré, R. (2004). Viscoelasticity of human alveolar epithelial cells subjected to stretch. *American Journal of Physiology Lung Cellular and Molecular Physiology, 287*(5), 1025-1034. https://doi.org/10.1152/ajplung.00077.2004

Van Erp, P. E. J., Boezeman, J. B. M. & Brons, P. P. T. (1996). Cell cycle kinetics in normal human skin by in vivo administration of iododeoxyuridine and application of a differentiation marker - Implications for cell cycle kinetics in psoriatic skin. *Analytical Cellular Pathology, 11*(1), 43-54.

Vicente-Manzanares, M. & Sanchez-Madrid, F. (2000). Cell polarization: A comparative cell biology and immunological view. *Developmental Immunology, 7*(2-4), 51–65. https://doi.org/10.1155/2000/70801

Weedon, D. (2010). *Weedon's Skin Pathology* (3. Edition). Elsevier.

Wilson, A. (2019*). A Brief Introduction to Supervised Learning*. Towards data science. https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590.

Xing, W. & Du, D. (2018). Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *Journal of Educational Computing Research, 57*(4), article number 073563311875701. https://doi.org/10.1177/0735633118757015

Yang, L., Ghosh, R. P., Matthew, F. J., Chen, S., You, C., Narayan, R. R., Melcher, M. L. & Liphart, J T. (2020). NuSeT: A deep learning tool for reliably separating and analyzing crowded cells. *PLoS Computational Biology, 16*(9), article number e1008193. https://doi.org/10.1371/journal.pcbi.1008193

Yousef, H., Alhajj, M. & Sharma, S. (2017, Jun.). *Anatomy, Skin (Integument), Epidermis*. StatPearls. https://www.statpearls.com/ArticleLibrary/viewarticle/21212

Yu, FX., Zhao, B., Panupinthu, N., Jewell, J., Lian, I., Wang, L., Zhao, J., Yuan, H., Tumaneng, K., Li, H., Fu, X., Mills, G. & Guan, KL. (2012). Regulation of the Hippo-YAP pathway by G-protein-coupled receptor signaling. *Cell. 2012, 150*(4), 780-791. https://doi.org/10.1016/j.cell.2012.06.037

Zhao, B., Li, L., Karen, T., Wang, CY. & Guan, K. (2010). A coordinated phosphorylation by Lats and CK1 regulates YAP stability through SCFβ-TRCP. *Genes & Development, 24*(1), 72-85. https://doi.org/10.1101/gad.1843810

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, D. J., Wang, C., Chinnaiyan, A. M., Lai, Z. & Guan, K. (2008) TEAD mediates YAP-dependent gene induction and growth control. *Genes & Development, 22*(14), 1962-1971. https://doi.org/10.1101/gad.1664408

Zhao, B., Wei, X., Li, W., Udan, R. S., Yang, Q., Kim, J., Xie, J., Ikenoue, T., Yu, J. & Li, L. (2007). Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes and Development, 21*(21), 2747-2761. https://doi.org/10.1101/gad.1602907

Zhou, J., Wang, J., Zhang, N., Zhang, Y. & Li, Q. (2015). Identification of biomechanical force as a novel inducer of epithelial-mesenchymal transition features in mechanical stretched skin. *American Journal of Translational Research, 7*(11), 2187-2198.

APPENDIXES

Appendix A – Materials

Table A.1: Overview of all materials used in this project.

Product name	Specification	Manufacturer
HaCaT cell line. Article nr: 300493	Human keratinocyte cell line (HaCaT). Cells referred to as wildtype (WT) did not containg fluorescent tags (Boukamp et al. 1988). Cells expressing fluorescently tagged mCherry-histone H2B protein from this cell line were also used (Lång et al. 2012)	CLS
Iscove's Modified Dulbecco's Medium (IMDM). Product nr: I3390	A modified DME containing 4,500 mg/L glucose, sodium pyruvate, sodium bicarbonate, selenium, and amino acids like L-	Sigma-Aldrich
Fetal Bovine serum (FBS)	Contains over 1000 compounds. These include proteins, lipids, carbohydrates, enzymes and hormones.	Thermo Scientific
PenStrep, Penicillin-Streptomycin. Catalog nr: DE17-602E	A mixture of penicillin and streptomycin (an effective antibiotic against gram positive bacteria, and against both gram negative and gram positive bacteria, respectively). 893975	Lonza
Phosphate buffered saline (PBS)	Water mixed with 8,0 g/L NaCl, 0,2 g/L KCl, 1,42 g/L Na_2HPO_4, 0,24 g/L KH_2PO_4.	Produced locally
Trypsin-EDTA. Catalog nr: CC-5012	0,05% Trypsin, and 0.53mM EDTA with phenol red indicator	Lonza

Product name	Specification	Manufacturer
Trypan blue stain. Catalog nr: 15250061	0.4% stain concentration. Stains dead cells and used for cell counting.	Thermo Fisher Scientific
Vectashield anti-fade mounting medium. Product nr: H-1200-10	The Medium contains DAPI, a fluorescent marker which binds to the DNA	Vector labs
Anti-Aurora B antibody. Product nr: ab2254	Primary polyclonal antibody produced in rabbit.	Abcam
Anti-YAP1 Antibody (63.7). Product nr: sc-101199	Primary monoclonal antibody produced in mouse	Santa Cruz biotech
Alexa Fluor 488 goat anti-mouse. Product nr: A11029	Secondary antibody	Thermo Fisher Scientific
Alexa Fluor 594 goat anti-rabbit. Product nr: A11072	Secondary antibody	Thermo Fisher Scientific
Paraformaldehyde, PFA	Cross-linking agent	Sigma-Aldrich
Triton X-100	Permeabilizing agent	Sigma-Aldrich
Bovine Serum Albumin fraction V (BSA)	Blocking solution	Sigma-Aldrich

Appendix B - Instruments

Tabell B.1: An overview of instruments used in this project.

Product name	Specifications	Manufacturer
T75 culture flasks	75 cm^2 growth area for cell culture	VWR
12- well glass-bottom plates	Working volume: 1-2 mL. 14 mm glass diameter coverslip	MatTek corporation
Falcon tube	15 and 50 mL volume tubes	Sarstedt
Eppendorf tube	1,5 mL volume	Sarstedt

Product name	Specifications	Manufacturer
Countess 3FL Automated Cell Counter	Used to automatically estimate the number and viability of cells in a solution	Thermo Fisher Scientific
Countess Cell Counting Chamber Slide	10 μL chamber volume	Thermo Fisher Scientific
Zeiss AxioObserver.Z1 wide-field microscope	Automated inverted wide-field microscope. It can acquire images of live image samples and fixed samples in multi-well plates or on glass slides. Equipped with a Colibri7 LED light source, an Orca Flash 4.0 V3 digital CMOS camera (Hamamatsu), and incubation chamber.	Zeiss
S1 pipet filler	An automated pipet fille	Thermo Fisher
Disposable pipets	Disposable sorological pipets used for the S1 pipet filler. The pipets are sterile and individually wrapped.	Thermo Fisher
Glass coverslips	12 mm diameter	VWR
Glass slides	Microscope superfrost adhesion slides. Dimentions: 76x26 mm	Thermo Fisher
Leica TCS SP8 STED microscope	A confocal microscope equipped with STED, UV and WL lightsources. The microscope is equipped with several lenses and the (HC PL APO CS2) 40x 1.30 NA oil immersion lense was used for experiments	Leica

Appendix C – Computer programs

Table C.1: An overview of computer programs used for image analysis..

Program	Specifications	Manufacturer
ZEN 3.1 pro software	Software used to control the Zeiss Axio observer.Z1 microscope	Zeiss
StarDist 2D	2D segmentation program. The library keras was used as a neural network.(Schmidt et al. 2018)	Uwe Schmidt, Martin Weigert
Fiji ImageJ	An open-source image processing software (Schindelin et al. 2012)	ImageJ
Leica software	A software used for establishing the settings of Leica imaging and image editing	Leica
Pycharm	A software used for computer programming using the python language	JetBrains

Appendix D – Raw data
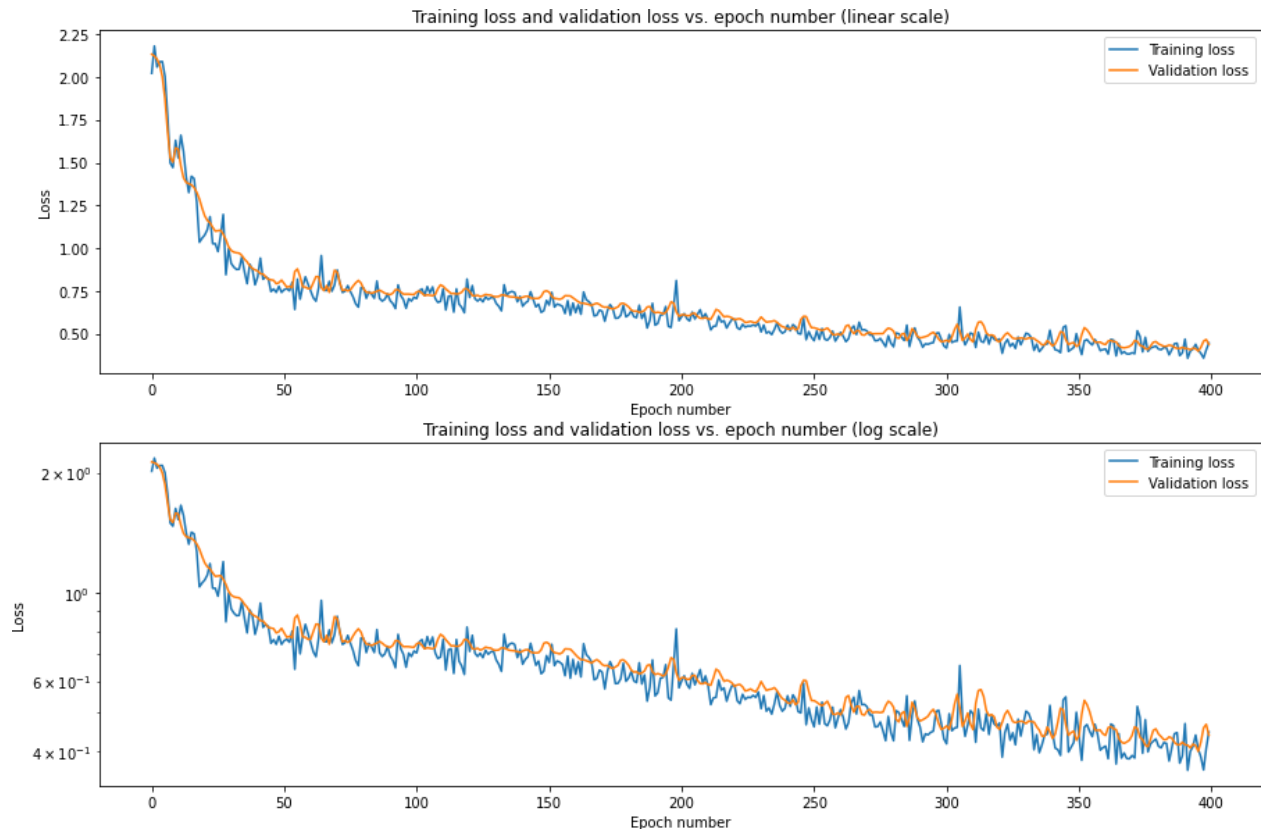
Figure D.1: The numeric and log scale training loss vs epoch number and validation loss vs epoch number for the '100' model.
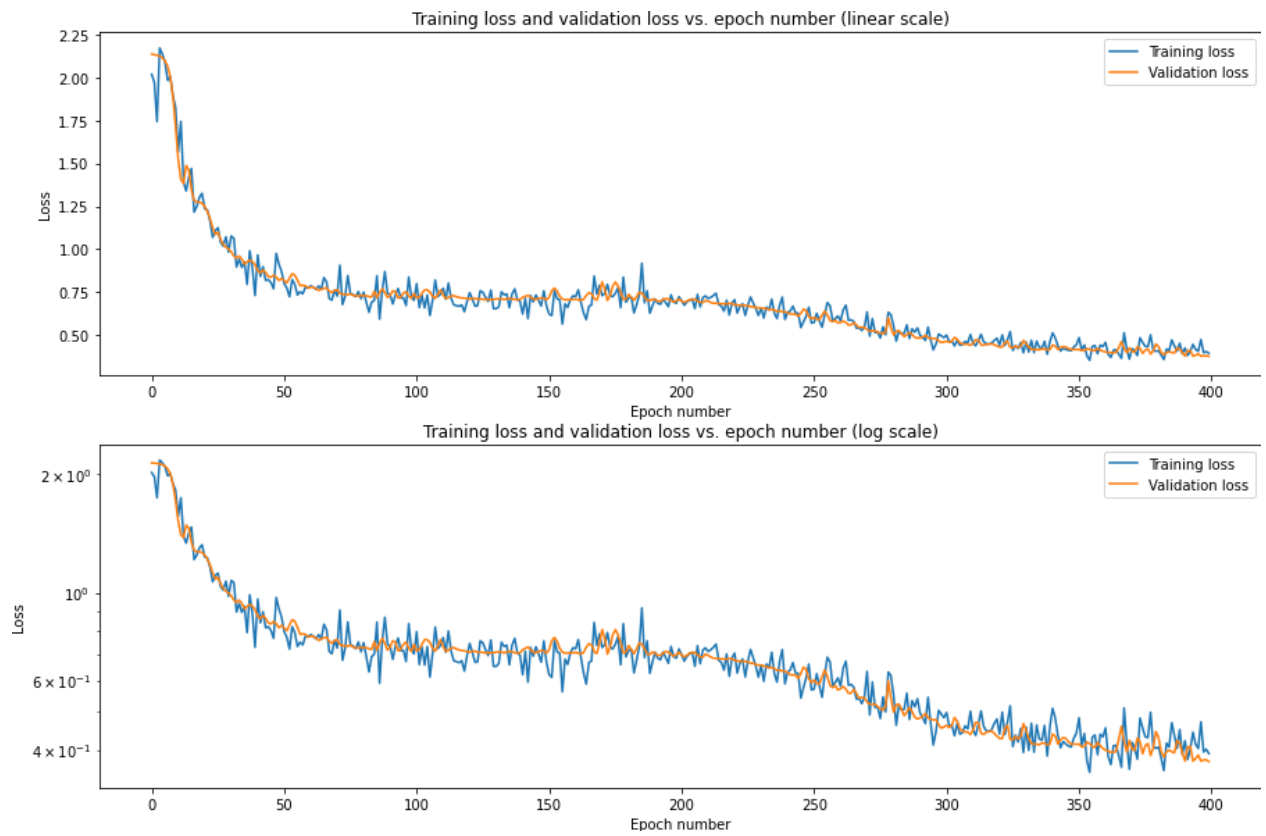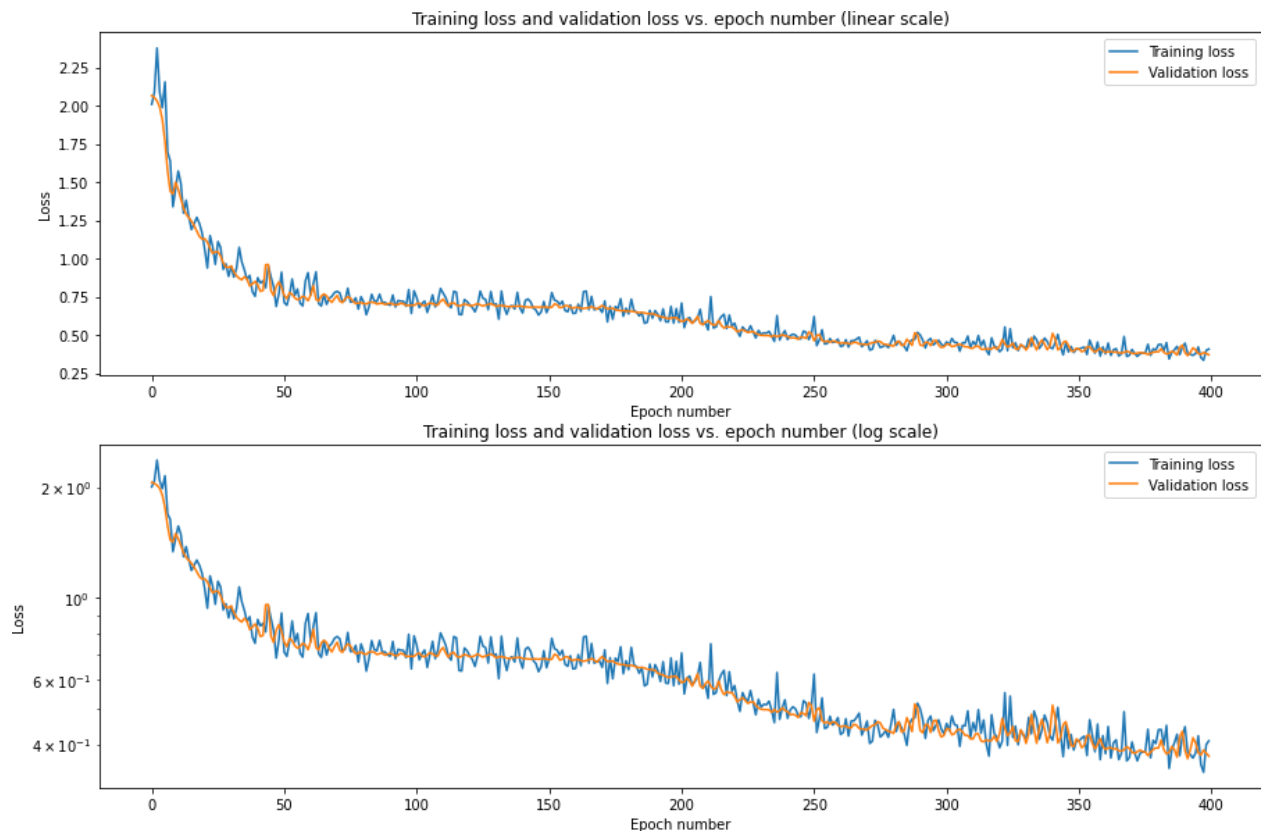
Figure D.3: The numeric and log scale training loss vs epoch number and validation loss vs epoch number for the '500' model.
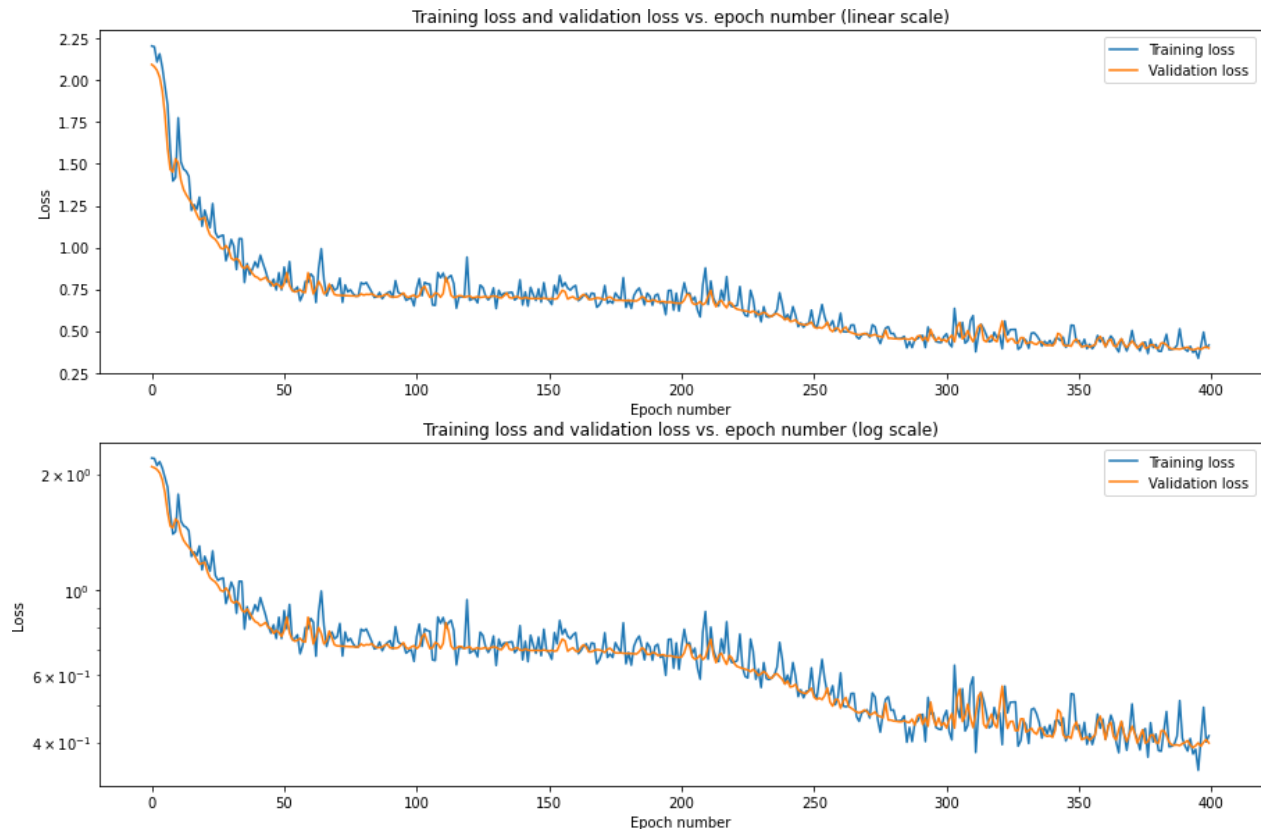
Figure D.4: The numeric and log scale training loss vs epoch number and validation loss vs epoch number for the '1000 1-step' model.
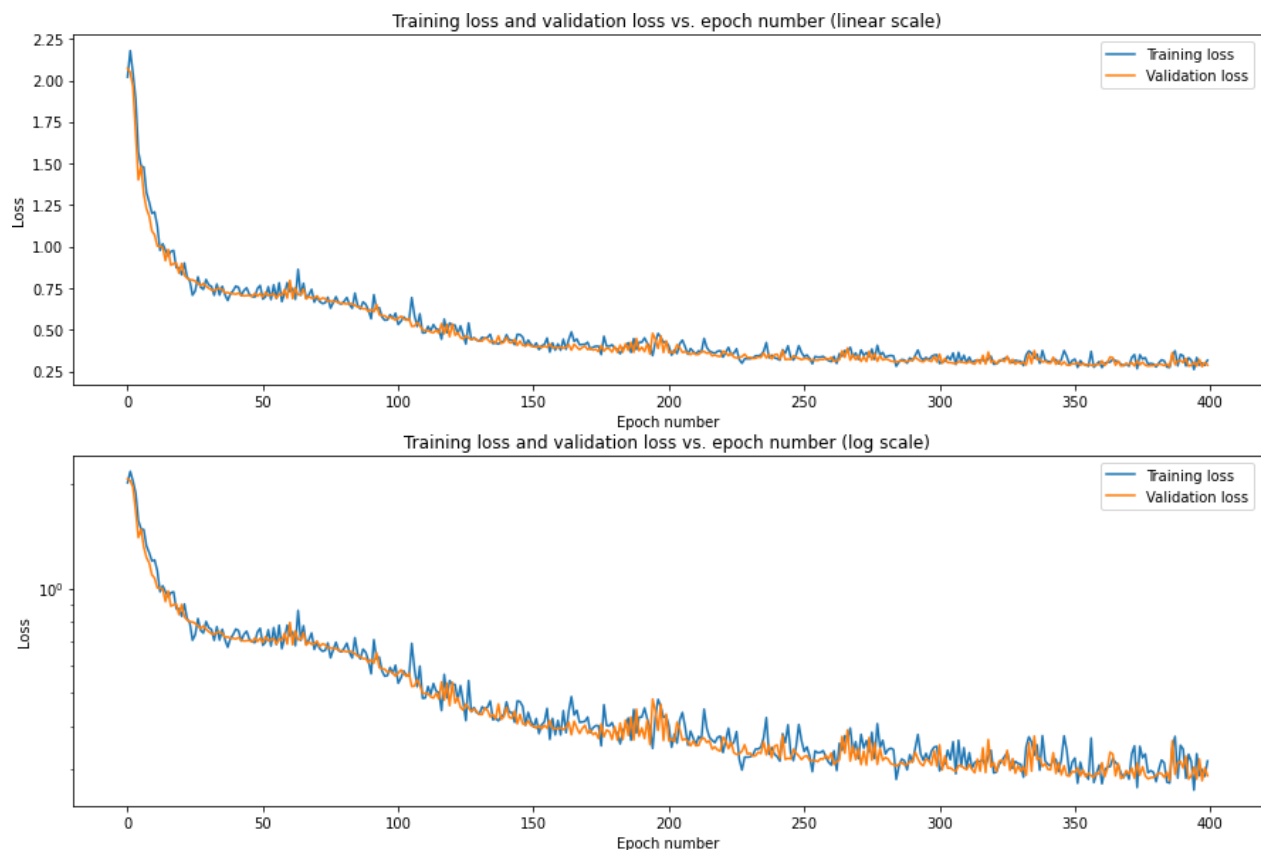

Figure D.5: The numeric and log scale training loss vs epoch number and validation loss vs epoch number from the first training session of the '1000 2-step' model.
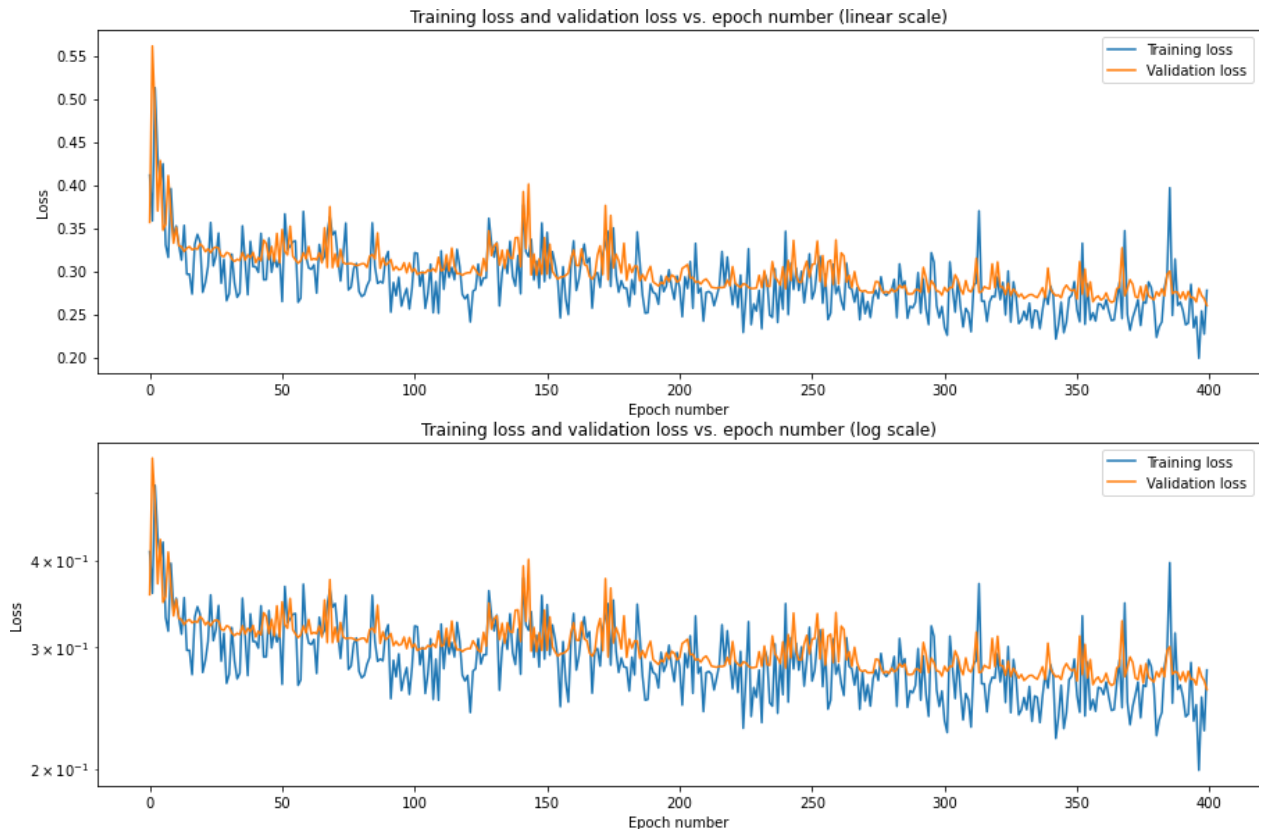
Figure D.6: The numeric and log scale training loss vs epoch number and validation loss vs epoch number from the second training session of the '1000 2-step' model.

D.7: Script for plotting data (perimeter, area)

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import os

path ="C:\\Users\\annlaa\\Desktop\\Øyvind\\Interphase\\"

names = []
for subdir, dirs, files in os.walk(path):
    for file in files:
        names.append(os.path.join(file))

list_of_data=[]

for name in names:
    df = pd.read_csv(path + name, sep=',', engine = 'python', usecols= [1, 7])
    a = df.groupby(['Slice'])['Area'].mean()
    result = a.drop(columns=['Slice'], axis=1)
    new = np.asarray(result)
    list_of_data.append(new)

mean=np.mean(list_of_data, axis=0)
std=np.std(list_of_data, axis=0)

x = np.arange(len(mean))
```

74

```
x = np.multiply(x, 4)
x = np.divide(x, 60)

fig, ax = plt.subplots(ncols=1, nrows=1, figsize = (6, 5))

plt.plot(x, mean, '-', color='r')
plt.fill_between(x, mean-std, mean+std, color='r', alpha=0.2)

plt.xlabel("Time (h)", fontsize = 13)
plt.xticks(fontsize = 13)
plt.ylabel("Cell area (µm^2)", fontsize = 13)
plt.yticks(fontsize = 13)
plt.ylim(0, 230)
plt.savefig("Mean_area_all_cells", dpi=300)
plt.show()
```

D.8: Script for plotting cell count

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import os

path1 = "C:\\Users\\annlaa\\Desktop\\Øyvind\\Mitosis\\"
path2 ="C:\\Users\\annlaa\\Desktop\\Øyvind\\Interphase\\"

names = []
for subdir, dirs, files in os.walk(path1):
    for file in files:
        names.append(os.path.join(file))

name2 = []
for subdir, dirs, files in os.walk(path2):
    for file in files:
        name2.append(os.path.join(file))

list_of_data=[]

for name in names:
    df = pd.read_csv(path1 + name, sep=',', engine = 'python')
    df2 = df["Slice"]
    df2.to_numpy()
    new_list = []

    for i in range(1, 600):
        n = (df2 == i).sum()
        new_list.append(n)
```

```python
        Result=np.asarray(new_list)
        list_of_data.append(Result)

average = np.average(list_of_data, axis=0)
std = np.std(list_of_data, axis=0)

list_of_data2=[]

for name in name2:
    df = pd.read_csv(path2 + name, sep=',', engine = 'python')
    df2 = df["Slice"]
    df2.to_numpy()
    new_list = []

    for i in range(1, 600):
        n = (df2 == i).sum()
        new_list.append(n)

    Result=np.asarray(new_list)
    list_of_data2.append(Result)

average2 = np.average(list_of_data2, axis=0)
std2 = np.std(list_of_data2, axis=0)

x = np.arange(len(average))
x = np.multiply(x, 4)
x = np.divide(x, 60)
fig, ax = plt.subplots(ncols=1, nrows=1, figsize = (6, 5))

#plt.plot(x, average, '-', color='b')
#plt.fill_between(x, average-std, average+std, color='b', alpha=0.2)

plt.plot(x, average2, '-', color='r')
plt.fill_between(x, average2-std2, average2+std2, color='r', alpha=0.2)

plt.xlabel("Time (h)", fontsize = 13)
plt.xticks(fontsize = 13)
plt.ylabel("Number of cells", fontsize = 13)
plt.yticks(fontsize = 13)
plt.ylim(0, 250)
plt.savefig("Average_cell number_all_cells", dpi=300)
plt.show()
```

D.9: Script for creating tracking files

```python
import cv2
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import os


def frame_by_frame(grouped):
    n = 1
    for name, group in grouped:
        print("working on " + str(name))
        frame = grouped.get_group(name)
        result = frame.drop(columns='Slice', axis=1)
        Results = result.to_numpy()
        img = np.zeros((187, 196, 3), np.uint8)
        radius = 4
        thickness = -1
        color = (255, 255, 255)
        for coordinates in Results:
            cv2.circle(img, tuple(coordinates), radius, color, thickness)
        plt.imshow(img)
        plt.xticks([]), plt.yticks([])
        print("Done with " + str(name))
        cv2.imwrite(os.path.join(path, "frame" + str(n) + ".tif"), img)
        n = n + 1


path = "C:\\Users\\annlaa\\Desktop\\Tracking_files\\Test2_IMDM\\T_files\\"

data = pd.read_csv("IMDM-test2_Results_no scale.csv", engine = 'python',
usecols=[3,4,7])

column_list = list(data.columns)

df = pd.DataFrame(data=data, columns=column_list)

df['X']=df['X'].astype(int)
df['Y']=df['Y'].astype(int)
```

```python
grouped = df.groupby(['Slice'])

frame_by_frame(grouped = grouped)
```