DISSERTATIONES
INFORMATICAE
UNIVERSITATIS
TARTUENSIS
52

MOHAMMED ABDULHAMEED SHAIF ALI

Deep Learning Methods for Cell Microscopy Image Analysis

DISSERTATIONES INFORMATICAE UNIVERSITATIS TARTUENSIS
52

DISSERTATIONES INFORMATICAE UNIVERSITATIS TARTUENSIS
52

MOHAMMED ABDULHAMEED SHAIF ALI

Deep Learning Methods for Cell Microscopy Image Analysis

UNIVERSITY OF TARTU
Press
1632

Institute of Computer Science, Faculty of Science and Technology, University of Tartu, Estonia.

Dissertation has been accepted for the commencement of the degree of Doctor of Philosophy (PhD) in Computer Science on 2nd of April, 2024 by the Council of the Institute of Computer Science, University of Tartu.

Supervisors

Dr. Dmytro Fishman
University of Tartu
Tartu, Estonia

Dr. Leopold Parts
Wellcome Sanger Institute
Cambridgeshire, UK &
University of Tartu
Tartu, Estonia

Opponents

Prof. Carolina Wählby
SciLifeLab and Dept. IT, Uppsala University
Uppsala, Sweden

Dr. Mohammed Al-masni
Department of Artificial Intelligence, Sejong University
Seoul, South Korea

The public defense will take place on 21st of May, 2024 at 10:15 in University of Tartu Delta building, Narva maantee 18, room 2046.

The publication of this dissertation was financed by the Institute of Computer Science, University of Tartu.

ISSN 2613-5906 (print) ISSN 2806-2345 (PDF)
ISBN 978-9916-27-514-6 (print) ISBN 978-9916-27-515-3 (PDF)

University of Tartu Press
http://www.tyk.ee/

To my family and friends

ABSTRACT

Analyzing microscopy images of cells is indispensable in modern biological research as it allows scientists to understand cellular activity better and develop new drugs. With the advent of advanced microscopes and the increasing amount of microscopy data, there is a growing demand for automated and accurate image analysis procedures. Deep learning technology has recently proven effective in analyzing images and various computer vision tasks, including image segmentation and object detection. We have explored and implemented various deep-learning methods to analyze cell microscopy images in a series of studies that comprise the core contributions of this thesis.

We first concentrate on the task of segmenting nuclei in microscopy images, which is often regarded as the preliminary step in such image analysis. Our work's initial phase involved evaluating state-of-the-art deep learning methods for nuclei segmentation, which proved challenging, particularly in the case of specific types of microscopy images known as brightfield images. Subsequently, we proposed a lightweight model, PP-U-Net, for the same purpose. PP-U-Net performs on par with the state-of-the-art with 20-fold fewer parameters. During investigating the causes of PPU-Net segmentation errors, we found that images often contain artifacts that substantially distort downstream analyses. To address this challenge, we developed a novel framework, ScoreCAM-U-Net, which leverages deep learning methods to identify and eliminate artifacts with minimal user input. This framework combines the performance advantages of pixel-level segmentation with the convenience of image-level annotations and is expected to become a standard step for all large-scale microscopy experiments. Finally, we distilled the knowledge and insights we gained from this work to support candidate drug discovery research. The methods developed in this study have been utilized by a variety of industrial and academic organizations, which have benefited from them in their research.

CONTENTS

LIST OF FIGURES

22. Cell fluorescence intensity dependence on M4 receptor ligand concentration determined with live cell fluorescence microscopy at the presence of 2 *nM* UR-CG072. Displacement curves of three different ligands are shown: pirenzepine (A, C), atropine (B) and arecholine (D). Three different artifact removal methods at the image analysis stage are compared (colors): manual artifact segmentation, ScoreCAM-U-Net segmentation and no artifact removal. For each combination of ligand and artifact removal method a regression analysis is performed with Hill equation (Hill coefficient fixed at -1) with the best fits shown as continuous lines. For each displacement curve, the $Log(IC_{50}) \pm$ SD is presented, where SD represents the standard deviation estimation of $Log(IC_{50})$. Each displacement curve was measured in duplicates with each data point representing the average fluorescence intensity of cells in each well. 62

LIST OF ABBREVIATIONS

AE	AutoEncoder
BF	BrightField
BCE	Binary Cross Entropy
CHO	Chinese Hamster Ovary
CNN	Convolutional Neural Network
DL	Deep Learning
DNA	Deoxyribonucleic Acid
FN	False Negative
FP	False Positive
GPCRs	G-protein Coupled Receptors
HeLa	Henrietta Lacks
hM4R	Human M4 Muscarinic Acetylcholine Receptor
IoU	Intersection over Union
$LogIC_{50}$	Logarithm of half maximal Inhibitory Concentration
LNCaP	Lymph Node Carcinoma of the Prostate
M4	Muscarinic acetylcholine receptor 4
mAChR	Muscarinic ACetylcholine Receptors
MCF-7	Michigan Cancer Foundation-7
MDCK	Madin-Darby Canine Kidney
ML	Machine Learning
NIH3T3	National Institutes of Health 3-day Transfer, Inoculum 3 × 105 cells
PaDim	Patch Distribution Modeling
PatchSVDD	Patch Support Vector Data Description
PPU-Net	Pyramid Pooling U-Net
ReLU	Rectified Linear Unit
RF	Random Forest
Score-CAM	Score-weighted Class Activation Mapping
TP	True Positive

LIST OF ORIGINAL PUBLICATIONS

Publications included in the thesis

I. **Mohammed A. S. Ali**, Oleg Misko, Sten-Oliver Salumaa, Mikhail Papkov, Kaupo Palo, Dmytro Fishman, and Leopold Parts. Evaluating Very Deep Convolutional Neural Networks for Nucleus Segmentation from Brightfield Cell Microscopy Images. **SLAS DISCOVERY: Advancing the Science of Drug Discovery**. 2021 Oct;26(9):1125-37.

Author's contribution: (leading author) Contributed to selecting studied models for segmenting nuclei; designed the PP-U-Net model; contributed to designing experiments and conducted most of them; Wrote the initial draft of the manuscript; contributed to figure design and drawing; participated in reviewing and refining the paper.

II. **Mohammed A. S. Ali**, Kaspar Hollo, Tõnis Laasfeld, Jane Torp, Maris-Johanna Tahk, Ago Rinken, Kaupo Palo, Leopold Parts, and Dmytro Fishman. ArtSeg—Artifact segmentation and removal in brightfield cell microscopy images without manual pixel-level annotations. **Scientific reports**. 2022 Jul 6;12(1):1-13.

Author's contribution: (leading author) Contributed to the designing and developing Score-CAM-U-Net together with co-authors; contributed to designing experiments and implementing some of them; contributed to figures drawing; drafted the initial version of the manuscript except for the sections in Methods and results related to pharmacological parameter estimates; participated in reviewing and refining the paper.

III. Maris-Johanna Tahk, Jane Torp, **Mohammed A. S. Ali**, Dmytro Fishman, Leopold Parts, Lukas Grätz, Christoph Müller, Max Keller, Santa Veiksina1, Tõnis Laasfeld, and Ago Rinken. Live-cell microscopy or fluorescence anisotropy with budded baculoviruses—which way to go with measuring ligand binding to M4 muscarinic receptors? **Open Biology**. 2022 Jun;12(6): 220019.

Author's contribution: Contributed to preparing the software used in the experiments; contributed to figures drawing related to the deep learning section; participated in reviewing and refining the written paper.

1. INTRODUCTION

A cell is the main building unit of all living organisms and the site for almost all biological activity. Cells manufacture proteins, produce energy, transport materials, and dispose waste. They can replicate themselves through mitosis and can interact with their environment through signaling pathways. Without cells, life would not exist. Understanding how cells function enables people to learn more about biological processes, diseases, and effective medications.

Cells are tiny. A normal mammalian cell, for example, has a diameter of around $10-20\mu m$ - roughly one-fifth the size of the smallest object the naked eye can see. As the human eye lacks the resolution to see objects at this scale, leave alone rigorously study, scientists utilize special equipment – a microscope – to go beyond their natural abilities. A *microscope* is a device that magnifies small objects that would otherwise be impossible to observe. In this thesis, we analyze images acquired from the widely-used *light microscope*- an instrument that utilizes light characteristics to enlarge microscopic objects.

With the advent of advanced microscopes, microscopy has recently been substantially automated. Consequently, the amount of microscopy data has dramatically increased as now millions of images can be automatically produced in a single experiment [1, 2, 3]. To investigate such enormous image collections, automated and accurate analysis procedures are crucial.

Image analysis techniques have been continuously evolving. Particularly, machine learning algorithms have proven effective and are thus widely used in the field, with numerous of them also made easily accessible via bioimage analysis tools [4] such as ImageJ/Fiji [5], CellProfiler [6], and Ilastik [7]. Despite their widespread adoption, these tools occasionally fail to produce accurate enough results as the conventional machine learning algorithms demanded careful engineering and domain expertise to extract features [8]. The stability and dependability of such approaches on a wide scale are further challenged by the variations in signal quality and differences between imaging modalities inherent to high-throughput cellular microscopy [9, 10].

Unlike conventional machine learning, which demands careful feature engineering, deep learning is a family of ML techniques that automatically extracts the relevant patterns from raw data by using computational models with multiple processing layers. A form of deep learning models known as Convolutional Neural Networks (CNNs) have demonstrated their ability to identify key features and patterns in an image through the process of sliding a filter across the image. CNNs have revolutionized the computer vision field with state-of-the-art performance in tasks such as image classification [11, 12], object detection [13, 14] and segmentation [15, 16]. The rapid development of deep learning has continuously provided new insights that could, in principle, advance microscopy image analysis but have not yet been fully leveraged for this purpose.

The work towards this thesis started with utilizing the recent research in deep

learning to segment nuclei, which often is one of the first steps in the cell microscopy image processing pipelines. Accurate segmentation of nuclei is paramount across a spectrum of biological applications. For instance, nuclear structure and morphology abnormalities are commonly associated with diseases such as cancer. Thus, segmenting nuclei and examining their properties contributes to the diagnosis and progression monitoring of cancer [17]. Another compelling application of efficient nuclei detection is its potential to facilitate cell tracking, thereby supporting detailed studies of cellular behaviors and the analysis of their responses to medication [18]. Brightfield images were the main focus since they are relatively simple to produce but difficult to inspect and analyze. We evaluated a set of state-of-the-art architectures (at the time of Publication I) for segmenting nuclei as well as designed a new one called *PPU-Net* [19] for the same purpose. The evaluated models showed variable performance in segmenting nuclei from brightfield images, with the PPU-Net performing on par with the state-of-the-art available at that time with 20-fold fewer values to optimize during the training process (i.e., parameters), making it lighter and less complex. To gain a deeper understanding, we investigated the causes of inconsistent performance across different cell types and individual images, the sufficient number of training images required, and the most common sources of error.

During investigating the causes of PPU-Net segmentation errors, we found that having artifacts (a signal that does not reflect what is expected) in the images is a major source of error. We, therefore, aimed to deeper understand the artifacts problem. We examined different datasets and found that artifacts are prevalent in different shapes and sizes and that they can significantly distort downstream analyses. With the goal of alleviating their impact, we proposed a framework that delineates and removes artifacts with minimal human labor by requiring only image-level annotations rather than the more burdensome pixel-level annotations[20]. As artifacts are challenging in nature and preparing the segmentation ground truth at the pixel level is laborious and time-consuming, we proposed using only image-level labels for this task. We first proposed to utilize the ScoreCAM [21], a technique designed to interpret the deep learning image classification algorithms by highlighting the parts of the images that significantly impact the particular decision. In the case of classifying an image as artifactual or not, the hypothesis states that the most influential parts of the image for making the decision are the artifacts, which has been empirically proven. Further, we used the output of ScoreCAM as pseudo-labels to train a segmentation model [20]. This way, we combine the performance advantages of using pixel-level segmentation and the convenience of image-level annotations. We called our proposed framework *ScoreCAM-U-Net*, and we foresee that the step of unwanted object removal is likely to become a standard step for all large-scale microscopy experiments.

Finally, we applied the acquired knowledge in a real-world context. We explored the value of applying deep learning methods for artifact removal and segmentation in drug discovery research. To do so, we collaborated with chemists

and biologists who study one of the most prominent cell membrane receptors-M4. Despite its expanding significance, it has proven challenging to create new medications that target such a receptor [22]. Our collaborators used high-affinity fluorescence ligands to study the binding interaction to the M4 receptor. To do so, one needs to delineate the cells and investigate the fluorescent signal within, which depends on the strength of affinity between the protein and ligand as well as the ligand concentration. However, the signal generated by the fluorescence ligands within those cells is not always enough to enable the model to detect a cell from the background reliably. Therefore, we decided to segment cells from the more challenging yet fluorescent-independent brightfield modality. We first employed deep learning to segment the cells' bodies from brightfield images. Next, we analyzed the fluorescent signal of the cells extracted from the corresponding fluorescent images using the cells' coordinates from the segmentation results. We then investigated the impact of removing artifacts from the brightfield images on the signal resulting from the receptor-ligand interaction. We showcased that removing artifacts reflected a more unbiased signal and that using our proposed model to remove them got near-optimal performance.

2. BACKGROUND

Cells are the fundamental building blocks of all living organisms. Therefore in order to understand how organisms function, we need to understand cells. This chapter introduces the essential principles of cell biology, presents microscopy as the primary tool to gather data from the cells, and gives an overview of methods to analyze and understand the images. The reader will then be equipped to delve into the remaining parts of the thesis and to place the work in the following chapters in a broader context.

2.1. Cells are the basic unit of life

A cell is a small enclosed volume filled with liquid substances and contains many other biomolecules. It is separated from the outside environment by a membrane. A cell is the central building block of all living organisms and a place where nearly all biological processes take place. The primary functions of the cell include manufacturing proteins, producing energy, transporting materials, and disposing of waste. Cells also have the ability to divide into multiple copies via mitosis or, in some cases, to generate new combinations of genetic material via meiosis. Life would not exist without this tiny yet vital component. In this thesis, we present methods for automatically analyzing cells in microscopy images with the goal of enabling the community to uncover more of their mysteries. In order to appreciate these techniques, we will first present some salient properties of the cell.

2.1.1. Structure of a mammalian cell

Cells can be broadly classified into eukaryotic, like those found in mammals, and prokaryotic such as bacteria. The former maintain their genetic information (i.e., Deoxyribonucleic acid (DNA)) inside the cell in a nucleus, while the latter has no designated place to store its genetic material. In eukaryotes, the nucleus is separated from the other structures in the cells (also known as intracellular organelles) by the nuclear envelope. Like all other compartments, nucleus is fueled by the cell's power plant - the mitochondrion. Right by the nucleus is another vital organelle, the endoplasmic reticulum. Part of the endoplasmatic reticulum is covered with spherical structures called ribosomes (they also exist in other places in the cell) in which proteins are synthesized. The newly made proteins are then sent to the Golgi complex, which packages and transmits them toward their final destination within or outside the cell. All inner compartments are embodied in the cytosol, which also plays a role in signal transduction between them, and is separated from the outside environment by the plasma membrane [23]. These and several other cell organelles are outlined in Figure 1. Importantly, cells can also communicate with their surroundings outside of the plasma membrane via signaling pathways.

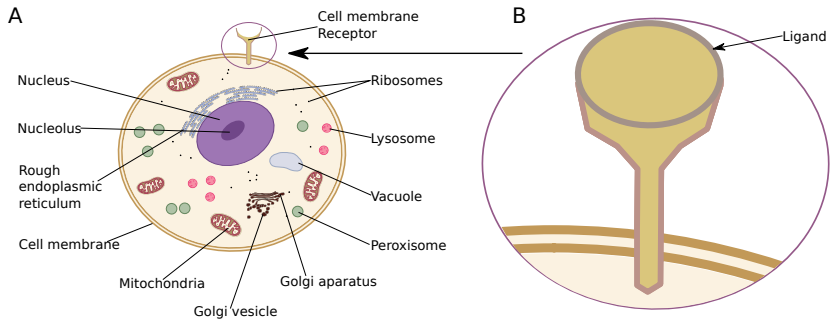

Figure 1. Graphical illustration of the prominent compartments of the eukaryotic cells (animal cell in this example). A) Cellular compartments and B) a ligand (circle-like) binding to a specific cell membrane surface receptor. Straight black lines point out each element from its corresponding name.

Cells can recognize and respond to external signals to approach nutrients, avoid danger, and communicate [24]. These signals are detected by receptors, which can be found within the cell or on its membrane (Figure 1). The most prominent family of cell membrane surface receptors are G-protein coupled receptors (GPCRs) [25]. Publication III of this thesis delves further into the ligand-binding properties of a type of these receptors.

2.1.2. Cell lines

To study cells, it is convenient to grow them outside their native environment (e.g., outside living animals) under controlled settings in plastic dishes ("cell culture"). It is relatively convenient to work with cells in a dish to be able to, e.g., look at the impact of novel drugs on cancer cell growth. Therefore, replicating cell cultures (sometimes indefinitely) called cell lines are used. Cell lines can be easy to work with, inexpensive, offer an infinite material supply, and avoid issues related to using human and animal tissue [26]. They have different properties. In what follows, we briefly introduce the cell lines that were used as a source of data in this thesis.

Human cancer cell lines are actively being established for all major cancer types and used in drug screening. The first immortal human cell line was created from the cervical carcinoma of Henrietta Lacks, a 31-year-old woman, at the beginning of 1951. This cell line is known as HeLa, the first two letters from the woman's first and last names. Cancer research has greatly relied on this line [27, 28]. Almost twenty years later in 1972, the human alveolar epithelial cell line (i.e., A549) was extracted from a pulmonary adenocarcinoma and identified as a representative of the Alveolar Type II pneumocytes of the human lung, making it a centerpiece of respiratory research [29]. In the same year, a cell line of human fibrosarcoma, HT1080, was isolated from a tumor biopsy sample taken from

a 35-year-old human male. The cells are a suitable model for cell invasion and migration research. Notably, the source patient had not received chemotherapy or radiation, which would have likely resulted in the undesirable mutations that these therapies frequently produce [30]. A year later, in 1973, Michigan Cancer Foundation-**7** (MCF-7) cell line was established. It plays a vital role in diagnostic, and drug discovery research for breast cancer, the most frequently diagnosed and the main cause of cancer death in women [31]. In 1975, Another human cancer cell line, liver hepatocellular carcinoma (i.e., HepG2), was desegregated from the liver tissue of a 15-year-old male with hepatocellular carcinoma, which is still one of the widely-spread cancer types. These cells are important in studying human liver disorders [32]. The highly androgen-sensitive Lymph Node Carcinoma of the Prostate (LNCaP) cell line was isolated from the lymph node of a 50-year-old man with metastatic prostate cancer in 1977 [33]. Altogether, these classical cell lines have been growing in cell culture for nearly half a century.

Mouse and dog cell lines have been around for a long time and are used in drug screening as well. Madin-Darby Canine Kidney (MDCK) are epithelial cells from the kidney tissue of adult dogs and were initially utilized in virology. These cells are permissive and promote the proliferation of the influenza virus, making them appropriate for virology research and vaccine manufacture [34]. National Institutes of Health **3**-day Transfer, inoculum $\mathbf{3} \times 10^5$ cells (NIH3T3) are fibroblast cells that were derived from an NIH Swiss mouse embryo and have proven to be particularly effective in DNA transfection investigations [35, 36]. Chinese Hamster Ovary (CHO) is the most frequently used mammalian cell line. Those cells demonstrated the capacity to grow in suspension culture, enabling massive production. CHO is utilized to produce the vast majority of recombinant proteins currently on the market [27]. Cells within those cell lines, and in general, are tiny in size. Therefore, a tool that allows the magnification of these small structures should be utilized in order to study them.

2.2. Microscopy

Cells are small in size, with a typical mammalian cell having a diameter of $10 - 20\mu m$, which is around one-fifth the size of the smallest object visible to the unassisted eye. Due to their tiny size, studying cells and their internal structures requires magnification, which is achieved by using a microscope, a tool commonly employed by researchers who study cells to extend their natural abilities.

2.2.1. Fundamentals of microscopy

A microscope (micro- = "tiny"; -scope = "to gaze at") is a device that allows the observation of small objects by magnifying them. The primary component of a microscope is a set of lenses that together facilitate the magnification of the item being studied. The light signal coming from the object being viewed (either refracted or emitted by the object) is collected by the objective lens and transmitted

through the tube lens and eyepiece lens before being projected onto the viewer's eye or being recorded by a camera mounted right after the eyepiece lens. The quality of the recorded images is a key element for successfully analyzing the viewed objects.

The major determinants of imaging quality are the magnification and resolution of the field of view. The field of view itself is the scene that the eye or the camera perceives (see Figure 2). This scene is usually a magnified version of the actual one. However, magnification alone will not enable us to clearly distinguish between nearby objects or between objects and the background. Magnification requires both good resolution and contrast in order to be effective [37]. When objects to be imaged are too close to each other, magnification alone is not enough to tell that they are separate. The resolution of a microscope is defined as the minimum distance between two objects at which they can still be recognized as separate entities; the smaller this distance, the higher the instrument's resolution. The ability to capture objects is also limited by both the uncertainty of the measured signal (i.e., noise) and the background signal. The relative difference in light intensity between the darker background and the created image of objects is referred to as contrast; the higher the contrast, the better the microscope [37].

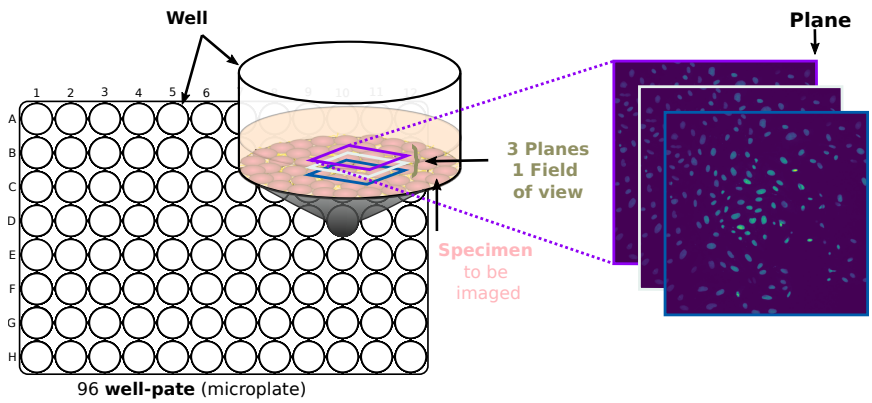

Figure 2. Graphical representation of typical high throughput microscopy imaging. Example of 96 well-plate structure with a zoomed view (cylinder) of one of its wells. The spherical red objects inside the well depict specimens to be imaged. One field of view was captured with three different focal planes (three colored rectangles), and we can see the corresponding generated pictures on the right-most side of the figure.

2.2.2. Light microscopy

Microscopy images have advanced drug discovery research, genetic perturbations understanding, and cell phenotyping [39, 40, 38]. For example, microscopy can be used to monitor the behavior of cancer cells in response to different drug candidates' treatments, allowing for identifying the most promising compounds for

further study [18]. Among the different available types of microscopes, the one that has been used to produce the images analyzed in this thesis is known as the light microscope. The light microscope is an instrument that enlarges microscopic objects by making use of light and its characteristics. Light microscopy techniques can be divided into two main categories, brightfield, and fluorescence. In brightfield, an image of a sample (i.e., biological specimen to be viewed) is obtained by capturing the light passing through the sample. This technique utilizes the impact of an object on the light traveling through it when such an object absorbs, disperses, or deflects the light. In this configuration, the objective lens and light source are placed on opposite sides of the sample [41]. Although brightfield microscopy allows us to observe cells that would not otherwise be visible, it only provides general outlines of them, leaving their interiors poorly visible, if not invisible. An example of brightfield images can be seen in Figure 3- B, and this type of image is used in chapters 3, 4, and 5.

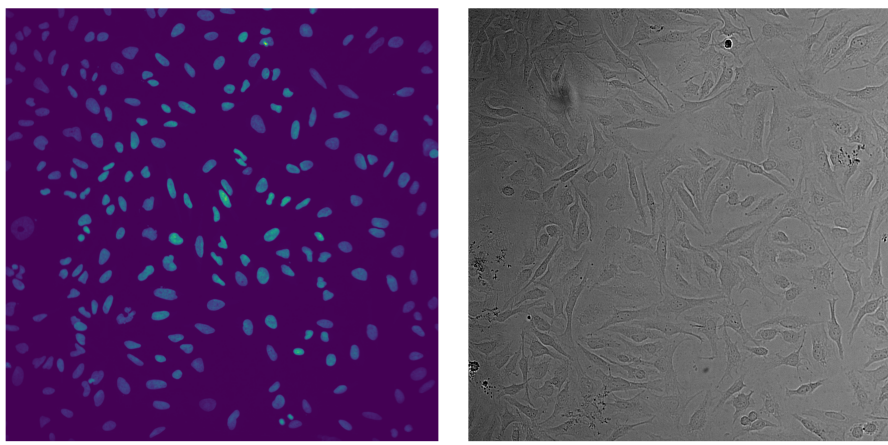

Figure 3. Examples of a fluorescent image showing stained cell nuclei (A) and its corresponding brightfield counterpart (B) with no visible stain. The images are cervical cancer cells (HeLa), captured using the Opera high-content screening system (Revvity) [2].

As an alternative, fluorescence microscopy is a powerful technique that can capture more details with high resolution and contrast. This type of microscopy utilizes fluorophore molecules that absorb photons at one wavelength and emit them at another (i.e., fluoresce). Since most molecules inside the cell do not fluoresce, experimenters apply fluorescently labeled tags or antibodies that bind to the molecules of interest. The fluorescence-bound molecules are then excited by light of their excitation wavelength, and they emit light of a different wavelength which is filtered, gathered, and used to form a picture of such molecules [41]. Different fluorophores emit at different wavelengths, which results in different pictures. Hence, using multiple fluorophores on the same sample allows for gathering multiple pictures (also known as channels). Staining the different sub-

cellular structures in the sample allows a collection of images in more than one color, giving better resolution and contrast. An example of fluorescent images of cell nuclei can be seen in Figure 3- A. In this thesis, fluorescent images are used in chapters 3 and 5.

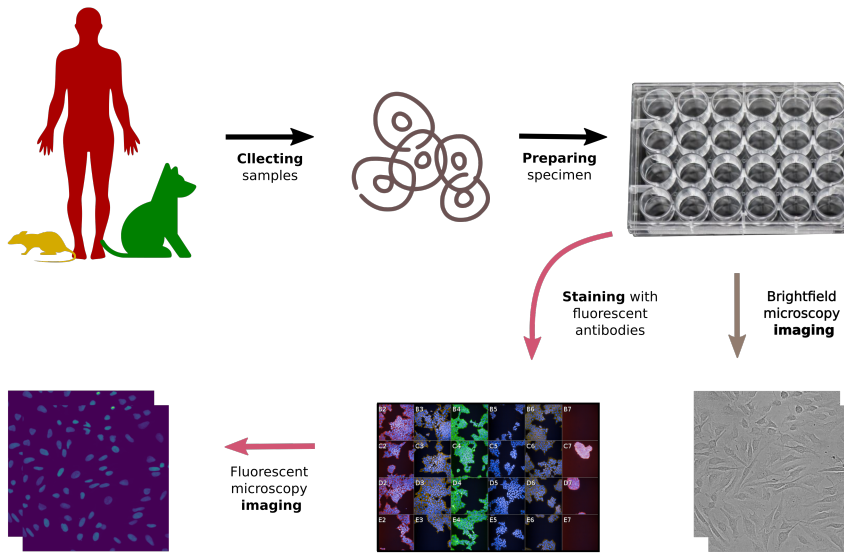

Figure 4. Fluorescent and brightfield microscopy image acquisition pipeline. First, the sample is collected (top left) and prepared by putting it into an appropriate solution (top right). By passing natural light through the sample using a microscope, one can already acquire brightfield images (branch indicated by the brown arrows; bottom right). To acquire fluorescent images, cells in the well-plate need to be stained with fluorescent antibodies which attach to specific compartments, and using a microscope, we can image those compartments (branch indicated by the pink arrows; bottom left).

.

Fluorescent microscopy is further distinguished into widefield and confocal microscopy. In the widefield, the illuminating light excites fluorophores in the entire focal volume of the sample, which causes blurriness from areas out of focus (i.e., above and below the focus plane). On the other hand, a confocal microscope scans a sample with a focused beam of light at a single in-focus point and collects those points to combine an entire image, creating a less blurred image [41].

To image cells using microscopy, the specimen (biological sample to be imaged) is seeded into a well-plate (also known as a microplate), which typically comprises 48 or 96 tiny compartments called wells filled with an appropriate solution (Figure 2). Samples in different wells might be treated identically or distinctively. Images are then collected in multiple channels (as explained above). Typically, the plane of focus is also moved through the z-direction (depth) to picture the sample in the third dimension. Figure 4 shows a generic visual represen-

tation of how pictures are acquired in light microscopes.

Microscopy imaging enables a range of studies. For instance, it enables compound screening which involves testing numerous compounds, each applied to a distinct cell group housed in a separate well, to ascertain their effects on specific cellular responses, such as cell death. We can then acquire the images and analyze them. Image analysis aids in identifying compounds that induce the desired effect - in this case, a reduction in cell count - narrowing down the compounds' options for further investigation. In further studies, these shortlisted compounds may be scrutinized for specific modes of action, such as DNA damage induction. In this scenario, again, each well accommodates a group of cells treated with a different compound [42]. Another example that microscopy imaging can facilitate is target identification studies; for example, those focusing on rectifying genetic anomalies. This involves testing the same compound on cells with different single-gene mutations to assess its ability to correct the gene damage. In this setup, each well hosts cells with a different mutation. The cell painting method (in which multiple cell compartments are stained and the images are acquired) might be used to generate images characterizing each mutation. If the treated cells' image features (fluorescence intensity, morphological characteristics, etc.) match those of non-mutated cells, it suggests that the corresponding compound potentially targets and rectifies the mutated gene [42].

High throughput microscopy. Microscopy has recently been revolutionized with the introduction of the high-content, high-throughput microscope, which enables fully-automated imaging pipelines. High-throughput microscopy represents a sophisticated technological methodology that allows for the rapid imaging and analysis of a multitude of biological samples. This automated technique is harnessed in various areas. For instance, high-throughput microscopy is utilized in drug discovery by facilitating the screening of myriad potential drug compounds and their varying concentrations upon cellular structures. For instance, each well will contain a different compound-concentration set. This, in turn, allows researchers to evaluate changes in cell such as morphology, proliferation, and death as a reaction to these compounds, thereby providing valuable insights into the viability and potential efficacy of new drugs target [43]. Another example in which high-throughput microscopy can be utilized is locating the cellular compartments housing fluorescently-tagged proteins. Proteins play essential roles in numerous cellular processes, and having a thorough understanding of their whereabouts within cellular microenvironments is vital for comprehending their functions and interactions [38].

High-throughput screening experiments are often performed on a large scale. For example, a range of drug compounds and concentrations in a drug discovery scenario necessitates a corresponding multitude of samples. Moreover, in order to ensure statistically significant results, it is imperative to replicate each experimental condition numerous times. This guards against the possibility that observed effects are merely coincidental rather than valid outcomes of the tested

condition. As a result, thousands or even millions of images can now be generated from a single experiment, dramatically increasing the amount of available microscopy data [1, 2]. For example, an experiment involving one plate [96 wells or 384 wells] with 5-20 fields of view from each well and 3-10 z-planes would generate between 1,500 and 76,800 images. The usual high-content screening experiments utilize tens to hundreds of plates. Such a massive number of images is prohibitively expensive to inspect or process manually. In this thesis, we focus on analyzing existing techniques for processing microscopy images and propose further advances in several key areas.

2.2.3. Microscopy image analysis

Cell microscopy has been actively involved in numerous health-related applications (e.g., drug discovery and disease diagnostics) [44]. For example, in experiments that aim to discover new drug targets, scientists typically evaluate thousands of chemicals looking for a potential drug candidate. They prepare different sets of cells, add different compounds to them, and take images using microscopes to observe which group responds in a particular way that indicates a favorable outcome. For example, we might be interested in looking for a specific chemical impact on breast cancer cell growth. Scientists then need to inspect the images to scrutinize the cells' reactions, which is prohibitively expensive to do manually across vast numbers of produced microscopy images. Therefore, automated and accurate image analysis techniques are essential.

The first step in analyzing cell microscopy images is often detecting nuclei [9]. For instance, outlining nuclei and then evaluating the pleomorphism score (the extent of abnormalities in the morphology of tumor cells' nuclei) is critical for diagnosing and grading breast cancer [45, 46]. Another important measure when analyzing microscopy images is studying fluorescence ligands' binding properties and cell receptors, which is vital when evaluating drug candidates. For example, concentration-response experiments can be conducted to help ascertain drug potency (the amount of drug needed to elicit a certain response) and efficacy (the maximum response achievable from a drug). These experiments involve varying the concentration or dose of a substance (e.g., fluorescence ligands) and measuring the effect on a biological system. More generally, automatic analyses of cell microscopy images are often disturbed by unwanted objects present in those photos (i.e., imaging artifacts) [47, 48], as seen in Figure 5. Therefore, detecting and factoring out the impact of such problematic areas would allow more accurate measurements, making artifact identification another vital albeit overlooked research area in cell biology. Later in this thesis, we present our contributions toward each of the areas above.

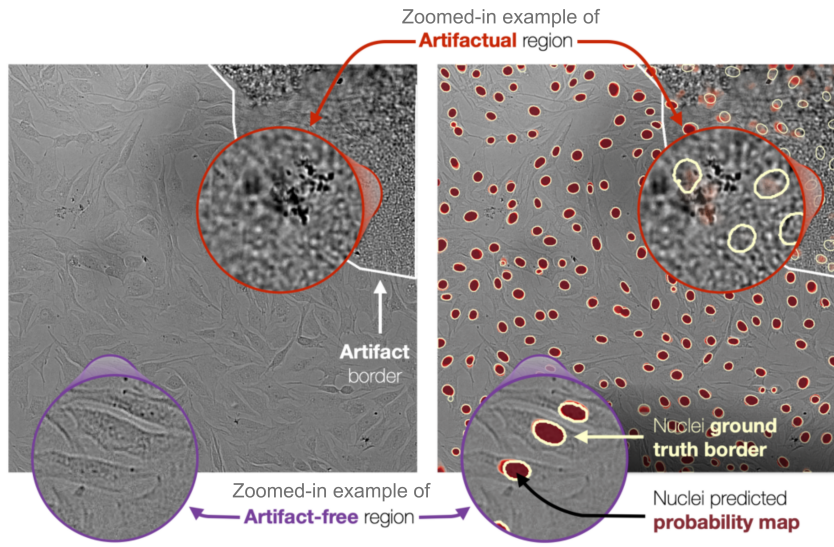

Figure 5. Examples of artifacts in microscopy images and their impact on nuclei segmentation. A brightfield image is shown together with automatically detected nuclei on top in dark red. Zoomed-in purple and light red circles show illustrations of artifact-free and artifactual areas, respectively. Artifacts' ground truths are surrounded by white borders, while yellow-ish borders encircle the ground truth of nuclei. Arrows and text are guides to corresponding regions and elements.

Computer vision algorithms are commonly employed to automate the analysis of microscopy images. These algorithms encompass various methods with varying objectives. Some of those methods can identify the presence of a specific object in an image, such as whether cell line cells exist in an image. These methods are known as classification techniques, as they categorize images into various classes. Another set of algorithms assigns a class to each pixel in an image based on the object those pixels belong to (e.g., cell and background), which will eventually identify and delineate image elements and is called segmentation. Segmentation is further categorized into three types. Semantic segmentation (which we used in this thesis), in which each pixel in an image is assigned a label based on predefined categories (e.g., cell and background); instance segmentation, which distinguishes instances in each category (e.g., pixels in each cell will be assigned a different label); and panoptic segmentation, which combines both of the previous types [50, 49]. Semantic segmentation (Publications I, II, and III) and classification (Publication II) are used in this work.

Classical approaches to analyze microscopy images. Many approaches have been proposed in the past to assist with automating the analysis of microscopy images. For example, Otsu's thresholding [51], followed by water-shedding [52, 53], is a popular method for nucleus segmentation because of its simplicity and computational efficiency. Despite widespread adoption, such techniques relied on

strong assumptions about the data that did not hold in most of the practical use cases, yielding considerable amounts of errors [54]. For instance, Otsu's thresholding makes an assumption that the intensity of pixels in an image has a bimodal distribution. This means that the intensity values are concentrated around two distinct peaks in a histogram of pixel intensities. While this assumption may hold in certain cases for fluorescent images, it is far from being true in brightfield microscopy images. Typically, brightfield images display uneven intensity levels and poor contrast [55]. As a result, the use of Otsu's thresholding in such cases can result in substantial segmentation errors, leading to inaccurate quantitative measurements. Alternatively, machine learning-based methods have been introduced to derive more accurate quantitative measurements from microscopy images [7].

Machine Learning (ML) is a subfield of artificial intelligence that deals with designing algorithms that learn connections between input data (e.g., an image of cell nuclei) and the desired outcome (e.g., corresponding mask to the input image where nuclei pixels are given a value of one while others are given zero) without being explicitly programmed to do so. Machine-learning-based techniques typically involve preparing the required targets (also known as annotations or labels) and then training an algorithm to predict those targets. By doing this, the algorithm learns how to produce targets for unseen yet alike data. For image classification, the labels are typically a class name or id, while segmentation labels take the form of masks in which the pixels of each element in an image are given one value. For example, in the problem of semantic segmentation of nuclei, we may assign ones for pixels belonging to nuclei and zeros for other (e.g., background) pixels. Examples of classification and semantic segmentation labels are in Figure 6.

Numerous machine learning methods are made easily accessible in biological imaging software tools [4]. The most widely used packages and open-source tools for cell microscopy analysis include ImageJ/Fiji [5], CellProfiler [6], and Ilastik [7]. ImageJ, a Java-based image processing tool, empowers users to examine, quantify, and visualize scientific image data. Its extensibility via plugins and macros, wide community backing, and ability to handle several image formats make it particularly noteworthy. The specific algorithms used in ImageJ can vary depending on the task, as it has many built-in functions for image manipulation and transformation, including filters, edge detection, and more. In a similar vein, Ilastik distinguishes itself with its interactive image analysis feature. It employs user-supplied annotations, allowing users to sketch labels or portions of labels and instantly observe the results while training models. With Random Forest [56] as a default algorithm, Ilastik provides access to a range of classifiers available in the sci-kit-learn library [57]. Lastly, CellProfiler, which has been outperformed by U-Net [10], is an open-source software designed for high-throughput cell image analysis. It uses various image processing techniques and algorithms like filtering and intensity thresholding to identify individual cells in microscopy images and measure their properties.

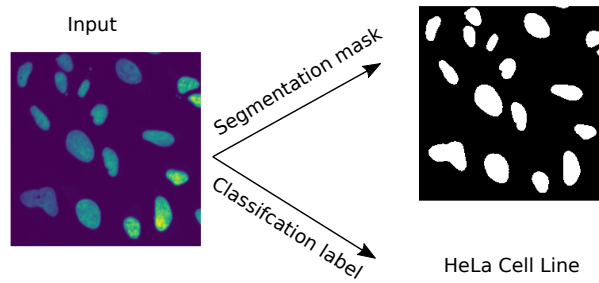

Figure 6. Examples of image classification and object segmentation labels. Left: a human cell nuclei input image to the ML algorithm. Right: ground truth annotation of the corresponding image for different tasks (arrows point out the annotations of each task); nuclei semantic segmentation (top), and classification of which cell line this image was derived from (bottom).

Traditional machine learning algorithms require careful design and expert knowledge to transform raw image pixel values into a representation that allows the learning subsystem to recognize patterns in the input data [8]. Moreover, high-throughput microscopy imaging experiments introduced variations in signal quality and microscopy modalities, which pose challenges to the robustness and reliability of those methods at a large scale [9, 10]. Therefore, more advanced microscopy image analysis methods had to be developed to handle all the increasing complexity and variability of data. Over the last decade, deep learning, a novel form of ML, has helped to solve many computer vision problems, including image classification [11], object detection [14], and segmentation [15]. The central theme of this thesis is investigating deep learning for analyzing microscopy images. The following section is devoted to outlining the fundamentals of deep learning, with a focus on the use of this technology in computer vision.

2.3. Deep learning

Deep Learning (DL) is a family of machine learning techniques that use computational models with several processing layers to automatically discover the representation (i.e., features) needed to solve a given problem from the raw data [8]. The learned representations are obtained by transforming the input data through a sequence of stacked modules (i.e., layers). Each layer builds on the previous one to create a more abstract representation starting with the raw input. For instance, the first layer would identify edges; the second layer would pinpoint particular arrangements of those edges; and the third layer may combine earlier patterns into more complex combinations and so on until the patterns start to resemble the whole objects [8]. Each layer is composed of several (typically hundreds or thousands) units called artificial neurons. A neuron output o from inputs $\mathbf{x}$ is typically given by:

$$o = f(\sum_j (w_j \times x_j) + b) \tag{2.1}$$

where f is an activation function, x_j is an input to the neuron, w_j is the weight corresponding to the j^{th} input, and b is the bias that shifts the activation function curve to the left or right. The activation function is a non-linear operation that transforms the output of the neuron. Outputs from neurons of the layer are then used as inputs for neurons of the next layer, and so on. There are several activation functions; the most common ones are sigmoid and Rectified Linear Units (ReLU), which, respectively, are defined as:

$$\begin{aligned} sig(z) &= \frac{1}{1+e^{-z}} \\ r(z) &= \max(0,z) \end{aligned} \tag{2.2}$$

where z is the total weighted output of $\sum_j (w_j \times x_j) + b$.

The first layer that takes the raw input is called the input layer. This raw input can take various forms depending on the specific task or application. For instance, in an image classification, the input layer may receive pixel values from an image. The final layer in the network is called the output layer, and all levels between input and output are called hidden layers. Such a collection of layers is called an artificial neural network or deep neural network. By stacking enough of these simple transformations (i.e., layers), which are mostly non-linear, arbitrarily complex functions can be represented [8]. The patterns (also known as features) that characterize the input are learned without needing to be handcrafted (i.e., manually designed), leading to potentially more generic and robust algorithms which are less susceptible to failure in the presence of unexpected inputs. The limitation of handcrafted features is due to assumptions of the feature designers which may not consider all the relevant aspects for the given task. In contrast, deep learning models learn the relevant features directly from the data without the need for manual feature engineering. Nonetheless, this is not without its caveats. It is important to note that the effectiveness of the neural network in learning patterns is contingent upon the quality and quantity of the training data, its diversity, and how well it represents the distribution that will be encountered at the time of using those networks.

2.3.1. Model training strategies

In DL and ML in general, supervised learning is a frequently used type of learning in which the system is explicitly exposed to the desired targets. For example, we may want to design an algorithm that can differentiate between images of yeast cells containing fluorescently-tagged plasma membranes from images with tagged nuclei. In supervised learning, we first collect yeast images in which proteins are expressed in the plasma membrane and in nuclei. The information about the location of expressed protein is recorded as a binary variable of either 1 (if expressed

in the membrane) or 0 (if expressed in the nuclei). This collection of samples is usually split into three sets (e.g.,70% for the first, 15% for the second, and 15% for the third set) and referred to as training data, validation data, and test data.

During the training phase, the algorithm is exposed to images in the training set and trained to predict their corresponding classes. For each input image, the algorithm produces probability values that indicate the likelihood of the image belonging to each class. If this process succeeds, the network outputs the highest score for the required class of each image in the vast majority of the training samples, which is unlikely to occur from the beginning of the training process. Therefore, the training process involves iteratively adjusting the parameters of the algorithm (i.e., weights and biases) to minimize the difference between the predicted probabilities and the true labels of the images (i.e., minimizing the loss function). The training process is typically repeated for multiple epochs, with all the samples in the training set being used in each epoch. Following each iteration of updating the parameters, the algorithm's performance is evaluated on a separate set of samples called the validation set, which is used to guide the adjustment of the algorithm's parameters to improve its generalizability to new data. To improve the performance, the algorithm optimizes the loss function. The most commonly used loss functions are binary cross-entropy, categorical cross-entropy, and mean squared error [58]. In this thesis (Publications I, II, and III), we used Binary Cross Entropy (BCE). Given x_j be an instance of the input samples, the true label for it be t_j, and the model used for prediction be $f(.)$. The BCE loss can be calculated as:

$$BCE(t_j, x_j) = -(t_j \cdot \log(p_j) + (1 - t_j) \cdot \log(1 - p_j)). \tag{2.3}$$

Where $p_j = f(x_j)$ is the prediction of the instance x_j by the model $f(.)$. The value of t_j in equation (2.3) can be either 1 (positive) or 0 (negative). The goal of optimizing the loss function during training is to minimize the discrepancy between the predicted and actual values by modifying the model parameters (i.e., weights and biases). This is achieved in deep neural networks through backpropagation [59] by utilizing the gradient descent algorithm [60]. Backpropagation is used to calculate the gradients of the loss function with respect to the weights and biases, while gradient descent is used to update the weights and biases in the direction that minimizes the loss as follows:

For each weight w in the weight matrix W and each bias b in the bias vector B, the update is performed as:

$$\begin{aligned} w_{k+1} &= w_k - \alpha \times \frac{\partial BCE(x_j, t_j)}{\partial w} \\ b_{k+1} &= b_k - \alpha \times \frac{\partial BCE(x_j, t_j)}{\partial b} \end{aligned} \tag{2.4}$$

Where k denotes the current iteration, α is the learning rate that scales the size of each update, and $\frac{\partial BCE(x_j,t_j)}{\partial w}$ and $\frac{\partial BCE(x_j,t_j)}{\partial b}$ indicate the partial derivative of the

loss function with respect to the given weight w and bias b, respectively, computed for input x_j and true label t_j. In practice, this update is often conducted for multiple input instances (i.e., mini-batch) at a time. The gradients for all instances in the mini-batch are then averaged and used for the update.

It is not uncommon, nevertheless, to have very few, noisy, or even no ground-truth annotations in the training data as it is sometimes labor-intensive, expensive, or even unfeasible to annotate enough data, particularly for object segmentation [61, 62]. In these cases, other types of learning, like weakly supervised or self-supervised learning [63], may be used instead of traditional supervised learning to compensate for incomplete or noisy labels. Different types of machine learning with respect to the completeness of labeled data are presented in Figure 7. Publication II in this thesis utilizes the weakly-supervised learning paradigm; hence, we briefly discuss it in what follows.

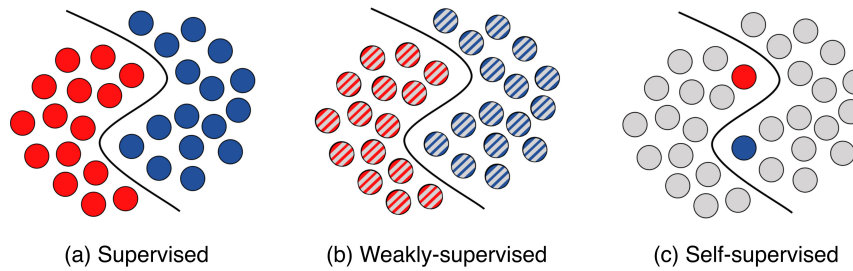

Figure 7. Types of machine learning strategies with respect to the completeness of labeled data. Red and blue circles: labeled data points for two different classes; grey circles: unlabeled data points; black lines: actual decision boundary; striped circles: data points with inaccurate or coarse-grain labels (e.g., image-level label for the task of segmentation). The figure is adapted from the paper in [62]

.

Weakly-supervised learning. Weakly-supervised learning refers to a machine learning paradigm where models are trained using incomplete or imprecise labels. There are two main categories of this paradigm: inaccurate and inexact supervision [64, 65]. Inaccurate supervision occurs when the training labels contain errors or noise (e.g., partial or noisy masks for an object segmentation task). One way to deal with this type of weak supervision is by utilizing more noise-robust neural networks or attempting to rectify possibly erroneous labels [64, 66]. On some occasions, however, we may not even have noisy training labels and may only possess information that cannot be directly used as labels for the given task; for instance, in an object detection task, if we only know the objects that are present in an image without their corresponding locations. In these cases, inexact supervision algorithms can be employed. Such algorithms attempt to deduce the missing information from the provided data. For example, in our framework, Score-CAM-U-Net, incorporated in publication II of this thesis, we utilize an in-

exact weakly supervised method called Score-CAM [67] that leverages lower-level features of a CNN image classification model to estimate the likelihood of an object being present in a particular image window. The visual illustration of Score-CAM is in Figure 8.

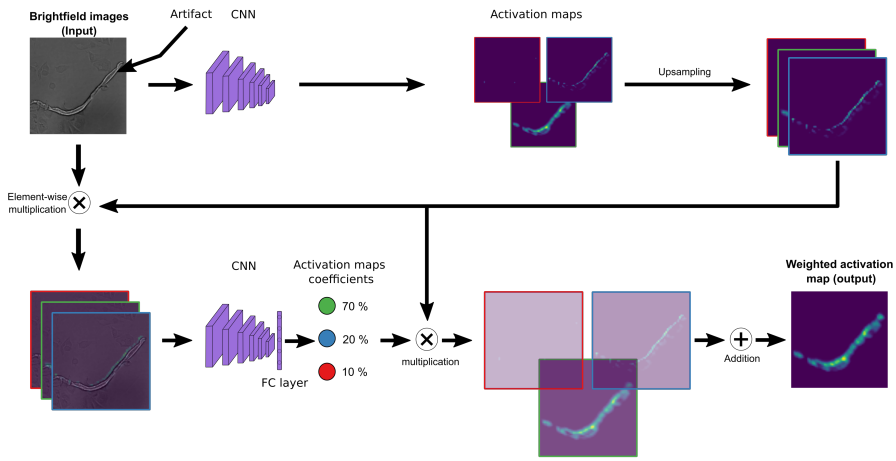

Figure 8. Graphical representation of Score-CAM[21]; an example of CAM-based methods which we used for weekly supervised segmentation. A brightfield image containing an artifact (top-left) was used as an input to a CNN classification model (stack of 3d purple rectangles). Images with colored borders represent different feature maps. Colored circles: coefficients correspond to activation maps of the same colors. Straight black arrows indicate the flow of processing.

In weakly-supervision and all other learning paradigms mentioned previously, the most popular type of artificial neural network are convolutional neural networks [68], which we also use in this thesis. We present a more thorough overview of such a type of deep learning models in what follows.

2.3.2. Convolutional neural networks

The CNN is a family of deep learning models that specializes in processing grid-structured data such as images. Their central building units are the convolutional layer, the pooling layer, and the non-linear activation function, which are usually grouped together to form a convolutional block (as shown in Figure 9). Typically, CNN comprises several blocks, with each block (except the output one) incorporating one or more convolutional layers, pooling, and an activation function (see Figure 9A). The output of each convolutional layer is connected to its input through a set of learnable weights called filters (also known as kernels) (Figure 9C). Considering the input to the convolutional layer is an image of one channel, the filter is a small matrix, typically with a size much smaller than the input image. It is overlayed with the input image at different possible regions in a sliding window fashion, and its values are multiplied element-wise with the image values.

These element-wise multiplications are then summed up to produce a single output value. The filter is shifted by a pre-defined number of pixels called the stride, and the process is repeated to generate the next output value, and so on, until it covers the entire input image. The resulting output is passed through an activation function such as ReLU [69] (Figure 9B), which produces a feature map (see Figure 9C). The feature map is then processed by a pooling layer, e.g., MaxPooling, that retains the highest values in the pooling window (this window moves in a sliding window fashion across the entire feature map), [8], effectively reducing the dimensionality of such feature maps. An example of MaxPooling is shown in Figure 9C.

Stacking multiple convolutional blocks is an effective method for detecting visual features in input images by training the weights in the filters. Notably, the weights of the convolutional filters are shared across the entire image, which is a key feature that makes CNN efficient for image processing tasks. This means that the same filter is used to extract features from every part of the image rather than using a separate filter for each location. In order to use shared weights to detect features from different parts of the image, CNNs leverage two properties: First, pixels located nearby tend to contain related or contextually linked information. Second, a pattern that appears in one image location might be present elsewhere [8]. The sharing of weights reduces the number of parameters that need to be learned, which makes the network easier to train and less prone to overfit the training data.

Overfitting. Overfitting is a common challenge in machine learning where a model becomes over-adapted to the training data, leading to poor performance when exposed to new or unseen data [70]. Overfitting presents a pervasive challenge for deep learning practitioners when training their models, as we have encountered in this thesis. Hence, we investigate the underlying causes and potential solutions we used in the thesis to tackle such an issue. Although it is challenging to comprehensively characterize the causes of overfitting, we nonetheless highlight the most dominant ones: a substantial shift between the distribution of the training and the unseen data, and neural network complexity (i.e., over parametrization) [71, 72]. Accordingly, to resolve the overfitting problem, solutions can target either the neural network design or the data. Here, we focused on the most popular methods, which are also used later in this thesis: data augmentation [72], transfer learning [73], label smoothing [74], and dropout [71].

Data augmentation. Data augmentation refers to a set of techniques that artificially expands the size of a training dataset by applying a range of transformations to existing data samples. Such methods aim to produce a more generalized deep learning model, which is better equipped to handle new and unseen data. Depending on the given application, various data augmentation strategies can be used. These may include geometric transformations (e.g., flipping, rotating, or translating the image), adjustments to the color space, mixing images, and random erasing [72]. In addition to using those strategies to increase the number of train-

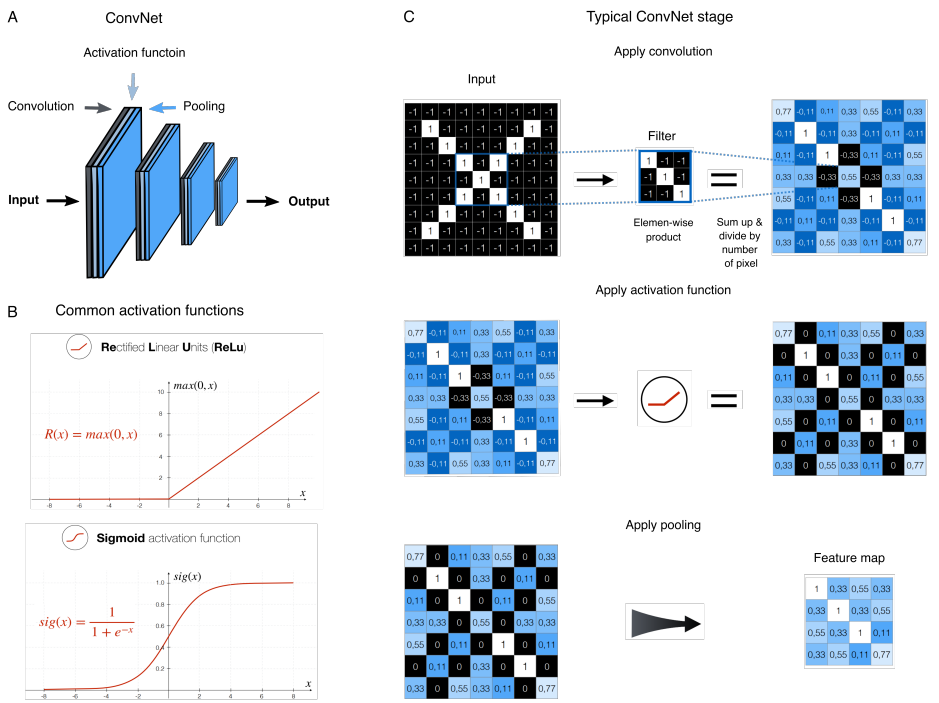

Figure 9. Typical CNN architecture and components. A) CNN are composed of a series of convolutional blocks (stages), with each block often containing a convolutional layer, an activation function, and a pooing layer. B) The most common activation functions employed in CNN, i.e., ReLU and Sigmoid. C) Components of a typical convolutional block and how they work. First, a convolution operation (top), then a ReLU activation function (middle), and finally, a MaxPooling operation with the pooling filter size of (2×2) and step size (i.e., stride) of 2 pixels at a time.

ing samples from the current task-specific data, it is also possible to leverage data from a different task through transfer learning.

Transfer learning. Transfer learning (also known as pre-training as in Publication I) is a machine learning technique that involves using a model trained on one task (i.e., source domain) as a starting point for a model on another related task (i.e., target domain). Transfer learning is grounded on the assumption that the low-level image features, such as shapes and colors, can be learned from examples in various contexts, which can be valuable in training a model for a novel task [72, 73, 75].

Label smoothing. Besides utilizing data augmentation and transfer learning to amplify signals in deep learning models, label smoothing is another way to mitigate overfitting. This involves setting the prediction targets to slightly different values than the usual ones. For example, in semantic segmentation, the targets for the pixels might be set to 0.1 and 0.9 rather than the traditional 0 and 1, respec-

tively [74]. By introducing a small amount of uncertainty into the training labels, the model is encouraged to consider other plausible classes besides the true one and thus avoid overfitting to the training set.

Dropout. In addition to the above techniques, which tackle overfitting from the data perspective, it is possible to address the problem by modifying the model itself. One such technique is the dropout, which involves randomly removing neurons from the model during each training epoch. This helps to prevent the neurons from becoming too dependent on each other, reducing the risk of a model (e.g., CNN model) overfitting the training data [71].

2.4. Evaluation metrics

Once the model is trained, one would need to decide how to effectively measure its performance. We evaluated the performance of all models that we have in our work using pixel-level metrics (used in Publications I, II, and III), object-level metrics (used in Publications I and II), and metrics that measure the impact of our models on the downstream analytics (used in Publications II and III). To report semantic segmentation at the pixel level, we used accuracy, precision, and recall, as well as the F1 score, which combines precision and recall. Accuracy (i.e., the proportion of correct predictions to the total number of predictions) is a straightforward measure but may not be suitable when the classes of pixels are unevenly distributed. For example, in nuclei segmentation, where we differentiate between the nuclei and the background classes, the majority of pixels typically belong to the background. Instead, precision and recall can help evaluate the model's performance in such cases. Precision measures the proportion of samples identified as positive that are actually positive, while recall measures the proportion of positive samples that were correctly identified as positive. These metrics are given by:

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F1-score &= \frac{2\cdot TP}{2\cdot TP+FP+FN} \end{aligned} \tag{2.5}$$

Where TP refers to true positive, which occurs when the algorithm correctly identifies a pixel as belonging to the target category (e.g., nuclei in nuclei segmentation) and the ground truth also labels the pixel as belonging to that category; FP refers to false positive, which occurs when the algorithm incorrectly identifies a pixel as belonging to the target category, but the ground truth labels the pixel as

belonging to a different category (e.g., background); and FN indicates false negative, which happens when the algorithm fails to identify a pixel as belonging to the target category, but the ground truth labels the pixel as belonging to that category.

While the metrics above are commonly used for evaluating pixel-level performance in machine learning [10], they do not quantify errors at the object level, such as merged or missed objects. To address this, we use metrics that rely on area coverage to identify correctly segmented objects. We first calculate the intersection-over-union (IoU) metric for pairs of segmented and ground truth objects as the ratio of the number of pixels in the intersection of the two objects to the number of pixels in their union. We then consider a ground truth object identified (i.e., true positive) at a given IoU threshold if there is a segmented object with an IoU value to it over this threshold. Figure 10-left showcases different examples where for both cases of IoU between the prediction and the ground truth being larger and smaller than a given threshold. To better quantify the quality of segmentation, we report the object-level F1 score across a range of IoU thresholds starting at $t = 0.5$ to $t = 0.95$ with a step $\Delta = 0.05$, as well as averaged over these thresholds, as established earlier [9]. This metric provides a comprehensive evaluation of segmentation performance that accounts for errors both at the object level and the pixel level by combining information on the proportion of correctly identified objects and the pixel-wise accuracy of their estimated masks. Moreover, an IoU threshold greater than 0.5 ensures that there is only one prediction match for each ground truth object and vice versa, ensuring a minimum level of segmentation quality.

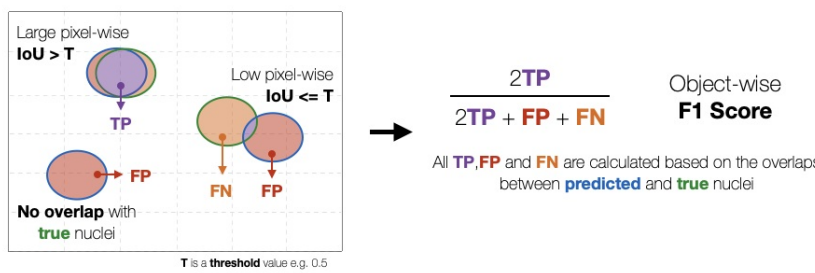

Figure 10. Visual explanation of the object-wise F1 score metric. TP: true positive; FP: false positive; and FN: false negative.

3. ACCURATELY SEGMENTING NUCLEI IN BRIGHTFIELD CELL MICROSCOPY IMAGES USING CONVOLUTIONAL NEURAL NETWORKS (PUBLICATION I)

Finding nuclei in cell microscopy images is often the first step in many analysis workflows; therefore, automating it would be of considerable practical interest [9]. However, such automation is generally hard due to the vast data volumes and complicated imaging modalities in high-throughput, high-content microscopy [76]. Moreover, since producing the challenging brightfield images is relatively fast and non-invasive, it would be even more rewarding, yet more difficult, to segment nuclei directly from them [10]. This chapter investigates the problem of nuclei segmentation in microscopy images, directing more attention to the challenging brightfield modality.

Deep learning methods, particularly CNN, have dramatically enhanced computer vision tasks across various areas, and microscopy images were no exception [77, 78, 79]. One of the earliest methods that employed convolutional neural networks to segment cell nuclei from fluorescence images was Deepcell [80]. Since then, several more have been proposed [81, 82, 83, 84, 85, 86, 87, 88]. The popular CNN model- U-Net [89], later outperformed by U-Net++ [90], has also been introduced as a plugin for bio-image analysis software ImageJ [91], making it accessible to a broader audience. Though attempts have been made to segment nuclei from brightfield images [10, 92], there is still a performance gap when compared to doing so from fluorescence ones. Meanwhile, advances in deep learning have continuously provided new ideas that could, in principle, improve nuclear segmentation performance yet have not been utilized for doing so [15, 90, 93, 94].

In this chapter, we address the problem of nuclear segmentation in fluorescence and brightfield images making use of the latest developments in deep learning. We evaluate Deeplabv3+ [95], Tiramisu [93], U-Net++ [90], a modified version of U-Net [10], as well as a new proposed PPU-Net architectures for nuclei segmentation. In order to get a deeper understanding, we also examine the roots of errors, the reasons for uneven performance among different cell lines and images, and the data requirements to train effective segmentation models.

3.1. Datasets

Eight distinct cell lines from two datasets were used in the work of this chapter. Their source and image acquisition process have been described earlier [10], and we briefly reiterate the main points here. First, the images of seven cell lines (i.e., HeLa, MDCK, HepG2, MCF7, NIH3T3, A549, and HT1080) were acquired using a $20x$ water immersion objective on an Opera Phenix^{TM} high-content screen-

ing system (Revvity) in the confocal mode for both fluorescence and brightfield modalities. The cells of each cell line were contained in 48 different wells, with nine fields of view captured from each well (432 combined). Hence, 3024 images of size $1080x1080px(1px = 0.59\mu m)$ were acquired for each modality. This data set is referred to as "seven cell lines".

Second, the images of the eighth cell line (i.e., LNCaP) were acquired by CellVoyager 7000 (Yokogawa) instrument using a $20x$ objective in the confocal mode for both fluorescent and brightfield modalities. For this cell line, a total of 784 images of $2556x2156$ pixels $(1px = 0.325\mu m)$ were captured for each modality, and we referred to this dataset as LNCaP.

Using Revvity's Harmony image processing software [96] and professional quality control to fine-tune the parameters, ground truth masks of the nuclei were created from fluorescence images in the seven cell lines as previously described [10]. Meanwhile, to generate the ground truth of the LNCaP dataset, we employed the U-Net++ model trained to segment fluorescence images from the seven cell lines images on the fluorescence images in this dataset.

3.2. Evaluated models

We assessed four CNNs representing effective architectural elements for image segmentation at the time of publication. These include dense block (a series of consecutive layers where each layer is directly connected to every other layer in a feed-forward manner); skip connections (connections that bypass one or more layers in a neural network); pyramid pooling (a module applies pooling operations at various scales (i.e., full image scale, as well as progressively smaller partitions such as dividing the image into four, etc.) and concatenates the output); and atrous convolution (convolutional operation that incorporates "holes" or gaps between the kernel's values) [10, 90, 93, 95]. In what follows, we describe the main components of the evaluated architectures and refer the reader to [19] for further details.

A successful U-Net model which was previously adjusted and adopted for brightfield nuclei segmentation serves as a baseline [10]. The model comprises a contraction path (i.e., encoder), an expansion path (i.e., decoder), a bottleneck, and skip connections between the encoder and the decoder Figure 11-A). The contraction path has 15 layers with 3x3 convolutions followed by ReLU activation. A 2x2 max-pooling layer and a skip connection to the decoder are integrated every third convolution. The expansion path mirrors this with 15 convolutions, ReLU, and upsampling every third convolutional layer. A bottleneck block of three convolutional layers connects the encoder and decoder (Figure 11-A). This architecture comprises a total of 1.3 million trainable parameters.

U-Net++ [90] enhances skip links between contraction and expansion paths by introducing convolutional layers (Figure 11-B). Both paths include four blocks each, comprising two convolution layers, batch normalization, and either 2x2 max-pooling (downsampling) or transposed convolution (upsampling). The bottleneck contains a similar block without max pooling. Skip connections from the first, second, and third contraction blocks to corresponding expansion blocks consist of three, two, and one blocks, respectively (Figure 11-B). Overall, U-Net++ consists of approximately 9 million trainable parameters.

Deeplabv3+ [95] features an asymmetric encoder-decoder architecture. The encoder consists of four parts: entry flow, middle flow, exit flow, and atrous spatial pyramid pooling (a technique used to extract multi-scale features from the image and reduce the resolution). Entry, middle, and exit flows have 3, 16, and 2 blocks, each consisting of two 3x3 convolutions and another one with a stride of two for downsampling. The input to each block is concatenated to its output. The feature maps input to the decoder is first bilinearly upsampled by a factor of four and then combined with the output of the encoder level with matching resolution. Following concatenation, 3×3 convolutions are applied, succeeded by another bilinear upsampling, this time by a factor of four, to restore the required output resolution (Figure 11-C). Deeplabv3+ features approximately 41 million

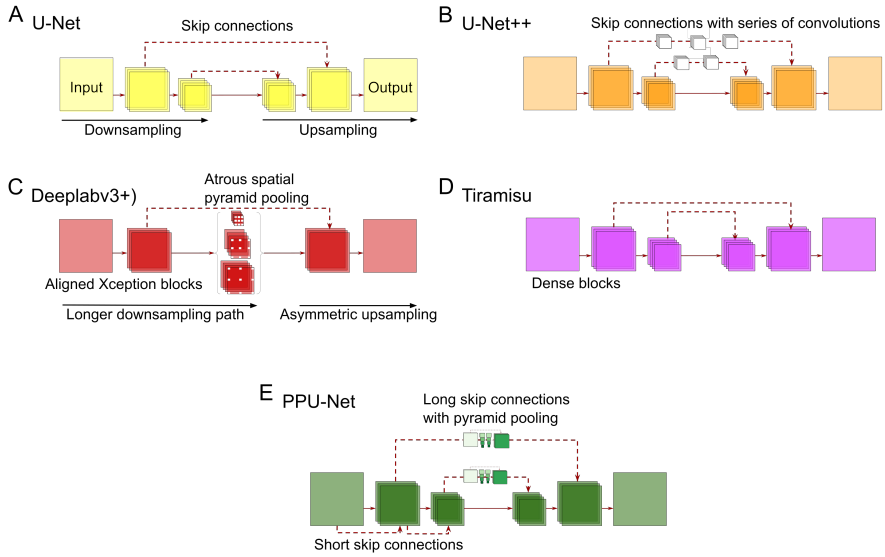

Figure 11. Several convolutional neural network architectures were compared: (A) U-Net has downsampling and upsampling paths with standard skip connections between them; (B) U-Net++ adds multiple convolutional layers in the skip connections; (C) Deeplabv3+ uses atrous convolution and spatial pyramid pooling in the upsampling path; (D) Tiramisu uses dense blocks as core building components; (E) The proposed PPU-Net has skip connections between downsampling blocks and deploys pyramid pooling modules in the long skip connections between the encoder and decoder.

parameters.

The Tiramisu [93] architecture employs dense blocks [97], where the output of each convolutional layer within the block is combined with the input of all subsequent convolutional layers in the same block. This architecture consists of an encoder with five dense blocks, each followed by batch normalization, ReLU activation, and 2x2 max pooling. Similarly, the decoder is composed of five dense blocks, each followed by a 3x3 transposed convolution with a stride of 2. The encoder's dense blocks contain (4, 5, 7, 10, 12) layers, while the decoder's dense blocks contain (12, 10, 7, 5, 4) layers. To connect the contraction and expansion paths, an additional dense block is applied in the bottleneck between them. Lastly, a convolutional layer is introduced at the start of the contraction path (Figure 11-D). Overall, this network comprises approximately 9.4 million parameters.

Inspired by the successful ideas in the field, we designed PPU-Net architecture to strike a sweet spot between model performance and size (Figure 11-D). PPU-Net uses both short and long skip connections, motivated by demonstrations that they improve performance and convergence. The short skips also stabilize training and mitigate vanishing gradients that hinders the effective update of parameters [98]. Moreover, to capture global and local context without losing spatial information, PPU-Net employs the pyramid pooling module [99] in the long skip

pathways.

The PPU-Net has contracting and expanding paths with a bottleneck between them, connected by skip pathways. The contracting path contains 10 blocks, each with two 3x3 convolutions followed by batch normalization and ReLU activation. There is a 2x2 maxpooling layer every second block. The expanding path mirrors the contracting one, with the maxpooling replaced by a 2x2 upsampling layer. In the skip pathways, the contracting path outputs are processed by a pyramid pooling module before concatenating to the expanding path input. This module integrates multi-scale information by average pooling at different scales (16x16, 8x8, 4x4, 2x2, 1x1). The output of each scale level is processed with 16 1x1 convolutions, batch normalization and ReLU activation before upsampling it bilinearly back to the original scale for concatenation with the input of the module. The PPU-Net architecture contains approximately 2.1 million trainable parameters.

3.3. Models performance on segmenting nuclei from brightfield images

We then evaluated the performance of the aforementioned models using the two described datasets. The reported object-wise F1 score is an average across IoU thresholds ranging from 0.5 to 0.95 in 0.05 increments and all results are based on pixel-level outputs binarized at a 0.5 cutoff unless detailed otherwise. Individual objects were identified by grouping interconnected positive pixels into distinct entities utilizing *measure.label* from the skimage package [57]. Following this, objects with a size less than 25 pixels were eliminated, and holes smaller than 25 pixels were filled using the functions *remove_small_objects* and *remove_small_holes* from the same package.

Analyzing the models' performance, we found that all models segmented nuclei from fluorescent images effectively (with pixel-wise F1 score averaged over all seven cell lines of 99% for U-Net, 99% for U-Net++, 98% for Deeplabv3+, 99% for Tiramisu, and 99% for PPU-Net), indicating that the signal in such images is sufficiently clear to be recognized independently of the model used. This suggests that fluorescence segmentation is a solved problem to a practical extent. Hence, the rest of the analysis in this chapter is dedicated to the more challenging brightfield images.

Unlike for fluorescence, models' performances on brightfield images have a wider range. The methods achieved pixel-wise F1 scores of 81.3%, 85.0%, 86.2%, 85.6%, and 86.5% for U-Net, U-Net++, Deeplabv3+, Tiramisu, and PPU-Net respectively in the seven cell lines dataset; and 84.0%, 86.2%, 87.7%, 85.8%, 87.4% in the LNCaP dataset, showing that newer architectures outperform the U-Net model. Qualitative and quantitative results of all models are presented in Figure 12.

We also found that PPU-Net, with 20 times fewer parameters than the comparatively accurate alternatives, marginally topped other methods in object-wise in-

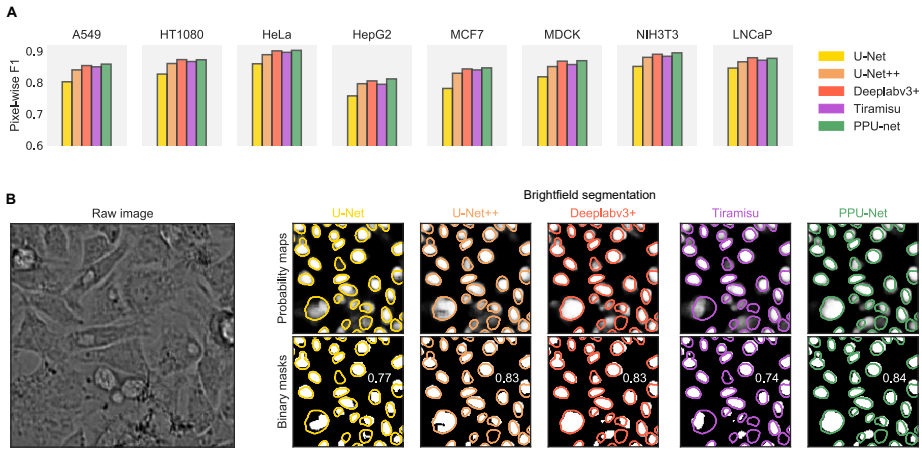

Figure 12. Models' performance on segmenting nuclei from brightfield images. A) Pixel-wise F1 scores (y-axis) of models (different colors) in the LNCaP (the rightmost sub-figure) and the seven cell lines dataset (the seven left sub-figures). B) An example of a brightfield image (leftmost panel) and the corresponding nuclei segmentation using different models (right panels). The panels in the top row to the right contain the probability maps, while the bottom row ones are the same maps after binarizing them using a threshold of 0.5. The colored contours are boundaries of ground truth nuclei, while the segmentation results are the white color inside those contours. Embedded numbers: the segmentation pixel-wise F1 score for the image using the corresponding model. The results presented here are derived from two distinct models. The first model was trained and tested using the seven cell lines combined, while the second model was trained and tested using the LNCaP dataset.

dicators(2.4) in the seven cell lines dataset and was a match for Deeplabv3+ in the LNCaP data (object-wise F1 scores of 48.8%, 55.4%, 58.1%, 58.2%, and 59.8% for U-Net, U-Net++, Deeplabv3+, Tiramisu, and PPU-Net, respectively, for seven cell lines; and 45.2 %, 51.0%, 54.9%, 51.4%, and 54.6% for the LNCaP). Notably, although the PPU-Net demonstrated promising results, it was not employed in the construction of the Score-CAM-U-Net framework in Publication II (Chapter 4), nor was it utilized in Publication III. This decision stemmed from the fact that the work for these publications had already commenced prior to the formal publication of the PPU-Net.

3.4. Dataset properties and training choices influence performance

In our quest to understand the criteria for training a successful brightfield segmentation model, we asked how images with different properties impact the performance. First, we examined the impact of cell density on model performance and

observed a negative correlation between the number of cells in an image and its pixel-wise F1 score across all models, as it can be seen in Figure 13, agreeing with the previous findings [10]. Notably, minor exceptions, such as in MDCK, can be attributed to the overall low density in the corresponding cell line. Second, we wondered whether object size influences prediction accuracy and found that all models segment larger nuclei more accurately, Figure 14-B.

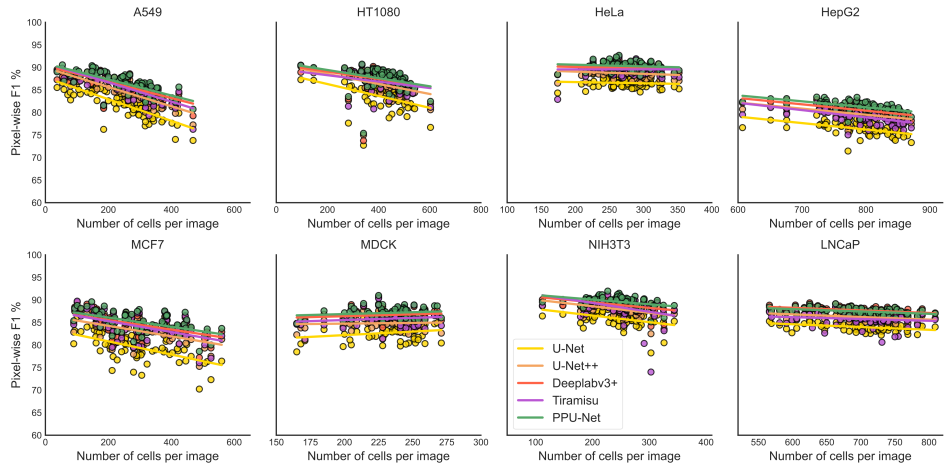

Figure 13. Model performance across varied nuclei densities. Pixel-wise F1 scores (y-axis) for different cell counts per image (x-axis), depicted by markers for each model (colors) and cell line (panels). Solid lines represent regression fits.

To get the reported results so far, thousands of labeled images were used to train models. In reality, ground truth masks are time-consuming to produce, and a limited number of annotated images are available. How many training photos are necessary to train a successful model is thus a vital practical question. To answer this, we first investigated whether data augmentation and label smoothing would enhance segmentation accuracy in the presence of little training data. We observed that using either of those strategies improves almost all models' performances compared to not using them, with data augmentation enhancing the results more than label smoothing. Notably, only 16 images with data augmentation were sufficient to train a model that achieves a performance within 95% of the performance accomplished by a model trained using the whole dataset of 286 images. Notwithstanding, the influence of the improvements introduced by both strategies tends to lessen with increasing training data volume (Figure 14-A).

Finally, we investigated whether segmentation performance might be enhanced by using more focal planes during training. We observed that adding one more plane improved the pixel-wise F1 score. Nevertheless, we noticed diminishing returns behavior by adding more focal planes (Figure 14-C).

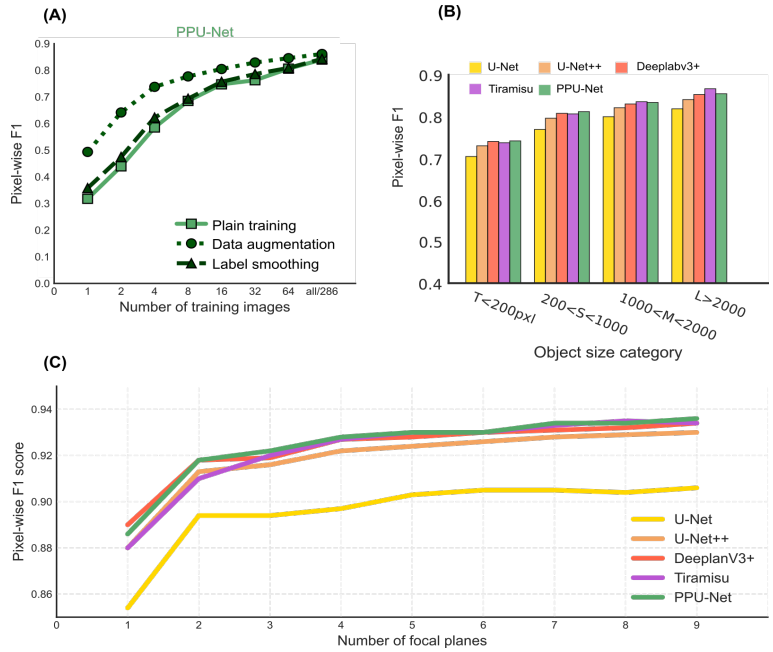

Figure 14. Accuracy improves with more relevant domain data and larger objects. (A) The A549 cell line's pixel-wise F1 score (y-axis) for models trained plainly (solid), with label smoothing (dashed), or data augmentation (dotted), against the quantity of training images (x-axis). (B) Intersection over Union ("IoU", y-axis) for nuclear sizes (x-axis) across different models (colors). Size labels: Tiny (T), Small (S), Medium (M), Large (L). (C) Pixel-wise F1 scores (y-axis) increase with more focal planes (x-axis) used in training.

3.5. Common errors in segmentation

In our quest to unravel the sources of segmentation errors, we visually inspected the segmentation output in an attempt to track the origins of such errors. We found that most errors can be ascribed to four primary sources. First, we observed a physical shift in the boundaries of some nuclei between brightfield and fluorescence modalities, for instance, if cells migrate due to improper adhesion to the plate (Figure 15-A). Since the ground truth masks were generated based on fluorescence modality, the unaligned nuclei may bias the brightfield model during training and, importantly, result in an error if they appear during model evaluation. Notably, this is not considered as an error from the model but a labeling issue due to modality misalignment. Second, there might be poor contrast between the signal in nuclei and the background, which could cause merging objects or entirely missing them (Figure 15-b). Third, cells that are out-of-focus, which may still be identifiable in fluorescence but challenging to spot in brightfield due to signal distortion (Figure 15-C). Finally, imaging artifacts, which can be biological contamination (e.g., bacterial colonies) or unwanted particle existence during image acquisition (e.g., dandruff, skin, or dust particles) (Figure 15-D). The latter

source turned out to be prevalent and diverse, and, therefore, we dived more into investigating and curing it in the next chapter of this thesis.

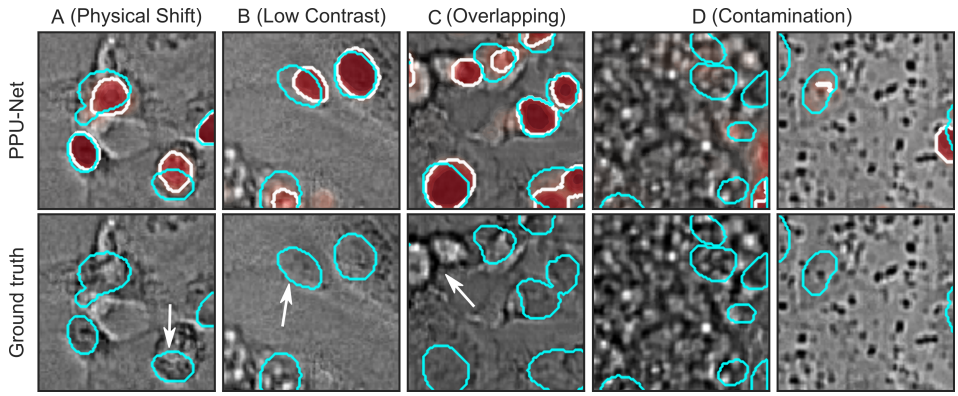

Figure 15. Segmentation error examples from visual inspection. Cyan, white, and red represent the ground truth contour, prediction contour, and prediction probability maps, respectively. The top images include both ground truth and PPU-Net predictions, whereas the bottom show only the ground truth. (A) Physical nucleus shift between modalities. (B) Low contrast. (C) Overlap of out-of-focus and in-focus nuclei. (D) Severe (left) and mild contamination (right).

3.6. Discussion

In this chapter, we have evaluated modern examples of cutting-edge segmentation models' families for the task of segmenting nuclei from brightfield and fluorescence cell microscopy images. We found that all evaluated models were able to detect nuclei from fluorescence with high accuracy, while the models ranged from moderate to well-performing in the case of brightfield images. Additionally, we proposed PPU-Net, a novel architecture for the same task. PPU-Net has fewer trainable parameters and shorter training and prediction times than the comparably accurate alternative model, setting PPU-Net, arguably, as the most practical model for the nuclei segmentation task.

In concordance with what has been observed before (e.g., [9, 10]), we found that providing more training data improves the models' capacity to detect nuclei; having more segmentation ground truth masks can be a considerable bottleneck, nevertheless. We then demonstrated that data augmentation substantially improves the model training under limited training data. Models' performances are also enhanced by using smoothed labels (i.e., permitting false-negative and false-positive rates in the truth labels) during training. Therefore, we recommend augmenting the training data when their volume is a constraint, but computational time and cost are not. We also found that the network design choices and size, training techniques, and data quality all affect the quality of the outcome.

Upon inspecting the errors, we realized the existing models could be improved further. Key sources of these errors include unwanted anomalous regions, potential physical shifts, noisy ground truth labels, and low contrast. Mainly, by minimizing or entirely eliminating these anomalous areas and optimizing label quality, we can push the models' performance. Moreover, the challenge of accurately segmenting small and densely packed nuclei, such as those in the HepG2 cell line, emerged as a notable issue. Addressing this in the future would likely necessitate the use of dedicated object delineation models and bespoke datasets. Another major avenue for future exploration is expanding this approach to segment entire cells, which holds promise for cytometry applications, especially in scenarios with dense cell cultures. More training data and innovative strategies for handling overlapping objects are required to progress in this direction.

4. WEAKLY SUPERVISED SEGMENTATION OF ARTIFACTS IN BRIGHTFIELD MICROSCOPY IMAGES (PUBLICATION II)

During our work to segment nuclei from brightfield microscopy images in the previous chapter, we noticed that some images contain visual artifacts that disturb accurate object segmentation, substantially contributing to the prediction error. We hypothesized that automatically removing such artifactual regions would yield more precise segmentation results and be of great practical interest. In this chapter, we study the problem of artifacts, their prevalence, and the impact of them and their removal on the downstream analyses.

4.1. Artifacts problem

An artifact is any signal that differs from what was expected [100]. The primary source of artifacts in cellular microscopy is the introduction of external objects during screening experiments, such as dust particles, debris of dead cells, bacterial contamination, etc. In this chapter, we are interested in identifying and removing such abnormalities [101, 102] in brightfield images, and we use the word artifact to refer to them.

Manual detection and removal of artifacts in cell microscopy is tedious, given the typically large number of images that would need to be processed [103, 104]. A popular alternative is a computer-aided solution to this problem, yet two complications make this task difficult. First is the stochastic nature of the occurrence of artifacts in microscopy images, which yields sparse data. Second, artifact features, such as texture and morphology, are usually very diverse and hence are difficult to model. This means, it is impractical to exhaustively collect representative samples of all potential artifact forms, making computational modeling hard.

Recently, deep learning has arisen as a promising technology to approach a wide range of computer vision tasks, including artifact detection [105]. Although fully supervised deep CNNs are state-of-the-art for the majority of computer vision applications, they are not the best choice for artifact segmentation from microscopy images [89, 91, 15, 14, 106]. A major bottleneck for this type of deep learning approach is their need for pixel-level labeling. The scarcity of pixel-level labeled data for artifacts, and the expert knowledge required to obtain such data, substantially increases the cost and time of data acquisition. Furthermore, given the arbitrary nature of artifacts' size, shape, and texture, models that detect artifacts would often require frequent retraining to accommodate new types of artifacts, making the reliance on pixel-level labeling even more resource-intensive. Alternatively, weakly supervised methods require only image-level annotations, substantially reducing the needed time for data preparation. In this chapter, we present empirical evidence that artifacts are prevalent in brightfield microscopy.

We then propose a framework that incorporates weak and fully supervised methods to segment and remove such artifacts involving only image-level labels.

4.2. Datasets

Three datasets are used in this chapter, covering nine different cell lines, two imaging instruments, live and fixed cells, and two plate formats. The provenance of two out of the three datasets (Seven cell lines, and LNCaP) was described in chapter 3, and we do the same for the third dataset here. Briefly, live CHO-K1-hM4R cells were seeded into 96 Well Black μ-Plate (Ibidi) $5-7$ hours before starting the imaging process. Four fields of view per well were obtained using a LUCPLFLN 20 × objective lens with a field of view size of $1224x904 pixels (1 pixel = 0.323 \mu m)$ and a density of 25000 cells per well. TAMRA-labeled fluorescence ligand was utilized to visualize cells expressing muscarinic M4 receptors in the fluorescence channel. In concentration-response experiments arecholine (Sigma), atropine, UR-SK75, and UNSW-MK259 were used as drug candidates. This dataset is referred to as ArtSeg-CHO-M4R, was produced by the GPCR group in the Institute of Chemistry, University of Tartu, and was made publically available [20].

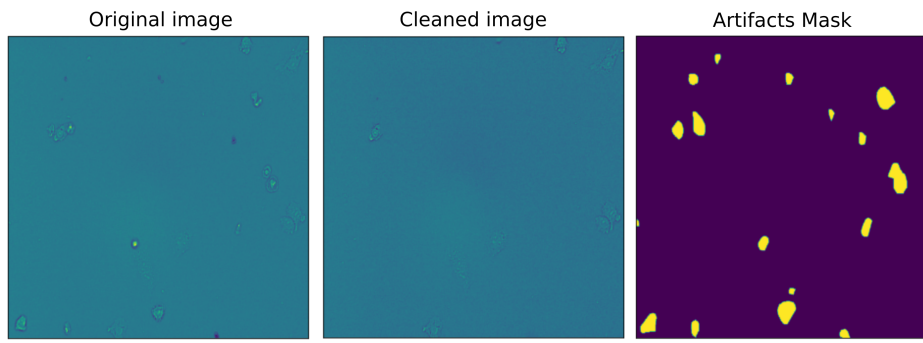

Figure 16. Example of image with artifacts that have been cleaned. (left) A brightfield image containing artifacts; (middle) the image post-artifact cleaning; (right) the artifacts masks indicating the location artifacts in the original image.

To prepare the datasets for artifact segmentation, the LNCaP and the seven cell lines were visually scrutinized, and 6.5% (i.e., $51/784$) and 11.4% (i.e., $344/3024$) of the images in LNCaP and the seven cell lines, respectively, were found to contain artifacts. Then, the same number of samples were randomly selected from each corresponding dataset to be used as modeling images without artifacts, together with the images with artifacts. As for the ArtSeg-CHO-M4R dataset, 99.2% (i.e., $(1171/1181)$) of the images were found to include artifacts. As almost all images in this dataset contain artifacts, the clean images were created by substituting the values of the pixel in the manually labeled artifacts with

the values of the matching pixels in the estimated background image. We fit the initial image with a two-dimensional second-order polynomial function [107] to estimate the background. Moreover, noise profile of the background pixels is added to the estimated background in order to emulate imaging noise. An example of an image containing artifacts and the corresponding cleaned image is shown in Figure 16. Importantly, the modified artifactual areas in the generated clean images were not detectable either by a trained model or by a human expert. To get the ground truth for all datasets, experts with prior training in cell biology and microscopy manually generated both the image-level and the pixel-level annotations. Each dataset is then partitioned into three splits with 482 training, 101 validation, and 104 testing images for seven cell lines; 1386 training, 404 validation, and 572 testing images for ArtSeg-CHO-M4R; and 70 training, 16 validation, and 16 testing images for LNCaP.

4.3. Artifacts are prevalent and diverse

Inspecting the datasets, we observed that the artifacts ranged in size from large that occupy 63%, 49%, and 10% of an image in the ArtSeg-CHO-M4R, seven cell lines, and LNCaP datasets, respectively, to tiny ones that cover only a few pixels in an image. The average artifacts' sizes were 4417 (0.38% of image size), 75933 (1.2% of image size), and 3450 (0.31% of image size) pixels in the seven cell lines, LNCaP, and the ArtSeg-CHO-M4R datasets, respectively. We also found that 10% of an image size was covered by artifactual regions in 16% of the images in the seven cell lines data. Notably, we found that a significant majority of images within the ArtSeg-CHO-M4R dataset contained artifactual areas, with a total of 13713 artifactual items in 1171 impacted images. This suggests the existence of even more severe situations wherein datasets are characterized by greater levels of uncleanliness. Artifacts also show heterogeneity in shape and texture, as it is shown in Figure 17. All these findings suggest that artifacts in cell brightfield microscopy are prevalent and heterogeneous in their properties.

4.4. Artifacts segmentation and removal with minimal human intervention

This section introduces our solution for artifacts segmentation from brightfield cell microscopy images with minimal human involvement during model development. The proposed framework incorporates the advantages of weakly supervised and strongly supervised methods for artifact segmentation involving only image-level labels. Precisely, the pipeline combines ScoreCAM-based weakly supervised model [21] with the U-Net fully supervised model [10], hence the name *ScoreCAM-U-Net*. To our knowledge, this is the first effort to segment artifacts from cell microscopy images in a weakly supervised mode. The graphical overview of the *ScoreCAM-U-Net* is depicted in Figure 18.

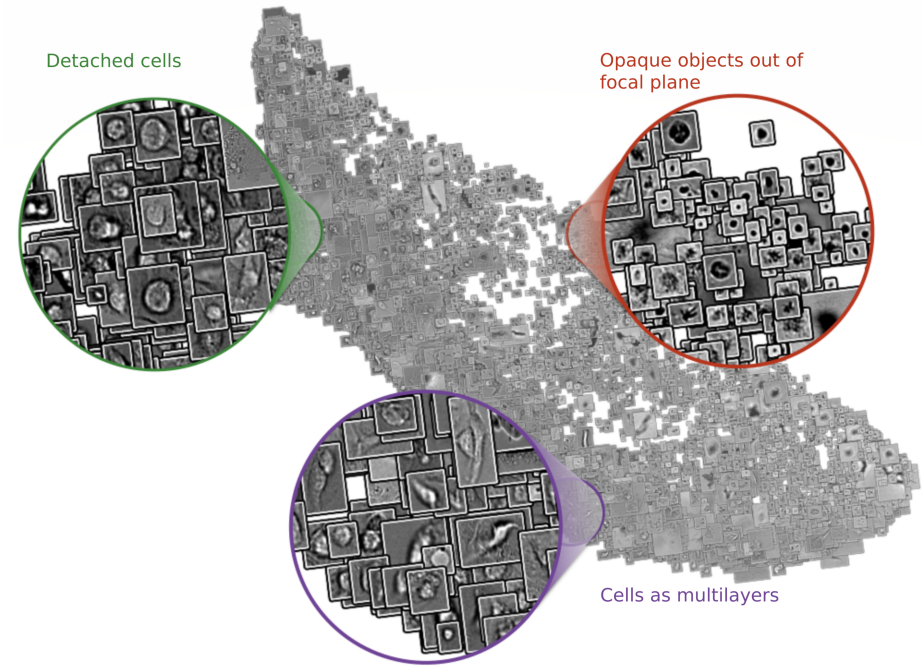

Figure 17. Artifacts exhibit diversity. A UMAP projection of artifacts found within the seven cell lines dataset. The UMAP takes as its input the pixels from each patch that holds an artifact, and it generates the first two features in the UMAP embeddings of each patch as outputs. These two features were then employed as 'x' and 'y' values, respectively, in order to create a 2D plot corresponding to the input patch

Score-CAM [21] is a method originally proposed to interpret image classification predictions made by deep learning models. The Score-CAM analyzes both the model's input and its output and indicates which areas in the input image influence the prediction decision the most. When predicting whether an image contains artifacts or not, the hypothesis is that the most influential pixels for this decision are the pixels of artifactual areas. The output of the Score-CAM is a set of activation maps that ideally have high probability values in the pixels that contain artifacts and low values otherwise. Those probability maps were then binarized and used as pseudo segmentation masks to train the strongly supervised model. As a strongly supervised, we used the U-Net [89] model whose architecture has already been customized for nuclei segmentation from brightfield image [10]. It is important to note that each dataset was used separately to train and evaluate each model.

Next, we evaluated the performance of the proposed ScoreCAM-U-Net method. We tested both the method's capacity to detect artifacts in an image and its ability to delineate them. The performance of ScoreCAM-U-Net was then compared with a number of alternative solutions. First, we compared the results we got

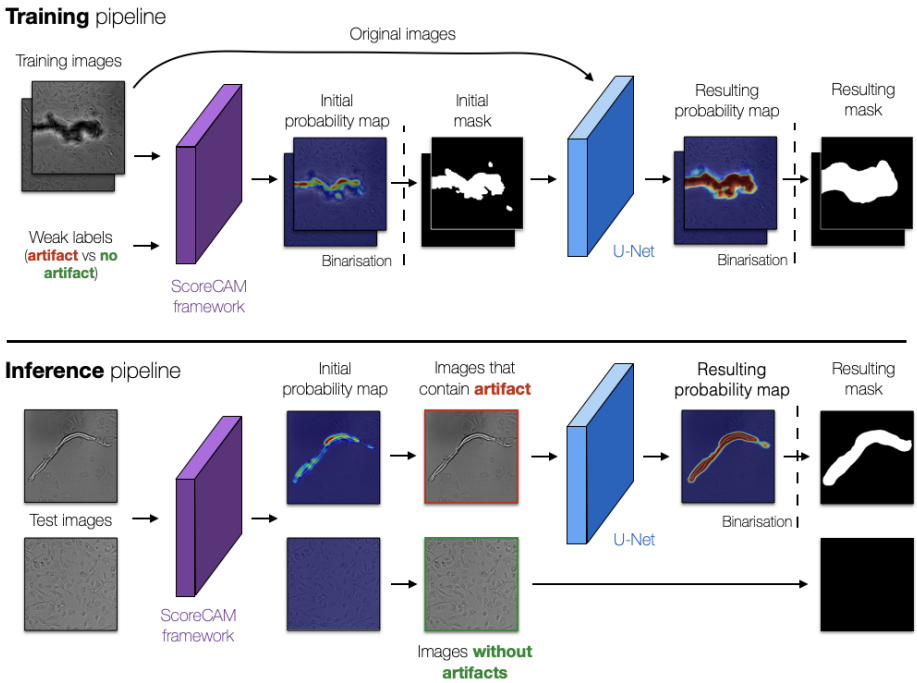

Figure 18. The proposed artifact segmentation pipeline (i.e., ScoreCAM-U-Net). Top: training phase in which the Score-CAM (purple) is used to produce probability maps with pixels belonging to artifact (if it exists), ideally, having higher values than the other pixels. The pixel values in the probability maps are then binarized (dashed vertical line) and used with the corresponding input images to train the segmentation model, U-Net (blue). Bottom: inference phase in which the trained segmentation model, U-Net, is used to delineate artifactual regions in the images that Score-CAM classified as having artifacts.

from the ScoreCAM-U-Net with the results of its two constituent individual models. Then, we compared the proposed approach against the then-state-of-the-art methods used to identify anomalies utilizing only image-level annotations in fields other than microscopy: an autoencoder-based approach(AE) [108], Patch Support Vector Data Description (PatchSVDD) [109], and Patch Distribution Modeling (PaDiM) [110]. To calculate the metrics, we initially binarized the probability maps generated by the models using cutoff values of 0.75 for AE, 0.3 for PaDIM, 0.0005 for PatchSVDD, 0.001 for ScoreCAM, ScoreCAM-U-Net, and 0.45 for U-Net. These specific thresholds were chosen to optimize pixel-wise Intersection over Union (IoU) performance on the validation data. Further, individual objects for all models' segmentation were identified by grouping interconnected positive pixels into distinct entities utilizing *measure.label* from the skimage package [57]. Subsequently, we employed the *remove_small_objects* function from the skimage package to eliminate objects smaller than 1000, 500, and 500 pixels in seven cell lines, ArtSeg-CHO-M4R, and LNCaP datasets, respectively. These size thresh-

olds were selected to maximize pixel-wise IoU of the majority number of models, and different sizes do not drastically change the performance of the models. We advise leveraging expert domain knowledge to determine the appropriate object size for the filtration process.

Evaluating the classification performance of the models, we found that both Score-CAM and ScoreCAM-U-Net were able to classify images accurately to artifactual or clean images with the same performance, with image classification F1 scores of 93.7%, 93.2%, and 90% in LNCaP, seven cell lines, and ArtSeg-CHO-M4R datasets respectively. They got the same detection performance as they both have the same classification model in the backend (i.e., ResNet). Compared to the ScoreCAM-based methods, the performance of the other approaches was inferior, except in the case of the fully-supervised U-Net in the LNCaP data (F1 score of 99.4%, and 93.7% for U-Net and ScoreCAM-U-Net, respectively), Figure 19B.

We then assessed the models' performance in segmenting artifacts. ScoreCAM-U-Net surpasses the non-strongly supervised alternatives, obtaining the highest average IoU and area under the precision-recall curve on the LNCaP and the seven cell lines datasets (see Figure 19B). Also when compared to the supervised U-Net model, ScoreCAM-U-Net secured the second-best IoU performance in both the seven cell lines (with ScoreCAM-U-Net at 49.5 and U-Net at 72.9) and the LNCaP datasets (where ScoreCAM-U-Net scored 39.9 as opposed to U-Net's 65.74) (see Figure 19B). We observe no clear outperforming weakly supervised approach in the ArtSeg-CHO-M4R data and that ScoreCAM-U-Net achieves the second largest IoU score compared to the fully supervised approach- U-Net- in LNCaP and seven cell lines datasets. Qualitative examples of all model segmentation results can be seen in Figure 19A.

It is noteworthy that despite the fact that the fully supervised method topped the weakly supervised alternatives, it took a relatively significant amount of time to produce the pixel-level labels demanded by the strongly supervised technique. While it typically took an expert 279 seconds to produce pixel-level labels for a single brightfield image, it only took them 2 seconds to tell whether or not the image had an artifact. This means, given the same dataset, the strongly supervised methods require roughly two orders of magnitude more of the experts' time than the weakly supervised ones to create pixel-level labels in the given case. Therefore, when choosing the method for artifact segmentation, it is fair to consider both the data size as well as the required annotation time.

4.5. Removing artifacts improves quality of nuclei segmentation

In chapter 3, we found that the presence of artifacts in the brightfield images substantially degrades nuclei segmentation quality in those images (see Figure 5 in chapter 2). Here, we test the impact of removing artifacts on the same task (i.e., nuclei segmentation from the brightfield images) in the seven cell lines dataset.

First, we separately evaluated nuclei segmentation performance in the artifactual and clean areas. We observed that the pixel-wise and object-wise F1 scores degraded, respectively, from 0.89 and 0.65 in the artifact-free areas to 0.60 and 0.28 in the artifactual regions. This is reflected in plain analyses that do not distinguish between artifact-free and artifactual areas, diminishing segmentation performance (0.87 pixel-wise and 0.61 object-wise F1 scores).

Second, we compared the impact of automatically removing artifacts using ScoreCAM-U-Net with manually removing them by an expert. Interestingly, we observed that the removal using ScoreCAM-U-Net has the same impact as the manual one, enhancing the segmentation accuracy to near-optimal (0.89 pixel-wise and 0.64 and object-wise F1 scores).

Subsequently, we examined the nuclear size and morphology properties, both with and without artifact correction. Nuclei within artifact-containing regions exhibited distinct properties, with a solidity of 0.92 and a size of 213 pixels, while artifact-free areas showed the same properties as 0.95 for solidity and 400 pixels for size. Interestingly, automatic artifact removal using ScoreCAM-UNet restored the expected nuclei size (397 pixels compared to 400 pixels for manual removal) and solidity (0.95, the same as manual removal value) in artifact-free regions, closely approaching the quality achieved through manual removal. These results suggest that automatic artifact removal effectively overcomes artifact-related issues, yielding results comparable to manual filtering. Notably, the results reported for nuclei segmentation within this section are based on the same preprocessing and post-processing procedures discussed earlier in Chapter 3 Section 3.3.

4.6. Discussion

To improve analyses from brightfield cell microscopy images, we proposed a deep-learning-based pipeline that demands minimal human involvement during model development for identifying artifacts. The *ScoreCAM-U-Net* pipeline combines the merits of weakly supervised learning (Score-CAM [21]) which requires only image-level labels, and strongly supervised learning which produces pixel-level resolution (U-Net [10]). The performance of ScoreCAM-U-Net in artifact segmentation surpasses the other non-strongly supervised alternatives. Also, the two ScoreCAM-based methods outperform all other approaches in identifying whether an image has artifactual areas. The only deviation is the performance of U-Net in the LNCaP data, which can be attributed to the limited number of images in the test split of the LNCap dataset. In fact, misclassification of only one image is sufficient for the observed performance difference. In general, ScoreCAM-U-Net nearly provides the performance of the strongly supervised method with an order of magnitude less time required to prepare the image-level annotations for training. Even under the assumption that weakly supervised models necessitate a larger volume (for example, larger by a factor of 10) of data to achieve performance parity with their strongly supervised counterparts, the time

discrepancy in ground truth preparation remains substantial. For example, If we were to train the fully supervised model using 200 images, an expert would still spend approximately 55,000 seconds annotating them at a pixel level, while they would spend only around 4,000 seconds annotating 2,000 images (10x images more than the ones used for supervised model) with image-level labels. This confirms the efficiency advantage of weakly supervised learning, despite under the assumption of its potential demand for larger datasets. Moreover, considering the unpredictable characteristics of artifacts, including their variable shapes, sizes, and textures, artifact segmentation models inherently necessitate retraining to accommodate the prospective emerging artifact types. This introduces a further demand for the more regular preparation of ground truth data, further highlighting the resource-intensive nature of this task under strongly supervised modeling. To our knowledge, this is the first effort to segment artifacts in cell microscopy brightfield images in a weakly supervised manner. It is important to clarify that our discussion does not intend to diminish the efficacy of strongly supervised approaches. Rather, we highlight the practicality and efficiency of weakly supervised techniques, particularly in the context of managing extensive datasets such as those encountered in large-scale screening campaigns or time-constrained scenarios. Embracing weakly supervised methods can serve as a strategic choice to optimize resource utilization while achieving commendable results. This nuanced perspective allows researchers and practitioners to make informed decisions when selecting the most suitable supervision strategy for their specific circumstances.

We believe that the impact of removing artifacts in downstream tasks is the most meaningful metric to reflect the practical advantages of removal. Our findings show that artifacts negatively affect nuclei segmentation and that eliminating artifacts improves nuclei segmentation performance both for recovery (e.g., pixel-wise and object-wise F1 scores) and estimating morphology properties (e.g., size and solidity of nuclei). Our model is of great practical interest in nearly any situation where large-scale cell microscopy brightfield images needed to be quantified.

The primary reward of ScoreCAM-U-Net is scalability. Using Score-CAM-U-Net will allow for acquiring clean images for large-scale screening experiments where it is prohibitively costly to inspect images manually. Moreover, both components of the ScoreCAM-U-Net can be replaced by instances of the same type, showcasing the flexibility of the proposed approach.

Despite its practical benefits, the proposed Score-CAM-U-Net approach has several notable limitations. Firstly, it lacks the ability to distinguish various artifact forms, such as differentiating fragments of dead cells from bacterial colonies. Therefore, a natural future step is to build up on our approach so that to distinguish different artifact types. Secondly, there still is a considerable discrepancy between results from automatic artifact removal compared to manual removal. Even in downstream applications, alignment between results after the removal using the two methods is sub-optimal. This highlights the necessity of more advanced models to shrink such discrepancies. Thirdly, the work does not investigate how

the model's performance varies with different artifact sizes. This raises another important avenue for future investigation. Finally, our current framework is not trained end-to-end, leaving room for further exploration and improvement in this area. Furthermore, as we focused our work on brightfield images, extending the work to other imaging modalities is of high practical interest.

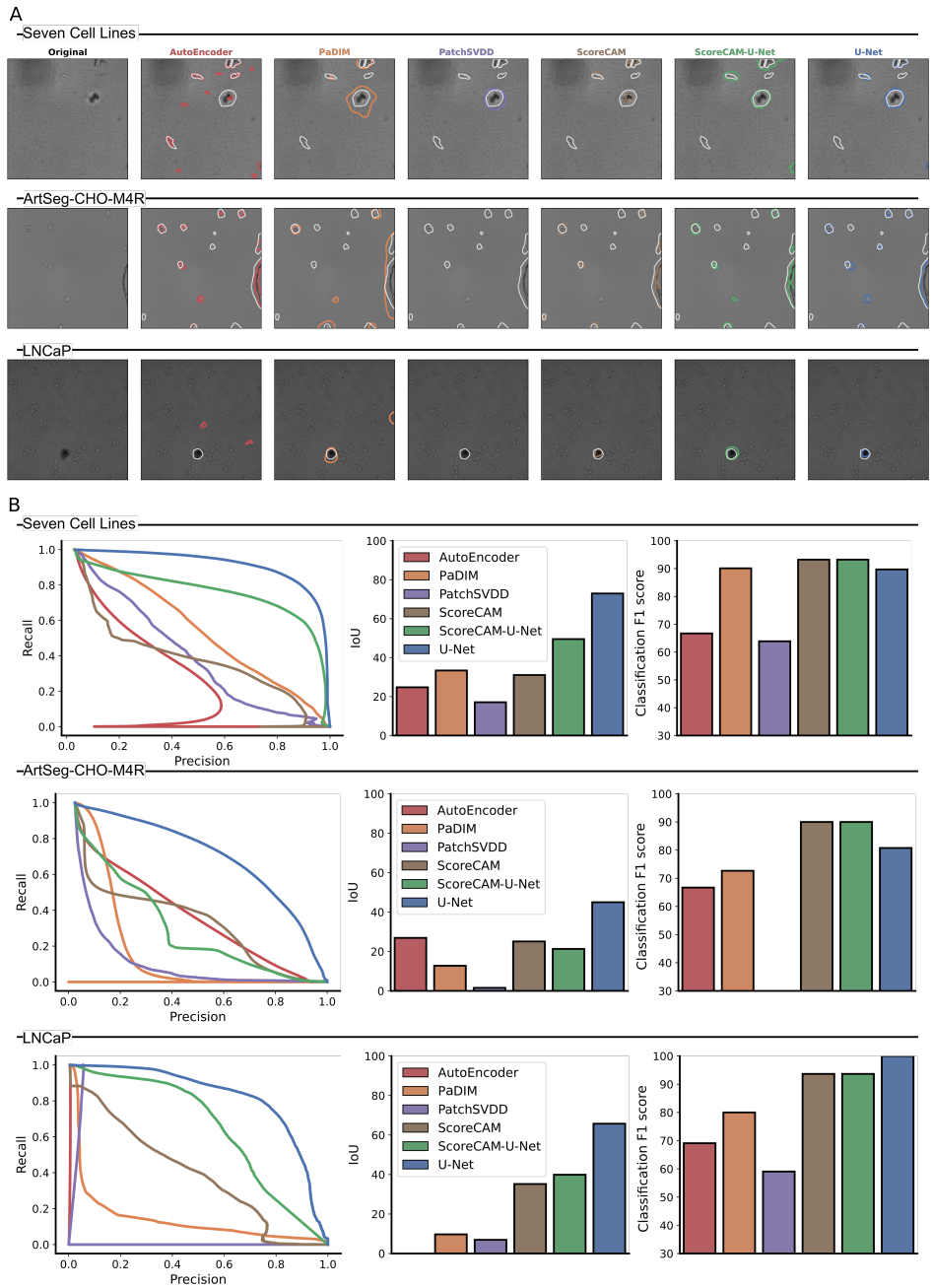

Figure 19. Qualitative and quantitative (image classification and artifact segmentation) results of the models (colors) on the three used datasets (i.e., LNCaP, ArtSeg-CHO-M4R, and seven cell lines). A) Example of segmenting artifacts in brightfield images (leftmost column) using all models (colors; columns) on different datasets (rows). Colored contours: border of segmented objects by different models. White contours: ground truth borders of the artifact objects. B) Quantitative results of all models (colors) on the three datasets (rows). Rightmost column: F1 score (x-axis of the column) results of the models' ability to classify images to either have artifacts or not. Middle column: average segmentation IoU (x-axis of the column) of the models. Leftmost column: the precision-recall curves of models artifact segmentation.

5. APPLYING NEURAL NETWORKS FOR DRUG DISCOVERY APPLICATIONS- MEASURING LIGAND BINDING AFFINITY TO M4 MUSCARINIC RECEPTOR AS AN EXAMPLE (PUBLICATIONS II, III)

In publication III, our collaborators from the chair of bioorganic chemistry in the Institute of Chemistry at the University of Tartu used high-affinity fluorescence ligands to investigate the ligand-binding properties to one of the most prominent cell membrane surface receptors- M4R. A group of G protein-coupled receptors (GPCRs) known as muscarinic acetylcholine receptors (mAChR), which comprises the M1–M5 subtypes, are essential for regulating numerous physiological processes, including heart and bladder operation, dopaminergic neurotransmission, and memory [111, 112, 112, 113]. The M4R has also been proposed as a possible drug target for treating neurodegenerative and neuropsychiatric illnesses like Alzheimer's disease, and schizophrenia [114, 115]. Moreover, the same receptor is recently linked to alcoholism and cocaine and heroin abuse, making M4 receptors even more diverse drug targets [116, 117]. Despite its expanding significance, it has proven challenging to create new medications that target the M4 receptor [22].

High-affinity fluorescence ligands are helpful tools for basic receptor research and have increased the possibilities for screening drug candidates. Using the fluorescence ligands as reporter probes will make it possible to determine the binding affinities of the unlabeled ligands. Fluorescence ligand concentration and the strength of interaction (i.e., affinity) between the ligand and the receptor determine the fluorescence intensity. Hence, we can figure out the M4 receptor-ligand affinity by analyzing the cells' fluorescence intensity. This scenario would require detecting cells and investigating the fluorescent signal within. With high-throughput microscopy, where a single screening experiment could generate terabytes of images, manually examining such a huge amount of data would be prohibitively expensive.

To automatically measure the protein-ligand interaction using fluorescence ligands, one would typically build a pipeline that involves cell segmentation and analyzing the fluorescence signal in the cell. However, the signal generated by the fluorescence ligands within cells may not always be sufficient to enable a reliable distinguish for a cell from the background. A varying number of ligands in different cells can cause inconsistent brightness, making some cells less or partially detectable than others. Additionally, photobleaching, where the fluorescent signal depletes over time, is another less common yet existing concern when training reliable models for cell segmentation from fluorescent images. Therefore, we opted for a more robust approach by segmenting cells from the more challenging yet fluorescent-independent brightfield modality. We first employed a deep learning model to segment the cells' bodies from brightfield images. Moreover,

as shown in publication II, one would need to clean the imaging artifacts, which brightfield images are often afflicted by, to get more realistic analysis readouts. In this chapter, we segmented the cells from microscopy images and studied the impact of artifacts and artifact removal on determining receptor-ligand interaction affinity. The pipeline used in this chapter is graphically represented in Figure 20. Importantly, despite this study having been applied to the M4R, it could easily be replicated and carried out on any receptor fluorescent ligand combination.

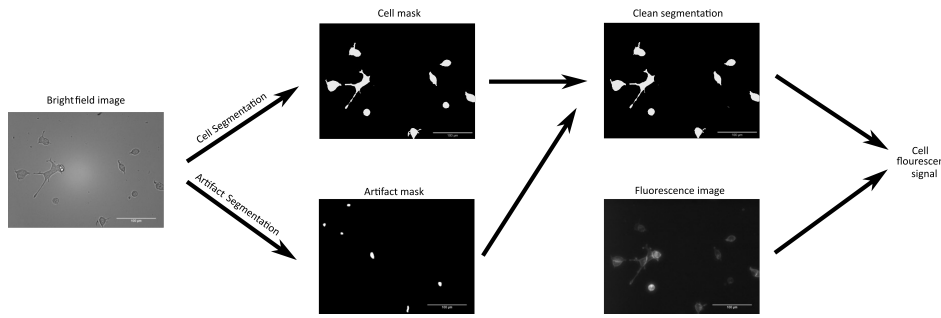

Figure 20. Pipeline of segmenting cells, removing artifacts, and measuring the fluorescence signal within the cell.

5.1. Datasets

Segmenting cells from fluorescent images can be straightforward when the fluorescent signal within the cell is sufficiently bright, while doing so from brightfield images is more challenging and often require more data to train the model effectively [19]. Since generating a substantial number of ground truth images for training a brightfield-based model is expensive and time-consuming, we opted for a strategy that generates pseudo-masks, utilizing a model trained on a limited set of fluorescence images that are labeled manually. Initially, ten in-focus RFP (red fluorescence protein) fluorescence channel images of CHO-K1-hM4R cells stained with DiI (Invitrogen, Eugene, Oregon, USA) dye were manually labeled, split into 60% training, 20% validation, and 20% test sets, and used to train a model for segmenting cells based on fluorescent images. Subsequently, this model was used to predict the masks for 191 DiI-labeled fluorescence images. These generated masks were subsequently divided into training, validation, and test sets, along with the corresponding in-focus brightfield images of the same fields of view (133, 29, and 29 images, respectively). This larger dataset was then used to train a brightfield-based model, providing a more cost-effective and efficient approach. Once the brightfield-based model was trained for cell segmentation, it was applied to segment cells in the ArtSeg-CHO-M4R dataset, which was detailed in Chapter 4. It is worth mentioning that the collaborators from the GPCR group at the Institute of Chemistry, University of Tartu, were responsible

for the datasets generation.

5.2. Segmenting cells from BF images using CNN

To analyze the signal that characterizes the ligand-binding properties, we first needed to segment cells. The U-Net-based CNN architecture was trained for cell segmentation from the brightfield images [10], and we called this model U-Net-BF to differentiate it from the model trained to segment cells from fluorescent during pseudo-masks preparation (see 5.1) that was called U-Net-FL. BF and FL stand for brightfield and fluorescent, respectively. A graphical representation of the architecture is shown in Figure 21. The architecture was trained for up to 200 epochs, and Adam optimizer [118] was used to optimize binary cross-entropy loss. The training was terminated completely if validation loss was not improving for 20 consecutive epochs. Batches of eight images of size 288 × 288 pixels were used to train the U-Net model based on the available processing budget [94].

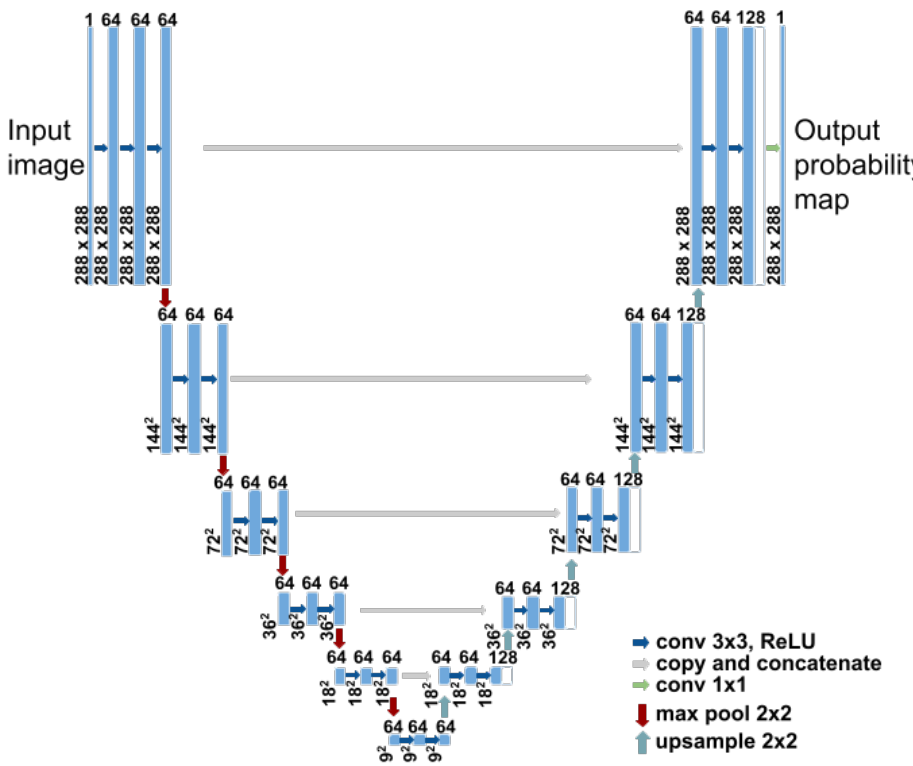

Figure 21. Graphical depiction of the segmentation architecture utilized to segment cells in Chapter 5.

Using bio-image analysis tools such as ilastik [7] is a de facto practice. Therefore, we compare the DL segmentation model results to a random forest-based

Table 1. Quality metrics of the DL and RF segmentation models.

Metric	DL-based model	RF-based model
Recall	0.86	0.72
Precision	0.93	0.74
F1 score	0.89	0.73

pixel classification (RF-based) model built on top of ilastik, where random forest is the default algorithm of pixel-classification workflow in Ilastik. To calculate the metrics, we initially binarized the probability maps generated by all models using a 0.5 cutoff value. Subsequently, individual objects were identified by grouping interconnected positive pixels into distinct entities. We found that DL-based model clearly outperforms the RF-based model with recall, precision, and F1 score of 86%, 93%, 89%, respectively, for the DL model and 72%, 74%, 73% respectively for the RF model as can be seen in table 5.2.

5.3. Pharmacological parameter estimates and the impact of removing artifacts

We next studied receptor-ligand interaction affinity and the impact of artifacts and artifact removal on the determination of such interaction. An overview of the general workflow is illustrated in Figure 20. Concentration–response experiments were conducted using arecholine (Sigma), atropine, UNSW-MK259, and UR-K75 as examples of drug candidates. In each experiment, cells were segmented from brightfield images using the trained model described above (i.e., U-Net-BF) with an average F1-score of 0.89. To test the impact of artifacts, the image analysis was applied with and without removing those artifacts. Additionally, to evaluate the effectiveness of the automatic detection of artifacts, the artifactual areas were removed manually by an expert as well as automatically using ScoreCAM-U-Net (see chapter 4). The average intensity of the pixels in the detected cells and the average intensity of the background were obtained from the red fluorescence images corresponding to the brightfield images. The values were averaged for all images from the same well. The background intensity in each well is subtracted from the cellular intensity in the same well to acquire the signal of the receptor-bound fluorescence ligands only. Then, regression analysis is used to obtain the fluorescence ligand's half-maximal displacement values ($LogIC_{50}$), which indicate the amount of ligand needed to bind to half of the receptors. In particular, the fluorescence intensity dependence on the competitive ligand concentration was fitted with the Hill equation using GraphPad Prism and the "log(inhibitor) vs. response" nonlinear regression model, which is equivalent to the logistic regression.

To quantify the impact of artifacts, we opted for the absolute difference between the $LogIC_{50}$ values obtained from manual artifact removal and the automated alternative (i.e., ScoreCAM-U-Net). Such differences reflect how precise pharmacological parameters can be acquired with and without eliminating arti-

facts. We found that manual anomaly elimination evidently impacts both the plateau positions in the displacement curves and the $Log(IC_{50})$ values, as shown in Figure 22. The mean absolute difference between $Log(IC_{50})$ calculated under manual artifact removal and either automatic or no artifact removal is 0.16 and 0.29 units, respectively, while the standard deviation between biological replicate experiments is 0.11. This suggests that removing artifacts is crucial to get an unbiased estimation of the pharmacological parameters. It also suggests that removing artifacts using ScoreCAM-U-Net substantially narrowed the difference compared to the removal by an expert. We also used Pearson's correlation coefficient, r, between well-average fluorescent intensities calculated after removing artifacts manually and automatically. We observed that the r value between manual and automatic removal is 0.98, while it is 0.93 between manual and no removal, suggesting the analysis readout similarity between manual-based and ScoreCAM-U-Net-based artifact removal.

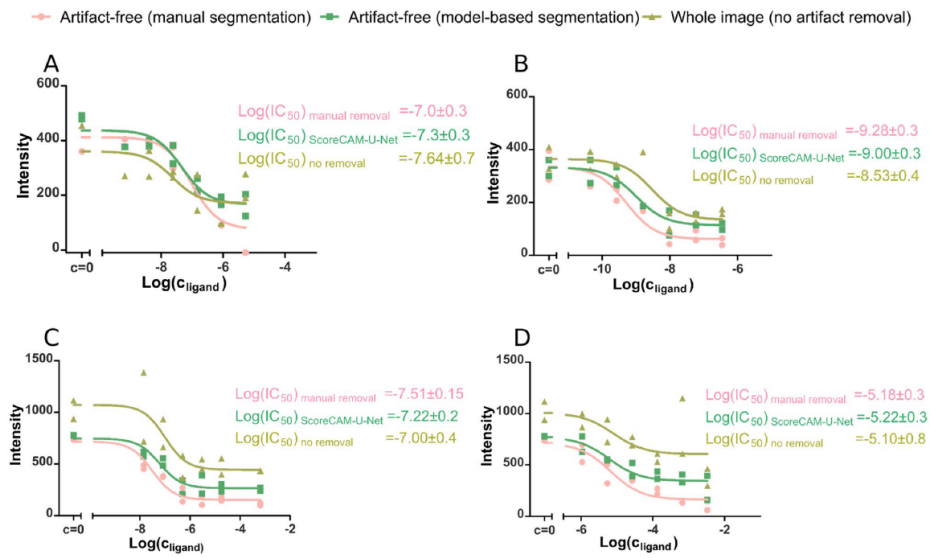

Figure 22. Cell fluorescence intensity dependence on M4 receptor ligand concentration determined with live cell fluorescence microscopy at the presence of 2 *nM* UR-CG072. Displacement curves of three different ligands are shown: pirenzepine (A, C), atropine (B) and arecholine (D). Three different artifact removal methods at the image analysis stage are compared (colors): manual artifact segmentation, ScoreCAM-U-Net segmentation and no artifact removal. For each combination of ligand and artifact removal method a regression analysis is performed with Hill equation (Hill coefficient fixed at-1) with the best fits shown as continuous lines. For each displacement curve, the $Log(IC_{50}) \pm$ SD is presented, where SD represents the standard deviation estimation of $Log(IC_{50})$. Each displacement curve was measured in duplicates with each data point representing the average fluorescence intensity of cells in each well.

5.4. Discussion

Due to its connections with numerous disorders and body functions, the M4 receptor is considered an appealing target for pharmaceutical research. Although fluorescence-based techniques have become more popular in drug screening nowadays, research on the M4 receptor has been considerably restricted by the scarcity of appropriate fluorescent ligands for this receptor. In this part of the thesis, a research group from the Institute of Chemistry at the University of Tartu used high-affinity fluorescence ligands to develop assays for M4 receptors, which involves a comprehensive analysis of microscopy images.

A deep learning (DL) model was trained to segment cells from the challenging brightfield images and was subsequently compared to an ilastik-based model. This comparison demonstrates that while the Ilastik pipeline models are simpler to retrain and adapt, the DL pipeline yields results of superior quality. Retraining is a pivotal step whenever a substantial domain shift is introduced into the data, such as changes in the cell line or microscope equipment, which is a general concern for any ML/DL method. However, it is crucial to note that retraining is typically a one-time requirement for a specific assay configuration. Even though the DL-based model's results were satisfactory for our current objectives, we did not delve into exploring additional models. Nonetheless, a more advanced model could further enhance performance and could be considered in the future.

We then assessed the drug-receptor binding affinities based on live-cell microscopy. We also examined the implications of eliminating image artifacts in this assessment. The ligand-receptor affinities calculated after cleaning images using ScoreCAM-U-Net exhibited a higher concordance with the affinities obtained after experts manually cleaned those images. The model-based calculations decrease the uncertainty in linear regression and the variability of the outcomes of independent experiments, showing both a better theoretical model fit and increased measurement repeatability. However, while removing artifacts using ScoreCAM-U-Net was closer to the experts' removal, it was not optimal. Such disparities highlight a potential avenue for future research, emphasizing the need for exploring an even more accurate artifact removal model. Furthermore, given that ScoreCAM-U-Net operates as two distinct models, exploring an integrated end-to-end framework could present an avenue for potential enhancements. Importantly, despite this study and analysis being applied to the M4 receptor, it could easily be replicated and carried out on any receptor fluorescent ligand combination.

6. CONCLUSIONS

Obtaining precise and trustworthy measurements from cell microscopy images is crucial in studying human biology. Although visually inspecting those images helps to develop the intuition, automatically processing them is the only way to analyze this massive data. The large data volume and complicated signal in high-content microscopy make it challenging to automate such analysis. This thesis proposes several novel techniques and experiments towards better understanding of mammalian cells from microscopy images.

Microscopes can produce different types of images. Generating microscopy brightfield images is relatively quick and noninvasive; however, they are challenging to work with. One routine step in the analysis of microscopy images is the identification of the cells' nuclei, which, for example, plays a significant role in cancer diagnosis and grading. Detecting and delineating nuclei from the noninvasive brightfield would, therefore, be of great practical value. In publication I, we comprehensively studied this problem. We assessed the utility of the then-state-of-the-art convolutional neural networks in segmenting nuclei from microscopy images, focusing on brightfield modality. We further investigated the causes of error, reasons for inconsistent performance between cell lines and images, the data needs, and the characteristics of the architecture for building effective segmentation models. We then proposed a novel lightweight architecture (PPU-Net) that makes use of the obtained insights. Our analyses show that foreign objects often found in microscopy images such as particle contamination during acquisition (e.g., dandruff or skin), or bacterial colonies account for a large part of model's mistakes. In the later publication where we studied these objects further, we refer to them as artifacts.

In publication II, we showed that artifacts are prevalent and observed in different shapes and sizes in brightfield images. Searching through a published body of research, we realized that the problem of artifacts is largely overlooked by microscopists and biomedical image analysts. Checking the impact of the artifacts, we found that they severely impact the measurements in the downstream analyses, such as nuclei segmentation, motivating the need for developing an approach that can identify and remove such artifacts. The most straightforward way to do so is to train a strongly supervised model that would learn to segment the artifacts from hundreds of manually labeled examples. However, this would take a lot of time for experts to annotate the necessary training data. Moreover, artifact characteristics, such as morphology and texture, are often very heterogeneous and hence challenging to define. This heterogeneous nature made collecting a one-fits-all dataset implausible. Hence, we decided to try another approach - develop a simple-to-use model that would require very little manual input yet be robust. Such an approach could be trained on demand in any lab for any dataset. One good alternative is to involve only image-level labels (also known as weak labels), which dramatically reduces the annotation time. For that, we adopted ScoreCAM, a technique ini-

tially proposed to explain predictions made by deep-learning image classification models. ScoreCAM analyzes both the model input and its corresponding output and shows components of the image that substantially contributed to the particular prediction. Furthermore, we decided to combine the performance advantages of the strongly supervised methods and the convenience of image-level annotations. To do so, the output of ScoreCAM was binarized and used as pseudo-masks for the U-Net-based model, which was subsequently trained to segment the artifacts. We showed that this combination, which we call ScoreCAM-U-Net, accurately detected artifacts with minimal human intervention.

Inspired by what we had learned, we decided to take the acquired knowledge into a more practical, real-world context. Specifically, we examined the usefulness of using DL in drug discovery research in publication III. We collaborated with a team of chemists and biologists who were developing assays to study the M4 cell membrane receptor. This receptor is highly important, but it has been difficult to create new medications that can effectively target it. Because high-affinity fluorescence ligands have been proven to be useful in fundamental receptor research, our collaborators used them to build assays for studying the interaction between the ligand and M4 receptor. To study this interaction, we needed to identify the cells and analyze the fluorescent signal within them, which is influenced by the affinity between the receptor and the ligand as well as the ligand concentration. Typically, detecting cells from fluorescent images is the way, as this is a straightforward task. However, sometimes the signal generated by the fluorescence ligands within the cells is not strong enough for the model to distinguish cells from the background, making this approach impractical. As a result, we decided to segment cells from brightfield images, which are more challenging but do not rely on fluorescent signals. We used DL to segment the bodies of cells from brightfield images and used the segmentation coordinates to extract the cells from fluorescent images. We then analyzed the fluorescent signal of those cells. Next, we examined the effect of removing artifacts from brightfield images on the signal resulting from the interaction between the receptor and ligand. We found that removing these artifacts resulted in a more unbiased signal, and our proposed model achieved near-optimal performance in doing so.

To conclude, in this work, we started by comparing the performance of various state-of-the-art segmentation models for segmenting nuclei from brightfield microscopy images and proposed our lightweight model for the same purpose. We realized that the presence of foreign objects in the images is one of the central sources of errors. To mitigate the impact of such objects, we proposed an approach to detect and remove them with minimal human input. Finally, we distilled the knowledge and insights we attained into helping in candidate drug screening research. The outcome of this work has benefited various industrial and academic projects and organizations, including Revvity (https://www.revvity.com/), a corporation that develops instruments, systems, and software for analyzing cells. Furthermore, our work has facilitated impactful research in labs at the University of

Tartu, Estonia, and the Institute of Molecular Biology, Mainz, Germany. Notably, our second publication has played a pivotal role in a patent application centered on computer tomography image segmentation, thus contributing to the advancement of the field.

BIBLIOGRAPHY

[1] Jeremy Simpson. *High content screening microscopy: It's all about numbers*. 2014. URL: https://www.physoc.org/magazine-articles/high-content-screening-microscopy-its-all-about-numbers/ (visited on 10/03/2022).

[2] *Opera Phenix Plus High-Content Screening System*. https://www.perkinelmer.com/product/opera-phenix-plus-system-hh14001000. Accessed: 2022-10-03.

[3] Chris Allan et al. "OMERO: flexible, model-driven data management for experimental biology". In: *Nature methods* 9.3 (2012), pp. 245–253.

[4] Kevin W Eliceiri et al. "Biological imaging software tools". In: *Nature methods* 9.7 (2012), pp. 697–710.

[5] Johannes Schindelin et al. "Fiji: an open-source platform for biological-image analysis". In: *Nature methods* 9.7 (2012), pp. 676–682.

[6] Anne E Carpenter et al. "CellProfiler: image analysis software for identifying and quantifying cell phenotypes". In: *Genome biology* 7.10 (2006), pp. 1–11.

[7] Christoph Sommer et al. "Ilastik: Interactive learning and segmentation toolkit". In: *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE. 2011, pp. 230–233.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.

[9] Juan C Caicedo et al. "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images". In: *Cytometry Part A* 95.9 (2019), pp. 952–965.

[10] Dmytro Fishman et al. "Practical segmentation of nuclei in brightfield cell images with neural networks trained on fluorescently labelled samples". In: *Journal of Microscopy* 284.1 (2021), pp. 12–24.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012).

[12] Jie Mei et al. "CoANet: Connectivity attention network for road extraction from satellite imagery". In: *IEEE Transactions on Image Processing* 30 (2021), pp. 8540–8552.

[13] Hao Zhang et al. "Dino: Detr with improved denoising anchor boxes for end-to-end object detection". In: *arXiv preprint arXiv:2203.03605* (2022).

[14] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[15] Liang-Chieh Chen et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs".

In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 834–848.

[16] Abdullah Rashwan et al. “Dilated SpineNet for semantic segmentation”. In: *arXiv preprint arXiv:2103.12270* (2021).

[17] Ishita Singh and Tanmay P Lele. “Nuclear Morphological Abnormalities in Cancer: A Search for Unifying Mechanisms”. In: *Nuclear, Chromosomal, and Genomic Architecture in Biology and Medicine*. Springer, 2022, pp. 443–467.

[18] Srinivas Niranj Chandrasekaran et al. “Image-based profiling for drug discovery: due for a machine-learning upgrade?” In: *Nature Reviews Drug Discovery* 20.2 (2021), pp. 145–159.

[19] Mohammed AS Ali et al. “Evaluating very deep convolutional neural networks for nucleus segmentation from brightfield cell microscopy images”. In: *SLAS DISCOVERY: Advancing the Science of Drug Discovery* 26.9 (2021), pp. 1125–1137.

[20] Mohammed AS Ali et al. “ArtSeg—Artifact segmentation and removal in brightfield cell microscopy images without manual pixel-level annotations”. In: *Scientific reports* 12.1 (2022), pp. 1–13.

[21] Haofan Wang et al. “Score-CAM: Score-weighted visual explanations for convolutional neural networks”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, pp. 24–25.

[22] David M Thal et al. “Crystal structures of the M1 and M4 muscarinic acetylcholine receptors”. In: *Nature* 531.7594 (2016), pp. 335–340.

[23] B. Alberts et al. *Molecular Biology of the Cell*. 4th. Garland, 2002.

[24] L David et al. “Lehninger principles of biochemistry”. In: (2000).

[25] Maris-Johanna Tahk et al. “Live-cell microscopy or fluorescence anisotropy with budded baculoviruses—which way to go with measuring ligand binding to M4 muscarinic receptors?” In: *Open Biology* 12.6 (2022), p. 220019.

[26] Gurvinder Kaur and Jannette M Dufour. *Cell lines: Valuable tools or useless artifacts*. 2012.

[27] Jennifer Dumont et al. “Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives”. In: *Critical reviews in biotechnology* 36.6 (2016), pp. 1110–1122.

[28] John R Masters. “HeLa cells 50 years on: the good, the bad and the ugly”. In: *Nature Reviews Cancer* 2.4 (2002), pp. 315–319.

[29] *Cell line profile, A549*. `https://www.culturecollections.org.uk/media/126404/a549-cell-line-profile.pdf`. Accessed: 2022-10-31.

[30] *At the Cell Counter: HT1080 Cells*. `https://www.moleculardevices.com/applications/cell-counting/cell-counter-ht1080-cells`. Accessed: 2022-11-02.

[31] Şerban Comşa, Anca Maria Cimpean, and Marius Raica. “The story of MCF-7 breast cancer cell line: 40 years of experience in research”. In: *Anticancer research* 35.6 (2015), pp. 3147–3154.

[32] *Cell line profile, Hep-G2*. `https://www.culturecollections.org.uk/media/134737/hep-g2-cell-line-profile.pdf`. Accessed: 2022-10-31.

[33] Mark A Castanares et al. “Characterization of a novel metastatic prostate cancer cell line of LNCaP origin”. In: *The Prostate* 76.2 (2016), pp. 215–225.

[34] Francesca Maria Capellini et al. “Characterization of MDCK cells and evaluation of their ability to respond to infectious and non-infectious stressors”. In: *Cytotechnology* 72.1 (2020), pp. 97–109.

[35] Christine Leibiger et al. “First molecular cytogenetic high resolution characterization of the NIH 3T3 cell line by murine multicolor banding”. In: *Journal of Histochemistry & Cytochemistry* 61.4 (2013), pp. 306–312.

[36] *NIH 3T3 CELL LINE*. `https://www.nih3t3.com/`. Accessed: 2022-11-03.

[37] Paul C Goodwin. “A primer on the fundamental principles of light microscopy: Optimizing magnification, resolution, and contrast”. In: *Molecular reproduction and development* 82.7-8 (2015), pp. 502–507.

[38] Tanel Pärnamaa and Leopold Parts. “Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning”. In: *G3: Genes, Genomes, Genetics* 7.5 (2017), pp. 1385–1392.

[39] Nicola Bougen-Zhukov et al. “Large-scale image-based screening and profiling of cellular phenotypes”. In: *Cytometry Part A* 91.2 (2017), pp. 115–125.

[40] Juan C Caicedo, Shantanu Singh, and Anne E Carpenter. “Applications in image-based profiling of perturbations”. In: *Current opinion in biotechnology* 39 (2016), pp. 134–142.

[41] Kurt Thorn. “A quick guide to light microscopy in cell biology”. In: *Molecular biology of the cell* 27.2 (2016), pp. 219–222.

[42] Paul Lang et al. “Cellular imaging in drug discovery”. In: *Nature Reviews Drug Discovery* 5.4 (2006), pp. 343–356.

[43] Vincent Blay et al. “High-throughput screening: today’s biochemical and cell-based approaches”. In: *Drug Discovery Today* 25.10 (2020), pp. 1807–1821.

[44] Andrew Bullen. “Microscopic imaging techniques for drug discovery”. In: *Nature reviews Drug discovery* 7.1 (2008), pp. 54–67.

[45] Tomohiro Hayakawa et al. “Computational nuclei segmentation methods in digital pathology: a survey”. In: *Archives of Computational Methods in Engineering* 28.1 (2021), pp. 1–13.

[46] Caner Mercan et al. “Deep learning for fully-automated nuclear pleomorphism scoring in breast cancer”. In: *npj Breast Cancer* 8.1 (2022), pp. 1–11.

[47] Jeanne Ayache et al. “Artifacts in transmission electron microscopy”. In: *Sample preparation handbook for transmission electron microscopy*. Springer, 2010, pp. 125–170.

[48] E Ann Ellis and Leona Cohen-Gould. “Recognizing and Preventing Artifacts in Microscopy: A Roundtable Discussion”. In: *Microscopy and Microanalysis* 22.S3 (2016), pp. 2074–2075.

[49] Daan De Geus, Panagiotis Meletis, and Gijs Dubbelman. “Panoptic segmentation with a joint semantic and instance segmentation network”. In: *arXiv preprint arXiv:1809.02110* (2018).

[50] Alexander Kirillov et al. “Panoptic segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 9404–9413.

[51] Nobuyuki Otsu. “A threshold selection method from gray-level histograms”. In: *IEEE transactions on systems, man, and cybernetics* 9.1 (1979), pp. 62–66.

[52] Serge Beucher. “Use of watersheds in contour detection”. In: *Proceedings of the International Workshop on Image Processing*. CCETT. 1979.

[53] Carolina Wählby et al. “Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections”. In: *Journal of microscopy* 215.1 (2004), pp. 67–76.

[54] Jinghua Zhang et al. “LCU-Net: A novel low-cost U-Net for environmental microorganism image segmentation”. In: *Pattern Recognition* 115 (2021), p. 107885.

[55] Jan Oscar Cross-Zamirski et al. “Label-free prediction of cell painting from brightfield images”. In: *Scientific reports* 12.1 (2022), pp. 1–13.

[56] Leo Breiman. “Random forests”. In: *Machine learning* 45 (2001), pp. 5–32.

[57] Fabian Pedregosa et al. “Scikit-learn: Machine learning in Python”. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[58] Laith Alzubaidi et al. “Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions”. In: *Journal of big Data* 8 (2021), pp. 1–74.

[59] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors”. In: *nature* 323.6088 (1986), pp. 533–536.

[60] Sebastian Ruder. “An overview of gradient descent optimization algorithms”. In: *arXiv preprint arXiv:1609.04747* (2016).

[61] Mingchen Gao et al. “Segmentation label propagation using deep convolutional neural networks and dense conditional random field”. In: *2016*

IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE. 2016, pp. 1265–1268.

[62] Lars Schmarje et al. “A survey on semi-, self-and unsupervised learning for image classification”. In: *IEEE Access* 9 (2021), pp. 82146–82168.

[63] Priya Goyal et al. “Self-supervised pretraining of visual features in the wild”. In: *arXiv preprint arXiv:2103.01988* (2021).

[64] Zhi-Hua Zhou. “A brief introduction to weakly supervised learning”. In: *National science review* 5.1 (2018), pp. 44–53.

[65] Longlong Jing and Yingli Tian. “Self-supervised visual feature learning with deep neural networks: A survey”. In: *IEEE transactions on pattern analysis and machine intelligence* 43.11 (2020), pp. 4037–4058.

[66] Xiang Wu et al. “A light CNN for deep face representation with noisy labels”. In: *IEEE Transactions on Information Forensics and Security* 13.11 (2018), pp. 2884–2896.

[67] Peng Tang et al. “Weakly supervised region proposal network and object detection”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 352–368.

[68] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[69] Vinod Nair and Geoffrey E Hinton. “Rectified linear units improve restricted boltzmann machines”. In: *Icml*. 2010.

[70] Xue Ying. “An overview of overfitting and its solutions”. In: *Journal of physics: Conference series*. Vol. 1168. 2. IOP Publishing. 2019, p. 022022.

[71] Nitish Srivastava et al. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[72] Connor Shorten and Taghi M Khoshgoftaar. “A survey on image data augmentation for deep learning”. In: *Journal of big data* 6.1 (2019), pp. 1–48.

[73] Fuzhen Zhuang et al. “A comprehensive survey on transfer learning”. In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.

[74] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

[75] Chuanqi Tan et al. “A survey on deep transfer learning”. In: *International conference on artificial neural networks*. Springer. 2018, pp. 270–279.

[76] Estibaliz Gómez-de-Mariscal et al. “Deep-learning-based segmentation of small extracellular vesicles in transmission electron microscopy images”. In: *Scientific reports* 9.1 (2019), pp. 1–10.

[77] William Jones et al. “Computational biology: deep learning”. In: *Emerging Topics in Life Sciences* 1.3 (2017), pp. 257–274.

[78] Erick Moen et al. “Deep learning for cellular image analysis”. In: *Nature methods* 16.12 (2019), pp. 1233–1246.

[79] Anant Madabhushi and George Lee. “Image analysis and machine learning in digital pathology: Challenges and opportunities”. In: *Medical image analysis* 33 (2016), pp. 170–175.

[80] David A Van Valen et al. “Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments”. In: *PLoS computational biology* 12.11 (2016), e1005177.

[81] Lipeng Xie et al. “Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images”. In: *Neurocomputing* 376 (2020), pp. 166–179.

[82] Sudhir Sornapudi et al. “Deep learning nuclei detection in digitized histology images by superpixels”. In: *Journal of pathology informatics* 9.1 (2018), p. 5.

[83] Yousef Al-Kofahi et al. “A deep learning-based algorithm for 2-D cell segmentation in microscopy images”. In: *BMC bioinformatics* 19.1 (2018), pp. 1–11.

[84] Korsuk Sirinukunwattana et al. “Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images”. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1196–1206.

[85] Peter Naylor et al. “Nuclei segmentation in histopathology images using deep neural networks”. In: *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE. 2017, pp. 933–936.

[86] Aarno Oskar Vuola, Saad Ullah Akram, and Juho Kannala. “Mask-RCNN and U-net ensembled for nuclei segmentation”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 208–212.

[87] Reka Hollandi et al. “nucleAIzer: a parameter-free deep learning framework for nucleus segmentation using image style transfer”. In: *Cell Systems* 10.5 (2020), pp. 453–458.

[88] Florian Kromp et al. “Deep Learning architectures for generalized immunofluorescence based nuclear image segmentation”. In: *arXiv preprint arXiv:1907.12975* (2019).

[89] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[90] Zongwei Zhou et al. “Unet++: A nested u-net architecture for medical image segmentation”. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[91] Thorsten Falk et al. “U-Net: deep learning for cell counting, detection, and morphometry”. In: *Nature methods* 16.1 (2019), pp. 67–70.

[92] Sajith Kecheril Sadanandan et al. “Automated training of deep convolutional neural networks for cell segmentation”. In: *Scientific Reports* 7 (2017).

[93] Simon Jégou et al. “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 11–19.

[94] Liang-Chieh Chen et al. “Rethinking atrous convolution for semantic image segmentation”. In: *arXiv preprint arXiv:1706.05587* (2017).

[95] Liang-Chieh Chen et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.

[96] *Harmony High-Content Imaging and Analysis Software*. `https://www.perkinelmer.com/product/harmony-4-8-office-hh17000001`. Accessed: 2022-11-23.

[97] Gao Huang et al. “Densely connected convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.

[98] Michal Drozdzal et al. “The importance of skip connections in biomedical image segmentation”. In: *Deep learning and data labeling for medical applications*. Springer, 2016, pp. 179–187.

[99] Hengshuang Zhao et al. “Pyramid scene parsing network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2881–2890.

[100] Douglas M Hawkins. *Identification of outliers*. Vol. 11. Springer, 1980.

[101] Lukas Ruff et al. “A unifying review of deep and shallow anomaly detection”. In: *Proceedings of the IEEE* 109.5 (2021), pp. 756–795.

[102] Faruk Ahmed and Aaron Courville. “Detecting semantic anomalies”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 04. 2020, pp. 3154–3162.

[103] Shu Chen et al. “Avoiding artefacts during electron microscopy of silver nanomaterials exposed to biological environments”. In: *Journal of microscopy* 261.2 (2016), pp. 157–166.

[104] Donna R Whelan and Toby DM Bell. “Image artifacts in single molecule localization microscopy: why optimization of sample preparation protocols matters”. In: *Scientific reports* 5.1 (2015), pp. 1–10.

[105] Guansong Pang et al. “Deep learning for anomaly detection: A review”. In: *ACM Computing Surveys (CSUR)* 54.2 (2021), pp. 1–38.

[106] Mingxing Tan and Quoc Le. “Efficientnetv2: Smaller models and faster training”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 10096–10106.

[107] Alessandro Lanza, Federico Tombari, and Luigi Di Stefano. “Accurate and efficient background subtraction by monotonic second-degree polynomial fitting”. In: *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE. 2010, pp. 376–383.

[108] Paul Bergmann et al. “Improving unsupervised defect segmentation by applying structural similarity to autoencoders”. In: *arXiv preprint arXiv:1807.02011* (2018).

[109] Jihun Yi and Sungroh Yoon. “Patch svdd: Patch-level svdd for anomaly detection and segmentation”. In: *Proceedings of the Asian Conference on Computer Vision*. 2020.

[110] Thomas Defard et al. “Padim: a patch distribution modeling framework for anomaly detection and localization”. In: *International Conference on Pattern Recognition*. Springer. 2021, pp. 475–489.

[111] Jongrye Jeon et al. “A subpopulation of neuronal M4 muscarinic acetylcholine receptors plays a critical role in modulating dopamine-dependent behaviors”. In: *Journal of Neuroscience* 30.6 (2010), pp. 2396–2405.

[112] Otto-Erich Brodde et al. “Presence, distribution and physiological function of adrenergic and muscarinic receptor subtypes in the human heart”. In: *Basic research in cardiology* 96.6 (2001), pp. 528–538.

[113] Tatsuya Haga. “Molecular properties of muscarinic acetylcholine receptors”. In: *Proceedings of the Japan Academy, Series B* 89.6 (2013), pp. 226–256.

[114] WY Chan et al. “Allosteric modulation of the muscarinic M4 receptor as an approach to treating schizophrenia”. In: *Proceedings of the National Academy of Sciences* 105.31 (2008), pp. 10978–10983.

[115] Christian C Felder et al. “Current status of muscarinic M1 and M4 receptors as drug targets for neurodegenerative diseases”. In: *Neuropharmacology* 136 (2018), pp. 449–458.

[116] Leigh C Walker et al. “Muscarinic M4 and M5 receptors in the ventral subiculum differentially modulate alcohol seeking versus consumption in male alcohol-preferring rats”. In: *British Journal of Pharmacology* 178.18 (2021), pp. 3730–3746.

[117] Leigh C Walker et al. “Acetylcholine muscarinic M4 receptors as a therapeutic target for alcohol use disorder: converging evidence from humans and rodents”. In: *Biological Psychiatry* 88.12 (2020), pp. 898–909.

[118] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[119] Guido Forni and Alberto Mantovani. “COVID-19 vaccines: where we stand and challenges ahead”. In: *Cell Death & Differentiation* 28.2 (2021), pp. 626–639.

[120] Michael Boutros, Florian Heigwer, and Christina Laufer. “Microscopy-based high-content screening”. In: *Cell* 163.6 (2015), pp. 1314–1325.

[121] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. PMLR. 2015, pp. 448–456.

[122] Deepak Pathak et al. “Context encoders: Feature learning by inpainting”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2536–2544.

[123] Zhirong Wu et al. “Unsupervised feature learning via non-parametric instance discrimination”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 3733–3742.

[124] Mathilde Caron et al. “Unsupervised learning of visual features by contrasting cluster assignments”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 9912–9924.

[125] Mathilde Caron et al. “Deep clustering for unsupervised learning of visual features”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 132–149.

[126] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. “" Why should i trust you?" Explaining the predictions of any classifier”. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1135–1144.

[127] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

[128] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. “A multidisciplinary survey and framework for design and evaluation of explainable AI systems”. In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11.3-4 (2021), pp. 1–45.

[129] Alexey Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020).

[130] Jingyun Liang et al. “Swinir: Image restoration using swin transformer”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 1833–1844.

[131] Salman Khan et al. “Transformers in vision: A survey”. In: *ACM Computing Surveys (CSUR)* (2021).

[132] Alana de Santana Correia and Esther Luna Colombini. “Attention, please! A survey of neural attention models in deep learning”. In: *Artificial Intelligence Review* (2022), pp. 1–88.

[133] Hila Chefer, Shir Gur, and Lior Wolf. “Transformer interpretability beyond attention visualization”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 782–791.

[134] Daniel M Rosenbaum, Søren GF Rasmussen, and Brian K Kobilka. “The structure and function of G-protein-coupled receptors”. In: *Nature* 459.7245 (2009), pp. 356–363.

[135] Edgar Jacoby et al. “The 7 TM G-protein-coupled receptor target family”. In: *ChemMedChem: Chemistry Enabling Drug Discovery* 1.8 (2006), pp. 760–782.

[136] Alexander Kamb, Susan Wee, and Christoph Lengauer. “Why is cancer drug discovery so difficult?” In: *Nature reviews Drug discovery* 6.2 (2007), pp. 115–120.

[137] Jung AB. *Imgaug*. URL: `https://github.com/aleju` (visited on 10/17/2022).

[138] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[139] Qun Liu and Supratik Mukhopadhyay. “Unsupervised learning using pretrained CNN and associative memory bank”. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2018, pp. 01–08.

ACKNOWLEDGEMENTS

I would like to begin by expressing my deepest gratitude to the almighty Allah for guiding me through this journey and granting me the strength and perseverance to complete my Ph.D. thesis.

I am incredibly thankful to my supervisors, Leopold Parts and Dmytro Fishman, for their invaluable guidance, support, and encouragement throughout the research process. Their expertise, insights, and constructive criticism have been instrumental in shaping my work and helping me grow professionally and personally. Special thanks and gratitude are to Jaak Vilo, who helped me find such amazing supervisors. I am also thankful to Ardi Tampuu for taking the time to read my thesis and providing insightful comments that significantly improved it. I extend my sincere appreciation to Carolina Wählby and Mohammed Al-masni for their dedicated efforts in reviewing my thesis and offering valuable feedback. Their constructive comments and suggestions have played a crucial role in enhancing the quality of this work.

I would also like to extend my heartfelt appreciation to my mother, father, wife, son, and all the extended family, who have been my rock throughout this journey. Their unwavering support and love have kept me going through challenging times, and I cannot thank them enough for everything they have done for me.

I would not forget to thank my friends and colleagues for their constant encouragement and support, which have been a source of strength and motivation for me.

I acknowledge the use of Large Language Models (LLMs) for some text rephrasing during the final writing phase of this thesis.

Finally, I would like to thank the University of Tartu, Institute of Computer Sciences staff for their assistance and support throughout my Ph.D. studies. The institute provided me with a supportive and inspiring research environment.

SISUKOKKUVÕTE

Sügavõppemeetodid rakumikroskoopia kujutise analüüsiks

Rakk on kõigi elusorganismide peamine ehitusüksus ja peaaegu kõigi bioloogiliste protsesside toimumiskoht. Rakud toodavad valke ja energiat, transpordivad materjale ja hävitavad jäätmeid. Nad saavad end mitoosi kaudu paljundada ja suhelda signaaliradade kaudu oma keskkonnaga. Ilma rakkudeta ei oleks elu olemas. Rakkude toimimise mõistmine võimaldab inimestel bioloogiliste protsesside, haiguste ja ravimite kohta rohkem teada saada.

Rakud on väga väikesed. Tavalise imetaja rakudiameeter on umbes $10-20\mu m$, mis on ligikaudu viiendik väikseimast objektist, mida inimene palja silmaga näha on võimeline. Kuna inimese silmal puudub võime näha sellise suurusega objekte, rääkimata nende põhjalikust uurimisest, kasutavad teadlased erivarustust – mikroskoopi. *Mikroskoop* on seade, mis suurendab väikseid objekte, mida muidu oleks võimatu uurida. Antud doktoritöös analüüsime pilte, mis on saadud laialdaselt kasutatava *valgusmikroskoobiga* - see on instrument, mis kasutab ära valgust ja selle omadusi, et suurendada mikroskoopilisi objekte.

Tänu mikroskoopia arengule on seda valdkonda suudetud viimasel ajal ka märkimisväärselt automatiseerida. Mikroskoopia andmete hulk on hoogsalt suurenenud, sest nüüd saab ühes eksperimendis automatiseeritud korras luua miljoneid pilte [1, 2, 3]. Selliste tohutute pildikogumite uurimiseks on hädavajalikud automatiseeritud ja täpsed analüüsiprotseduurid.

Pildianalüüsi tehnikad on pidevalt arenenud. Eriti tõhusaks on osutunud masinõppe algoritmid ja seetõttu kasutatakse neid laialdaselt ka selles valdkonnas. Paljud algoritmid on kergesti kättesaadavad erinevate bioloogiliste piltide analüüsiks mõeldud tööriistade [4] kaudu - ImageJ/Fiji [5], CellProfiler [6], ja Ilastik [7]. Hoolimata selliste tööriistade laialdasest kasutuselevõtust, ei pruugi need vahendid alati anda piisavalt täpseid tulemusi, kuna traditsiooniliste masinõppe algoritmide kasutamine eeldab nii insener-tehnilisi oskusi kui ka valdkonnateadmisi [8]. Selliste lähenemisviiside stabiilsust ja usaldusväärsust mõjutavad ka signaali kvaliteedi varieeruvus ning erinevused pildistamismeetodites, mis on iseloomulikud kõrge läbilaskevõimega rakumikroskoopiale [9, 10].

Erinevalt traditsioonilistest masinõppemeetoditest, mis nõuavad hoolikat sisendandmete eeltöötlust, eraldavad sügavõppe meetodid automaatselt asjakohaseid mustreid toorandmest, kasutades selleks mitmekihilisi arvutuslikke mudeleid. Üheks sügavõppe meetodite liigiks olevas konvolutsioonilises närvivõrgus liigutatakse filtreid üle pildi, et tuvastada pildi peamisi tunnuseid ja mustreid. Konvolutsioonilised närvivõrgud saavutavad tipptasemel tulemusi ülesannetes nagu pildi klassifitseerimine [11, 12], objekti tuvastamine [13, 14] ja segmenteerimine [15, 16]. Sügavõppe kiire areng on pidevalt loonud uusi teadmisi, mida saaks mikroskoopiaga seotud pildianalüüsis ära kasutada, kuid mida pole sellel eesmärgil veel täielikult rakendatud.

Doktoritöö algas süvaõppe uusimate meetodite rakendamisega rakutuumade segmenteerimiseks, mis on sageli üks esimesi samme raku mikroskoopiapiltide töötlemise protsessides. Täpne rakutuumade segmenteerimine on ülioluline paljude bioloogiliste rakenduste kontekstis. Näiteks on rakutuuma struktuuri ja morfoloogia kõrvalekalded sageli seotud haigustega nagu vähk. Seega aitab tuumade segmenteerimine ja nende omaduste uurimine kaasa vähidiagnoosimisele ja haiguse kulu jälgimisele [17]. Tuumade tuvastamist rakendatakse ka rakkude jälitamiseks, mille abil on omakorda võimalik uurida rakusüsteemide käitumist ning selle muutumist erinevate ravimite toimel [18]. Peamiseks fookuseks olid heleväljа mikroskoobipildid, kuna neid on suhteliselt lihtne toota, kuid raske kontrollida ja analüüsida. Me katsetasime tuumade segmenteerimiseks mitmeid tipptasemel mudeliarhitektuure ja lõime ka uue arhitektuuri, *PPU-Net* [19]. Hinnatud mudelid saavutasid varieeruvaid tulemusi tuumade segmenteerimisel heleväljapiltidelt - PPU-Net saavutas aga tollel hetkel olemasolevate tipptaseme mudelitega võrdseid tulemusi, kasutades selleks 20 korda vähem treenitavaid parameetreid, mis muutis selle konkurentidest kergemaks ja vähem keerukamaks. Samuti uurisime mudelite ebastabiilsete tulemuste põhjuseid erinevatel rakutüüpidel ja üksikutel piltidel, vajalike treeningpiltide arvu ja kõige sagedamaid vigade allikaid.

PPU-Net segmenteerimisvigade põhjuste uurimisel leidsime, et anomaaliate olemasolu piltidel (signaal, mis ei kajasta oodatut) on suur veaallikas. Seega taht-sime paremini mõista anomaaliate probleemi. Uurisime erinevaid andmekogumeid ja leidsime, et anomaaliad esinevad erinevates kujudes ja suurustes ning need võivad oluliselt moonutada järgnevaid analüüse. Nende mõju leevendamise eesmärgil lõime raamistiku, mis määratleb ja eemaldab anomaaliad minimaalse inimtööga, tehes ainult pilditasemel märgendeid, mitte töömahukaimaid pikslitasemel märgendeid [20]. Kuna anomaaliad on oma olemuselt keerulised ja nende märgendamine pikslitasemel on töömahukas ja aeganõudev, siis pakkusime välja ainult pilditasemel märgendite kasutamise selle ülesande jaoks. Esiteks pakkusime välja ScoreCAMi nimelise meetodi kasutamise [21], mille eesmärk on tõlgendada sügavõppe pildiklassifitseerimise algoritme, tuues esile pildi osad või tunnused, mille põhjal langetab sügavõppe meetod oma otsuse. Kui sügavõppe meetod klassifitseerib, kas pildil esineb anomaaliaid või mitte, siis hüpoteesi kohaselt on sügavõppe mudeli otsuse tegemisel kõige mõjukamaks pildi osaks või tunnuseks just anomaalia - see hüpotees tõestati hiljem empiiriliselt. ScoreCAMi väljundit kasutati hiljem pseudo-märgenditena segmenteerimismudeli treenimiseks [20]. Sel moel ühendame pikslitasemel segmenteerimise kvaliteedi ja pilditasemel märgendite hankimise mugavuse. Nimetasime oma väljapakutud raamistiku *ScoreCAM-U-Netiks* ning näeme ette, et tulevikus muutub soovimatute objektide eemaldamine tõenäoliselt kõigi suuremahuliste mikroskoopiaeksperimentide standardprotsessi osaks.

Lõpuks rakendasime omandatud teadmisi pärismaailma kontekstis. Uurisime sügavõppe meetodite kasutamise väärtust anomaaliate eemaldamiseks ja segmenteerimiseks ravimite avastamise uurimistöös. Selleks tegime koostööd keemikute

ja bioloogidega, kes uurivad ühte silmapaistvamat rakumembraani retseptorit - M4. Vaatamata selle kasvavale tähtsusele, on osutunud keerukaks luua uusi ravimeid, mis oleksid suunatud sellele retseptorile [22]. Meie koostööpartnerid kasutasid kõrge afiinsusega fluorestseeruvaid ligande, et uurida sidumisinteraktsiooni M4 retseptoriga. Selleks on vaja eraldada rakud ja uurida nendes olevat fluorestsentssignaali, mille tugevus sõltub valgu ja ligandi vahelise afiinsuse tugevusest ning ligandi kontsentratsioonist. Fluorestseerivate ligandide poolt rakkudes tekitatud signaal ei osutunud piisavaks, et mudel saanuks eristada rakku taustast. Seega otsustasime segmenteerida rakke keerukamast, kuid fluorestseerivusest sõltumatutest heleväljapiltidest. Esiteks kasutasime sügavõpet rakukehade segmenteerimiseks heleväljapiltidelt. Järgmisena analüüsisime fluorestseeruvat signaali rakkudest, mis eraldati vastavatelt fluorestseeruvatelt piltidelt segmenteerimistulemustest saadud rakukoordinaatide abil. Seejärel uurisime anomaaliate eemaldamise mõju heleväljapiltide signaalile, mis tulenes retseptori-ligandi interaktsioonist. Näitasime, et anomaaliate eemaldamine muutis signaali rohkem erapooletutuks ja et meie loodud mudeli kasutamine nende eemaldamiseks andis peaaegu optimaalse tulemuse.

PUBLICATIONS

CURRICULUM VITAE

Personal data

Name:	Mohammed Abdulhameed Shaif Ali
Date of birth:	01/01/1987
Citizenship:	Yemeni
Contact:	moh.alsamei@gmail.com

Education

2019–2024	Ph.D. in Computer Scince. University of Tartu, Estonia.
2013–2016	M.Sc. in Information Technology. Cairo University, Egypt.
2007–2011	B.Sc. in Information Technology. Cairo University, Egypt.

Employment

2023 – present	Computer vision scientist, Syngenta, United Kingdom
2021–2023	Junior Research Fellow of Artificial Intelligence, Institute of computer science, University of Tartu, Tartu, Estonia
2012–2012	Oracle Developer, Hayel Saeed Anam & co, Taiz, Yemen

Teaching

Teaching assistant for Machine Learning course at University of Tartu, Estonia in fall 2022
Teaching assistant for Machine Learning course at University of Tartu, Estonia in fall 2021
Teaching assistant for Machine Learning course at University of Tartu, Estonia in fall 2020
Teaching assistant for Computer Graphics course at Alhikma University, Yemen in fall 2012

Supervised theses

Kaspar Hollo (MSc): Exploring the Value of Weakly-Supervised Deep Learning Approaches for Artefact Segmentation in Brightfield Microscopy Images.
Danver Värv (BSc): Mother-bud detection and classification in yeast cells

Scientific work

Main fields of interest:

- Health Image Processing
- Machine Learning and Deep Learning

ELULOOKIRJELDUS

Isikuandmed

Nimi:	Mohammed Abdulhameed Shaif Ali
Sünniaeg:	01/01/1987
Kodakondsus:	Jeemeni
Kontakt:	moh.alsamei@gmail.com

Haridus

2019 – 2024	Informaatika doktorikraad. Tartu Ülikool, Eesti.
2013 – 2016	Infotehnoloogia magistrikraad. Kairo Ülikool, Egiptus.
2007 – 2011	Infotehnoloogia bakalaureusekraad. Kairo Ülikool, Egiptus.

Töökogemus

2023 – praegune	arvutinägemisteadlane, Syngenta, Ühendkuningriik
2021 – 2023	Tehisintellekti nooremteadur, arvutiteaduse instituut, Tartu Ülikool, Tartu, Eesti
2012 – 2012	Oracle'i arendaja, Hayel Saeed Anam & co, Taiz, Jeemen

Õpetamine

Masinõppe kursuse õpetamise assistent Tartu Ülikoolis Eestis sügisel 2022
Masinõppe kursuse õpetamise assistent Tartu Ülikoolis Eestis sügisel 2021
Masinõppe kursuse õpetamise assistent Tartu Ülikoolis Eestis sügisel 2020
Arvutigraafika kursuse õpetamise assistent Alhikma ülikoolis Jeemenis sügisel 2012

Juhendatud lõputööd

Kaspar Hollo (MSc): Nõrgalt juhendatud süvaõppe mudelite tõhusus helevälja helevälja mikroskoopiapiltidelt anomaaliate segmenteerimisel.
Danver Värv (BSc): Ema-tütarrakkude tuvastamine ja klassifitseerimine pärmirakkudes.

Teadustöö

Peamised huvivaldkonnad:

- Tervise pilditöötlus
- Masinõpe ja sügavõpe

DISSERTATIONES INFORMATICAE PREVIOUSLY PUBLISHED IN DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS

19. **Helger Lipmaa.** Secure and efficient time-stamping systems. Tartu, 1999, 56 p.
22. **Kaili Müürisep.** Eesti keele arvutigrammatika: süntaks. Tartu, 2000, 107 lk.
23. **Varmo Vene.** Categorical programming with inductive and coinductive types. Tartu, 2000, 116 p.
24. **Olga Sokratova.** Ω-rings, their flat and projective acts with some applications. Tartu, 2000, 120 p.
27. **Tiina Puolakainen.** Eesti keele arvutigrammatika: morfoloogiline ühestamine. Tartu, 2001, 138 lk.
29. **Jan Villemson.** Size-efficient interval time stamps. Tartu, 2002, 82 p.
45. **Kristo Heero.** Path planning and learning strategies for mobile robots in dynamic partially unknown environments. Tartu 2006, 123 p.
49. **Härmel Nestra.** Iteratively defined transfinite trace semantics and program slicing with respect to them. Tartu 2006, 116 p.
53. **Marina Issakova.** Solving of linear equations, linear inequalities and systems of linear equations in interactive learning environment. Tartu 2007, 170 p.
55. **Kaarel Kaljurand.** Attempto controlled English as a Semantic Web language. Tartu 2007, 162 p.
56. **Mart Anton.** Mechanical modeling of IPMC actuators at large deformations. Tartu 2008, 123 p.
59. **Reimo Palm.** Numerical Comparison of Regularization Algorithms for Solving Ill-Posed Problems. Tartu 2010, 105 p.
61. **Jüri Reimand.** Functional analysis of gene lists, networks and regulatory systems. Tartu 2010, 153 p.
62. **Ahti Peder.** Superpositional Graphs and Finding the Description of Structure by Counting Method. Tartu 2010, 87 p.
64. **Vesal Vojdani.** Static Data Race Analysis of Heap-Manipulating C Programs. Tartu 2010, 137 p.
66. **Mark Fišel.** Optimizing Statistical Machine Translation via Input Modification. Tartu 2011, 104 p.
67. **Margus Niitsoo**. Black-box Oracle Separation Techniques with Applications in Time-stamping. Tartu 2011, 174 p.
71. **Siim Karus.** Maintainability of XML Transformations. Tartu 2011, 142 p.
72. **Margus Treumuth.** A Framework for Asynchronous Dialogue Systems: Concepts, Issues and Design Aspects. Tartu 2011, 95 p.
73. **Dmitri Lepp.** Solving simplification problems in the domain of exponents, monomials and polynomials in interactive learning environment T-algebra. Tartu 2011, 202 p.

74. **Meelis Kull.** Statistical enrichment analysis in algorithms for studying gene regulation. Tartu 2011, 151 p.
77. **Bingsheng Zhang.** Efficient cryptographic protocols for secure and private remote databases. Tartu 2011, 206 p.
78. **Reina Uba.** Merging business process models. Tartu 2011, 166 p.
79. **Uuno Puus.** Structural performance as a success factor in software development projects – Estonian experience. Tartu 2012, 106 p.
81. **Georg Singer.** Web search engines and complex information needs. Tartu 2012, 218 p.
83. **Dan Bogdanov.** Sharemind: programmable secure computations with practical applications. Tartu 2013, 191 p.
84. **Jevgeni Kabanov.** Towards a more productive Java EE ecosystem. Tartu 2013, 151 p.
87. **Margus Freudenthal.** Simpl: A toolkit for Domain-Specific Language development in enterprise information systems. Tartu, 2013, 151 p.
90. **Raivo Kolde.** Methods for re-using public gene expression data. Tartu, 2014, 121 p.
91. **Vladimir Šor.** Statistical Approach for Memory Leak Detection in Java Applications. Tartu, 2014, 155 p.
92. **Naved Ahmed.** Deriving Security Requirements from Business Process Models. Tartu, 2014, 171 p.
94. **Liina Kamm.** Privacy-preserving statistical analysis using secure multi-party computation. Tartu, 2015, 201 p.
100. **Abel Armas Cervantes.** Diagnosing Behavioral Differences between Business Process Models. Tartu, 2015, 193 p.
101. **Fredrik Milani.** On Sub-Processes, Process Variation and their Interplay: An Integrated Divide-and-Conquer Method for Modeling Business Processes with Variation. Tartu, 2015, 164 p.
102. **Huber Raul Flores Macario.** Service-Oriented and Evidence-aware Mobile Cloud Computing. Tartu, 2015, 163 p.
103. **Tauno Metsalu.** Statistical analysis of multivariate data in bioinformatics. Tartu, 2016, 197 p.
104. **Riivo Talviste.** Applying Secure Multi-party Computation in Practice. Tartu, 2016, 144 p.
108. **Siim Orasmaa.** Explorations of the Problem of Broad-coverage and General Domain Event Analysis: The Estonian Experience. Tartu, 2016, 186 p.
109. **Prastudy Mungkas Fauzi.** Efficient Non-interactive Zero-knowledge Protocols in the CRS Model. Tartu, 2017, 193 p.
110. **Pelle Jakovits.** Adapting Scientific Computing Algorithms to Distributed Computing Frameworks. Tartu, 2017, 168 p.
111. **Anna Leontjeva.** Using Generative Models to Combine Static and Sequential Features for Classification. Tartu, 2017, 167 p.
112. **Mozhgan Pourmoradnasseri.** Some Problems Related to Extensions of Polytopes. Tartu, 2017, 168 p.

113. **Jaak Randmets.** Programming Languages for Secure Multi-party Computation Application Development. Tartu, 2017, 172 p.
114. **Alisa Pankova.** Efficient Multiparty Computation Secure against Covert and Active Adversaries. Tartu, 2017, 316 p.
116. **Toomas Saarsen.** On the Structure and Use of Process Models and Their Interplay. Tartu, 2017, 123 p.
121. **Kristjan Korjus.** Analyzing EEG Data and Improving Data Partitioning for Machine Learning Algorithms. Tartu, 2017, 106 p.
122. **Eno Tõnisson.** Differences between Expected Answers and the Answers Offered by Computer Algebra Systems to School Mathematics Equations. Tartu, 2017, 195 p.

DISSERTATIONES INFORMATICAE UNIVERSITATIS TARTUENSIS

1. **Abdullah Makkeh**. Applications of Optimization in Some Complex Systems. Tartu 2018, 179 p.
2. **Riivo Kikas**. Analysis of Issue and Dependency Management in Open-Source Software Projects. Tartu 2018, 115 p.
3. **Ehsan Ebrahimi**. Post-Quantum Security in the Presence of Superposition Queries. Tartu 2018, 200 p.
4. **Ilya Verenich**. Explainable Predictive Monitoring of Temporal Measures of Business Processes. Tartu 2019, 151 p.
5. **Yauhen Yakimenka**. Failure Structures of Message-Passing Algorithms in Erasure Decoding and Compressed Sensing. Tartu 2019, 134 p.
6. **Irene Teinemaa**. Predictive and Prescriptive Monitoring of Business Process Outcomes. Tartu 2019, 196 p.
7. **Mohan Liyanage.** A Framework for Mobile Web of Things. Tartu 2019, 131 p.
8. **Toomas Krips.** Improving performance of secure real-number operations. Tartu 2019, 146 p.
9. **Vijayachitra Modhukur.** Profiling of DNA methylation patterns as biomarkers of human disease. Tartu 2019, 134 p.
10. **Elena Sügis.** Integration Methods for Heterogeneous Biological Data. Tartu 2019, 250 p.
11. **Tõnis Tasa.** Bioinformatics Approaches in Personalised Pharmacotherapy. Tartu 2019, 150 p.
12. **Sulev Reisberg.** Developing Computational Solutions for Personalized Medicine. Tartu 2019, 126 p.
13. **Huishi Yin.** Using a Kano-like Model to Facilitate Open Innovation in Requirements Engineering. Tartu 2019, 129 p.
14. **Faiz Ali Shah.** Extracting Information from App Reviews to Facilitate Software Development Activities. Tartu 2020, 149 p.
15. **Adriano Augusto**. Accurate and Efficient Discovery of Process Models from Event Logs. Tartu 2020, 194 p.
16. **Karim Baghery.** Reducing Trust and Improving Security in zk-SNARKs and Commitments. Tartu 2020, 245 p.
17. **Behzad Abdolmaleki.** On Succinct Non-Interactive Zero-Knowledge Protocols Under Weaker Trust Assumptions. Tartu 2020, 209 p.
18. **Janno Siim.** Non-Interactive Shuffle Arguments. Tartu 2020, 154 p.
19. **Ilya Kuzovkin.** Understanding Information Processing in Human Brain by Interpreting Machine Learning Models. Tartu 2020, 149 p.
20. **Orlenys López Pintado.** Collaborative Business Process Execution on the Blockchain: The Caterpillar System. Tartu 2020, 170 p.
21. **Ardi Tampuu.** Neural Networks for Analyzing Biological Data. Tartu 2020, 152 p.

22. **Madis Vasser.** Testing a Computational Theory of Brain Functioning with Virtual Reality. Tartu 2020, 106 p.
23. **Ljubov Jaanuska.** Haar Wavelet Method for Vibration Analysis of Beams and Parameter Quantification. Tartu 2021, 192 p.
24. **Arnis Parsovs.** Estonian Electronic Identity Card and its Security Challenges. Tartu 2021, 214 p.
25. **Kaido Lepik.** Inferring causality between transcriptome and complex traits. Tartu 2021, 224 p.
26. **Tauno Palts.** A Model for Assessing Computational Thinking Skills. Tartu 2021, 134 p.
27. **Liis Kolberg.** Developing and applying bioinformatics tools for gene expression data interpretation. Tartu 2021, 195 p.
28. **Dmytro Fishman.** Developing a data analysis pipeline for automated protein profiling in immunology. Tartu 2021, 155 p.
29. **Ivo Kubjas.** Algebraic Approaches to Problems Arising in Decentralized Systems. Tartu 2021, 120 p.
30. **Hina Anwar.** Towards Greener Software Engineering Using Software Analytics. Tartu 2021, 186 p.
31. **Veronika Plotnikova.** FIN-DM: A Data Mining Process for the Financial Services. Tartu 2021, 197 p.
32. **Manuel Camargo.** Automated Discovery of Business Process Simulation Models From Event Logs: A Hybrid Process Mining and Deep Learning Approach. Tartu 2021, 130 p.
33. **Volodymyr Leno.** Robotic Process Mining: Accelerating the Adoption of Robotic Process Automation. Tartu 2021, 119 p.
34. **Kristjan Krips.** Privacy and Coercion-Resistance in Voting. Tartu 2022, 173 p.
35. **Elizaveta Yankovskaya.** Quality Estimation through Attention. Tartu 2022, 115 p.
36. **Mubashar Iqbal.** Reference Framework for Managing Security Risks Using Blockchain. Tartu 2022, 203 p.
37. **Jakob Mass.** Process Management for Internet of Mobile Things. Tartu 2022, 151 p.
38. **Gamal Elkoumy.** Privacy-Enhancing Technologies for Business Process Mining. Tartu 2022, 135 p.
39. **Lidia Feklistova.** Learners of an Introductory Programming MOOC: Background Variables, Engagement Patterns and Performance. Tartu 2022, 151 p.
40. **Mohamed Ragab.** Bench-Ranking: A Prescriptive Analysis Approach for Large Knowledge Graphs Query Workloads. Tartu 2022, 158 p.
41. **Mohammad Anagreh.** Privacy-Preserving Parallel Computations for Graph Problems. Tartu 2023, 181 p.
42. **Rahul Goel.** Mining Social Well-being Using Mobile Data. Tartu 2023, 104 p.

43. **Anti Ingel.** Algorithms using information theory: classification in brain-computer interfaces and characterising reinforcement-learning agents. Tartu 2023, 142 p.
44. **Shakshi Sharma.** Fighting Misinformation in the Digital Age: A Comprehensive Strategy for Characterizing, Identifying, and Mitigating Misinformation on Online Social Media Platforms. Tartu 2023, 158 p.
45. **Kristiina Rahkema.** Quality Analysis of iOS Applications with Focus on Maintainability and Security Aspects. Tartu 2023, 182 p.
46. **Ivan Slobozhan.** Studying Online Social Media Engagement in CIS Countries during Protests, Mass Demonstrations and War. Tartu 2023, 81 p.
47. **Nurlan Kerimov.** Building a catalogue of molecular quantitative trait loci to interpret complex trait associations. Tartu 2023, 248 p.
48. **Pavlo Tertychnyi.** Machine Learning Methods for Anti-Money Laundering Monitoring. Tartu 2023, 117 p.
49. **Abasi-amefon Obot Affia.** A Framework and Teaching Approach for IoT Security Risk Management. Tartu 2023, 180 p.
50. **Raimond-Hendrik Tunnel.** Video Game Design and Development Bachelor's Curriculum for Estonia. Tartu 2024, 137 p.
51. **Ahto Salumets.** Bioinformatics analysis of various aspects in immunology. Tartu 2024, 198 p.