Saad Ullah Akram

CELL SEGMENTATION AND TRACKING VIA PROPOSAL GENERATION AND SELECTION
SAAD ULLAH AKRAM

CELL SEGMENTATION AND TRACKING VIA PROPOSAL GENERATION AND SELECTION

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in the Arina auditorium (TA105), Linnanmaa, on 30 November 2017, at 12 noon

UNIVERSITY OF OULU, OULU 2017
Abstract

Biology and medicine rely heavily on images to understand how the body functions, for diagnosing diseases and to test the effects of treatments. In recent decades, microscopy has experienced rapid improvements, enabling imaging of fixed and living cells at higher resolutions and frame rates, and deeper inside the biological samples. This has led to rapid growth in the image data. Automated methods are needed to quantitatively analyze these huge datasets and find statistically valid patterns. Cell segmentation and tracking is critical for automated analysis, yet it is a challenging problem due to large variations in cell shapes and appearances caused by various factors, including cell type, sample preparation and imaging setup.

This thesis proposes novel methods for segmentation and tracking of cells, which rely on machine learning based approaches to improve the performance, generalization and reusability of automated methods. Cell proposals are used to efficiently exploit spatial and temporal context for resolving detection ambiguities in high-cell-density regions, caused by weak boundaries and deformable shapes of cells. This thesis presents two cell proposal methods: the first method uses multiple blob-like filter banks for detecting candidates for round cells, while the second method, Cell Proposal Network (CPN), uses convolutional neural networks to learn the cell shapes and appearances, and can propose candidates for cells in a wide variety of microscopy images. CPN first regresses cell candidate bounding boxes and their scores, then, it segments the regions inside the top ranked boxes to obtain cell candidate masks. CPN can be used as a general cell detector, as is demonstrated by training a single model to segment images from histology, fluorescence and phase-contrast microscopy.

This work poses segmentation and tracking as proposal selection problems, which are solved optimally using integer linear programming or approximately using iterative shortest cost path search and non-maximum suppression. Additionally, this thesis presents a method which utilizes graph-cuts and an off-the-shelf edge detector to accurately segment highly deformable cells.

The main contribution of this thesis is a cell tracking method which uses CPN to propose cell candidates, represents alternative tracking hypotheses using a graphical model, and selects the globally optimal sub-graph providing cell tracks. It achieves state-of-the-art tracking performance on multiple public benchmark datasets from both phase-contrast and fluorescence microscopy containing cells of various shapes and appearances.

Keywords: biomedical image analysis, cell proposals, cell segmentation, cell tracking, deep learning, joint detection and tracking, microscopy image analysis

Tämä väitöskirja esittää uusia menetelmiä solujen segmentoinnissa ja seurantaan keskittyen koneoppimiseen perustuvaan lähestymistapaan, jotka parantavat automaattisten menetelmien suorituskyvyn ja uudelleenkäytettävyyttä. Spatiaalista ja ajallista kontekstia tehokkaasti hyödynnäviä soluheodotelmiä käytetään ratkaisemaan solujen heikoksi erottuvan parhaan ja joustavien soluista esimerkiksi historiaa tuntemattomia soluita vastaan käytettyä apuna. Tämä soluteknologia on tärkeää ehdottamaan solujehdotelmia, jotta voidaan suunnattavasti analysoimaan suuria solukissoon kuva-aineistuja. CPN regressoi ensin solukandidaatteja ympäröivän suorakaiteen kohokohteen sisällöstä ja niiden pistemääristä, jonka jälkeen se segmentoi alueet parhaiten sijoittuvien suorakaiteiden ympäröimistä apuna. CPN:ää voidaan käyttää yleisenä soluilmastimena erityyppisillä kuvantamistekniikoilla tuotetuissa kuvisissa, joissa käytetyn menetelmän kerrotaan laajentaa automatisointia käyttämällä yleinen soluilmastimena erityyppisillä kuvantamistekniikoilla tuotetuissa kuvisissa.
Acknowledgements

The research work presented in this thesis was conducted at Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, during the years 2013-2017.

First of all, I would like to express my deepest gratitude to my supervisors Prof. Janne Heikkilä and Prof. Juho Kannala for their guidance during my PhD. Their feedback, encouragement and support has been crucial for completing this thesis. They have provided me a lot of freedom, and have enabled me to pursue various research interests, enriching my PhD journey. I am also grateful to my supervisor Prof. Lauri Eklund, who gave valuable feedback on the biological relevance and content of my work.

I would also like to thank pre-examiners Prof. Jan Kybic and Prof. Jussi Tohka for their valuable constructive feedback on the initial draft of this thesis, which has enabled me to improve the final version. I am grateful to Prof. Michal Kozubek and Prof. Jussi Tohka for kindly accepting to act as opponents at my doctoral defense. I acknowledge Prof. Guoying Zhao, Prof. Esa Rahtu, and Dr. Veli-Pekka Ronkainen for their feedback on studies and research during follow-up group meetings.

I would like to acknowledge colleagues at Biocenter Oulu, especially Mika, Audrey, Ehsan, Ulla, Fumi, Veli-Pekka, Ilya, Seppo and Lauri who provided the image data for me to work on, and at CMVS, especially Neslihan, Sami and Hannakaisa for their guidance at the start of my PhD journey. I also want to thank all friends and acquaintances for the unforgettable memories.

I gratefully acknowledge Tekes, UniOGS, and Infotech Oulu and Biocenter Oulu doctoral programs for providing the funding for the work presented in this thesis. ABCData project funded by Tekes initiated my journey in microscopy image analysis field, and Infotech Oulu and Biocenter Oulu funded the last four years of my work.

Finally, I would like to thank my parents and family for their consistent and unconditional support throughout my life.

Oulu, October 2017
Saad Ullah Akram
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>FC</td>
<td>Fully Connected Layer</td>
</tr>
<tr>
<td>FCN</td>
<td>Fully Convolutional Network</td>
</tr>
<tr>
<td>RPN</td>
<td>Region Proposal Network</td>
</tr>
<tr>
<td>CPN</td>
<td>Cell Proposal Network</td>
</tr>
<tr>
<td>SSD</td>
<td>Single Shot MultiBox Detector</td>
</tr>
<tr>
<td>FPN</td>
<td>Feature Pyramid Network</td>
</tr>
<tr>
<td>DSSD</td>
<td>Deconvolutional Single Shot Detector</td>
</tr>
<tr>
<td>MNC</td>
<td>Multi-task Network Cascades</td>
</tr>
<tr>
<td>DCAN</td>
<td>Deep Contour-Aware Network</td>
</tr>
<tr>
<td>CTC</td>
<td>ISBI Cell Tracking Challenge</td>
</tr>
<tr>
<td>NMS</td>
<td>Non-Maximum Suppression</td>
</tr>
<tr>
<td>ILP</td>
<td>Integer Linear Programming</td>
</tr>
<tr>
<td>KSP</td>
<td>k-Shortest Path Algorithm</td>
</tr>
<tr>
<td>BB</td>
<td>Bounding Box</td>
</tr>
<tr>
<td>IoU</td>
<td>Intersection over Union</td>
</tr>
<tr>
<td>MOT</td>
<td>Multiple Object Tracking</td>
</tr>
<tr>
<td>VOT</td>
<td>Visual Object Tracking</td>
</tr>
<tr>
<td>MSER</td>
<td>Maximally Stable Extremal Region</td>
</tr>
<tr>
<td>SE</td>
<td>Structured Edge</td>
</tr>
<tr>
<td>H&amp;E</td>
<td>Hematoxylin and Eosin</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>DIC</td>
<td>Differential-Interference Contrast Microscopy</td>
</tr>
<tr>
<td>PhC</td>
<td>Phase-Contrast Microscopy</td>
</tr>
<tr>
<td>Fluo</td>
<td>Fluorescence Microscopy</td>
</tr>
<tr>
<td>EM</td>
<td>Electron Microscopy</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spread Function</td>
</tr>
<tr>
<td>LoG</td>
<td>Laplacian of Gaussian</td>
</tr>
<tr>
<td>DoG</td>
<td>Difference of Gaussian</td>
</tr>
<tr>
<td>gLoG</td>
<td>generalized Laplacian of Gaussian</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian Mixture Model</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale-Invariant Feature Transform</td>
</tr>
<tr>
<td>HOG</td>
<td>Histogram of Oriented Gradients</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>GT</td>
<td>Ground Truth</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>US</td>
<td>Under-Segmentation</td>
</tr>
<tr>
<td>OS</td>
<td>Over-Segmentation</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the Curve</td>
</tr>
<tr>
<td>TRA</td>
<td>Tracking Precision</td>
</tr>
<tr>
<td>SEG</td>
<td>Segmentation Accuracy</td>
</tr>
<tr>
<td>2D</td>
<td>2-Dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>3-Dimensional</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>HCS</td>
<td>High-Content Screening</td>
</tr>
<tr>
<td>HTS</td>
<td>High-Throughput Screening</td>
</tr>
<tr>
<td>G</td>
<td>Proposal Graph: Graph Representing Alternative Tracking Hypotheses</td>
</tr>
<tr>
<td>g</td>
<td>Tracking Graph: A Tracking Hypothesis</td>
</tr>
<tr>
<td>g*</td>
<td>Tracking Graph: Optimal Solution</td>
</tr>
<tr>
<td>P_G</td>
<td>Set of Cell Proposals in G</td>
</tr>
<tr>
<td>E_G</td>
<td>Set of Event Edges in G</td>
</tr>
<tr>
<td>S</td>
<td>Source Node</td>
</tr>
<tr>
<td>T</td>
<td>Sink Node</td>
</tr>
<tr>
<td>p_i</td>
<td>Proposal i</td>
</tr>
<tr>
<td>e_{i,j}</td>
<td>Edge from Node i to Node j Representing Event c</td>
</tr>
<tr>
<td>C_x</td>
<td>Cost of the Node/Edge x</td>
</tr>
<tr>
<td>I</td>
<td>Set of all Frames in a Video</td>
</tr>
<tr>
<td>I_i</td>
<td>Frame Containing Proposal p_i</td>
</tr>
<tr>
<td>I_{i,j}</td>
<td>Frames Containing Proposal p_i and p_j</td>
</tr>
<tr>
<td>b_i</td>
<td>Binary Selection Variable For Proposal p_i</td>
</tr>
<tr>
<td>b_{i,j}</td>
<td>Binary Selection Variable For Event e_{i,j}</td>
</tr>
</tbody>
</table>
List of original publications

This dissertation is based on the following articles, which are referred to in the text by their Roman numerals (I–V):


The writing and experiments for these articles were done by the author of this dissertation, while the co-authors gave valuable suggestions and feedback during the discussions about the methods, results and text. The ideas presented in these articles were formulated and refined in these group discussions.

Microscopy videos used in Paper I were provided by M. Kaakinen.
Contents

Abstract
Tiivistelmä
Acknowledgements
Abbreviations
List of original publications
Contents
1 Introduction
  1.1 Background and motivation
  1.2 Analysis pipeline
  1.3 Applications
  1.4 Scope of the thesis
  1.5 Contributions
  1.6 Summary of the original articles
  1.7 Outline of the thesis
2 Background
  2.1 Computer vision
  2.2 Deep learning
  2.3 Object detection
  2.4 Image segmentation
  2.5 Object proposals
  2.6 Object tracking
  2.7 Microscopy image analysis
3 Cell proposals
  3.1 Related works
  3.1.1 Object proposals
  3.1.2 Cell proposals
  3.2 The proposed methods
  3.2.1 Blob proposals
  3.2.2 Cell proposal network
  3.2.3 Proposal conflicts
  3.3 Discussion

13
1 Introduction

1.1 Background and motivation

Biology and medicine depend heavily on imaging to understand cellular processes and how they can be influenced to obtain the desired response. For example, in immunology, this can range from targeted killing of harmful cells to increasing proliferation rate of cells to combat foreign particles such as viruses and bacteria. In recent decades, there have been rapid advances in microscopy imaging, dyes for labeling cellular structures, automation of microscopes, computational methods for recovering the underlying data from single blurred image or multiple images captured from different positions, and enclosures that can tightly control the environmental conditions, e.g. temperature, humidity, pH, etc. Recent systems are even dynamically adjusting the imaging settings by analyzing the captured images to improve spatial resolution and signal strength (Royer et al. 2016). This has expanded the range of tools in the biologist’s arsenal, which are being used to observe cells at higher resolutions, higher frame rates and in cultures better resembling in-vivo conditions. This in its turn has enabled experiments which are more biologically relevant and expanded the range of questions which can potentially be answered.

The biological processes often are very complex and influenced by hundreds of genes, proteins and external influences (Zimmer et al. 2006), making it difficult to understand them from the study of few cells or samples. It is therefore necessary to analyze a large number of cells in multiple samples to extract biological insights which are statistically valid. Due to the rapid advances in automation of microscopes, it is now possible to capture images from multiple spatial locations in a large sample (Hilsenbeck et al. 2016) and multiple samples (Neumann et al. 2010) from a large-scale experiment (Eliceiri et al. 2012) over a long period (few days) with little or no human intervention. The image data generated by these experiments can be huge, with single experiments in developmental biology capturing the development of embryos of model organisms (e.g. *c. elegans*, *d. rerio*, *drosophila*, *m. musculus*) producing tens of TBs of imaging data (Liu & Keller 2016), and screening studies imaging tens of thousands of wells producing hundreds of thousands of images/sequences, requiring multiple TBs of storage space (Laufer et al. 2013, Neumann et al. 2010).
It is extremely challenging to manually analyze this image data because of its size, high dimensionality (making it difficult to view 4D (3D+t) cells on 2D screens), and need to find subtle patterns which are influenced by multiple factors not always controllable. In addition, low contrast cells and boundaries of unstained cells can be difficult to detect (Huth et al. 2010). It can require viewing multiple neighboring frames to accurately associate cells; it can be challenging to estimate centers of irregularly shaped cells; and the large number of cells which are often required to be annotated, leads to imprecise manual cell center and boundary annotations. All these factors make manual analysis not just tedious and time-consuming but also lead to results which are often only qualitative, very subjective and have lower accuracy. For example, imprecise annotation during manual tracking of cells can lead to over-estimation of cell speeds by up to 410% when the cell movements between adjacent frames are considerably smaller than the cell sizes (Huth et al. 2010). In pathology, there can be large disagreement in the classification of nuclei between different pathologists (42%) or even the same pathologist (21%) at different times (Fuchs & Buhmann 2011). As a result of the above-mentioned limitations of manual analysis, the need and importance of automated microscopy image analysis methods, to enable extraction of reliable and quantitative insights from huge microscopy datasets, is increasing.

1.2 Analysis pipeline

This section presents a brief overview of common microscopy image analysis pipeline from imaging to quantitative analysis. A more detailed overview along with software used in each stage is available in (Eliceiri et al. 2012). Specific analysis pipelines used for cell migration assays and high-content screening (HCS) experiments can be found in (Masuzzo et al. 2016) and (Boutros et al. 2015), respectively.

First, the sample is prepared, placed on the microscope stage, and imaged for a period to record the dynamic cell behavior. The imaging process is highly automated and current microscopes can capture large datasets without any human intervention. Some microscopy techniques, such as light sheet microscopy which images the sample from different angles, capture image data which is not easy to view or analyze. These microscopy techniques use computational methods to combine multiple captured images and obtain a higher quality image. Section 2.7 provides a brief introduction to various microscopy techniques relevant to this thesis.
Once the images have been captured, preprocessing may need to be performed to enhance the signal-to-noise ratio (SNR) and make the automated or manual analysis easier. The preprocessing steps vary depending on the experimental setup and microscopy modality. Phase-contrast microscopes, often used to observe unstained live cells in thin samples, produce images which sometimes suffer from contrast and illumination problems, and are typically processed using background subtraction. Fluorescence microscopy, often used to observe specific 3D structures within fluorescently stained cells from thick samples, produces images which suffer from noise originating from out-of-focal planes. Deconvolution (Sarder & Nehorai 2006) is commonly used for removing this noise and recovering sharp images. Despite recent advances in imaging
techniques, 3D images still suffer from anisotropic resolution, with axial resolution often being 2-8x lower than the spatial resolution. This is usually dealt with by interpolation.

The next stage in the analysis is cell detection, segmentation and tracking, which locates cells in individual frames and associates them in neighboring frames, to enable analysis of dynamic cellular behavior. Accurate cell detection, segmentation and tracking are needed to perform single-cell-level analysis, which is very important when different cell populations within a sample may have heterogeneous behavior (Gawad et al. 2016). This part of the analysis chain is the focus of this thesis and the following chapters will present methods to solve problems encountered in this stage.

The final stage of the analysis utilizes the cell detections, segmentation masks and lineage trees to compute some biologically relevant quantities (e.g. cell velocity, size, displacement, intensity, compactness, texture, etc.) which capture cell shape, appearance and dynamic behavior. The patterns in these quantities within a population at any given time or over time are analyzed at individual-cell, sub-population or whole-population level, and biologically meaningful insights are extracted.

Some population or sub-population level quantifications can be obtained without segmenting and tracking cells (e.g. optical flow can be used to quantify the cell population velocity); however, to obtain detailed and precise individual-cell-level quantifications, requires accurate cell segmentation and tracking.

1.3 Applications

This section mentions few application areas within biology where automated microscopy image analysis can be used to either reveal, quantify or verify some biologically relevant characteristics.

Cell detection and segmentation

Cell shape is influenced by the interactions between the cell’s cytoskeleton, membrane, membrane-bound proteins and extra-cellular environment (Pincus & Theriot 2007), including cell-cell interactions (Stegmaier et al. 2016). It is often used as an indicator for cell’s physiological states, including differentiation, apoptosis, mitosis and lymphocyte activation (Pincus & Theriot 2007); its sociology (Meijering 2012); and morphogenesis of tissues, organs and entire embryos (Stegmaier et al. 2016). Quantification of cell morphology also plays an important role in high-content screens (HCS) (Boutros et al. 2018).
2015, Zanella et al. 2010) used in drug discovery and biology to study the influence of small-molecules, RNA interference and mutations on the cell behavior.

In histopathology, nuclei detection and segmentation enables quantification of nuclei categories, counts and morphology which are important measures for the diagnosis and grading of various cancers. The accuracy of nuclei detection and segmentation in these images is very important as it has significant influence on the performance of diagnosis and grading (Irshad et al. 2014).

**Cell tracking**

Automated cell tracking enables the quantification of cell proliferation, division, death, differentiation and migration, which are central to many fundamental biological processes, such as immune response, wound healing, tissue homeostasis, morphogenesis and spread of many diseases, including cancer (Trepat et al. 2012).

Cell tracking can be used to analyze the (morpho)dynamic properties of cell populations at single-cell-level and understand how cells function and interact with their environment. Changes in cellular motion and morphology, which reflect organizational and physiological states of cells, can be quantified and used to assess the outcomes of different treatments, such as drugs or hormones, on biological tissues or cells (Pincus & Theriot 2007). This can indicate the clinical potential of a drug (Zimmer et al. 2002) and its cytotoxicity (Zanella et al. 2010).

In cancer research, cell tracking can shed light on how the cancer cells spread and can help in the development of treatments. Cell lineage trees can be used for locating cellular origin of solid tumors in mouse models for various human cancers, including skin, lung, breast and prostate cancer (Blanpain 2013). Cell tracking has been used to quantify dose dependent pro-migratory effect of chemicals (e.g. hyaluronic acid) on brain cancer cells in 3D culture (Adanja et al. 2011), identification of different modes of tumor cell migration and infiltration, and how significantly cell migration and infiltration depends on the cell type and the microenvironment around cells (Demou & McIntire 2002).

In developmental biology, complex patterns of cell migration are essential for proper tissue formation. Automated reconstruction of cell lineage trees of complex organisms and organotypic cultures (e.g. zebrafish, kidney, etc.) is one central goal (Amat et al. 2014) of developmental biology as it enables understanding of how embryos and organs develop, and how embryonic cells differentiate into specialized cells. These cell lineage
trees enable visualization of parent-daughter relationships, (a)synchronous divisions, symmetry breaking events, and gene expression levels (Meijering et al. 2009).

In recent years, automated cell tracking has also enabled insights which were not feasible with manual analysis either due to subtlety (Zimmer et al. 2006) or scale (Neumann et al. 2010). It has enabled the discovery of small periodic membrane deformations of crawling amoeba (Zimmer et al. 2006) and the use of high-content screening (HCS), which produced ∼190,000 sequences, to identify individual genes involved in various biological functions, including cell division, migration and survival (Neumann et al. 2010).

1.4 Scope of the thesis

Cell segmentation and tracking from microscopy images and videos is a challenging problem. It is also a very broad research area as the challenges involved differ significantly depending on the shapes and appearances of cells, which are heavily dependent on the imaging setup, sample preparation and cell type, etc. This thesis primarily tries to solve one of the biggest challenges in cell detection and segmentation, i.e. the separation of individual cells within cell clusters. The methods developed in this thesis are mostly tested on live-cell microscopy images from fluorescence and phase-contrast microscopes; however, some experiments have been conducted on histological images as well. Fig 3 (Chapter 2) shows a small region from these datasets to better highlight the scope of this thesis.

Since the challenges encountered in automated analysis of images from different experiments can differ drastically, and most of the research is focused on answering the biologically relevant question, the methods in general focus on obtaining the best possible performance exploiting all assumptions which can be met in the current set of images, and the generality and reusability of methods are often sacrificed. This has led to the development of many methods which often rely on long chains of simple image processing operations (Khan et al. 2014) involving many parameters, which are usually tuned manually. These methods can exploit the dataset-specific properties to obtain very good performance, but it can be challenging and time-consuming to apply them to other sequences. Often, the code is not published; even when it is released, it can require some expertise to adjust the parameters and sometimes it is necessary to alter the chain of image processing operations to obtain acceptable results.
Due to the issues mentioned above, most commercial and open-source microscopy image analysis software packages rely on simple detection (e.g. blob detection), segmentation (e.g. thresholding, watershed) and tracking (e.g. distance-based nearest-neighbor association) methods (Wiesmann et al. 2015), as they can achieve acceptable results for a wider range of images and are more intuitive for non-experts to tune. In recent years, software packages (e.g. Ilastik, CellProfiler, ImageJ) have started providing simple user-interfaces for biologists to manually annotate a small set of training examples and use machine learning based methods for classifying and segmenting cells from more challenging images.

The focus of this thesis is to develop methods which can learn from the given training data and require little or no manual adjustment of parameters and models. It is hoped that instead of developing customized solutions for analysis of each new set of images, effort can be better spent on annotating data. This despite being more time-consuming in the short-term, enables the development of methods which are more general, reusable and often outperform handcrafted methods. In addition, these annotations can be re-used by newer methods to improve their performance and to track progress in this research area.

Often, only the nuclei are imaged, which are used as a proxy for cells in various quantifications (e.g. cell counting, cell proliferation, cell movement, etc.). In this thesis, the word *cells* is sometimes used loosely to refer to nuclei, as is common in the automated microscopy image analysis literature (Maška et al. 2014, Schiegg et al. 2015a).

### 1.5 Contributions

The main contributions of this thesis are listed below:

- A *Cell Proposal Network* (CPN) which proposes cell candidate masks and their scores using convolutional neural networks (CNNs), which learn the cell shapes and appearances from the given training data (Paper III, Paper IV and Paper V).
- A method for cell proposal generation (BLOB) based on multiple filter banks containing gLoG filters (Kong et al. 2013), which can be used for analysis of round or elliptical cells/nuclei (Paper II).
- A graph-cut based method to segment cells of highly deformable shapes, which utilizes distinctive features at cell boundaries to create a barrier between touching cells to detect individual cells (Paper I).
A method for cell tracking which iteratively finds the shortest cost path in the proposal
graph and builds cell tracks (Paper II).

A cell tracking method which uses Integer Linear Programming (ILP) to select the
globally optimal cell proposals and their associations, and achieves state-of-the-art
performance on multiple microscopy datasets (Paper V).

A single CPN model for proposing candidates and segmentation masks for phase-
contrast, fluorescence and histological images, demonstrating the possibility of a
*general* cell detector (Paper IV).

### 1.6 Summary of the original articles

This thesis is based on the five articles in which the contributions mentioned in
Section 1.5 were originally published. The summary of these articles is given below, and
these articles are reprinted in the appendix of this thesis.

In Paper I, a cell segmentation method is proposed, which uses an edge detector
to create a wedge between touching cells and graph-cut to detect individual cells.
A state-of-the-art edge detector (Dollár & Zitnick 2013), which utilizes multi-scale
structure around each pixel to predict directional edges, is used for creating a barrier
between touching cells. Then, graph-cut is used to segment the image, and each
connected component is used as a cell detection. These detections are utilized as markers
by watershed algorithm to obtain cell segmentation. When processing 3D images, each
slice is segmented independently, and then segments in adjacent slices are connected to
obtain the 3D cell segmentation.

Paper II presents a cell candidate generation method, which uses multiple blob-like
filter banks to detect round and elliptical cells/nuclei. These proposals are utilized by
a cell tracking method, which selects a locally optimal subset of proposals and their
association edges, by iteratively finding the shortest cost path in the proposal graph.
This method is evaluated on three fluorescence microscopy datasets and achieves better
performance than previous methods.

In Paper III, the idea of posing object proposal generation as an object bounding box
regression problem (Ren et al. 2015) is used to propose cell candidates. As part of this
method, a *Cell Proposal Network* (CPN) is presented, which learns to propose better
cell candidates and achieves higher performance than previous methods, which rely on a
very strong cue of cell shape. This network is adapted from ZF model (Zeiler & Fergus
2014) to tackle two main issues: small cells and lack of training data. The proposed
model is trained using augmented annotations; i.e. cell segmentation masks which are obtained using an automated method supervised with the ground truth cell markers. The cell proposal masks are obtained by thresholding the cell proposal bounding boxes predicted by CPN.

Paper IV extends the method in Paper III by adding a second network, which proposes cell segmentation masks. The new network has a similar structure as Fully Convolutional Network (FCN) (Long et al. 2015) and utilizes both low-level and high-level features for predicting segmentation masks. This network is tested on three datasets from histology, phase-contrast and fluorescence microscopy, and achieves better segmentation performance than previous methods. As part of this work, it is shown that it is possible to train a single model for multiple microscopy modalities, without compromising performance significantly.

Finally, in Paper V, the segmentation network of Paper IV is replaced with a new network, which first gradually merges upscaled high-level features with low-level features, like U-Net (Ronneberger et al. 2015), and then uses fully connected layers to predict individual cell segmentation masks. This network requires less computational memory and is better able to handle confounding cells in the proposed bounding boxes. In this article, cell tracking is posed as a proposal selection problem and Integer Linear Programming (ILP) is used to select the optimal subset of non-conflicting cell and edge proposals. This method achieves state-of-the-art tracking performance on multiple sequences from four fluorescence and phase-contrast microscopy datasets containing cells of various shapes and appearances.

### 1.7 Outline of the thesis

This thesis consists of an overview of the original articles attached at the end. Here, the contents of individual chapters are summarized.

Chapter 2 introduces computer vision field and few research areas within it which are relevant to the contents of this thesis. These research areas include: object proposals, object detection, image segmentation, object tracking and deep learning. It also introduces microscopy image analysis field and presents the datasets used to evaluate the proposed methods.

In Chapter 3, an overview of object and cell proposal methods is given, and two cell candidate generation methods proposed as part of this thesis are introduced.
Chapter 4 provides a brief introduction to cell detection and segmentation, describes the main challenges involved, and explains how cell proposals can be used for cell detection and segmentation. It also presents the graph-cut based cell segmentation method proposed in Paper I.

In Chapter 5, an overview of cell tracking is provided along with few key challenges. This chapter also introduces two cell tracking methods based on the idea of proposal selection, originally presented in Paper II and Paper V.

Chapter 6 discusses the limitations of current automated cell detection, segmentation and tracking methods, and future lines of research which deserve attention. Finally, Chapter 7 concludes this thesis by summarizing the main contributions.
2 Background

2.1 Computer vision

Humans rely heavily on their eyes to sense, and brains to interpret the world, and they can do this task with little effort and thought. Computer vision is the field with the aim of imparting this very same capability to machines. It is a very challenging task as it requires computers to extract high-level information from digital images and videos, which are in essence just multi-dimensional grids of numbers.

Computer vision is a very large and diverse field spanning multiple inter-disciplinary research areas. Some popular research lines deal with image formation, image processing, image classification, object detection, instance-segmentation, motion analysis, scene reconstruction, image rendering, etc.

The ultimate goal of computer vision is to enable a computer to recover the whole 3D scene; partition it into regions spanning multiple levels of understanding, given an image; and understand how these regions move, interact and deform over time, given a video sequence. However, for most real-world applications, this complete understanding of images is unnecessary, as they deal with a specific issue and have much more relaxed requirements and priorities. In consumer cameras, for example, it is sufficient for the device to detect human faces and track moving objects, and the need for recognizing the identity of the people or type of the objects is not very important; whereas in biometric applications, it is important for the device to recognize individuals with very high accuracy and to be able to differentiate between human faces and pictures/videos of them.

There are multiple levels of understanding that computer vision researchers have been tackling over the years, and they can differ depending on the application. Fig. 2 shows four main levels of understanding that are touched in various parts of this thesis. All these levels have received large research focus in general as well as in many specific applications. Image classification is the coarsest level of understanding and is concerned with labeling an image with the tag of the objects/concepts present in that image. The second coarsest level of understanding is object detection, which deals with detecting and localizing the objects present inside a given image. The next higher level of understanding is semantic segmentation, which partitions an image into regions with each region containing one type of object. If individual instances of objects can be
there are other lower and higher levels of image understanding as well. Lower levels of image understanding include extracting some image statistics, image matching, optical flow, etc. Higher levels of image understanding in their turn include inferring some properties of objects, how they interact with each other, etc. These higher levels of
information are particularly useful in robotics and augmented reality applications, where they can be used to interact and manipulate objects in real world.

This thesis is concerned with methods for automated analysis of cell behavior in microscopy images and videos; it focuses on object detection, instance-level object segmentation and object tracking for a specific class of objects, i.e. biological cells.

2.2 Deep learning

In this section, a brief overview and recent history of deep learning is presented. More detailed history and overview can be found in (Schmidhuber 2015, LeCun et al. 2015) and its applications in biomedical and medical image analysis can be found in (Litjens et al. 2017, Ching et al. 2017).

Machine learning, the field concerned with enabling computers to learn, has played a major role in object detection, segmentation and tracking over the years. Until very recently, computer vision and microscopy image analysis relied heavily on domain expertise and careful engineering to design features for transforming raw pixel values into suitable internal representations. These representations were then utilized either directly or by a machine learning method to perform the desired task. These internal representations played a very important role in image understanding, and in general were difficult to design due to the large variations in the appearances and shapes of objects.

Multiple levels of representations are necessary to capture object characteristics at various levels of hierarchy to enable computer vision methods to be robust to various challenges (e.g. occlusion, scale, pose) encountered in real-world images. However, until very recently, shallow representations which focused only on few levels of representations were common, as they were more intuitive to design and easier to learn from available annotated data. The lower level representations were dominated by gradient based feature descriptors, which included SIFT (scale-invariant feature transform) and HOG (histogram of oriented gradients), and the higher level representation were dominated by deformable part models, which represented objects (e.g. car) using a collection of parts (e.g. wheel, window, etc.). These shallow representations had limited success but were not able to handle challenging real-world images, which included large occlusions and cluttered backgrounds.

In last five years, deep learning has transformed the computer vision field by enabling rapid advances in many research areas and has brought many tasks (e.g. object detection),
which were thought to be intractable in near term just few years ago (Szeliski 2011, p.657), to levels where they can be considered solved for many practical applications.

Deep learning is a particular class of machine learning methods, which allows computers to learn multiple levels of representations by concatenating simple non-linear modules. Each of these modules can transform its input data into a higher and more abstract level. By concatenating enough of these simple modules, very complex functions can be learned. First few modules in the chain extract low-level representations equivalent to edges, corners, and other simple structures; while the middle layers utilize low-level representations to detect mid-level structures, e.g. wheels, eyes, faces, etc.; and finally, the last few layers extract high-level representations detecting complex objects or structures, e.g. cars, cats, people, etc., while ignoring irrelevant variations in these objects due to lighting, pose, scale, occlusion, etc.

Since these models can learn these representations from the training data, they can determine which of them are useful for a given task, which is the main reason for their success. The research areas (e.g. object detection) where it has traditionally been difficult to design these representations have had large improvements but areas where manually designed representations were good enough as the problems were much more constrained (e.g. mitosis detection, cell detection), the gains in performance have been smaller.

Current wave of deep learning success has been enabled by the availability of large annotated datasets and faster GPUs, which have allowed the researchers to train large models which outperform handcrafted methods. Initial successes of current deep learning methods in computer vision were in the research areas of character recognition, traffic sign recognition (Stallkamp et al. 2011) and mitosis detection (Ciresan et al. 2013) in 2011 and 2012. The first major success came with AlexNet (Krizhevsky et al. 2012), which not only won the ImageNet challenge about image classification but also had a large margin (top-5 error: 16% vs 26%) between it and the second best method. Since then, deep learning has won many computer vision challenges ranging from classification, detection, segmentation and captioning.

In recent years, there have been many advances, which have enabled better training of deeper and wider networks resulting in improved performance. Few important network models or modules have been: R-CNN (Girshick et al. 2014) for object

1http://image-net.org/challenges/LSVRC/2012/results.html
2https://camelyon17.grand-challenge.org/results/
3http://mscoco.org/dataset/#detections-leaderboard
detection, FCN for segmentation (Long et al. 2015), Faster R-CNN (Ren et al. 2015) for object proposals, Inception module (Szegedy et al. 2015) for wider networks and ResNet module (He et al. 2016) for training extremely deep networks. At present, very large and wide networks can be trained quickly (Goyal et al. 2017) by sharing computations on multiple servers and reducing training time almost linearly. In the near future, the gains achieved by increasing the network depth and width are likely to saturate as already extremely deep (1000+ layers) (He et al. 2016) and wide (32 parallel paths) (Xie et al. 2017) networks are being used. However, there is still potentially some room for improvement in image classification, object detection and object segmentation by increasing the dataset size and better utilizing weak labels which can be obtained much more readily (Sun et al. 2017).

Deep learning has had some successes in microscopy image analysis as well, but the gains in this field have been moderate (Cireşan et al. 2013, Ronneberger et al. 2015) compared to general computer vision. This is due to the fact that for many microscopy image analysis tasks, simple representations can already achieve very good performance for all but the most extreme cases and therefore there is not much room for improvement. In addition, lack of large public datasets has limited the robustness and generalization ability of the trained models and has hindered major advances in the field so far.

Nevertheless, deep learning has been used for mitosis detection, cell classification, cell segmentation, cell tracking, deblurring and upsampling 3D images, and cancer grading. In histological image analysis, deep learning has been widely used as large annotated datasets have become available. Here, deep learning has had a much bigger impact and recent challenges (CAMELYON17 4, Her2 5, TUPAC16 6) have been dominated by deep learning based methods. For live-cell microscopy image analysis, the impact of deep learning has been limited so far. Cells in general have huge variability in their shapes and appearances across images from different experiments but the variability within images from a single experiment is somewhat limited. As a result, the common approach thus far has been to train the networks on a small (often private) dataset for the target application. Since these trained models do not generalize to new images captured under different conditions and often have to be re-trained, they have limited impact.

U-Net (Ronneberger et al. 2015), a semantic segmentation network, is by far the most popular microscopy image analysis network.

---

4https://camelyon17.grand-challenge.org/results/
5http://www2.warwick.ac.uk/fac/sci/dcs/research/tia/her2contest/outcome
6http://tupac.tue-image.nl/node/62
2.3 Object detection

Object detection (Fig. 2b) refers to the task of finding individual instances of the objects of interest (e.g. people, animals, vehicle) in an image. It often refers to localizing objects in an image using bounding boxes but in some cases it may also refer to localizing objects using points close to their centers. A lot of object detection research has focused on the detection of few particular categories of objects (e.g. face, pedestrian, biological cell) due to the importance of their practical applications.

General object detection is an ill-posed problem as the definition of an object can vary depending on the task at hand and it can be very challenging and expensive to annotate all potential image regions which can be considered an object (e.g. person, face, eye, iris). Most object detection methods use the term objects to refer to a small set of clearly defined “things”, which have a specific size and shape (Forsyth et al. 1996). The set of “things” considered as objects has varied over time depending on the dataset being used to train and evaluate the detection method.

General object detection has received a lot of attention in the last decade; first, with the development of PASCAL VOC datasets (Everingham et al. 2015), and more recently, with the ImageNet challenge (Russakovsky et al. 2015) and COCO datasets (Lin et al. 2014). The annotated objects in these datasets contain a diverse set of things, e.g. household items, vehicles, animals, etc. PASCAL VOC has 20 objects, COCO has 91 objects and ImageNet challenge has 200 objects. These datasets have been driving the research progress and current object detection methods (He et al. 2017) are capable of detecting large, non-occluded, objects quite reliably.

Object detection has many applications, including image retrieval, object counting, retrieval of items from warehouses, mail sorting, video surveillance, autonomous driving, robotics, detecting apparel and detecting logos of popular brands. Object detection also plays a vital role in other computer vision research areas, such as object tracking, object segmentation, caption generation and visual reasoning.

2.4 Image segmentation

Image segmentation (Fig. 2c, 2d) is the process of partitioning an image into a set of pixel-accurate regions, where each region contains only one meaningful entity. This entity can either be a “thing” which has a specific size and shape, and is often referred to as an object (e.g. car, pedestrian) or part of an object (e.g. wheel, hand) or be some
"stuff" (e.g. grass, forest, sky), which has a homogeneous or repetitive pattern (texture) but lacks any specific and distinctive spatial extent and shape (Forsyth et al. 1996).

General object segmentation is an ill-defined problem because depending on the application, desired segmentation can differ. In some applications (e.g. surveillance, people counting, pedestrian detection), it is sufficient to segment each human as separate, but in other applications (e.g. human-computer interaction, virtual reality, sign-language recognition), the desired segmentation may require segmenting individual body parts (e.g. hands, face, legs, etc.) as separate regions.

Image segmentation spans multiple lines of research: semantic segmentation, instance-level object segmentation and cosegmentation (Zhu et al. 2016). Semantic segmentation (Fig. 2c) deals with partitioning an image into non-overlapping meaningful regions. So far, it has been the most popular segmentation problem because the previous segmentation methods were not very capable of distinguishing between multiple instances of an object and had to heavily rely on object detectors, which themselves had limited performance. However, in last few years, with rapid improvements in object detection performance, instance segmentation is gaining more attention.

Instance segmentation (Fig. 2d) aims to partition an image into regions with each region containing only one instance of an object. These regions can then be used to extract meaningful information about the objects depending on the application. Current methods (He et al. 2017) can segment large, thick, un-occluded objects quite reliably. However, their performance drops for small, thin, elongated, and occluded objects.

Cosegmentation refers to the task of segmenting objects from background in a given image using a model which is trained on images with weak annotation of just the presence or absence of the object of interest (Rother et al. 2006).

2.5 Object proposals

An object proposal is a candidate for an object detection and/or segmentation. Object proposals enable the subsequent analysis stages to focus on a small set of image regions. They need to have high recall, with a corresponding candidate for as many objects as possible, without increasing the total number of proposals too much. They have played a significant role in object detection methods during last decade by replacing sliding window approach and enabling the use of more advanced classifiers (Alexe et al. 2010, Uijlings et al. 2013).
Object proposals are used extensively in many other research areas, including instance-level segmentation, action recognition and object tracking as a preprocessing step and are one of the key component in current top performing object detection and instance-level segmentation methods. In some research areas (e.g. multi-object tracking), the term object detections is usually used instead of object proposals as the proposals are the predictions of an object detector. However, they serve a similar purpose as object proposals, i.e. to focus attention of the methods to regions in an image more likely to contain the object. In the context of tracking, object proposals enable efficient representation of multiple hypotheses for video patches by significantly reducing the number of detection and association alternatives.

2.6 Object tracking

Object tracking refers to the process of locating a moving object (e.g. ball, animal, person, etc.) or objects (e.g. football players, pedestrians, vehicles, etc.) over time in a video. It is an important research area within computer vision and has two main branches: visual object tracking (VOT) and multiple object tracking (MOT). VOT research deals with tracking a single instance of an object (e.g. person, animal, vehicle, etc.) which is annotated in the first frame of a video, often using a bounding box. It focuses on the appearance and motion models which have to be learned on-line (Smeulders et al. 2014). MOT in its turn deals with tracking multiple instances of a single type of object, with objects entering and leaving the field of view and it primarily focuses on accurate association of detections in adjacent frames.

Object tracking has a central role in many application areas, including vehicle navigation, where an autonomous car has to track other cars and pedestrians in its surrounding to avoid collisions; surveillance, where tracking people can enable recognition of suspicious activity; sports analytics, where player statistics can be used for adjusting game plan; human-computer interaction; traffic monitoring; and motion based recognition.

2.7 Microscopy image analysis

Microscopy imaging is a vital tool to observe dynamic cellular behavior, which plays a central role in development and immune response of living organisms. To obtain biologically relevant quantifications, especially at individual-cell-level which is often
needed to analyze heterogeneous cellular behavior from image data captured during experiments, accurate cell detection, segmentation and tracking are essential.

Cell detection, segmentation and tracking can be considered as the specific application area of object detection, instance-level segmentation and multi-object tracking, respectively, where the objects of interest are biological cells. However, the techniques used to analyze microscopy images differ considerably from those used to process natural images due to the significantly different characteristics of images. Direct application of a method developed for natural images to microscopy images or vice versa often does not perform well. Microscopy images have some unique constraints (e.g. some cell types and nuclei have rigid round/tubular shapes) which are often exploited and challenges (e.g. cells can divide, similar appearance and shapes of cells, low frame rate) which have to be solved, to adapt general computer vision method to this application area.

Most of the research in cell detection, segmentation and tracking is done either inside or in close collaboration with the biology research labs and the focus of the developed methods is to answer the immediate biological questions. The novelty, generality and usability of methods to answer other research questions are often secondary priorities. As a result, the literature is filled with methods which use long chains of basic image processing operations (e.g. thresholding, watershed, dilation, erosion, blob detection) to achieve good performance on the particular set of images they are designed for but do not generalize to other sets of previously unseen images.

**Microscopy modalities**

This section gives a brief introduction of microscopy techniques that are relevant to this thesis. A brief overview of various other microscopy modalities used in cell biology can be found in (Thorn 2016).

In **bright-field microscopy**, a sample is imaged by measuring its effect on the light passing through it. Since cells are thin and transparent, only their outlines can be seen. However, phase shift in the light induced by the cells can be visualized to obtain detailed higher contrast images using either phase-contrast (PhC), which makes cells appear dark on light background or differential-interference contrast (DIC), making cells appear shaded and defocused (Thorn 2016). Bright-field microscopy is often used in applications where cells cannot be stained as the dye introduced may alter the cell behavior or when cells have to be imaged for long duration.
Fluorescence microscopy (Fluo) uses fluorescent dyes, that emit light of a particular wavelength when excited by a light source, to image cells. Multiple fluorescent dyes can be used to label different structures inside the cells, including nuclei, membranes and/or cytoplasms, with cytoplasm labeled images being the most challenging for automated analysis. Fluorescence microscopy typically produces images with high contrast in which only the labeled cell structures are visible and hence are usually easier to analyze using automated methods. However, during live-cell imaging experiments, fluorescence microscopes expose a biological sample to high energy excitation light over long period, which can lead to photo-bleaching (loss of fluorescent markers) reducing the image quality with time and/or photo-toxicity (damage to cells) impacting the normal behavior of cells by altering their physiology and normal development (Liu & Keller 2016). As a result, often the intensity of excitation light, image resolution and frame rate have to be finely balanced when imaging cells for long duration (Keller 2013).

There are many fluorescence microscopy techniques: epifluorescence, confocal, light-sheet, etc. Confocal microscopy captures an image by scanning the sample over a regular grid. A single point is imaged by illuminating it with a focused laser beam and detecting the emitted light from that point after it passes through a pinhole, which blocks light emitted by out-of-focal planes. This leads to sharper images and enables deeper penetration of biological samples.

In histology (Hist), cells and tissues in a specimen are imaged with bright-field microscopy after the specimen has been sectioned, stained (e.g. with hematoxylin and eosin (H&E)) and mounted on a glass slide (McCann et al. 2015).

Datasets

General computer vision has a long tradition of releasing annotated datasets publicly for most interesting problems and using these datasets to evaluate new methods and track progress. This allows the research community to objectively compare different methods and speeds up research.

Recently, there have been few efforts to benchmark live-cell image analysis methods and few large public annotated datasets have been created. Two of the more popular efforts among these include ISBI Cell tracking challenge (CTC) (Ulman et al. 2017) and Broad Bioimage Benchmark Collection (BBBC) (Ljosa et al. 2012). Even though these datasets have been available for a few years now, it is still uncommon to find a new cell detection, segmentation, classification or tracking method, which reports its performance
on these datasets. One of the potential reasons for the limited adoption of these datasets as a standard is that even though these datasets include a wide variety of microscopy sequences and images, they still do not cover many application areas in which cell detection, segmentation or tracking is used. In addition, the number of samples from each type of experiment is limited and any method trained on these datasets is likely to overfit and is unlikely to transfer well to other images/videos and vice versa.

Other than Paper I, all methods proposed in this thesis are evaluated on the seven publicly available datasets listed in Table 1. Fig. 3 shows a sample region from each of these datasets. The first five datasets (Fluo-N2DL-HeLa, Fluo-N2DH-GOWT1, PhC-C2DH-U373, PhC-C2DL-PSC and Fluo-N2DH-SIM+) are from CTC (training set) (Ulman et al. 2017), while the PhC-HeLa-Ox dataset is from (Arteta et al. 2012) and Hist-BM is from (Kainz et al. 2015).

Three datasets each are from fluorescence (Fluo) and phase-contrast (PhC) microscopy, while one dataset is from histology (Hist). Fluo-N2DL-HeLa contains H2b-GFP expressing HeLa cells, Fluo-N2DH-GOWT1 GFP transfected GOWT1 mouse embryonic stem cells, Fluo-N2DH-SIM+ simulations of HL60 cells stained with Hoescht, PhC-C2DL-PSC Glioblastoma-astrocytoma U373 cells, PhC-C2DL-PSC Pancreatic Stem Cells, PhC-HeLa-Ox cervical cancer cells from HeLa cell line and Hist-BM H&E stained histological slides with healthy human bone marrow cells.

The annotations of microscopy datasets differ depending on the problem for which they are annotated. In cell detection and tracking research, marker annotation is common as hundreds of cells have to be annotated in many frames. A marker is a pixel or a set of connected pixels inside the body of a cell, often close to its center. In practice, as annotators have to mark hundreds if not thousands of cells, the marker position may not always be near its center and can be very close to cell boundaries for small or elongated cells. If a dataset is annotated for segmentation, it may have a binary mask separating cell pixels from background or it may have a unique label for each cell mask allowing evaluation of instance-level segmentation. In cases where images contain multiple types of cells, their category (type) might also be annotated.

Hist-BM and PhC-HeLa-Ox contain only marker annotations. Augmented annotations were used to train models for these datasets and Fluo-N2DL-HeLa. Augmented annotations refer to the segmentation masks produced by an automated segmentation method guided by weak manual annotation (cell markers). PhC-HeLa-Ox cell markers are expanded to cell segmentation masks by using simple heuristics to select non-conflicting MSER regions (Matas et al. 2004). Hist-BM cell segmentation masks
Fig. 3. Datasets: A small region (300x175) from each dataset is shown along with the ground truth cell markers (a) and the outlines of ground truth cell masks.
Table 1. Public datasets used to evaluate the methods presented in this thesis. Few statistics about each dataset are listed.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Images</th>
<th># Ground Truth</th>
<th>Resolution</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Markers</td>
<td>Masks</td>
<td>Tracks</td>
</tr>
<tr>
<td>Fluo-N2DL-HeLa</td>
<td>184</td>
<td>34059</td>
<td>1386</td>
<td>939</td>
</tr>
<tr>
<td>Fluo-N2DH-GOWT1</td>
<td>184</td>
<td>4373</td>
<td>265</td>
<td>77</td>
</tr>
<tr>
<td>PhC-C2DH-U373</td>
<td>230</td>
<td>1459</td>
<td>211</td>
<td>23</td>
</tr>
<tr>
<td>PhC-C2DL-PSC</td>
<td>852</td>
<td>51761</td>
<td>518</td>
<td>1137</td>
</tr>
<tr>
<td>Fluo-N2DH-SIM+</td>
<td>215</td>
<td>5978</td>
<td>5978</td>
<td>202</td>
</tr>
<tr>
<td>PhC-HeLa-Ox</td>
<td>22</td>
<td>2228</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hist-BM</td>
<td>11</td>
<td>4202</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

are obtained by assigning the label of each cell marker to the pixels near it, and using multi-label graph-cut to label remaining pixels. Fluo-N2DL-HeLa cell markers are used to obtain segmentation masks by using marker-controlled-watershed for frames which lack segmentation masks.

Software tools

There are many software tools and toolkits available for analyzing microscopy images, brief reviews of which can be found in (Wiesmann et al. 2015, Hilsenbeck et al. 2016, Meijering et al. 2012, Eliceiri et al. 2012). Most of these software tools can perform cell detection, segmentation and tracking, but the exact features may differ. The techniques available are often very simple, e.g. thresholding for segmentation, blob detection or watershed for splitting cell clusters and distance/size based association for cell tracking. More advanced methods are also available, but they are not used frequently as they require tuning multiple parameters, which are not always intuitive (Wiesmann et al. 2015). A recent comparison of features offered by various free software tools can be found in (Wiesmann et al. 2015).

ImageJ7 (Schneider et al. 2012) is by far the most popular software tool for microscopy image analysis and provides functionality for researchers to develop plugins for extending its capabilities and many analysis methods have been released as its plugin. Icy8 (de Chaumont et al. 2012) is another popular software tool which enables the

7https://imagej.net/ImageJ2
8http://icy.bioimageanalysis.org/
development of new methods through plugins. Ilastik$^9$ (Sommer et al. 2011) is more focused towards interactive learning based cell segmentation and tracking. CellProfiler$^{10}$ (Carpenter et al. 2006) is more suited for high-throughput analysis and is closely associated with CellProfiler Analyst, which is used for data exploration and classification of cells, and CellProfiler Tracer, used to visualize and inspect tracking results. Many microscope manufacturers also provide some analysis capabilities with their imaging software package or as a separate software tool.

There also exist few toolkits/libraries, targeted at computer scientists, which include the implementation of common image processing and analysis methods used within biomedical and medical image analysis fields. Few popular toolkits are: ITK$^{11}$ for segmentation and registration, VTK$^{12}$ for visualization, OpenCV$^{13}$ for general image processing and computer vision, and OpenGM$^{14}$ for graphical models.

---

$^9$http://ilastik.org/

$^{10}$http://cellprofiler.org/

$^{11}$https://itk.org/

$^{12}$http://www.vtk.org/

$^{13}$http://opencv.org/

$^{14}$http://hciweb2.iwr.uni-heidelberg.de/opengm/
3 Cell proposals

A cell proposal is a candidate detection or segmentation for an individual biological cell. It usually refers to the candidate segmentation mask for a cell; however, bounding boxes, ellipses or other representations may also be used to represent cell proposals. In this thesis, the words proposals and candidates are used interchangeably.

Cell proposals play a similar role in cell detection, segmentation and tracking as object proposals do in object detection, segmentation and tracking, i.e. to focus attention of succeeding analysis stages to a small set of image regions. Unlike general computer vision, where object proposals often span multiple categories of objects and the main goal is to separate object regions from the background, in microscopy image analysis, samples often either contain only one type of cell or only one type of cell is labeled and the background region is usually homogeneous, and the main purpose of cell proposals is to separate individual cells from within cell clusters.

Fig. 4. (a) A challenging image region with ground truth (GT) cell markers shown. (b) Outlines of cell proposals representing alternative hypotheses for the image in (a). (c) Cell segmentation via selection of a subset of non-conflicting proposals from (b).
Some regions in microscopy sequences can be very challenging and ambiguous and it can be very difficult to determine how many cells there are in these regions and where their boundaries lie. In many of these situations, information from adjacent frames can be very helpful in resolving most of the ambiguities. Even when analyzing a single microscopy image, it can be beneficial to use more complex features, classifiers and inference to resolve these ambiguities. Cell proposals are one way of efficiently utilizing temporal information, advanced features and inference by reducing the number of hypotheses for a given image or sequence. Fig. 4 shows an example of how cell proposals enable representation of alternative hypotheses in ambiguous regions and may improve performance for a single image and Fig. 7 (Chapter 5) shows an example of how cell proposals can be used to represent alternative tracking hypotheses for a small image sequence.

Cell proposal generation methods have to deal with similar challenges as cell detection and segmentation methods, which are discussed in Section 4.1.

3.1 Related works

This section gives a brief background for the cell proposal generation methods proposed in this thesis. Section 3.1.1 presents an overview of deep learning based object proposal methods and Section 3.1.2 provides an overview of cell proposal methods.

3.1.1 Object proposals

The sliding window approach was the dominant paradigm for detection of objects in the 2000s. It scanned an image with bounding boxes, which could span various sizes and aspect ratios, and evaluated a classifier to determine whether a given window contained an object or not for each bounding box at pixels (at fixed intervals) densely covering the whole image. The number of windows which had to be evaluated was in order of millions for typical images and prevented the use of computationally demanding features and classifiers.

Object proposals were proposed to reduce the number of windows which had to be evaluated (Endres & Hoiem 2010). They enabled extraction of a small number of windows from an image such that most objects were well represented by at least one window. This allowed the object detection methods to focus on a small set of regions
and made it possible to use more advanced and complex features and classifiers, which in its turn enabled better detection performance.

Object proposal methods focus on category-independent proposals and are inspired by the humans’ ability to localize objects without recognizing them (Endres & Hoiem 2010). These methods assume that there are some shared characteristics between objects of various categories (e.g. cat, horse, car, bus, table, chair) which distinguishes them from background regions (e.g. wall, grass, road, sky) and these characteristics can be used to locate objects in images. These methods initially relied on low-level features (color, texture, gradients or saliency); however, more recently, convolutional neural network (CNN) features are being used for proposing object candidates.

The object proposal methods can be divided into two categories: grouping methods, e.g. selective search (Uijlings et al. 2013) and window scoring methods, e.g. Faster R-CNN (Ren et al. 2015). Grouping methods generate proposals by merging superpixels or solving multiple graph-cut problems using diverse seeds. Window scoring methods first generate a large set of windows and then score them using various cues. To achieve good localization, these methods need to either sample windows densely or refine the generated windows. Most current popular object proposal methods are based on window scoring methods.

Earlier object proposal methods had good performance but still required more than a thousand proposals to achieve recall above 0.8 (@ intersection over union overlap (IoU) = 0.5) (Hosang et al. 2016). Initially, the detection classifiers were more expensive but with Fast R-CNN (Girshick 2015) which significantly reduced the time required for the classification by sharing the feature computations for all proposals in an image, object proposal generation started taking relatively more computation time. Faster R-CNN (Ren et al. 2015) was proposed to speed up object detection process. It used a CNN based region proposal network (RPN) for object proposal generation. RPN shared convolutional layers with Fast R-CNN and as a result it required very little additional computations and significantly reduced the overall detection time and made a significant step towards the development of real-time object detectors. It also greatly improved the proposal quality, with it requiring almost an order of magnitude fewer proposals to achieve similar recall as the more traditional object proposal methods. Faster R-CNN has been followed by Single Shot MultiBox Detector (SSD) (Liu et al. 2016), which does away with the object proposals entirely and directly output the detection results. Initial Faster R-CNN and SSD have been improved upon by Feature Pyramid Network (FPN) (Lin et al. 2017) and Deconvolutional Single Shot Detector (DSSD) (Fu et al.
2017), respectively, both of which use feature maps from lower layers to improve the performance for smaller objects. Faster R-CNN has also been extended to obtain instance-level segmentation masks first by Multi-task Network Cascades (MNC) (Dai et al. 2016) and more recently by Mask R-CNN (He et al. 2017).

Another approach for proposing object proposals was taken by DeepMask (Pinheiro et al. 2015) followed by SharpMask (Pinheiro et al. 2016), which proposed segmentation masks instead of bounding boxes. However, since these methods predict segmentation masks densely in the image, they are considerably slower than Faster R-CNN based methods which first regress a small set of bounding boxes and then predict masks for only these boxes. Faster R-CNN based methods have also been able to outperform these methods in detection and segmentation performance so far.

3.1.2 Cell proposals

Cell proposal methods can be grouped into four main categories: multiple settings, shape matching, superpixel merging and deep learning. In many instances, a cell segmentation method can be found to segment almost all cells if its parameters are optimally tuned for local image regions. However, it can be challenging to automatically select these optimal parameters. This insight is used by some cell proposal methods, which repeatedly segment images using different parameters of a segmentation method (thresholding, graph-cut (Jug et al. 2014), watershed (Liu et al. 2014b)), ensuring that they have a good proposal for most if not all individual cells.

Shape matching methods can also be considered as a subset of multiple settings category. In these methods, the varying parameters explicitly control the cell shape and size that the detection method is aiming to detect. As many cell lines and most nuclei have round or elliptical shapes, these methods are targeted towards images containing them. These methods either fit multi-scale ellipses to the contours of binary segmented regions (Türetken et al. 2017) or use multi-scale blob detection (Paper II) to detect potential cells. They work quite well as long as the cell shapes are rigid and the number of cells deviating from these shape constraints is low.

Superpixel merging methods (Funke et al. 2015) first generate a set of over-segmented superpixels, with each superpixel containing pixels from just one cell. The image used to generate superpixels can be the raw intensity image or cell probability map obtained using a pixel classifier (Schiegg et al. 2015a). These methods do not make any assumption about the shapes of cells and have the potential to be used for
processing images containing non-rigid cells. However, since these methods assume that either the intensity or the cell probability map has a strong gradient at the border between touching cells, and merge superpixels based on the strength of gradient at their shared boundary, they can fail in regions where this assumption is not satisfied, which can be common in more challenging images.

Most cell proposal methods belonging to multiple settings, shape matching or superpixel merging categories do not have an inherent way of scoring their proposals. So, these methods represent the generated proposals using some handcrafted shape and appearance features. These features can either directly be used as the score of regions (Kale & Aksoy 2010) or a classifier can be trained to score cell proposals. The shape and appearance features used by these methods are usually simple, e.g. area, mean intensity, histograms of intensity, etc., but some methods (Arteta et al. 2012) use more advanced features as well, e.g. histogram of proposal boundary.

More recently, convolutional neural networks (CNN) have also been used to propose cell candidates. The Cell Proposal Network (Paper III) uses a CNN to regress the bounding boxes of cells relative to the reference bounding boxes in a given image (Ren et al. 2015). Another approach is taken by (Yurchenko & Lempitsky 2017), who first synthetically generate a cell gallery, use this gallery to generate training samples representing typical cells in various configurations, and then use a CNN to learn a descriptor for the cell in the middle of a given image patch. During testing, a descriptor is obtained using the CNN for each input image patch and this descriptor is used to retrieve the likely cell candidates from the cell gallery. This method produces cell shapes and configurations for a given image, which include the recovery of occluded parts of cells, and can be particularly useful in applications where cells have frequent occlusions and their shapes need to be analyzed. This method is used for cells which have relatively simple tubular shapes and synthetic images can be generated easily; it might be difficult to obtain a large cell gallery and generate training samples for other cell types.

3.2 The proposed methods

This thesis presents two cell proposal generation methods. The first method BLOB is based on a popular cell detection technique, i.e. blob detection, and was presented in Paper II as part of a cell tracking method. The second method Cell Proposal Network (CPN) is based on the Region Proposal Network (RPN) (Ren et al. 2015), and its various
versions were presented in Paper III, Paper IV and Paper V. Here, both these proposal generation approaches are briefly described.

### 3.2.1 Blob proposals

Blob detection has been a popular cell detection method as some cells and nuclei have round shape, allowing blob detection methods from general computer vision field to be used to accurately detect these cells/nuclei. An overview of various blob detection methods is given in Section 4.2.1.

Cell proposal generation method BLOB presented in Paper II is inspired from the success of blob detection based cell detectors. It first obtains binary segmentation using a dataset dependent segmentation method. The goal of this initial segmentation is to separate cell pixels from background without missing many cells. It then uses filter banks which contain generalized Laplacian of Gaussian (gLoG) filters (Kong et al. 2013) for detecting cell candidates. gLoG filters are anisotropic versions of the more commonly used Laplacian of Gaussian (LoG) filters and can better detect elliptical cells present at various angles in the image. Since this method does not need to make hard detection decisions, it can use the peaks in responses of gLoG filters in each filter bank as the candidate detections. The cell candidate detections of each filter bank are utilized by watershed algorithm to obtain segmentation masks for these candidates. Multiple filter banks can detect the same set of cell candidates and lead to proposals which have large intersection over union (IoU) overlap with each other. Non-maximum suppression (NMS) is used to remove these duplicate cell candidates. This approach produces a large set of cell candidates, with very high individual cell recall; which are used within the tracking framework described in Section 5.3.2 to select the subset of cell proposals which minimizes the combined cost of the selected proposals and their associations.

The method presented in Paper II extracts a small set of basic shape and appearance features from each cell proposal and trains a support vector machine (SVM) classifier to score the proposals. Replacing these features with more advanced shape and appearance features (Arteta et al. 2012) and using a random forest classifier for scoring increases the performance slightly, as shown in Fig. 2 of Paper III.
3.2.2 Cell proposal network

Cell Proposal Network (CPN) was proposed to address the shortcomings of BLOB proposals, which could only be used if the cells had round blob like shapes. The goal was to have a method which could propose cell candidates after learning the cell shapes and appearances from the available training data. Despite the lack of training data, use of convolutional neural network (CNN) for proposing cell candidates was considered due to the huge success of Region Proposal Network (RPN) proposed as part of Faster R-CNN (Ren et al. 2015).

The main idea behind CPN is to use a CNN to extract features from rectangular windows, and then use these features to regress bounding boxes and their score (representing the probability of them being actual cells). Since these extracted features are translation-invariant, anchor bounding boxes are used to localize the features in the image, and the regressed boxes are relative to these anchors. Anchors are bounding boxes, of various sizes and aspect ratios, centered at the center of the above-mentioned rectangular windows. Multiple anchor sizes are used to prevent excessive loss when cell sizes can vary considerably, with each anchor targeted towards detecting cells within a narrow range. This feature extraction and bounding box regression is performed in a fully convolutional manner.

Original ZF (Zeiler & Fergus 2014) and VGG (Simonyan & Zisserman 2015) models used by RPN network cannot be directly used for proposing cell candidates as they have very large strides and receptive fields. So, the ZF model is modified by changing the number of feature maps in each layer and reducing the strides of convolutional layers. The anchor sizes are also adapted for the microscopy datasets. A larger model is used in the method presented in Paper IV by doubling the number of feature maps in each layer to improve the performance for histological and phase-contrast images, which contain more challenging cell shapes and appearances. The version presented in Paper V selects anchor bounding box sizes automatically using k-means clustering. The automated selection of anchor boxes removes a small manually tuned component, which can have some influence on the performance, especially when there is no appropriate anchor for some subset of cells.

The proposed method in Paper III uses intensity thresholding within the proposal bounding boxes to obtain the cell segmentation masks and works well for nuclei or cytoplasm labeled fluorescence microscopy images. However, this approach is not suitable for challenging images from other microscopy modalities, e.g. histology and
phase-contrast. To extend this method to these and other microscopy images, a second CNN is added in Paper IV to predict segmentation masks for individual cells. This CNN extracts features from a given image, combines feature maps from four different depths similar to FCN (Long et al. 2015), adaptively pools the feature map region within a given bounding box to a fixed size using ROI-Pooling (Girshick 2015) and predicts the segmentation mask for the cell in the center of the box. This network can be used for any microscopy image and leads to slight performance improvement for fluorescence microscopy images. However, it still sometimes fails to ignore parts of other cells in the proposed bounding boxes.

This issue is resolved by the network proposed in Paper V, which uses fully connected layers after the ROI-Pooling to better ignore other cells. The feature computation part of this network is inspired by U-Net (Ronneberger et al. 2015) and can be considered as a smaller version of U-Net as it has fewer convolutional layers and fewer feature maps in each layer. It gradually merges upscaled high-level features with low-level features to produce a feature map with the same size as the input image. Then, ROI-Pooling is used to extract the features within a given bounding box and two fully connected layers are used to predict cell masks. This allows the network to better ignore pixels belonging to other cells at the cost of slightly coarser masks.

3.2.3 Proposal conflicts

Multiple proposals can span the same image region as they provide alternative hypotheses; at most one proposal among these competing proposals can be selected for each region. If the proposals are generated in a hierarchical manner, then they can be sorted in trees and at most one proposal can be selected from each branch. However, if the proposals are generated independently, then it is simpler to restrict selection of at most one proposal from a pair of conflicting proposals (Paper V). In the case of independently generated proposals, it can also be challenging to obtain the conflicting relationship as the proposals for different cells can have some overlap with each other. This can be dealt with by allowing some overlap between selected proposals.

The method presented in Paper II first sorts proposals in trees and then restricts the selection of at most one proposal from each branch. This has a performance penalty as some good proposals are removed while they are being sorted in trees as they conflict with multiple branches of the trees. The simpler approach used in Paper V is more
suited for both BLOB and CPN proposals and leads to slightly better performance for both of them.

3.3 Discussion

BLOB proposal generation method presented in Paper II was evaluated on three datasets containing round/elliptical cells. It is able to separate individual cells from all but the most challenging cell clusters and has high recall rate (Table 1 in Paper II and Fig. 2 in Paper III). Using these proposals for cell tracking leads to considerable improvement for two datasets (Fluo-N2DH-GOWT1 and Fluo-N2DH-SIM+) and slight improvement for the third dataset (Fluo-N2DL-HeLa) compared to previous methods. However, it has a few limitations. First, multiple parameters have to be optimized to achieve the best performance which can be time-consuming. Second, it can only propose good candidates for round/elliptical cells, which limits its applicability to images satisfying this assumption.

CPN was evaluated on all datasets (except the simulated Fluo-N2DH-SIM+) mentioned in Table 1. Its proposed cell candidates have a high average precision (AP) (Fig. 3 in Paper IV and Fig. 4 in Paper V) indicating that they can achieve both high recall and good scoring for all six datasets containing cells of very different shapes, sizes and appearances (Fig. 3). However, the performance across datasets varies considerably with fluorescence datasets having better performance than phase-contrast datasets. There are multiple reasons for this difference in performance: phase-contrast datasets have larger variations in their cell shapes and appearances, the number of training samples is small (Table 1), and training data does not cover all the shapes and appearances seen in the testing images, making it challenging for the network to propose good candidates for some difficult regions.

Many under-segmentation and over-segmentation errors (Fig. 5a- 5b) are in image regions which resemble either a single cell or multiple touching cells, respectively; and without enough temporal context, proposing accurate cell candidates is extremely challenging. However, some other errors are due to few inherent limitations of the CPN. Firstly, CPN uses bounding boxes to represent cell candidates, which is not an appropriate representation when cells have elongated thin shapes and are densely packed. It leads to situations where bounding boxes of multiple cells can have large overlap with each other and even if the network can propose good candidates for these cells, they might be removed during non-maximum suppression (NMS) stage. Representing cell
candidates using a rotated bounding box can potentially mitigate this issue and improve the performance slightly. However, training such a network may introduce additional challenges, e.g. determining how to mitigate the loss for round cells as the orientation loss for these cells can be very sensitive to slight changes in cell shapes.

Secondly, CPN sometimes fails to propose good candidates for small dark cells when they are in close proximity of bigger and brighter cells (Fig. 5c). This issue can potentially be solved by utilizing low-level features (Lin et al. 2017) and/or increasing training samples representing these situations.

Thirdly, in some cases, it proposes multiple proposals for an ambiguous region but since the proposals are generated independently, they do not always explain the whole ambiguous region. In a bright ambiguous region, for instance, which might contain either a single cell or two clumped cells, CPN may just propose one large candidate covering the whole region and one small candidate covering part of this region (Fig. 5d). During the proposal selection, if the smaller candidate is more likely, it will be selected, leaving rest of the ambiguous region wrongly labeled as background despite it clearly being part of some cell.

The datasets used in this thesis are relatively small, as mentioned in Table 1, and as a result, the network architectures tested were relatively small. U-Net has been trained on some of these datasets but it considers each pixel as a single sample and as a result has many more samples in the training data. Whereas, CPN considers a cell bounding box as a single sample and has just few hundred samples and is more challenging to
train. Using auxiliary supervision may improve its performance. To properly train these networks and improve the performance, augmented annotations (Section 2.7) were necessary. Simulated datasets or better augmented masks also have the potential to improve the performance further.

In Paper IV, a single CPN model is trained simultaneously on three datasets from histology, phase-contrast and fluorescence microscopy. This model has slightly lower performance than the models trained individually on single datasets (Fig. 3 and Table 1 in Paper IV). It demonstrates the possibility of a single model to predict candidates for cells which have very different shapes and appearances and opens up the possibility of more general models which are trained on multiple datasets. When similar datasets are used, it may even improve performance by learning from a larger number of samples or it might be challenging to retain the same level of performance as it may ignore dataset-specific features which may be needed for very high performance. Due to lack of annotated datasets, this idea has not been fully explored and the true potential and limitations of a general cell detection/segmentation model remain unanswered.
4 Cell detection and segmentation

Cell detection is the process of locating each individual biological cell in a microscopy image. Often, each detection is represented using a pixel or a set of connected pixels close to the cell center.

Cell segmentation is the process of labeling each pixel in an image with the label uniquely identifying the cell present at that location, with the goal of partitioning the image into multiple regions, with each region representing one individual cell. Many cell segmentation methods use a preprocessing step to remove background regions. In this thesis, this is referred to as **binary segmentation** and is equivalent to semantic segmentation in natural images.

Cell detection/segmentation and object detection/segmentation have different constraints and challenges; as a result, methods used in these research areas despite sharing some underlying techniques differ significantly. In object detection/segmentation, where methods have to perform well in uncontrolled environments, it is difficult to utilize simple cues about object shape or appearance to detect/segment objects from natural images due to the wide variety of poses, appearances, shapes and occlusions that objects can have and the presence of highly variable background. However, cell detection and segmentation methods are often developed for analyzing a small set of images or videos of usually just one type of cell, imaged under highly controlled conditions. This often results in most cells and background satisfying few very strong assumptions. These assumptions can include cells having rigid shape, homogeneous appearance and characteristic intensity profile at their boundaries (Bensch & Ronneberger 2015), and background having homogeneous and distinct intensity. This makes it possible to develop detection and segmentation methods for the target application using long chains of simple image processing operations (thresholding, watershed, morphological filtering, etc.) (Meijering 2012). Most of these methods require tuning of multiple parameters to obtain the best performance, which severely limits their applicability to a small set of microscopy images and as a result most of these methods are not widely used beyond the experiment for which they are designed.

Cell detection and segmentation despite being one of the oldest problems in microscopy image analysis and having received a lot of attention over the years remain largely unsolved. They are often among the first few stages in the single-cell analysis of
microscopy images and their performance has a significant influence on the accuracy of analysis stages downstream (Matula et al. 2015) and impacts the reliability and robustness of the derived biological insights. So, it is vital that cell detection and segmentation methods are as robust and accurate as possible.

4.1 Challenges

Automated cell detection and segmentation are very challenging problems due to large variability in the appearances and shapes of cells. Some factors behind this huge variability include microscopy modality, staining, cell type, cell density, matrix composition, blurring, imaging artifacts and imaging settings (e.g. resolution, excitation light intensity, etc.). Fig. 3 (Chapter 2) and Fig. 6 show sample regions containing various cell types from multiple microscopy modalities highlighting the variations in cell shapes and appearances.

Few important challenges are discussed in detail in individual sections below. Some less critical challenges include the presence of background structures (e.g. parts of dead cells), which may be visible in images and may not always be trivial to ignore; imaging artifacts, such as spectral unmixing errors (Al-Kofahi et al. 2010) or stripes (Lou et al. 2012); and blurring, which makes it difficult to determine where exactly the cell boundaries lie.

Microscopy modality

Cell shapes and appearances in the microscopy images depend heavily on the sample preparation technique and the imaging setup, including the microscopy modality. Bright-field images have low contrast between cells and background, highly textured and inhomogeneous cell bodies, and patches in the cell boundaries which have similar appearance as the background, making it challenging to accurately detect and segment cells.

The difficulty of analyzing fluorescence images highly depends on the staining quality and imaging settings. In some cases, the targeted structures are either not homogeneously stained or lose fluorescent markers over time (photo-bleaching). This can lead to the intensity/color varying significantly within one cell and result in stronger gradients inside a cell body compared to those at the border between touching cells. In some experiments, the intensity of excitation light, image resolution and frame rate have
Fig. 6. Few samples showing the variation in cell shapes and appearances for various microscopy techniques. Cells in each sample are shown at three different scales. (a)-(c) Fluorescence microscopy with cell membrane, nuclei and cytoplasm labeled, respectively. (d) Phase-contrast microscopy. (e)-(f) Differential-interference contrast (DIC) microscopy. (g) Pap smear. Some additional samples are shown in Fig. 3. (a) is from (Saarela et al. 2017), (c) and (e) are from (Ljosa et al. 2012), (f) is from (Ulman et al. 2017), and (g) is from (Lu et al. 2017).

to be finely balanced to minimize photo-toxicity, which can lead to images with low signal-to-noise ratio, making it difficult to detect cells accurately.

In histopathology, where section of a 3D tissue sample is imaged, sectioning can damage nuclei distorting their shapes. In addition, imaging may capture nuclei at odd angles, partially overlapping or superposed on each other due to the thickness of the 2D section (Al-Kofahi et al. 2010). These images often contain nuclei of multiple types (mitotic, non-mitotic), with some nuclei in each category having similar texture, appearance and shape, making it difficult to accurately detect the target nuclei.
**High cell density**

Cellular interactions are relevant for many biological processes. In such models, cell density is high and cells often come in contact with each other and form cell clusters. It can be very challenging to infer how many cells there are in these clusters and where their boundaries lie because cells can have irregular shapes, and appearance within a single cell may vary more than between different cells. Sometimes, cells being imaged have a rigid shape (round or tubular) and this shape constraint can be used to resolve the location of individual cells in these clusters.

**3D: Anisotropy and blurring**

Fluorescence microscopy is increasingly being used to image cells in 3D to observe their more realistic behavior. Images acquired using confocal and light sheet microscopes are often anisotropic with axial (z) resolution 2-8x lower than their lateral (xy) resolution (Weigert et al. 2017). In addition, the light from structures in out-of-focal planes results in higher blurring in the axial direction for confocal and many other 3D microscopy modalities. Deconvolution (Sarder & Nehorai 2006) is commonly used to reduce this blurring, but often even after deconvolution images have higher blurring in axial direction, and in some instances where accurate point spread function (PSF) cannot be obtained, deconvolution can even introduce additional artifacts.

Recently, a deep learning based method (similar to image super-resolution) has been proposed to recover sharp isotropic 3D images from raw blurred anisotropic 3D images (Weigert et al. 2017), and it has the potential to significantly improve cell detection and segmentation for 3D images. However, it is not clear if the sharp images obtained correspond to the underlying 3D structures or are just plausible reconstructions.

**4.2 Related works**

This section gives a brief overview of cell detection and segmentation methods. There is some overlap between cell detection and cell segmentation methods, with cell detection methods playing a significant role in many cell segmentation methods. Here, they are presented separately as in most cases cell detection methods can be replaced in the cell segmentation methods. This section also provide a very brief overview of recent deep
learning based semantic and instance-level segmentation, especially the methods which have high relevance to this thesis.

Detailed reviews of semantic segmentation covering traditional methods can be found in (Zhu et al. 2016) and more recent deep learning based methods in (Garcia-Garcia et al. 2017). An overview of cell segmentation field can be found in (Meijering 2012, Bajcsy et al. 2015), and detailed reviews of cell/nuclei detection and segmentation methods can be found in (Irshad et al. 2014, Xing & Yang 2016).

4.2.1 Cell detection

Cell detection methods despite having detection in their names are often very different from object detection methods. Objects can be broken down into parts, which could then be used to detect objects; however, in the case of cells, at typical resolutions in microscopy images, such decomposition is often not possible. As a result, the methods used for cell detection have more in common with interest point detectors than object detectors. In fact, many interest point detectors (e.g. DoG, LoG, MSER) are often used for cell detection.

Most cell detection methods are designed for round cells and use a combination of simple image processing operations to detect individual cells. These methods can be broadly divided into the following categories based on the main operation used for cell detection: distance transform, blob detection, contour analysis and supervised learning. Some methods combine techniques from multiple categories (Al-Kofahi et al. 2010) to improve performance.

One of the earliest (Malpica et al. 1997) and simplest method for separating individual cells is based on the distance transform of the binary segmented image, which computes the distance of each foreground pixel from the background (Daněk et al. 2009). If the cells have smooth round boundaries and small overlap with each other, the peaks in distance transform image correspond to cell centers. However, when cell boundaries are not smooth, cells have large overlap with each other or cell shapes are irregular, it can lead to detection of too many cells. Cell centers are often brighter than their boundaries, so intensity image alone or in combination with distance transform can also be used to detect individual cells (Wählby et al. 2004).

Blob detection is a popular cell detection approach, which can exploit both the shape and intensity cues simultaneously. These methods use Laplacian of Gaussian (LoG) or Difference of Gaussian (DoG) filters to detect cells, which appear as peaks in the
response of these filters when applied to a microscopy image. Initial methods used a single filter for detecting specific-sized cells (Byun et al. 2006). However, more recent methods use multi-scale filters for increasing robustness to cell size variations. The responses for these filters are either combined into a single image and local peaks in that image are used as detections or local peaks in the scale-space response map are used as detections. These multi-scale methods can detect cells quite reliably; however, their performance can degrade in challenging conditions, where multiple cells are in contact, have very similar intensities and cell sizes vary considerably. In addition, there can be some irregularly shaped cells in an image which are often over-segmented by these methods. Adjusting the scale of filters in local image regions using distance transform (Al-Kofahi et al. 2010) can provide additional cues and better handle the above-mentioned challenges. Often, isotropic filters are used for cell detection; using gLoG filters, which are elliptical version of LoG filters (Kong et al. 2013), can improve detection performance in cases where cell shapes are more irregular.

Contour analysis based methods detect concave arcs on the boundary of the binary segmented regions to find the points where cells touch, and use these points to split cell clusters (Farhan et al. 2013). Hough transform (Ortiz-de Solórzano et al. 2001, Dimopoulos et al. 2014) and ellipse fitting (Türetken et al. 2017) are few other techniques used to separate touching round/elliptical shaped cells.

Supervised learning based methods have been used widely for nuclei detection in histological image analysis but their use has been limited for live-cell microscopy images. Initial methods classified windows in an image in a sliding window manner (Cireşan et al. 2013); however, more recent methods regress a heatmap of potential cell locations (Xie et al. 2015b,a). There are also some methods which use supervised learning to score cell candidates and then select a subset of these candidates to provide cell detection. These methods are covered in Section 3.1.2.

Cell detection methods can sometimes lead to too many detections for each cell. This issue can be mitigated by morphological filtering of binary masks to make cell boundaries smooth before applying distance transform, using h-maxima transform to remove local maxima and/or clustering nearby detections (Al-Kofahi et al. 2010). Morphological operations (repeated erosion) of binary segmented image can itself be used to separate touching cells in some cases.
4.2.2 Cell segmentation

One of the most common approaches for cell segmentation is cell segmentation by detection. It first detects individual cells (Section 4.2.1), and then uses these detections to delineate boundaries of individual cells. This delineation can be obtained by either splitting clumped regions in the initial binary segmentation (Al-Kofahi et al. 2010) or by clustering pixels/superpixels around cell detections. Another cell segmentation approach is based on proposal selection and it has become more prominent in the last decade. These methods, first, generate a large set of cell segmentation candidates (Section 3.1.2) and then select a subset of non-conflicting candidates (Section 4.2.4) to obtain the segmentation. There are other methods which do not fit into the above-mentioned categories. For example, some methods, especially the recent deep learning-based methods, segment an image into two classes and then use connected components to find individual cell masks (Bensch & Ronneberger 2015, Ronneberger et al. 2015).

Intensity thresholding was the first method used for cell segmentation and it still remains very popular (Meijering 2012); especially in software targeted at biologists due to its simplicity, reasonable performance in many applications and lack of any other general user-friendly method. It makes two assumptions; first, the cells have different intensities than the background, second, most cells are not in contact with each other. The first assumption is reasonable for many microscopy images, however, the second assumption is often not met in images with high cell densities and post-processing is typically used to separate individual cells. Many ways of selecting thresholds have been presented over the years, with OTSU thresholding (Otsu 1979) which maximizes inter-class variance being the most popular method. It is still sometimes used as a baseline. Many local adaptive thresholding methods (Kenong Wu et al. 1995) have been presented over the years to better handle the local variations in intensities due to noise and non-uniform illumination. The newer methods, which still use thresholding, use it as a preprocessing step to separate cell pixels from background, followed by splitting cell clusters using some cell detection method.

Watershed transform is another popular cell segmentation method. It considers the image as a topological map with intensity of the relief image representing the elevation of each pixel. It starts flooding this relief image from each local minimum with the flow spreading the labels of these local minima to all pixels in an image. The pixels which contain the flow of same label and form one connected component are treated as the segmentation of the marker at the local minima with the same label. Watershed is usually
applied to gradient, edge probability, cell probability, distance transform, gray-scale intensity or any combination of these images. One major drawback of watershed is its sensitivity to noise, which often leads to over-segmentation. However, many methods have been presented which use carefully designed preprocessing and post-processing stages to achieve good performance on specific microscopy images. These stages include marker-controlled watershed (which floods an image only from the provided marker locations), region merging after watershed, and using smoothing or h-minima transform to get rid of small local minima in the relief image.

Graph-based methods represent the image as a weighted graph, with each pixel or superpixel represented by a node, and edges between neighboring nodes representing their similarity. These methods try to find a node labeling which minimizes some energy function, which typically includes a unary and a pairwise term. The unary term represents the local node evidence, which can be the image intensity or the output of a classifier, and the pairwise term represents the similarity of connected nodes, usually using the output of some edge detector. Additional terms can be included in the unary and pairwise terms to incorporate additional prior information. For example, distance transform (Daněk et al. 2009) or a Gaussian model of cell shape (Al-Kofahi et al. 2010) can be used to compute the shape prior for round cells. This can be included by adding a term in the unary term which heavily penalizes other labels close to a cell marker (Daněk et al. 2009) or in the pairwise term with heavy penalty for different labels for connected pixels if not aligned with the vector field originating from cell markers (Lou et al. 2012). Multi-label graph-cuts are often used to separate touching cells, and binary-label graph-cuts are often used for separating cells from background as a preprocessing step, followed by cell cluster splitting. However, some methods use binary-label graph-cuts to segment individual cells by utilizing characteristic intensity changes at cell boundaries (Bensch & Ronneberger 2015) to ensure that each cell is segmented as a separate connected component. Intensity profile at cell boundaries can also be learned (Dimopoulos et al. 2014) and used to obtain the cell boundary probability map, which can be used as a pairwise term when segmenting individual cells.

Deformable models are also popular in cell segmentation research. Common models are defined either explicitly (active contours) or implicitly (level-sets). These methods optimize some energy function similar to the graph-cut based methods, which typically consists of an appearance term based on image intensity, gradient or texture, and a shape term based on boundary length, surface area or curvature of boundary (Meijering et al. 2009).
Another popular class of methods includes supervised learning based methods. A cell classifier is trained, often using Ilastik (Sommer et al. 2011), from a small set of manually labeled pixels and this classifier is then used to obtain a cell probability map. This cell probability map is then used instead of intensity image in other segmentation methods (e.g. watershed) or as a unary image in graph-cuts.

Most cell segmentation methods (watershed, thresholding, graph-cut, etc.) can be easily extended to 3D. However, due to anisotropic nature of typical 3D microscopy images, straight-forward extension to 3D may not always work. One common approach for dealing with this challenge is to interpolate slices in axial direction to obtain isotropic 3D image (Wählby et al. 2004). This can deal with anisotropic resolution but it might still be challenging to deal with anisotropic blurring and special adjustments might still be needed to ensure that the segmentation algorithm works as expected. Another popular approach is to process a 3D image as a set of 2D images, followed by a post-processing stage. These methods segment each slice in the 3D image independently as if it is a 2D image. Then, the segmented regions in neighboring slices are merged to obtain 3D segmentation (Stegmaier et al. 2016, Vázquez-Reina et al. 2011).

4.2.3 Deep learning: Segmentation

One of the first applications of convolutional neural networks (CNNs) for segmentation was neuron segmentation from electron microscopy (EM) images (Jain et al. 2007). It was also one of the initial successes of deep learning for segmentation, when in 2012, a deep learning based method (Cireşan et al. 2012) won the challenge about neuron segmentation from 2D EM images (Arganda-Carreras et al. 2015). However, initial deep learning based object segmentation methods (Girshick et al. 2014, Hariharan et al. 2014) used CNNs only for scoring proposals. These methods used an object proposal method for generating a small set of object candidate regions, scored them using a CNN, and then selected a subset of them using non-maximum suppression (NMS).

Another major success for deep learning for segmentation was Fully Convolutional Network (FCN) (Long et al. 2015), which merged features from multiple depths in the network and learned to predict the label for each pixel in an image in a fully convolutional manner. It is still the dominant semantic segmentation approach, and many variants with bigger networks, dilated convolutions (Chen et al. 2015), gradual merging of features (in an encoder-decoder manner) and conditional random fields (CRFs) for post-processing have been presented (Garcia-Garcia et al. 2017).
Among the instance-level segmentation methods, DeepMask (Pinheiro et al. 2015) was the first major success which used a CNN to propose individual object candidate masks, which could then be used for instance-level segmentation. However, it was soon followed by Faster R-CNN based segmentation methods, MNC (Dai et al. 2016) and its successor Mask R-CNN (He et al. 2017), which is at present the best performing method. These methods segment each candidate box proposed by region proposal network (RPN) and then use NMS to select the non-conflicting masks and obtain instance-level segmentation.

Recently, few deep learning based cell segmentation methods have been proposed (Song et al. 2017, Kraus et al. 2016). The most popular among them is U-Net (Ronneberger et al. 2015), which has an encoder-decoder structure. Its first few convolutional layers extract increasingly coarse high-level features, which are then upsampled using deconvolutional layers and gradually merged with finer low-level features from lower layers to obtain increasingly higher resolution feature maps, which contain both low-level and high-level information. Finally, the last feature map is used to predict the probability of each pixel belonging to some cell. Since U-Net is in essence a semantic segmentation network and only outputs binary class probabilities, it gives higher priority to pixels near cell-cell borders during training to separate individual cells as distinct connected components. U-Net has been extended to 3D (Çiçek et al. 2016) and has been applied to many other segmentation tasks in biomedical and medical image analysis. It is able to produce very accurate cell masks but it has limitations when it comes to separating touching cells. Deep Contour-Aware Network (DCAN) (Chen et al. 2017) is another method which uses FCN (Long et al. 2015) like architecture, but instead of just predicting the cell segmentation, it also predicts cell boundary probabilities, which are then used to create a wedge between touching cells, and this allows it to better separate individual cells from within cell clusters.

Most deep learning based cell segmentation methods are typically evaluated on a single set of images. (Van Valen et al. 2016) is one of the few recent works which have investigated the performance of deep learning for multiple diverse datasets.

### 4.2.4 Proposal selection

In object detection, the detectors produce high scores even when an object is partially visible as a side effect of having to deal with occlusions and pose variations or when it is not centered in the window to cope with localization errors in the window. These
factors often lead to multiple detections near each object of interest, which have varying overlaps with the actual object. These duplicate detections are removed in a post-processing step referred to as non-maximum suppression (NMS). The standard method (Greedy-NMS) used for NMS, repeatedly selects the highest scored detection and removes any unselected detection which has an overlap greater than a threshold with the selected detection. Greedy-NMS is widely used due to its simplicity and good performance. It uses a fixed overlap threshold for discarding detections, which can lead to removal of objects which are close to each other and overlap between their bounding boxes is greater than the selected threshold.

In recent years, some methods based on clustering (Rothe et al. 2014) and re-scoring detections (Hosang et al. 2017) have been presented which are able to make better selection but have not gained popularity due to their complexity and minor performance improvements. One recent simple and promising approach is Soft-NMS (Bodla et al. 2017), which instead of removing a detection lowers its score depending on the overlap and as a result can detect high scored detections separately even when they have moderate overlap.

In object detection, the main challenge is to select one detection among the many produced for each object and it is not very common to find situations (excluding some specific objects, e.g. people) where there might be confusion about whether an image region contains one instance or multiple instances of an object. As a result, the detections for each object can be considered independent of those for other instances of that object. However, in microscopy images, the main challenge lies in determining the number of cells in ambiguous regions, which themselves can often be separated from background quite reliably. In these cases, selecting one cell in a region can provide strong evidence for the presence/absence of other cells. As a result, more advanced methods are used which can consider alternative hypotheses for these regions using either Integer Linear Programming (ILP) (Bise & Sato 2015, Liu et al. 2014a) or dynamic programming (Arteta et al. 2016) when selecting cell detections.

### 4.3 The proposed methods

This thesis presents two cell segmentation methods. The first method presented in Paper I utilizes an edge detector to create a barrier between touching cells and then uses graph-cut to detect individual cells. This method is briefly presented in Section 4.3.1. The second method is based on the proposal selection approach and its various versions
were presented in Paper III and Paper IV. It can use the cell candidates generated by methods presented in Paper II or Paper V as well and is briefly introduced in Section 4.3.2.

4.3.1 Graph-cut

The cell segmentation method presented in Paper I is designed to segment cells with highly deformable shapes from fluorescence microscopy images (Fig. 9 in Paper I). This method makes two main assumptions; first, cells are more or less homogeneously stained and have higher intensity than background; second, when cells come in contact with each other, there exist some distinctive features at their mutual boundary compared to regions inside cell bodies.

This method uses a common three step approach used by cell segmentation methods; first, cell pixels are separated from the background; second, markers for individual cells are detected; and finally, detected markers are used to split cell clusters and obtain cell masks. The main novel contribution of this method is the marker detection stage, which first uses a state-of-the-art edge detector *Structured Edge* (SE) (Dollár & Zitnick 2013) to obtain directional edge probability maps, uses these maps to create a wedge between touching cells within cell clusters, and then finds binary segmentation using graph-cut (Kolmogorov & Zabih 2004). Each connected component in this binary segmentation corresponds to a seed for one cell and is expanded to obtain the segmentation for that cell using watershed.

This method can be directly extended to 3D but due to anisotropic nature of 3D images, the direct extension does not always work well. As a result, this method is applied to 3D images by first segmenting each slice in the 3D image independently, and then connecting the segmented regions in adjacent slices to obtain the 3D segmentation. The stitching is guided by the maximum intensity image - peak intensity for each pixel among all slices. First, the maximum intensity image is segmented and the best matching masks among all slices are found for each segment in it; then, those best matched masks are grown in axial direction by connecting them with masks in adjacent frames if they have similar appearance and high overlap.
4.3.2 Proposal selection

The cell segmentation methods presented in Paper III and Paper IV are based on proposal selection approach. First, candidates for individual cells are generated using the cell proposal methods described in Section 3.2. Then, cell segmentation is obtained by selecting a subset of non-conflicting proposals, either from each image independently or from the whole sequence jointly. Cell tracking methods presented in Paper II and Paper V use the whole sequence to select the greedy locally optimal and globally optimal subset of cell proposals, respectively. These proposal selection methods are described briefly in Section 5.3. Here, only the methods used in Paper III and Paper IV, which do not utilize temporal context, are presented.

Both these methods use greedy non-maximum suppression (Greedy-NMS) for selecting the proposals. The method presented in Paper III requires setting two manual parameters, intersection over union overlap (IoU) and score thresholds, and achieves good performance. Ideally, IoU threshold ought to be 0; but, since there are some mislabeled pixels in dense regions, using a low IoU threshold often leads to the removal of candidates for some cells when they happen to be in contact with other cells. The main reason for mislabeled pixels is the small training set, which contains many instances of well separated cells but few instances of them in contact with each other. This leads to the network often predicting the occluded cell body as part of the proposal (Fig. 7c shows an example). As a result, the IoU values depend on the cell shapes and their overlap with each other. The pixels which are common between the selected proposals are assigned the label of the proposal with the highest score. Ideally, proposal score threshold of 0.5 should be good. However, since the sampling of positive and negative samples during training is biased towards positive samples, it leads to high scores for many candidates in background regions. The detection method in Paper IV learns both these parameters (IoU and score threshold) from the training data, which improves the performance slightly.

4.4 Discussion

The main contribution of the segmentation method presented in Paper I is its ability to handle arbitrary cell shapes. It relies heavily on the performance of the edge detector and given good edge probability maps, it can produce very accurate cell boundaries for some very challenging deformable cells (Fig. 9 in Paper I). Its main shortcoming is that
it requires manually tuning multiple parameters to obtain good performance. In addition, since it relies on the maximum intensity image to fuse the individually segmented slices, it can only be applied to images containing thin section of the 3D sample, such that there are not many instances of cells overlapping each other in the axial direction.

This method uses an off-the-shelf edge detector, which is trained on natural images. In the last few years, better deep learning based edge detectors have been proposed and incorporating them in this method should lead to improved performance. These better edge detectors can also be trained on microscopy images, as has been done by (Liu et al. 2015) or the edge probabilities from a method such as DCAN (Chen et al. 2017) can be used, and they should improve the performance of this method.

The cell segmentation methods presented in Paper III and Paper IV are used only to compare the performance of the cell proposals based approach with previous methods. The use of NMS without any post-processing already produces more accurate cell masks and better detections compared to the baseline methods (Table 1 and Fig. 3 in Paper IV), so better proposal selection and post-processing have not been thoroughly investigated. Performance of these methods can be slightly improved by using a multi-label graph-cut to refine the final segmentation and better resolve labels for overlapping pixels. The cell proposal probability maps and distance transform can be used to set unary edge costs and difference in neighboring pixel probabilities can be used to set smoothness costs.
Cell tracking is the process of locating a moving cell over time and detecting cellular events (move, mitosis, death, entering and exiting the field of view) in a given microscopy video. It involves detecting cells in each frame, associating them in neighboring frames, linking parent and daughter cells, initiating new tracks for daughter cells and cells which move in to the field of view, and terminating tracks for cells which die or move out of the field of view.

Cell tracking is a specific application area within multiple object tracking (MOT) field. MOT research has mostly focused on tracking people (e.g. pedestrians, sports players, etc.) and vehicles, with very few methods for tracking other types of objects (e.g. animals). Even though the main task of general MOT and cell tracking is the same, i.e. to associate objects between adjacent frames, the methods used for this are usually different. This is mostly due to the different nature of objects (cells vs humans) and imaging conditions which results in different constraints and challenges (e.g. cell division, cell death).

Most cell tracking research is focused on answering the immediate biological questions, and methods are optimized for the target videos. These methods often make stringent assumptions about the microscopy videos, e.g. cells in adjacent frames have some overlap with each other, links to only a few detections in adjacent frames are sufficient to cover most cell moves (Schiegg et al. 2015a), and cells cannot enter or leave the field of view (Magnusson & Jaldén 2012). These assumptions are often not satisfied and this makes it difficult to apply these methods to new microscopy videos. In addition, often multiple parameters have to be manually set, which can be challenging and require substantial effort to obtain best results.

5.1 Challenges

Cell tracking is a very challenging problem due to low frame rate; sudden changes in cell velocity, shape and appearance; similar appearances and shapes of most cells; detection of cellular events, etc. This section briefly discusses few important challenges.

The number of cells being tracked can vary from frame-to-frame due to cells entering/leaving the field of view, cell deaths and cell divisions. Cells can enter and exit...
from the axial direction; in a 2D video, it means that the tracks can initiate or terminate at any location in an image.

Cells can divide (mitosis) or die at any time in the microscopy videos. The detection of these cellular events is important for cell tracking as it allows the tracks to be initiated and terminated at appropriate time and position. Under basal conditions, cell mitosis and death are often relatively rare events in a microscopy video, which means that often there is a very limited number of samples of these events in the training data, increasing the difficulty of learning an accurate model.

In long term experiments, phototoxicity and photo-bleaching concerns for fluorescence imaging limits the acquisition frequency and signal-to-noise ratio of the captured images, which makes it challenging to reliably segment and track cells (Hilsenbeck et al. 2016). Many cell tracking methods assume that the frame rate is high enough so that most cells have at least some spatial overlap with their position in the last frame. However, this assumption does not always hold true and a general tracking method has to cope with large movements between frames.

Cells can display various motility (movement) modes. They can move individually in several morphological migration modes (mesenchymal, amoeboid) based on the cell type, cytoskeletal structure, and protease production (Masuzzo et al. 2016). They (epithelial and mesenchymal cells) can also move in groups retaining cell-cell contacts with shared responsibilities (Theveneau & Mayor 2013). However, often the frame rate is low, with the time between adjacent frames being in order of minutes, which results in cells frequently and abruptly changing their movement direction and speed. This makes it difficult to model these properties and predict the location of cells in the next frame, which in turn increases the difficulty of the association task. As a result, cell tracking methods often have to assume Brownian motion. However, there are some limited scenarios, e.g. in "wound healing" - migration assays, where cells have a more directed motion and simple motion models can be used to improve tracking performance (Kaakinen et al. 2014).

Cells in a microscopy video often have similar shapes and appearances, and are sometimes indistinguishable from other cells in their vicinity. Tracking is further complicated by the fact that cells can be highly deformable and their shapes and appearances can change considerably over time depending on multiple factors, including their internal state. This makes it difficult to correctly associate detections in neighboring frames as there may be other confounding detections in these frames.
5.2 Related works

This section gives a brief overview of the cell tracking field. Reviews of multiple object tracking (MOT) can be found in (Luo et al. 2014), visual object tracking (VOT) in (Smeulders et al. 2014) and object tracking, in general, in (Yilmaz et al. 2006). There is no extensive recent review of cell tracking field, however, some brief overviews can be found in (Meijering et al. 2009, 2012) and reports from recent challenges about particle tracking and cell tracking in (Chenouard et al. 2014, Maška et al. 2014, Ulman et al. 2017).

Cell tracking is the most popular application area within multiple object tracking (MOT) field, primarily due to the ease of cell detection compared to other objects (e.g. people), which has enabled cell tracking methods to be used for the analysis of cell behavior at population level from microscopy videos for decades (Levine et al. 1980). These methods are often heavily engineered to achieve the acceptable performance level for the target application. People tracking on the other hand is the most actively researched application area in the tracking community (Milan et al. 2016) and has received much more focus when it comes to the development of new and improved association, appearance, motion and interaction models.

Cell tracking has received a lot of attention in the last couple of decades (Hilsenbeck et al. 2016, Eliceiri et al. 2012, Meijering et al. 2012, Ulman et al. 2017, Wiesmann et al. 2015) and many cell tracking methods have been presented. Most of these methods can be grouped into two broad categories: tracking by model evolution and tracking by assignment.

Tracking by model evolution methods perform segmentation and tracking jointly by fitting cell models to the image data. These methods represent cells using either a parametric (Dufour et al. 2011, Amat et al. 2014) or implicit (Dzyubachyk et al. 2010) model. Parametric models use active meshes (Dufour et al. 2011), active contours (Zimmer et al. 2002), or Gaussian Mixture Models (GMM) (Amat et al. 2014) to represent cells and require separate steps for handling topological changes, i.e. cell division, cells touching and cells entering/leaving the field of view. Implicit methods on the other hand represent the cell contours using the levelset of a function and can naturally handle topological changes but are computationally expensive. The main advantage of model evolution methods is that they can provide accurate cell boundaries even in noisy regions, which can be very beneficial when the morphological changes in cells have to be analyzed (Bajcsy et al. 2015).
Most cell tracking methods belong to the tracking by assignment category. These methods decouple cell tracking from cell detection, and in principle can be used with any other cell detector. These methods have two main stages; in the first detection stage, cells are detected using a sequence-specific method and these detections are represented using features which capture their shapes and appearances; in the second stage, detections in neighboring (often only adjacent) frames are linked with each other to create a graphical model representing track hypotheses, and cell tracks are obtained by either considering the whole graph (Magnusson & Jaldén 2012, Lou et al. 2014) or just neighboring frames (Padfield et al. 2011) or a combination of both of these using tracklets (Kanade et al. 2011). Using larger temporal context when associating detections often leads to more robust tracks as local detection failures can be better handled; however, due to computational reasons, initial methods used only neighboring frames. Tracklet based methods, which first create small high confidence track fragments by only considering neighboring frames and then associate these tracklets using the whole sequence to obtain final tracks, allow the use of larger temporal context efficiently and improve performance. In recent years, methods are increasingly considering all detections in the whole sequence when associating detections in the neighboring frames for all but the densest and longest sequences.

Recent cell tracking methods (Bise et al. 2011, Schiegg et al. 2013, Magnusson et al. 2015, Haubold et al. 2016) often include special handling of detection errors in the tracking graph during association stage and are able to resolve false positives, false negatives and under-segmentation errors. However, to resolve these errors, these methods have to introduce additional complexities, which can make the tracking graph complicated and they can still fail when the detection errors span multiple frames.

Recently, few cell proposal selection based methods (Jug et al. 2014, Schiegg et al. 2015a, Türetken et al. 2017) have been proposed for cell tracking. These methods utilize cell proposals to represent potential cell hypotheses and defer the detection decision till the tracking stage. It has few advantages; first, shape and appearance features of each cell candidate instead of a detection (which may contain 0 or more cells) can be used for better scoring of alternative detection and association hypotheses in ambiguous regions; second, in many sequences, the number of ambiguous regions is relatively small, which means that inference can still be fast; third, these methods are simpler than those which incorporate error correction in the tracking graphs; fourth, these methods can enable rapid interactive error correction during track proof-reading as alternative detections can
be viewed and often only the selection of proposals and association edges in local video patches have to be modified to incorporate the user feedback.

The proposal based methods have three main stages; first, they generate a large set of cell proposals covering multiple hypotheses for ambiguous regions; second, they connect proposals in a spatio-temporal graph to represent cellular events and prevent selection of conflicting proposals; third, they select a subset of cell proposals and edges connecting them, providing cell tracks. These proposal based tracking methods are computationally more expensive than the detection based methods as the number of proposals and edges linking proposals can be considerably higher for very challenging sequences. However, using better proposal generation methods and/or tracklets can reduce the computational demands significantly. Another alternative is to restrict the possible transitions between frames to speed up inference for long sequences (Schiegg et al. 2015a), which may lead to lower performance for sequences with low frame rate and high cell density.

5.3 The proposed methods

As part of this thesis, two cell tracking methods, presented in Paper II and Paper V, are developed. These methods utilize cell proposals to represent alternative hypotheses in ambiguous image regions and pose cell tracking as a proposal selection problem. The main observation motivating these methods is the fact that most of the cell tracking mistakes originate from errors in cell detection (Matula et al. 2015). Detection errors can be corrected during or after tracking as a post-processing step, but this makes the tracking methods more complicated and less likely to generalize. A better and more fruitful approach is to focus on the detection itself. However, cell detection is a very difficult problem due to various challenges mentioned in Section 4.1. Cell proposals decompose cell detection into two simpler tasks: cell proposal generation (Chapter 3) and cell proposal selection; this approach has few advantages, as mentioned in Section 5.2.

Section 5.3.1 describes how a proposal graph \((G)\) can be used to represent the alternative cell tracking hypotheses, Section 5.3.2 explains an iterative proposal selection method presented in Paper II and Section 5.3.3 presents an Integer Linear Programming (ILP) based proposal selection method used in Paper V.
5.3.1 Tracking graph

Given a microscopy video (Fig. 7a), the goal of a cell tracking method is to recover the underlying cell tracks, which can be represented using a directed graph (Fig. 7b). Each cell detection is represented by a node in this graph, and two special nodes (S and T) are used to initiate and terminate cell tracks. Five types of directed edges are needed to represent all cellular events (move, division, death, enter and exit). Fig. 7b shows the four types of edges that we use, death edges are not used as the microscopy videos we work with do not contain annotations for cell death events.

Given a set of cell proposals \( P_G \) (Fig. 7c), alternative hypotheses for cell tracks are represented using a graph, \( G \) (Fig. 7d). Each proposal \( p_i \) is represented by a node, which has a cost associated with it \( C_{p_i} \), computed using the probability of that node being an actual cell, \( P(p_i = 1|I_i) \).

\[
C_{p_i} = -\log\left(\frac{P(p_i = 1|I_i)}{1 - P(p_i = 1|I_i)}\right).
\] (1)
Undirected edges between conflicting cell proposals are used to prevent the selection of conflicting proposals. The cellular events are represented by the directed edges \( (E_G) \) in the graph \( G \). All cell proposals are linked with \( S \) and \( T \) nodes using enter and exit edges. Each cell proposal is linked, using move edges, with all proposals in the next frame, which are within a fixed gating threshold from it. All cell proposals are also linked with a few daughter proposal pairs using mitosis edges. The cost \( (C_{e_{ij}}) \) of an edge \( (e_{ij}) \) corresponds to the probability of the cellular event represented by it. The details of how these probabilities are computed for all cellular events are provided in Section 4.3 of Paper II and Section IV-B of Paper V.

Each lineage tree with its root connected to \( S \) node and all its leaves connected to \( T \) node represents one cell lineage tree hypothesis. The nodes in these lineage trees can have two child nodes only at a parent proposal, if the edges to a daughter proposal pair are selected; otherwise, each node can have only one child node. Each set of these non-overlapping trees represents one hypothesis of the underlying cell tracks \( (g = \{P_g, E_g\}) \) for a given video.

Cell tracking can then be formulated as the selection of the sub-graph \( (g^*) \) (solid nodes/edges in Fig. 7d) from the proposal graph \( (G) \) which minimizes the cost (maximizes the probability) of the selected sub-graph.

\[
g^* = \text{arg min}_{g \subseteq G} (\text{cost}(g)).
\] (2)

Assuming the probabilities of cell proposals and cellular events satisfy Markov assumption, and are conditionally independent given the images, (2) can be written as (3) (see Appendix 1) (Berczla et al. 2011):

\[
g^* = \text{arg min}_{g \subseteq G} \left( \sum_{p_i \in P_g} C_{p_i} + \sum_{e_{ij} \in E_g} C_{e_{ij}} \right).
\] (3)

In the absence of cell division events, (3) can be solved optimally in polynomial time using min-cost flow solvers (Haubold et al. 2016). However, when the cells can divide, the optimization becomes NP-hard as the flow in multiple paths (daughter tracks) becomes coupled (Haubold et al. 2016). In these cases, approximate flow-based solutions can be obtained in linear time (Haubold et al. 2016) or Integer Linear Programming (ILP) (Türetken et al. 2017) can be used to find the optimal solution for small graphs.
5.3.2 Iterative shortest path

The method presented in Paper II optimizes the cost function (3) by extending the idea of iteratively building cell tracks (Magnusson et al. 2015), by replacing cell detections with cell proposals. It iteratively finds the shortest cost path in the proposal graph \((G)\) and adds the track found to the tracking graph. The shortest cost path at each iteration is optimal with respect to the remaining proposal graph but not the original proposal graph \((G)\) as some associations and proposals are excluded due to the previously selected paths and this can lead to lower performance. However, by enabling alterations to the previously selected tracks using the swap nodes (Magnusson et al. 2015), which allow current and previous tracks to be split and swapped, can mitigate this issue. However, in the case of cell proposals, the selected proposals cannot be reset and this limits the performance of this method.

The proposal graph used in Paper II differs slightly from the proposal graph introduced above. It includes move edges between proposals in non-adjacent frames to allow it to recover from false negative (FN) cell detections. If a move edge between proposals in non-adjacent frames is selected, then linear motion is assumed to locate the cell’s position in skipped frames, and the proposal in the frame before the jump is placed at those locations.

5.3.3 Integer linear programming

The optimal solution to the cost function (3) can be obtained using Integer Linear Programming (ILP) (Türetken et al. 2017), as presented in Paper V. A binary variable is introduced for each node (except nodes \(S\) and \(T\)) and directed edge in the proposal graph \((G)\). This variable indicates whether the node/edge is part of the solution (True) or not (False).

In order to restrict the solution to feasible tracks, four types of constraints are used:

1. Only one cell proposal can be selected from a set of conflicting proposals to prevent multiple cells occupying the same image region.
2. Each selected proposal must have one incoming edge so that it is part of some track.
3. Each selected proposal must have one out-going edge (the edge from a parent to its second daughter cell is not considered as an outgoing edge).
4. If a mitosis edge is selected, then the mitosis edges from the parent to both daughter cells must be selected.
The ILP program representing the cell tracking function (3) can be very large with millions of variables and constraints. However, it can still be solved efficiently using off-the-shelf solvers such as Gurobi\textsuperscript{15} as it can be broken down into smaller subproblems, which can be solved quickly because in many regions there are only a few good edges/proposals and these local selections are often not influenced by the selection far away in both space and time.

5.4 Discussion

Cell tracking methods presented in Paper II and Paper V achieve better tracking performance than existing methods on a few benchmark sequences (see Table III in Paper V for breakdown of all tracking errors), with the combination of CPN proposals and ILP optimization (Paper V) performing best. The CPN based tracker considerably reduces the false negative detection and all association errors for all fluorescence sequences. However, for some phase-contrast sequences with low cell density, it produces coarse cell masks and has higher detection errors.

The method presented in Paper II finds an approximate solution, which cannot make changes to the cell proposals selected by the previous paths. A better approximate solution can be found using the network flow approach (Haubold et al. 2016) and allowing two types of flows. The first type of flow, originating from $S$ node and terminating at $T$ node, can represent the cell tracks as is common in network-flow based tracking methods. The second type of flow can be used to block or unblock cell proposals which conflict with selected or unselected cell proposals, respectively, by flooding or un-flooding all connected (conflicting) proposals through black undirected edges (Fig. 7d). However, since the focus of this thesis is not on reducing the computational demands of the algorithms and the globally optimal solution (using ILP) for all sequences used in this thesis can be found in few minutes, this has not been explored.

Despite using cell proposals, two types of detection errors, under-segmentations (US) and false positives (FP), dominate the tracking results. Many FPs are challenging to deal with as they are due to small errors in the cell segmentation masks, presence of low contrast structures which resemble cells, and over-segmentation of cells with uncommon shapes (Fig. 5 in Chapter 3 shows examples). These errors can be reduced by better modeling the probability of cells entering/exiting the field of view and improving

\textsuperscript{15}http://www.gurobi.com/products/gurobi-optimizer
cell proposals. Both of these tasks are challenging primarily due to the limited number of samples in the training data.

Many under-segmentation errors are due to cell clusters which have similar shapes and appearances as single cells (Fig. 5a in Chapter 3 shows examples) and it can be very challenging to develop a cell proposal method which can propose individual cells for these clusters using only a single frame. To handle this challenge, it is important for a cell tracking method to incorporate the ability to fix these errors in the cell proposals. One method for resolving these errors is to allow multiple cells within each cell proposal and move edge (Schiegg et al. 2013). This requires a slight change in the ILP program used for tracking. One possibility could be to introduce a binary variable for different number of cells in each proposal. These variables can then be used to dynamically adjust the number of in-coming and out-going edges at each cell proposal node. The main challenge with this approach is the computation of the probabilities of these additional hypotheses which might not be trivial.

One additional component of the method in Paper V which has the potential to be improved is the handcrafted feature set used to represent cell shapes and appearances. This not only limits the performance, but also might not transfer well to other microscopy images. The features from the first fully connected layer in the CPN-SEG network (bottom half of Fig. 1 in Paper V) were used instead of handcrafted features to represent cell proposals, but they had slightly lower performance. This might be due to the fact that CPN-SEG focuses on segmentation and is not trained to differentiate cells from background or other cells, and as a result, the learned features are only good for representing cell shapes and do not represent the appearance variations. Incorporating a small sub-network in CPN which can learn to differentiate cells can potentially provide features suited for representing cell shapes and appearances, and remove another handcrafted component.
6 Discussion and future outlook

This chapter discusses the limitations of the current cell detection, segmentation and tracking methods, and gives an outlook of the future research directions which require more attention. More specific discussion about the methods presented in this thesis can be found in last sections of Chapter 3-5.

6.1 Generalization

The cell detection, segmentation and tracking research areas are dominated by methods which consist of long chains of simple image processing operations. These methods have limited usability beyond the experiments for which they are developed and often require substantial manual effort to tune their parameters to achieve good performance on new images (Meijering et al. 2016). As a result, simple techniques are still widely used by biologists to analyze microscopy images (Wiesmann et al. 2015, Boutros et al. 2015). This limits the research questions which can be answered, and the reliability and robustness of the results obtained suffers consequently.

One important focus of the methods presented in this thesis was their reusability and generalization capability, so these methods were evaluated on multiple datasets. However, these datasets contain only a small fraction of challenges that might be encountered during analysis of new microscopy images. It is important that more focus is invested on the reusability, generalization and robustness of the developed methods, without which this field is unlikely to make rapid progress in near-term and niche solutions will continue to dominate in new application areas, with very little contribution to the field as a whole.

6.2 Datasets

Public datasets play a vital role in the computer vision field, with ImageNet challenge (Russakovsky et al. 2015) enabling the current resurgence of interest in this field. These datasets enable 1) training of large models, 2) objective comparison between methods, and 3) drive research focus and interest to particular tasks. In recent years, Pascal VOC (Everingham et al. 2015), ImageNet challenge and Microsoft COCO (Lin et al. 2014) datasets have been the driving force for the rapid progress in image classification,
object detection, instance-segmentation and many other adjacent research problems. Challenges and public datasets have also accelerated progress for some research areas within biomedical and medical image analysis field, e.g. Camelyon challenges have contributed to the rapid improvement in the whole-slide-image classification and tumor localization tasks.

Recently, ISBI Cell Tracking challenges (CTC) (Ulman et al. 2017) have provided thirteen (including two simulated) datasets of 2D and 3D sequences from multiple microscopy experiments. This has enabled an objective comparison of various cell segmentation and tracking methods. However, each dataset contains only a couple of sequences in its training and testing set, which cannot cover all the challenges encountered in images from real-world microscopy experiments, and any method trained on these datasets is likely to overfit and require modifications to work on new microscopy videos. As a result, it is still common for new cell tracking methods to report their results only on some private annotated dataset. If these private datasets can be gathered and openly shared, it can enable the creation of a large repository of diverse datasets, which in turn can allow the development of more robust methods and can speed-up progress in cell detection, segmentation and tracking fields.

However, sharing of data publicly is not as common among biologists as it is in the computer vision community. There is also an expectation, among some biologists, that data generators have an "exclusive right" to the problem they are studying, and others using their data (referred to as research parasite by some) ought not to study the same problem, even if they can do it better and/or faster (Longo & Drazen 2016). This closed culture is changing as some biologists see the value of novel analysis and re-testing of the conclusions, using publicly shared data (Greene et al. 2017), and realize that the value of data arises mostly from the knowledge that can be extracted from it, and publicly sharing data can enable others to verify the conclusions, rapidly detect mistakes, and can speed-up research (Ching et al. 2017).

6.3 Annotation

Deep learning has taken the central stage in most automated microscopy image analysis problems, including cell detection (Xie et al. 2015a), segmentation (Ronneberger et al. 2015, Çiçek et al. 2016, Chen et al. 2017) and tracking (Paper V). Training these deep and wide networks with millions of parameters requires a huge amount of annotated data, which is in limited supply due to the issues mentioned above and the need of
experts for annotations. Unlike annotating natural images, where a low-wage person can be assigned the annotation task through on-line crowd-sourcing platforms (e.g. Amazon Mechanical Turk), many problems in biomedical and medical domain require some domain knowledge for accurate annotation and it can be expensive to hire people with this kind of expertise.

To enable large models to be learned, in addition to large annotated datasets, it is also important that the annotations cover a diverse set of samples. When developing methods for cell tracking, where temporal history is rarely used, annotating small windows in both space and time (containing densely packed cells) from multiple sequences instead of annotating a whole sequence is a more efficient strategy. It provides a richer set of cells in various configurations and enables the methods to learn how the cells may look like in these challenging regions.

There are at least four complementary approaches for increasing the amount of training data when resources are limited. Firstly, elastic deformation can be used to augment training data and it is already being used by many methods (Çiçek et al. 2016) extensively. It can improve performance on some tasks (e.g. segmentation) but is not very helpful for other tasks (e.g. cell separation) where deformations alone cannot cover the various cell configurations encountered in images.

Secondly, training data can be simulated and used to train the models (Yurchenko & Lempitsky 2017). However, these computer generated simulations (Ulman et al. 2016) may not always be a good substitute for complex real data.

Thirdly, crowd-sourcing can be used to obtain annotations for some easier tasks such as cell segmentation and tracking (Schlesinger et al. 2017, Sameki et al. 2016). Even though these annotations may be somewhat noisy, they can still be beneficial in the absence of any other annotations and can be used to train models which are able to overcome the occasional errors in annotations.

Fourthly, weak annotations (markers) can be used to obtain augmented annotations (segmentation) (Manen et al. 2017) (Paper V) which can save considerable manual annotation effort. These augmented annotations despite containing occasional mistakes can provide better supervision during training and were used successfully for augmenting training data during the training of some Cell Proposal Networks presented in Paper III, Paper IV and Paper V.

---

16 annotation (segmentation) obtained using automated segmentation methods guided by weak manual annotation (markers).
Since one of the biggest challenges involved in cell segmentation is the separation of individual cells from within cell clusters, and markers often provide sufficient supervision to detect individual cells, a large number of these weak annotations combined with few mask annotations can be used by weakly-supervised and self-supervised methods (Khoreva et al. 2017) to reduce the annotation burden.

6.4 Manual verification

In many biological applications, especially in developmental biology where the whole pedigree of a cell is important, the accuracy and robustness requirements for cell tracking methods can be very high. A detection or an association error at any stage in the tracking renders the entire corresponding subtree of a lineage invalid. Since current state-of-the-art methods are not error-free, cell tracks often must be carefully edited manually to fix these errors (Meijering et al. 2009, Bray & Carpenter 2015). A method (Schiegg et al. 2015b) which can highlight the low confidence associations or detections can greatly reduce the manual effort required to correct tracking errors.

In many cell tracking methods, it is not easy to directly incorporate cell histories. Using a second post-processing stage which considers the whole track trajectory can potentially better highlight potential sites of erroneous detections and associations. One promising method is to use a hidden Markov model (HMM) or long short-term memory model (LSTM) to compute the probabilities of various types of errors at each detection and association in each track using information from the whole track.
7 Conclusions

Current cell detection, segmentation and tracking methods lack robustness and generalization. They can only be applied to images which satisfy few strict assumptions, and often require tuning their parameters to achieve good performance. This thesis has proposed a few general cell detection, segmentation and tracking methods, which make either a few (Paper I and Paper II) or no (Paper V) assumptions about the cell characteristics, to address the above-mentioned shortcomings.

Since cell detection is a major bottleneck in cell segmentation and tracking methods, and it can be extremely challenging to detect individual cells in high-cell-density microscopy images, cell proposals are extensively used in this thesis to efficiently utilize spatial and temporal context to better resolve detection ambiguities.

The focus of this work has been on the development of general and reusable methods, which can learn the cell shapes and appearances from the given training data, and that can be applied to other microscopy images without compromising too much on the performance. To achieve this very goal, recent methods from object detection and instance-level segmentation research have been applied to microscopy images with success.

The first method developed as part of this thesis (Paper I) was designed to segment cells of highly deformable shapes. It utilized a state-of-the-art edge detector to create a barrier at common borders between multiple cells inside cell clusters and used graph-cut to detect individual cells.

The rest of the methods in this work relied on two cell proposal generation methods: BLOB and CPN. They posed cell segmentation and tracking as cell proposal selection problems. BLOB proposals presented in Paper II make blob detection more robust by reducing its sensitivity to various parameters by leveraging a cell classifier learned from the training data. These proposals when combined with an iterative shortest cost path based proposal selection method presented in Paper II lead to better tracking performance than previous methods on datasets containing mostly round cells.

Cell Proposal Network (CPN) presented in Paper V utilizes convolutional neural networks (CNNs) to learn the shapes and appearances of a wide variety of cells from the given training data, and proposes cell candidates, which when combined with an integer linear programming based proposal selection method lead to state-of-the-art
cell segmentation and tracking performance on multiple publicly available microscopy datasets. This method has the potential to be applied to other more challenging microscopy images as well.

In Paper IV, a single CPN model, trained on datasets from multiple microscopy modalities, with competitive performance demonstrated the possibility of a single network to be used as a general cell detector. Similar methods which have good performance, generalization capabilities and are easy to use can have a huge impact on many research areas within biology. However, due to lack of large and diverse annotated datasets, the true potential and limitations of such an approach have not been fully explored yet.

The demo\textsuperscript{17} for Paper III and Paper IV, and code\textsuperscript{18} for Paper V are shared publicly to facilitate comparison with other methods, and to enable other researchers to build on top of these methods.

\textsuperscript{17}https://github.com/SaadUllahAkram/CellProposalNetwork
\textsuperscript{18}https://github.com/SaadUllahAkram/CellTracker
References

Adanja I, Megalizzi V, Debeir O & Decaestecker C (2011) A new method to address unmet needs for extracting individual cell migration features from a large number of cells embedded in 3D volumes. PLoS ONE 6(7): 1–12. 19


81


86


Medical Image Computing and Computer Assisted Intervention (MICCAI). 56
Appendix 1 Cell tracking: Optimization function

The optimization function used in Paper II and Paper V for cell tracking is:

\[ g^* = \arg \max_g \left( P(g|I) \right). \]  

(A1.1)

(A1.2) follows from (A1.1) as the solution space is restricted to the cell proposals and their associations.

\[ g^* = \arg \max_{g \subseteq G} \left( P(g|G) \right). \]  

(A1.2)

Selection of the cell \((p_i)\) and edge \((e_{i,j}^c)\) proposals in the solution graph can be represented by the binary decisions variables \(b_i\) and \(b_{c_{i,j}}\) respectively. Assuming probabilities of cell and edge proposals satisfy Markov assumption, and are conditionally independent given the images, (A1.2) can be written as (A1.3):

\[ g^* = \arg \max_{g \subseteq G} \left( \prod_{p_i \in P_G} P(p_i = b_i | I_i) \prod_{e_{i,j}^c \in E_G} P(e_{i,j}^c = b_{c_{i,j}} | I_{i,j}) \right). \]  

(A1.3)

Taking negative logarithm of (A1.3):

\[ g^* = \arg \min_{g \subseteq G} \left( -\sum_{p_i \in P_G} \log(P(p_i = b_i | I_i)) - \sum_{e_{i,j}^c \in E_G} \log(P(e_{i,j}^c = b_{c_{i,j}} | I_{i,j})) \right). \]  

(A1.4)

Since same operations are valid for both cell and edge proposal terms, the edge proposal term in (A1.4) is replaced with \(\Lambda\) in the following steps.

\[ g^* = \arg \min_{g \subseteq G} \left( -\sum_{p_i \in P_G} \log(P(p_i = b_i | I_i)) - \Lambda \right). \]  

(A1.5)

Since all decision variables in the optimization are binary, (A1.5) can be expanded to:

\[ g^* = \arg \min_{g \subseteq G} \left( -\sum_{p_i \in P_G} b_i \log(P(p_i = 1 | I_i)) - (1 - b_i) \log(P(p_i = 0 | I_i)) - \Lambda \right). \]  

(A1.6)

\[ g^* = \arg \min_{g \subseteq G} \left( -\sum_{p_i \in P_G} b_i \log \left( \frac{P(p_i = 1 | I_i)}{P(p_i = 0 | I_i)} \right) + \log(P(p_i = 0 | I_i)) - \Lambda \right). \]  

(A1.7)
The middle term in (A1.7) is a constant and can be ignored in the minimization.

\[ g^* = \arg \min_{g \in G} \left( - \sum_{p_i \in P_g} b_i \log \left( \frac{P(p_i = 1|I_i)}{1 - P(p_i = 1|I_i)} \right) - \Lambda \right). \quad (A1.8) \]

From (A1.8), the cost of selecting a cell and edge proposal are (A1.9) and (A1.10), respectively.

\[ C_{p_i} = -\log \left( \frac{P(p_i = 1|I_i)}{1 - P(p_i = 1|I_i)} \right). \quad (A1.9) \]

\[ C_{e_{i,j}} = -\log \left( \frac{P(e_{i,j} = 1|I_{i,j})}{1 - P(e_{i,j} = 1|I_{i,j})} \right). \quad (A1.10) \]

Simplifying (A1.8):

\[ g^* = \arg \min_{g \in G} \left( \sum_{p_i \in P_g} b_i C_{p_i} + \sum_{e_{i,j} \in E_g} b_{e_{i,j}} C_{e_{i,j}} \right). \quad (A1.11) \]

\[ g^* = \arg \min_{g=(P_g,E_g) \in G} \left( \sum_{p_i \in P_g} C_{p_i} + \sum_{e_{i,j} \in E_g} C_{e_{i,j}} \right). \quad (A1.12) \]

The cell and edge proposal probabilities in (A1.9) and (A1.10) are computed using various classifiers, including support vector machines (SVM), random forests and convolutional neural networks (CNNs). The details are given in the relevant articles, Paper II and Paper V.
Appendix 2 Errata

1. Equation (2) in Paper II has a mistake. The correct equation is:

\[ C_k = \sum_{e \in S} -\log\left( \frac{Pr(e)}{1 - Pr(e)} \right) \]

2. In Table 1 of Paper II, the results of KTH and LEID methods for GOWT1 dataset are swapped.
617. Iljana, Mikko (2017) Iron ore pellet properties under simulated blast furnace conditions: investigation on reducibility, swelling and softening


619. Schuss, Christian (2017) Measurement techniques and results aiding the design of photovoltaic energy harvesting systems


622. Li, Xiaobai (2017) Reading subtle information from human faces

623. Luoto, Markus (2017) Managing control information in autonomic wireless networking


626. Lappi, Tuomas (2017) Formation and governance of a healthy business ecosystem


630. Palosaari, Jaakko (2017) Energy harvesting from walking using piezoelectric cymbal and diaphragm type structures


Saad Ullah Akram

CELL SEGMENTATION AND TRACKING VIA PROPOSAL GENERATION AND SELECTION