Towards Increased Efficiency and Automation in Fluorescence Micrograph Analysis Based on Hand-Labeled Data

Der Technischen Fakultät der Universität Bielefeld
vorgelegt zur Erlangung des akademischen Grades
Doktor der Ingenieurswissenschaften

M.Sc. Christian Held
Department for Biomedical Image Processing
Fraunhofer Institute for Integrated Circuits
Erlangen, Germany

May 2013

Gedruckt auf alterungsbeständigem Papier gemäß ISO 9706

Acknowledgement

This work was carried out in the biomedical image processing group, headed by Dr. habil. Thomas Wittenberg, at the Fraunhofer Institute for Integrated Circuits in Erlangen. It was supported by a grant from the DFG to the Collaborative Research Center 796 - "Reprogramming of host cells by microbial effectors".
First of all, I would like to thank my supervisor, Dr. habil. Thomas Wittenberg, for giving me the opportunity to work on this interesting topic as well as to complete my PhD thesis at the Fraunhofer Institute for Integrated Circuits. His constant support and immense knowledge in biomedical research as well as his motivation and enthusiasm for this project make him an excellent supervisor.
I would also like to thank my doctoral thesis supervisor, apl. Prof. Tim W. Nattkemper, for his motivation, fruitful discussions and inspiring ideas, which have been a valuable basis for this work. I am very grateful to Dr. Ralf Palmisano, for his patience in explaining complex biological details. With his input, gapping the bridge between informatics and biology became a pleasant experience.
The Department of Biomedical Engineering at the Fraunhofer Institute, particularly the METEAN, were enjoyable places to work. I would thus like to thank my colleges at METEAN for the great atmosphere and the good time spent together.
The SFB 796 Graduate School offered an interesting interdisciplinary forum, providing an insight into microbiological research. I would like to thank Jens Wenzel, Prof. Roland Lang, Rike Webel, Prof. Manfred Marschall, Prof. Armin Ensser and Martin Drucker for their cooperation. Without their knowledge, motivation and patience discussing interdisciplinary questions this work would not have been possible.
Finally, I would like to thank my family and Anne for their constant support.

Originality of Work

The work presented in this thesis was done by myself under supervision of Dr. habil. Thomas Wittenberg und apl. Prof. Tim W. Nattkemper. Annotation of hand-labeled data was performed by Tobias Zielger and Andreas Fleischmann.

List of Publications

Parts of this thesis have been published in:

Journals

- C. Held, R. Palmisano, L. Häberle, M. Hensel, T. Wittenberg. Comparison of parameter-adapted segmentation methods for fluorescence micrographs. Cytometry Part A, 79(11):933-945, 2011.

- C. Held, V.Wiesmann, J. Wenzel, R. Palmisano, R. Lang, T. Wittenberg. Enhancing automated micrograph-based evaluation of LPS-stimulated macrophage spreading. Cytometry Part A, 83(4),409-418, 2013.

- C. Held, T. Nattkemper, R. Palmisano, T. Wittenberg. Approaches to automatic parameter fitting in a microscopy image segmentation pipeline: An exploratory parameter space analysis, Journal of Pathology informatics, 4:5, 2013.

Proceedings with Peer Review

- C. Held, J. Wenzel, R. Palmisano, R. Lang, T. Wittenberg. Segmentierung von Makrophagen in Fluoreszenzbildern mittels Fast Marching Level Set Verfahren. In: H. Handels, J. Ehrhardt, TM. Deserno, H-P. Meinzer, T. Tolxdorff: Bildverarbeitung für die Medizin 2011. Berlin, Heidelberg: Springer Berlin Heidelberg, 129-133, 2011.

- C. Held, J. Wenzel, R. Webel, M. Marschall, R. Lang, R. Palmisano, and T. Wittenberg. Using multimodal information for the segmentation of fluorescent micrographs with application to virology and microbiology. Conf Proc IEEE Eng Med Biol Soc. 2011, 6487-6490, 2011.

- C. Held, J. Wenzel, R. Lang, R. Palmisano, T. Wittenberg. Comparison of Methods for Splitting of Touching and Overlapping Macrophages in Fluorescent Micrographs. In Proceedings of the 9th international conference on Image Analysis and Recognition, ICIAR'12, page 456-464, Berlin, Heidelberg, 2012. Springer-Verlag.

- C. Held, V. Wiesmann, R. Palmisano, A. Ensser, T. Wittenberg. Automated high throughput image analysis of the ND10 complex of KSHV cells. Biomedizinische Technik, 57(1): 452-455, 2012.

Proceedings

- C. Held, R. Palmisano, T. Wittenberg. (Semi-) automated image analysis with application to fluorescence microscopy. In Proc's First Int. Conference - SFB 796: Mechanisms of viral host cell manipulation: From plants to humans, page 63. 2.-4.10.2011 in Bamberg, 2011.

Further contributions

- R. Webel, J. Milbradt, S. Auerochs, V. Schregel, C. Held, C. Nöbauer, E. Razzazi-Fazeli, C. Jardin, T. Wittenberg, H. Sticht, M. Marschall: Two isoforms of the protein kinase pUL97 of human cytomegalovirus are differentially regulated in their nuclear translocalization. Journal of General Virology, 92(3):638-649, 2011.

- J. Wenzel, C. Held, R. Palmisano, S. Teufel, J-P. David, T. Wittenberg, R. Lang. Measurement of TLR-induced macrophage spreading by automated image analysis: differential role of Myd88 and MAPK in early and late responses. Frontiers in Systems Physiology, 2(0), 2011.

- V. Wiesmann, C. Held, R. Palmisano, T. Wittenberg. Segmentation of HeLa cells in phase-contrast images with and without DAPI stained cell nuclei. Biomedizinische Technik, 57(1): 519-522, 2012.

Kurzfassung

In den letzten Jahren hat der Grad an Automatisierung im Bereich der Fluoreszenz Mikroskopie stark zugenommen. Dies umfasst sowohl die Bildaufnahme als auch die Probenpräparation und führt zu einer stark zunehmenden Menge an Bilddaten, welche für ein Experiment ausgewertet werden müssen. Die für die Durchführung eines Experiments benötigte Zeit wird infolgedessen maßgeblich durch die Dauer der Bildauswertung bestimmt, insbesondere bei manuell ausgewerteten Experimenten. Darüber hinaus sind manuelle Auswertungen meinst nur schwer reproduzierbar und häufig subjektiv.

Automatisierte Bildanalysesoftware kann dazu dienen, die experimentelle Auswertung zu beschleunigen und deren Reproduzierbarkeit und Objektivität zu verbessern. Die meisten Bildanalyseansätze basieren auf einer Segmentierung des Bildes. Hierfür wird häufig eine Kombination mehrerer Segmentierungsverfahren angewandt. Sowohl durch Auswahl und Anordnung der Methoden als auch durch das Einstellen freier Parameter wird eine Anpassung der Segmentierung an unterschiedliche Aufgabenstellungen ermöglicht. Der Hauptaufwand für den Nutzer besteht folglich in der Auswahl der zu verwendenden Methoden und in deren Parametrisierung. Beide Schritte stellen insbesondere für Anwender mit geringen Bildverarbeitungskenntnissen ein Hindernis dar. Infolgedessen werden im Forschungsalltag viele Experimente, welche von einer automatisierten Bildanalyse profitieren könnten, manuell ausgewertet.

Um den Anpassungsaufwand für die automatisierte Bildanalyse zu verringern, wird in dieser Arbeit ein neuartiges Konzept implementiert und evaluiert, welches die Segmentierung anhand von Nutzereingaben automatisiert optimiert. Statt Segmentierungsmethoden auszuwählen und deren Parametrisierung anzupassen, umrandet der Nutzer manuell oder interaktiv eine repräsentative Auswahl von Zellen. Die Optimierung des Segmentierungsablaufs kann dann automatisch und ohne weitere Nutzereingaben erfolgen.

Hauptherausforderung ist es, ein System mit akzeptabler Laufzeit und zugleich guter Segmentierungsleistung zu entwickeln, welches für verschiedene Fluoreszenz Datensätze angewandt werden kann. Hierfür wird eine sogenannte Segmentierungspipeline implementiert, welche aus austauschbaren Methoden besteht. Diese Methoden realisieren die Segmentierungsschritte Vorverarbeitung, Vordergrund-Hintergrundtrennung und Zelltrennung. Für die Implementierung der einzelnen Methoden werden sowohl auf dem Stand der Technik basierende als auch neue, im Rahmen dieser Arbeit entwickelte Verfahren implementiert und evaluiert.

Die Diskretisierung der Parameter muss so gewählt werden, dass eine möglichst große Bandbreite an Bilddatensätzen unterstützt wird. Um dies bei möglichst geringem Rechenaufwand zu gewährleisten werden sowohl Parameter mit konstanten als auch mit nichtlinear zu- oder abnehmenden

Schrittweiten implementiert. Neben der Diskretisierung der Segmentierungsparameter hängt die Qualität der Optimierung stark von der Auswahl des Optimierungsverfahrens ab. Um ein geeignetes Verfahren auszuwählen werden die Parameterräume verschiedener Segmentierungsverfahren untersucht. Dieses Experiment zeigt, dass die zu Grunde liegenden Parameterräume zwar großteils monoton sind, jedoch mit lokalen Extrema gerechnet werden muss. Der Vergleich verschiedener Optimierungsverfahren bestätigt, dass das Coordinate Descent Verfahren gut geeignet ist, um die Segmentierungspipeline an die Bilddaten anzupassen.

Damit der manuelle Annotationsaufwand für den Nutzer möglichst gering gehalten werden kann, wird der Zusammenhang zwischen Segmentierungsleistung und der Anzahl manuell annotierter Zellen untersucht. Dieses Experiment zeigt, dass für einen Großteil der Datensätze bereits zehn manuell annotierte Zellen eine gute Kalibrierung der Segmentierungspipeline ermöglichen.

Zusammenfassend lässt sich sagen, dass aufgrund der geringen Anzahl von Zellen, welche für die Kalibrierung des entwickelten Verfahrens nötig sind, und der kurzen Laufzeit der Optimierung, welche über Nacht durchgeführt werden kann, eine Verbesserung der Automatisierung und somit der Effizienz bei der Auswertung fluoreszenzmikroskopischer Bilder ermöglicht wird.

Abstract

In the past decade, automation in fluorescence microscopy has strongly increased, particularly in regards to image acquisition and sample preparation, which results in a huge volume of data. The amount of time required for manual assessment of an experiment is hence mainly determined by the amount of time required for data analysis. In addition, manual data analysis is often a task with poor reproducibility and lack of objectivity.

Using automated image analysis software, the time required for data analysis can be reduced while quality and reproducibility of the evaluation are improved. Most image analysis approaches are based on a segmentation of the image. By arranging several image processing methods in a so-called segmentation pipeline, and by adjusting all parameters, a broad range of fluorescence image data can be segmented. The drawback of available software tools is the long time required to calibrate the segmentation pipeline for an experiment, particularly for researchers with little knowledge of image processing. As a result, many experiments that could benefit from automated image analysis are still evaluated manually.

In order to reduce the amount of time users have to spend in adapting automated image analysis software to their data, research was carried out on a novel image analysis concept based on hand-labeled data. Using this concept, the user is required to provide hand-labeled cells, based on which an efficient combination of image processing methods and their parameterization is automatically calibrated, without further user input.

The development of a segmentation pipeline that allows high-quality segmentation of a broad range of fluorescence micrographs in short time poses a challenge. In this work, a three-stage segmentation pipeline consisting of exchangeable preprocessing, figure-ground separation and cell-splitting methods was developed.

These methods are mainly based on the state of the art, whereas some of them represent contributions to this status. Discretization of parameters must be performed carefully, as a broad range of fluorescence image data shall be supported. In order to allow calibration of the segmentation pipeline in a short time, discretization with equidistant as well as nonlinear step sizes was implemented. Apart from parameter discretization, quality of the calibration strongly depends on choice of the parameter optimization technique. In order to reduce calibration runtime, exploratory parameter space analysis was performed for different segmentation methods. This experiment showed that parameter spaces are mostly monotonous, but also show several local performance maxima. The comparison of different parameter optimization techniques indicated that the coordinate descent method results in a good parameterization of the segmentation pipeline in a small amount of time.

In order to minimize the amount of time spent by the user in calibration of the system, correlation between the number of hand-labeled reference

samples and the resulting segmentation performance was investigated. This experiment demonstrates that as few as ten reference samples often result in a good parameterization of the segmentation pipeline.

Due to the low number of cells required for automatic calibration of the segmentation pipeline, as well as its short runtime, it can be concluded that the investigated method improves automation and efficiency in fluorescence micrograph analysis.

Contents

1. Introduction

Research in microbiology, immunology, virology and similar disciplines often requires the investigation of interactions in biological systems, ranging from whole organism level down to single molecule level. In the field of systems biology, a variety of new biomedical imaging techniques have been developed in the last few decades which enable the investigation of such interactions (Megason and Fraser, 2007).

In this context, a new field of research known as bioimage informatics has been established, which deals with the development of novel software tools for image processing, data mining, database and visualization (Peng, 2008). With the increasing number of methods based on fluorescence microscopy, as well as advancing automation in sample preparation and image acquisition, the amount of image data that must be analyzed is strongly growing. For most applications, assessment of the data poses a bottleneck, particularly if evaluation is performed manually. Typical data analysis tasks require the counting of cells or nuclei, measuring the size of cells or quantifying intensity distributions. For some high throughput experiments, the assessment of several 100,000s of cells is required. Manual interpretation of such data is a repetitive and error-prone task with low reproducibility which can easily take several weeks.

Additionally, a specific outcome for an experiment is often expected and hypothesized by the experimenter. Such assumptions can lead to a biased interpretation of the data. As a result, the outcome of manually performed evaluations may differ if the data is repeatedly interpreted by two or more independent observers (interobserver variance) or if the dataset is repeatedly analyzed by a single observer (intraobserver variance).

To improve objectivity and reproducibility of the evaluations and to reduce the amount of time researchers have to spend with data analysis, automated image analysis software can be utilized. A coarse workflow on which most image analysis tools are based is depicted in Figure 1.1. After acquisition of the image data, each image is partitioned into regions representing cells or the image background. Based on this partitioning, also known as image segmentation, automated image analysis is performed. Thereby, low quality segmentation results usually lead to erroneous measurements. Hence, high quality image segmentation methods are crucial for the subsequent automated image analysis step.

To enable the segmentation of fluorescence image data from different application domains, publicly available tools like the CellProfiler (Carpenter

Figure 1.1.: Overview of the image analysis workflow applied for analysis of many fluorescence microscopy based experiments.

et al., 2006) or MBF ImageJ (Collins et al., 2007) offer an extensive collection of image processing and image segmentation methods. Adapting the available methods to a given type of micrograph requires the user to select a combination of segmentation methods, arranging them in a so-called segmentation pipeline, and to adjust all free parameters. This step, in this work referred to as calibration of the segmentation pipeline, offers flexibility for adaptation to data from different fluorescence imaging domains. However, with the flexibility of such a system, the time required for selection of a good combination of methods and for adjusting all parameters increases. Pretorius et al. (2011) report that a typical pipeline of the Cell Profiler consists of up to 150 parameters, and manual selection and adjustment of these parameters can easily take several days. If the default cell detection routine of the CellProfiler named "IdentifyPrimaryObjects" is used, 19 parameters have to be adjusted manually. Apparently, using state of the art image analysis software can improve objectivity and reproducibility of an evaluation. However, the amount of time researchers have to spend in the data analysis with such software is still high, particularly for non-image processing experts. That is why many experiments that could benefit from automated image analysis are still performed manually.

After calibrating the segmentation pipeline, methods and parameters must be re-adjusted if the acquisition parameters (microscope type, camera type, magnification, etc.) or parameters of the experimental setup (investigated cell type, staining protocol) change. Particularly in microbiological research, sample preparation or data acquisition can vary for each experiment. As a result, the amount of time researchers are willing to spend for manual calibration of the segmentation pipeline is very low for non-recurring experiments.

In order to improve acceptance of automated image analysis software, efficiency and automation of the software must be improved. This corresponds to reducing the amount of hand labelings and the time that is required for calibrating the segmentation pipeline for a specific experiment. Therefore, this work addresses the development of a novel software tool for analysis of fluorescence micrographs which enables automated calibration of the seg-

mentation pipeline based on hand-labeled data. It will be demonstrated that this system can be applied for segmentation of a broad range of fluorescence microscopy image data. For the improvement of runtime and automation, a strong focus was put on the selection and discretization of image processing methods, optimization of the corresponding parameter spaces, as well as on minimizing the amount of required hand-labeled reference data.

In the following, the concept of a segmentation pipeline calibrated based on hand-labeled data is illustrated. Next, a brief introduction into the technique of fluorescence microscopy and the resulting image data is provided, as well as an overview of this work.

1.1. Segmentation of fluorescence micrographs based on hand-labeled data

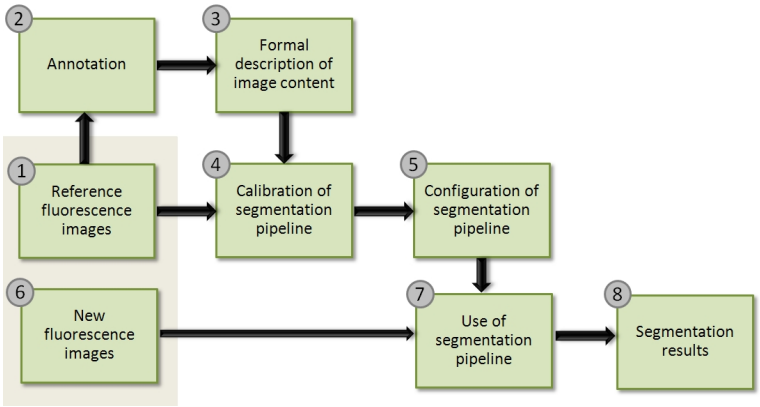
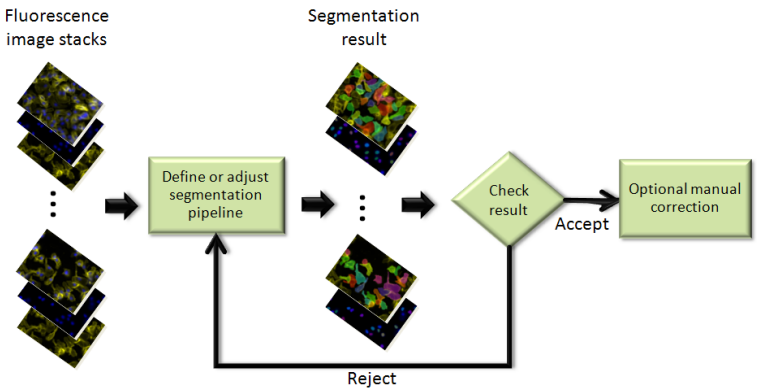


Figure 1.2.: Workflow of segmentation based on hand labelings. Reproduced with courtesy of Wittenberg et al. (2009).

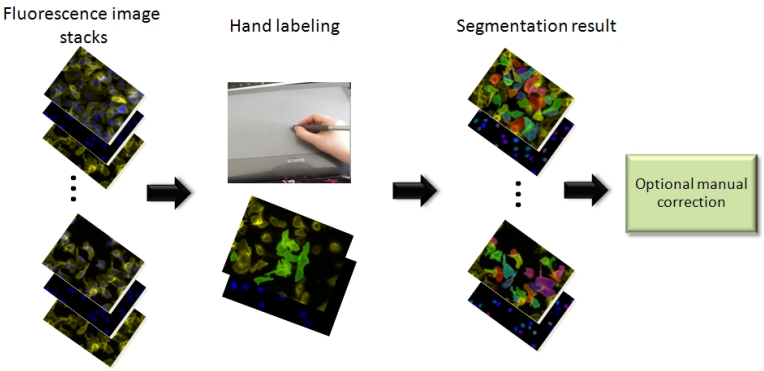
To reduce the amount of time required for the calibration of automated image analysis software, research was carried out on a segmentation pipeline applicable to a broad range of fluorescence imaging domains (but not limited to them) as described by Wittenberg et al. (2009). An overview of this concept is provided in Figure 1.2. Using this framework, the image dataset is split into a set of reference images (1) and a one of test images (6) for which segmentation is still unknown. The set of reference images is then manually annotated (2) by the user, and a formal description of the image content (3) is obtained. The annotated reference images are then used for calibration of the segmentation pipeline (4), which corresponds to selecting a good combination of parameterized image processing methods and adjusting all parameters. After calibration (5), test images (6) can be segmented

by application of the segmentation pipeline (7) and a segmentation of the complete data (8) is obtained.

Using the described image segmentation concept for a new dataset, the user is required to provide hand labelings of a representative subset of the images instead of selecting a combination of segmentation methods and adjusting their parameters. For a comparison of the manual image analysis workflow to the one using hand-labeled data see Figure 1.3.



(a) Manual workflow



(b) Segmentation based on hand-labeled data

Figure 1.3.: Comparison of two conceptually different workflows for the segmentation of fluorescence image data. Using the conventional manual workflow (a), calibration of the segmentation pipeline is performed by the user. Hence, selection of a combination of methods, variation of the parameters and manual inspection and evaluation of the preliminary segmentation results require user input. Calibrating the segmentation pipeline based on hand-labeled data (b), no user interaction is required apart from providing the hand-labeled data.

1.2. From optical microscopy to fluorescence microscopy

Microscopy enables the investigation of objects that cannot be seen with the naked eye. Its origin is a matter of debate, but mostly Hans Janssen is credited for the invention of the first compound microscope about 400 years ago. It consisted of a bi-convex eyepiece lens and a plano-convex objective lens and enabled 60x magnifications. Jan Swammerdam (1637-1703) was the first to use this microscope for the investigation of insects and the observation of red blood cells (Robenek, 1999).

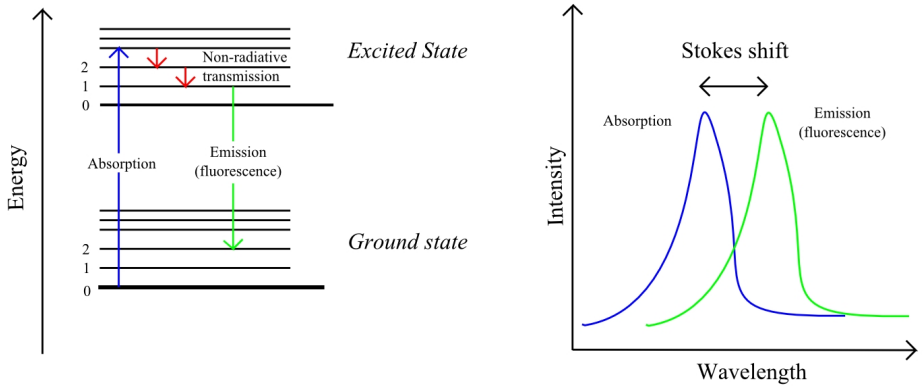


Figure 1.4.: Jablonski diagram illustrating fluorescence (left). After photon absorption, an electron is excited to a higher energy state. After a short period of time, the electron relaxes back to the ground state by emission of a lower energy photon. Fluorescence causes characteristic absorption and emission spectra for each fluorescent material. The distance between positions of the band maxima of absorption and emission spectra is known as Stokes shift (right).

Two hundred years after invention of optical microscopy, the effect of fluorescence was discovered and first described by George Gabriel Stokes, in 1852. The term fluorescence refers to the excitation process of an atom or a molecule by an outside source of energy (Ghiran, 2011). Due to photon absorption, electrons are transferred to an excited state. Photon emission is then caused by relaxation back from the excited to the ground state after a very short period of time, which is usually less than 1 μs (Ghiran, 2011). Due to non-radiative transmission (Figure1.4, left), wavelength of emitted light is longer then that of absorbed light. The distance between peaks of absorption and emission spectra is known as Stokes shift (Figure 1.4, right).

Several decades after discovering the effect of fluorescence, between 1911

and 1913, the first commercial fluorescence microscope was developed by Otto Heimstädt and Heinrich Lehmann and produced by Reichert and Zeiss (Rost, 1995). Most fluorescence microscopes use a light source emitting a continuous spectrum. Light of a specific wavelength is then selected by an excitation filter and applied to the sample. Hence, fluorophores contained in the sample are excited and fluorescence light with a longer wavelength is emitted. After removing excitation light by an emission filter, the fluorescence signal can be observed. For a schematic drawing of a fluorescence microscope, see Figure 1.5.

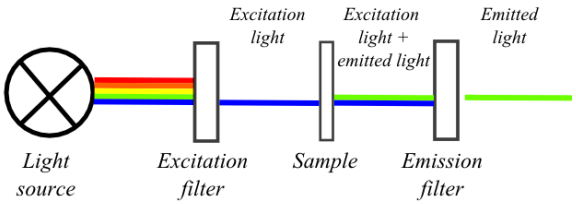
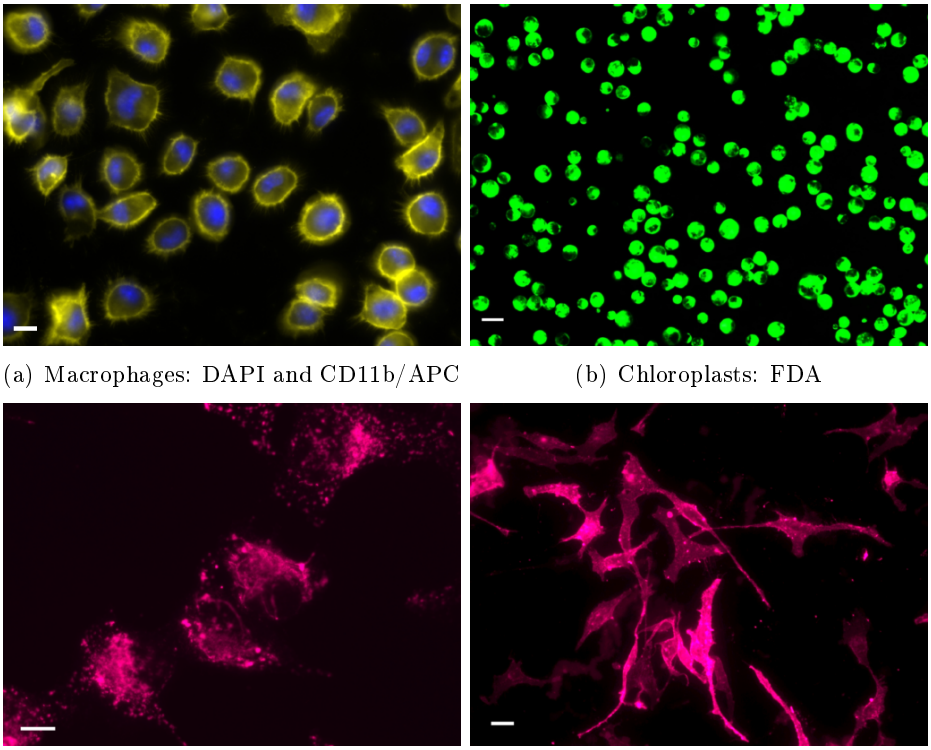


Figure 1.5.: Schematic illustration of a fluorescence microscope.

1.3. Fluorescence image data

Before 1941, fluorescence microscopy was only applicable for investigation of molecules showing auto-fluorescence. Describing a technique for labeling antibodies with fluorescent dyes, Coons et al. (1941) enabled a revolution in immunology and biology. Applying this technique, fluorescent markers can be attached to proteins or molecules and used for visualization and localization of reagents. In the past decades, various dyes have been developed for staining of specific targets. For a broad overview of available dyes, see (Day and Davidson, 2009; Chudakov et al., 2010). Figure 1.6 provides some exemplary images depicting different types of staining.

Analyzing fluorescence micrographs from different application domains shows that this kind of image data is very multifarious. Appearance of the objects of interest depends on the type of experiment, the applied staining and the experimental setup. Some experiments even pose a challenge for manual interpretation, as cell boundaries cannot be recognized unambiguously. In the following, a categorization of fluorescence micrographs is performed according to characteristics determining difficulty of the data analysis task:



(a) Macrophages: DAPI and CD11b/APC

(b) Chloroplasts: FDA

(c) HeLa: Phalloidin Alexa 568

(d) HeLa: DiD

Figure 1.6.: Representative micrographs depicting different types of fluorescent staining. (a) macrophages stained with the dyes DAPI (blue), for visualization of the cell nuclei, and CD11b/APC (yellow), for visualization of the macrophage cells. Chlorophyll inside chloroplasts stained with FDA (b, green) enables viewing non-apoptotic chloroplasts. The actin cytoskeleton of HeLa cells labeled with the probe Phalloidin Alexa 568 (c). Nuclei of HeLa cells labeled with DAPI (d, blue) and cytoplasmic membrane of HeLa cells labeled with DiD (d, violet). Scale bars correspond to $10\mu m$ (a, b, c) or $20\mu m$ (d).

- **Irregularity:** Irregularity of the cell boundary. Objects with low irregularity show a round, oval or convex shape (nuclei in Figure 1.6 (a)). Objects with a very irregular boundary are often multipolar (HeLa cells, Figure 1.6 (d)).

- **Overlapping:** Simple data contains mainly isolated cells (macrophages, Figure 1.6 (a)), whereas more complex data consists of touching or overlapping cells (HeLa cells, Figure 1.6 (d)). Interpretation of data containing overlapping cells is often an ambiguous task for machine as

well as for the human observer.

- **Intensity distribution:** The intensity distribution inside each object is based on the selected staining and on the investigated object. Objects showing homogeneous intensity distribution (nuclei, Figure 1.6 (a)), increasing or decreasing intensities from the core to the boundary of the object are easier to segment than objects showing a more complex intensity distribution including several local intensity minima and maxima (HeLa, Figure 1.6 (c)).

A categorization of the data that will be used in this work according to the mentioned characteristics is provided in Figure 1.7.

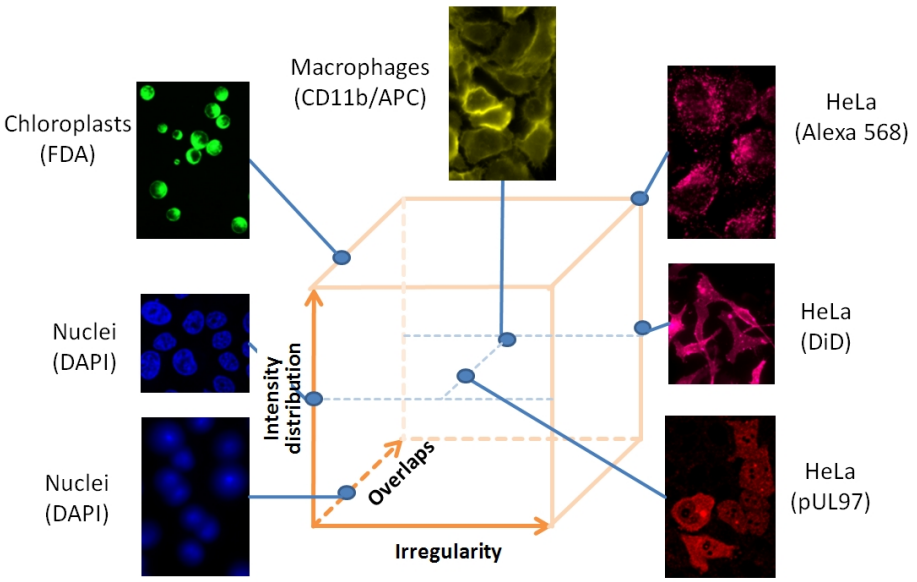


Figure 1.7.: Categorization of fluorescence image data used in this work according to irregularity, overlaps and intensity distribution. Objects showing irregular cell boundaries, strong overlaps and non-homogeneous intensity distribution often pose a challenge to automated image segmentation.

1.4. Overview

Following the introduction, Chapter 2 provides an overview of publications related to the segmentation of fluorescence micrographs. Additionally, publicly available software tools that can be used for analysis of fluorescence micrographs are described. Based on this review, contributions of this work to the state of the art are summarized in Chapter 3.

In Chapter 4, image processing and evaluation methods that constitute a basis for this work are summarized. These methods are applied for implementation and parameterization of the segmentation pipeline, which is described in detail in Chapter 5. This segmentation pipeline can be configured based on hand-labeled data without requiring any further user interaction, and improves automation in fluorescence micrograph analysis. Chapter 6 deals with the question of how to configure the segmentation pipeline based on the hand-labeled data in a short amount of time. After providing an overview of image data used in this work (Section 6.1), different objective functions for comparison between automated segmentation and hand-labeled data will be compared (Section 6.2). Based on this, different segmentation algorithms are compared to investigate if the described segmentation pipeline is applicable for segmentation of a broad range of fluorescence image data (Section 6.3). In this context, a novel graph cut based method was developed for figure-ground separation of micrographs. Based on studying the effect of artifacts, such as noise and out of focus cells (Section 6.4), it was shown that graph cut methods can improve segmentation for noisy data and out of focus cells.

Next, an exploratory parameter space analysis was performed (Section 6.5) to identify an optimization method able to identify a good parameterization of the segmentation pipeline in a short time. The resulting measurements also allow estimation of the error implemented by using non-greedy optimization techniques. Apart from the optimization runtime, efficiency of the described segmentation techniques strongly depends on the amount of required hand-labeled data. That is why research was carried out on the resulting segmentation performance for varying numbers of reference cells, as well as for different strategies for selecting representative cells (Section 6.7).

Based on the described experiments, two more runtime-efficient segmentation pipelines are explored in Chapter 7. The first version performs balancing between segmentation performance an runtime and can be configured, based on a reference image, in less than 8 hours, whereas the more runtime-efficient version can be configured in less than 9 minutes using a standard single core CPU (one core of an Intel Core 2 Duo CPU with 2,66 GHz).

2. Related work

In this chapter, an overview is provided on publications and software tools related to the field of fluorescence micrograph segmentation and analysis. First, methods from the literature are reviewed which have been applied for segmentation of fluorescence image data. Then, available freeware tools that can be used for micrograph segmentation and analysis are reviewed.

2.1. Fluorescence micrograph segmentation

A lot of research has been carried out on the segmentation of fluorescence micrographs. The most frequently covered topics are the segmentation of cell nuclei, the segmentation of cells or similar objects and the multi-channel segmentation of cells and nuclei. Multi-channel cell segmentation refers to incorporating information on size and extension of the nuclei for segmentation of the cells.

2.1.1. Shading correction

Some fluorescence images show poor image quality and exhibit low-frequency illumination artifacts; these can be removed by shading correction. Leong et al. (2003) describe shading correction based on difference imaging for the segmentation of digital microscope images. First, the image background is estimated by convolving the input image with a strong Gaussian filter kernel. Then, the estimated background is subtracted from the input image. A different method for shading correction is described by Wang et al. (2008). They used a cubic B-spline for approximation of the background intensities. Comparison of different retrospective shading correction methods for various images was carried out by Tomaževič et al. (2002). Their experiments show that performance of most shading correction methods depends on the image content, and especially on the size of the objects. Only the shading correction method using entropy minimization shows good results for all test images. Parameters of each method were thereby adjusted manually to obtain the best possible results for each method.

2.1.2. Segmentation of fluorescence nuclei

All publications dealing with the segmentation of fluorescent nuclei focus on a specific application domain. Hence, none of the publications in this

section handles the applicability for different types of nuclei staining. Basic approaches for nuclei segmentation with focus on object classification use interactive threshold selection methods or a fixed threshold value for segmentation of fluorescent nuclei. Research carried out by Wang et al. (2008) addressed support vector machine based cell cycle identification, done with basis on a segmentation of the cell. For segmentation of the FISH labeled nuclei, a cubic B-spline shading correction method is applied before utilizing a fixed threshold value. In (Wang et al., 2010b), this method was improved by using an interactive threshold value.

Most other publications apply fully automated methods for the segmentation of fluorescent nuclei, whereof the watershed transform is most frequently used. Fully automatic means that no user interaction is required after selecting segmentation methods and adjusting parameters. Malpica et al. (1997) proposed the hybrid watershed transform that combines a distance transformed binary image and a gradient magnitude image for splitting of clustered nuclei. Lin et al. (2003) utilized a histogram based threshold and applied a hybrid watershed transform for segmentation of 3D nuclei. Based on the segmentation, a Gaussian mixture model is trained for merging of touching or overlapping nuclei based on the features shape, texture, volume, convexity, circularity, area, eccentricity and mean radius. Wählby et al. (2004) described a segmentation scheme based on seeded watershed transform, applicable to 2D or 3D fluorescent nuclei data. First, morphological filters are applied for foreground and background seed identification. These seeds are then used as input for a gradient magnitude based watershed transform. For reduction of over-segmentation artifacts, a gradient magnitude based merging step is implemented. Chen et al. (2006) applied Otsu's threshold selection algorithm for identification of fluorescent nuclei. Additionally, an improved watershed transform was used for splitting of touching nuclei. For reduction of over-segmentation, an additional merging step based on size and circularity was implemented. Nandy et al. (2009) proposed the watershed transform for segmentation of FISH-stained nuclei. First, a wavelet based preprocessing routine is applied. Then, threshold selection is performed using a combination of isodata and triangle thresholding algorithms. For refinement of nuclei contours, a level set contour evolution method is applied. Splitting of touching nuclei is then performed based on the watershed transform. Jeong et al. (2009) performed threshold selection for fluorescent nuclei based on a Gaussian mixture model. Nuclei were then assigned to the classes "overlapping" or "isolated" by a Bayesian network. Next, the watershed transform was applied to the gradient image for splitting of touching nuclei. A further watershed transform based scheme was described by Cheng and Rajapakse (2009), which incorporates seeds extracted by an adaptive H-minima transform.

As an alternative to the watershed transform, the level set method can be applied for segmentation of fluorescent nuclei. Ersoy et al. (2009) utilized a

version, improved in terms of speed, of the graph partitioning active contours method for the segmentation of fluorescent nuclei. An algorithm based on Chan-Vese level set for segmentation and tracking of 3D nuclei was proposed by Dzyubachyk et al. (2010a). In (Dzyubachyk et al., 2010b), several extensions to the level set based segmentation methods are described and the improvement in terms of segmentation performance is demonstrated. Chinta and Wasser (2011) used multiple level sets for the segmentation of 3D nuclei. The level set is thereby initialized with the output from a combination of Gaussian mixture model and k-means clustering.

Apart from the large number of approaches applying the watershed transform or level sets for segmentation of fluorescent nuclei, some alternative methods have also been proposed. Ortiz De Solorzano et al. (2001) applied a seeded region growing method based on gradient curvature flow for the segmentation of fluorescent nuclei. A two-stage graph cut method for segmentation of fluorescence nuclei was used by Danek et al. (2009). Here, cells are first separated from the image background by a minimum cut based on edge capacities and histogram analysis. The resulting binary image is then distance transformed, and a second cut combining information on the distance image and gradient magnitude is applied for splitting of the nuclei. A completely different approach based on fast ellipse fitting for segmentation and splitting of fluorescent nuclei was proposed by Brüllmann et al. (2012).

Harder et al. (2006, 2007) described a comparison of different figure-ground separation methods for fluorescent nuclei. This comparison demonstrates that Otsu's global threshold selection method shows poor results, even after applying background correction techniques. Furthermore, two adaptive threshold selection methods are described and compared. One of these methods utilizes a sliding window based, locally adaptive threshold combined with a global one that is used if variance is below a user-defined threshold. The second adaptive threshold selection method computes one threshold value for each window, and applies this threshold for every pixel inside the window to reduce computational time of the algorithm. Harder et al. showed that using adaptive threshold selection segmentation quality for nuclei segmentation can be improved.

2.1.3. Segmentation of fluorescence cells

The segmentation of fluorescently labeled cells poses a challenge, as cells often touch or even overlap each other. An interactive approach for cell segmentation was published by Baggett et al. (2005). Using the described method, the user selects one point at the center of the cell and another point at the cell boundary. Then, a polar transform around the user-defined center point is performed. Based on a gray-weighted distance transform, the optimal cell border is extracted.

Other research groups have applied morphological operators for segmen-

tation of fluorescence cell images. Metzler et al. (1999) described a morphology based method for splitting of mouse fibroblasts. In (Metzler et al., 2000), this method was extended to a morphological multiple-scale method. Differently, Wang et al. (2010a) performed a segmentation of fluorescence cells using iterative erosions and dilations. Further, Zhang et al. (2010a) proposed a mathematical morphology based approach for the extraction of cell boundaries. Fluorescent puncta are then identified by means of a window based contrast measurement. Mech et al. (2011) proposed a threshold selection method combining histogram information and shape information for analysis of host-pathogen interactions. Further improvements have been achieved by removing objects that deviate in terms of roundness, area or intensity.

Apart from nuclei segmentation, the watershed transform is also applied for segmentation and splitting of fluorescence cells. Wählby et al. (2002) described an algorithm for the segmentation of Chinese hamster ovary (CHO) cells. For removal of the image background, a difference imaging scheme based on cubic B-spline is utilized. Segmentation of the cells is then performed by a watershed transform. In order to reduce over-segmentation, small objects are merged with touching neighbor objects. Karvelis et al. (2006) also used the watershed transform for segmentation of chromosomes based on gradient magnitude information. Differently, Hodneland et al. (2009) described and compared two methods for the segmentation of fluorescently stained cells. For preprocessing, a Gaussian filter and a ridge enhancement technique are applied. For segmentation and splitting of touching cells, a watershed transform based method is compared to a level set based one. Obtained results have demonstrated that both methods enable a good segmentation and are comparable. Zhu et al. (2010) also applied the seeded watershed transform for segmentation of fluorescent cells. Thereby, seeds are determined based on local intensity maxima.

As for segmentation of cell nuclei, the level set method is also applied for segmentation of fluorescent cells. Dufour et al. (2005) proposed a method based on multiple active surfaces, with or without edges, for the segmentation of 3D cells. A Chan-Vese level set based cell segmentation method was described by Bunyak et al. (2006). Srinivasa et al. (2007) used a multiscale active contour transformation framework for segmentation of different types of cells. In (Srinivasa et al., 2009), an active mask based framework was proposed. Möller et al. (2010) described a segmentation scheme based on coupled active contours that is applicable for the segmentation of fluorescently labeled cells as well as for the segmentation of particles and dots.

In the following, alternative methods to the watershed transform and the level set method that are applicable for cell segmentation are described. Nattkemper et al. (2000) utilized a competitive layer model (CLM) for segmentation of fluorescent micrographs using Gestalt-based feature binding. In (Nattkemper, 2002), a neural network was applied for detection of flu-

orescence cell positions. Cell boundaries were then determined by using a recurrent neural network model. Differently, Pham et al. (2004) performed a segmentation of fluorescently labeled vesicles by an improved fuzzy c-means clustering algorithm. Yet Qian et al. (2009) used a self-organizing map based multi-threshold technique for the identification of background pixels, foreground tissue as well as seeds. Based on this information, a seeded watershed transform was applied for segmentation of the cells. Bradbury and Wan (2010) used a spectral k-means approach for the segmentation of muscle cells in bright field images. This algorithm was based on spectral graph partitioning, splitting a graph into two pieces based on the eigenvectors and eigenvalues of the Laplacian matrix of the graph. Further, a shape based method determining the best fitting ellipse for segmentation of the eggshell surrounding embryo cells was described by Blanchoud et al. (2010). Zhang et al. (2010b) reported an image segmentation method based on the Laplacian of a Gaussian edge detector. In this application, straight lines are regarded as artifacts and removed by a combination of eccentricity filters and morphological operations. Yet Zaritsky et al. (2011) proposed the MultiCellSeg algorithm to separate cellular from background regions. This non-parametric algorithm classifies local patches in an image by a cascade of Support Vector Machines (SVMs) based on reference data. For refinement of the resulting regions, postprocessing based on a graph cut is implemented. Further, segmentation of cells guided by an atlas was proposed by Qu et al. (2011). This so-called simultaneous recognition and segmentation method uses a 3D target atlas guided voxel classification system. The atlas is thereby smoothly deformed until it best matches the image. The authors of this study further demonstrated the applicability of tis method for detection of varying cell types.

2.1.4. Multi channel segmentation of cells and nuclei

Segmentation, and particularly automated splitting of nuclei, is usually a much less complex task than segmentation and splitting of fluorescently labeled cells. Hence, information on position and extension of the nuclei can be combined for segmentation and splitting of the cells.

An interactive scheme that incorporates information on location and extension of the nuclei for segmentation of the cells was described by Palmieri et al. (2010). The proposed interactive method first requires the user to define a threshold level in order to select all of the stained nuclei (Palmieri et al., 2010). Cells are then segmented by a seeded watershed method.

Based on work of Wählby et al. (2002), Lindblad et al. (2004) described a multi-channel segmentation scheme applicable to CHO cells. First, a segmentation of the nuclei using contrast based threshold selection is performed, which maximizes the contrast between image background and cell nuclei. Touching or overlapping nuclei are separated by applying the watershed

transform to the distance transformed image. Information on the nuclei is then used for segmentation of the CHO cells by a seeded watershed transform based on intensity information. Bengtsson, E. and Wahlby (2004) also used information on the nuclei as input for an intensity based seeded watershed method.

Apart from methods based on the seeded watershed transform, level set based methods are also proposed for segmentation of multi-channel micrographs. In this context, Yan et al. (2008) performed an automatic RNA interference screening, for which a modified watershed transform was used for extraction of the nuclei. Furthermore, these authors demonstrated that independent level sets perform a low-quality segmentation of the cells. Owing to this, an interaction model was included which improves segmentation of touching cells. Yu et al. (2009) used a method based on topological analysis for evaluation of neuronal cells. After segmenting the nuclei, neuronal cells are also segmented. Therefore, a level set method is combined with topology preserving constraints that prevent cells from merging or splitting. This method was later improved by using generalized Voronoi diagrams (Yu et al., 2010).

2.1.5. Parameter adjustment for fluorescent micrograph segmentation

Most of the above described methods can be calibrated for different image data by adjusting several parameters. Nevertheless, manual adjustment of these parameters is very time-consuming. In this sense, a new concept to support the user in the adjustment of these parameters was provided by Pretorius et al. (2011). Instead of adjusting parameters and reviewing results in an iterative process, segmentation results are computed offline for several different combinations of parameters. A plugin for the CellProfiler software displays results for different combinations of parameters. The user can then select high-quality and low-quality segmentation results. Based on these selections, the optimal parameter combination is identified. In contrast to iterative calibration, the user is able to change parameters in real time. A disadvantage hereby is that pre-computation and storing of the segmentation results for different combinations of parameters is time-consuming and requires a large volume of memory. A concept for automated calibration of a segmentation pipeline based on hand-labeled data was proposed by Wittenberg et al. (2009). Carrying out research on applicability of this concept for segmentation of fluorescence microscopy image data is part of this work.

2.2. Available freeware tools for automated analysis of fluorescence micrographs

A variety of freeware as well as commercial tools allow the segmentation and automated analysis of fluorescent micrographs. For this section, focus is put on non-commercial tools. Of these tools, the CellProfiler has shown to be the most flexible software for biologists with only little knowledge of image processing. However, its full power can only be revealed by image processing experts. For experienced users, ImageJ (Abràmoff et al., 2004) is also an option. In order to support this hypothesis, a short summary will be provided for the tools ImageJ, CellProfiler, ICY, Blob-Finder, OME, Bisque and BioIMAX.

ImageJ is an open source, Java based image analysis tool used for many imaging and image processing applications ranging from skin analysis to neuroscience (Abràmoff et al., 2004). ImageJ, which is currently at version 1.45, provides a huge functionality. Basic image editing functionalities like changing brightness, image contrast or the image size are offered as well as routines for stack viewing, and stack editing. A rich collection of binary filters exists that consists of morphological erosion, dilation, opening, closing, skeletonization, hole filling, watershed transform, distance transform, Voronoi maps, etc. Furthermore, a collection of filters such as mean, median, Gaussian, maximum, minimum, unsharp masking, variance, edge detection, Fourier band pass filters and many others is provided. For image analysis and quantification, ImageJ offers region of interest (ROI) based measurement of area, minimum, maximum, mean, median and standard deviation of gray level, among many other measurements. Furthermore, functionality of ImageJ can be extended by creating custom plugins. On the ImageJ website, currently more than 400 plugins are available for image analysis, color image analysis, image filtering, segmentation, visualization, the handling of image stacks, etc. As each field of research uses different functionalities of ImageJ, some publicly available collections of plugins and macros for the program have been created. The MBF ImageJ collection of plugins and macros (Collins et al., 2007) was developed for the analysis of microscopic images. It additionally offers support for many commonly used file formats, such as the Zeiss .zvi extension and the Leica .lei and .lif image formats. Further functionalities offered are deconvolution, co-localization analysis, 3D surface rendering, particle analysis, among others. For a complete list of file formats and functionalities, see (Collins et al., 2007).

The CellProfiler (Carpenter et al., 2006; Lamprecht et al., 2007), currently available in version r10997, is a commonly used freeware and open source tool for the analysis of fluorescent micrographs. It was developed by the group of Anne Carpenter at the Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard University. Using the CellProfiler, the

user defines an image processing pipeline consisting of several methods. This pipeline can then be used for image segmentation and analysis. For users with only little image processing knowledge or for simple segmentation tasks, a method named "IdentifyPrimaryObjects" is offered which performs a segmentation of objects in a single image channel. The "IdentifySecondaryObjects" method performs a multi-channel segmentation by additionally using seeds from a different channel (e.g. cell nuclei). Both methods require adjustment of several parameters. For less experienced users, it is very helpful that a default value is assigned to each parameter. Apart from this, the CellProfiler offers several further image processing routines, and is able to execute ImageJ plugins and macros.

The software package ICY (de Chaumont et al., 2011) was developed at the Quantitative Image Analysis Unit at the Institute Pasteur, and is a further image processing software with good usability and plugin support. As of May 22, 2013, 187 different image processing plugins for ICY could be downloaded. Furthermore, ICY is able to execute ImageJ plugins.

Daime (Daims et al., 2006) is a further fluorescent image analysis tool, though its scope is tailored to the analysis of fluorescent in situ hybridization probes. It offers methods for background removal and noise reduction, as well as different threshold selection methods ranging from manual thresholds to locally adaptive thresholds and edge detection algorithms.

An additional tool, though with a very dedicated scope, is the Blob-Finder (Allalou and Wählby, 2009), which is designed for the evaluation of subcellular structures and uses a less flexible segmentation pipeline, with only few parameters. For segmentation of the nuclei, Otsu's threshold selection method is applied. Touching nuclei are then split by applying the watershed transform to the distance transformed image. Cytoplasm is estimated based on the nuclei by a distance transform with a user-defined radius. For the segmentation of subcellular structures, local maxima in the image are assumed to represent individual objects. These objects are separated from the image background by interactive thresholds.

Apart from tools with a strong focus on image processing and visualization, software solutions with a focus on data management or web based annotation are available that also offer some segmentation routines. The open microscopy environment OME (Goldberg, 2005), for instance, is a framework for analysis as well as management of image data. The OME offers routines for manual annotation as well as a wrapper to run ImageJ routines.

The Bio-Image Semantic Query User Environment (Bisque) software was developed by Kvilekval et al. (2009) at the Center for Bio-Image Informatics at University of California. It was developed for the exchange and analysis of biological 5D image data. Users can write custom plugins or download them. Currently available are, for example, a 3D nuclei counter and an ImageJ wrapper.

A further tool for bioimage analysis and exploration is the BioIMAX tool

(Loyek et al., 2011), developed by the Biodata Mining Group at the Bielefeld University, Germany. It provides methods for the annotation of image data as well as different clustering techniques for image segmentation.

2.3. Summary

In this chapter, an overview on publications dealing with the segmentation of fluorescent micrographs, as well as an overview on publicly available image analysis software for fluorescence image data, were provided. This review demonstrated that the task of fluorescence micrograph segmentation can be subdivided into three areas of application: The segmentation of fluorescence nuclei, segmentation of fluorescence cells and the multi-channel segmentation of cells and nuclei. For all three areas of application, methods based on the watershed transform as well as level sets were most commonly used.

Most publications dealing with the automated segmentation or analysis of fluorescence micrographs have focused on a specific experimental setup. Only publications from Tomaževič et al. (2002), Srinivasa et al. (2009), Möller et al. (2010) and Zaritsky et al. (2011) evaluated algorithms based on image data from multiple application domains. Configuration of the segmentation pipeline has usually been done manually. Exceptions to this were machine learning based methods from Nattkemper et al., who applied a competitive layer model (Nattkemper et al., 2000) and a neural network architecture (Nattkemper, 2002) for automatic segmentation of fluorescence micrographs. Differently, Lin et al. (2003) and Jeong et al. (2009) utilized machine learning techniques, including Gaussian mixture models and Bayesian networks, for segmentation of fluorescence nuclei.

For evaluation, 10 of the 55 studies compared results to hand-labeled data, whereas 27 studies counted the number of correctly and erroneously segmented objects. Remaining publications did not perform a quantitative evaluation.

In addition to an overview of segmentation methods, publicly available software solutions were also reviewed. It was shown that tools like the CellProfiler or ImageJ allow the user to define an image processing workflow based on macros or a pipeline concept. However, the selection of appropriate methods is very time-consuming and may pose an obstacle for researchers with no or only little knowledge of image processing. Alternative tools like the Blob-Finder have been developed with respect to a specific application, and hence cannot be applied for analysis of a broad range of fluorescence image data.

2.4. Conclusion

Much progress has been made towards the development of algorithms and software solutions enabling automated and objective image analysis for various fluorescence microscopy based experiments. Nevertheless, there is still room for improvement, in particular for the following topics:

- Segmentation of overlapping cells: Most state of the art methods enable the segmentation of touching cells. Resolving overlaps is possible for simple convex cells or nuclei (Brüllmann et al., 2012). However, methods for resolving overlaps for cells with irregular boundaries and non-homogeneous intensity distribution have not been developed thus far.

- Development and evaluation of segmentation methods based on data from different fluorescence imaging domains: Most methods focus on a particular experimental setup. In order to enable the segmentation of a broad range of fluorescence image data, algorithms must be adjustable for segmentation of various staining and cell types. For evaluation, most publications do not perform any qualitative comparison to hand-labeled data and often use only a single dataset.

- Improving automation for software tools: Existing tools either focus on a particular experimental setup or require the user to select a segmentation workflow and to manually adjust the parameters. The manual adjustment of the segmentation workflow to the data is often very time-consuming and requires expert knowledge of image processing. Hence, novel software tools are required aiming to reduce the amount of time that the user has to spend optimizing the segmentation workflow.

3. Summary of own contributions

Below, an overview of contributions of this thesis to the field of bioimage informatics is provided:

1. Design and implementation of a parameterized segmentation pipeline that is calibrated based on hand-labeled data. This system enables the segmentation of fluorescence micrographs from various application domains and reduces the time required for evaluation. From experimental data, it can be concluded that a small number of manually annotated cells (often <10) is sufficient for calibration of the segmentation pipeline. As configuration of the segmentation pipeline can be performed in less than 8 hours, or in less than 5 minutes if a more runtime-efficient setup is used, efficiency and the degree of automation in fluorescence micrograph analysis are improved. This is a novel approach that has not been implemented for micrograph segmentation before.

2. In order to improve figure-ground separation, a novel graph cut based method was developed. It could be demonstrated that this method outperforms other state of the art methods for several application domains, particularly for blurred and noisy micrographs.

3. Applicability of this system is illustrated based on data from varying fluorescence imaging domains, including simple nuclei as well as multipolar and overlapping cell data. For evaluation, the segmentation results were compared to manually hand-labeled data. This includes measurements of inter- and intra-observer variability. Most state of the art methods are evaluated based on a single dataset without manual hand labelings. Furthermore, parameters are often optimized manually, which might introduce a bias. To the author's knowledge, this is the first time that algorithms based on watershed transform, level sets and graph cut are compared using fluorescence microscopy image data from different application domains, based on manual hand labelings and automated parameter optimization.

4. In order to reduce the time required for calibration of the segmentation pipeline, time-efficient strategies for determining the optimal parameterization of the segmentation pipeline are compared. In this context, parameter spaces of different segmentation methods are investigated

for varying types of data. Apart from the present publication (Held et al., 2012), this is the first time that parameter spaces are investigated for different image processing methods based on a broad range of fluorescence microscopy image data.

4. Overview of image segmentation, optimization and evaluation methods

In this thesis, methods from different fields of research are used. The most important segmentation, optimization and evaluation methods which are relevant for this thesis are briefly described and discussed in this chapter.

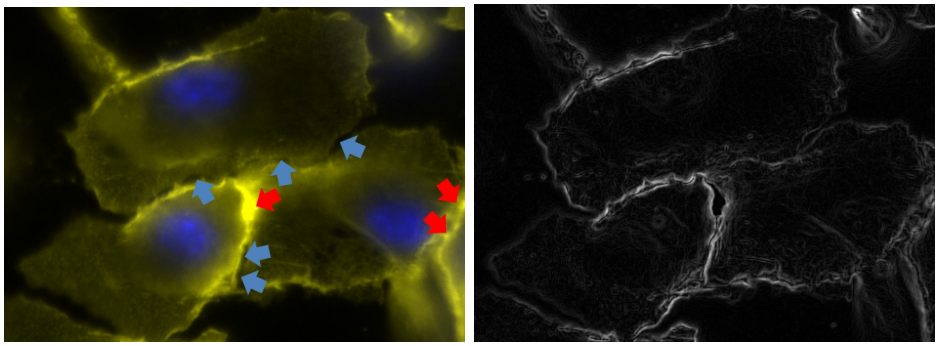
4.1. Image segmentation

Image segmentation provides a basis for most automated image analysis tasks. In the following, a brief description of segmentation methods that will be adapted for fluorescence micrograph segmentation in this work is provided.

4.1.1. Watershed transform

The watershed transform (WT) (Roerdink and Meijster, 2000) is a commonly used mathematical morphology based image segmentation method, and is often applied for splitting of touching cells or nuclei. Interpreting the image intensity profile as a surface, the WT simulates a flooding process. Thereby, each local intensity minimum is interpreted as a source of water. Instead of using local intensity minima, alternative sources (e.g. based on morphology) can be used as sources of water, also known as seeds. Based on the seeds, flooding is performed until regions from different sources touch each other. Corresponding touching points are then interpreted as the watershed lines that separate touching objects.

Applying the WT for splitting of touching cells or nuclei based on intensity information, each local intensity minimum leads to an individual cell or nucleus. Due to the flooding, touching cells or nuclei are split at the location of maximum intensity. As a result of this, a high quality splitting is only obtained if the input image shows monotonously increasing intensity values from the core of an object to the boundary. For micrographs showing more complex intensity distribution, erroneous segmentation results are obtained if the WT is applied to the intensity image. An example of such fluorescence image data is provided in Figure 4.1(a). In order to extend the scope of the WT to a broader range of image data, alternative input images are required.



(a) (b)

Figure 4.1.: When using the watershed transform for splitting of macrophages based on intensity (a), splitting must be carried out at locations with maximum intensities (red arrows), or at locations with minimum intensities (blue arrows). Instead, gradient magnitude can be used for splitting of the cells as splitting for this example can always be performed at location of maximum gradient magnitude (b).

Instead of image intensity, cells can be split at locations of strongest gradient magnitude (compare Figure 4.1(b)). For splitting of round, oval or convex objects, the inverted distance transformed image can be used as input for the WT. A typical application is the splitting of touching nuclei based on a binarization of these nuclei, as can be seen in Figure 4.2. After distance transform and inversion, each local intensity minimum (Figure 4.2(e)) represents an individual nucleus. Furthermore, intensities of the inverted distance image increase from the core of an object to the boundary, which is ideal for the watershed transform. As a result, using the gradient magnitude image or the distance transformed image enables splitting of cells with convex shape or non-complex intensity distribution. Note that an object with a complex intensity distribution shows several local intensity minima and maxima.

For objects with irregular boundaries and complex intensity distribution, additional information is required to enable robust splitting. For some applications, information on position and extension of the cell nuclei is available and incorporated for segmentation of the cells. Seeds based on local intensity minima can then be replaced by information on the nuclei. The resulting method is known as seeded watershed transform (SWT).

From an implementation point of view, the WT and SWT can be efficiently implemented based on a heap. As each pixel must be be visited only once, this results in complexity $O(n \log n)$.

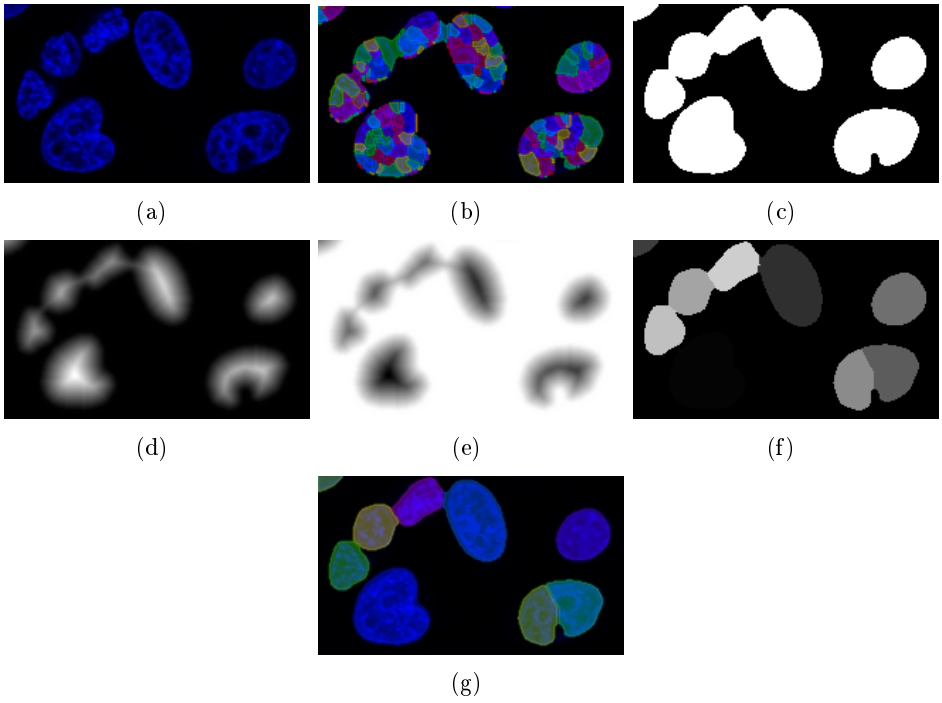


Figure 4.2.: Applying the watershed transform to the image intensities (a) results in a strong over-segmentation (b). Instead, the binarized image (c) can be used for splitting of the nuclei based on their convex shape. After performing a distance transform (d), each local maximum is assumed to represent an individual nucleus. Applying the watershed transform to the inverted distance image (e), touching nuclei are split (f) and over-segmentation is reduced. An additional overlay image is shown in (g).

4.1.2. Level sets and fast marching level sets

In the past years, the level set method (Sethian, 1999) has become a very popular framework for image segmentation. This method can be nicely illustrated by a wave front moving according to a speed function $F(x, y)$, defined for each pixel (x, y). The level set method performs tracking of a given initial wave front in time. Thereby, splitting and merging of different fronts is possible. Hence, topology of the described objects can change. Tracking the level set Φ in time requires solving the initial value problem:

$$\Phi_t + F(x, y)|\nabla\Phi| = 0,$$

where Φ_t denotes partial derivative of the level set function and $\Gamma = \{(x, y)|\Phi(x, y) = 0\}$ the initialization. Solving this initial value problem

requires an iterative solution scheme updating each image pixel at each iteration, which is very time-consuming. Instead, update of the level set function can be restricted to points located near the current boundary. This procedure speeds up the computation and is known as narrow band level set method (for details, see Sethian (1999)).

By restricting motion of the front to positive values $F(x, y) \geq 0$, the fast marching level set method is obtained. Due to this restriction, arrival time of the front $T(x, y)$ can be described by a boundary value problem:

$$|\nabla T(x, y)|F(x, y) = 1,\ T(x, y) = 0 \text{ on } \Gamma.$$

For solving the fast marching (FM) level set equation, Sethian (1999) provides a very efficient discretization scheme, which can be solved with complexity $O(n \log n)$.

In literature, level sets are applied for figure-ground separation as well as for the separation of touching cells incorporating information on the cell nuclei. Figure 4.3 shows an example of cells split based on a seeded fast marching level set (SFM) technique.

4.1.3. The graph cut method

The graph cut algorithm was first applied for image binarization by Greig et al. (1989). Several years later, a more efficient scheme was developed by Boykov et al. (1999). In this work, Boykov's implementation is used as described in (Boykov and Funka-Lea, 2006; Boykov and Kolmogorov, 2003, 2004; Boykov et al., 2001).

In contrast to the level set method that locally optimizes an energy function, the graph cut algorithm globally optimizes an energy function by interpreting the image as a graph. This graph is then cut into different regions, in such way that an energy function is globally minimized. For each pixel p contained in a set of Pixels P and a neighborhood system N represented by a set of neighbor pixels $\{p, q\}$, the cost function $E(A)$ is minimized. Thereby, A describes a labeling of the image. In this application, P contains every pixel in the input image and N denotes a 4-neighborhood system. $E(A)$ is then defined as:

$$E(A) = E_{\text{data}}(A) + E_{\text{smooth}}(A) = \lambda \sum_{p \in P} R_p(A) + \sum_{\{p,q\} \in N} B\{p, q\}\delta(A_p, A_q),$$

where R_p denotes a penalty for assigning pixel p to a label l out of a set of labels L.

The graph cut method can be applied for figure-ground separation as well as for splitting of touching cells. For figure-ground separation, a two-label graph cut is used. In this case labels represent foreground or background

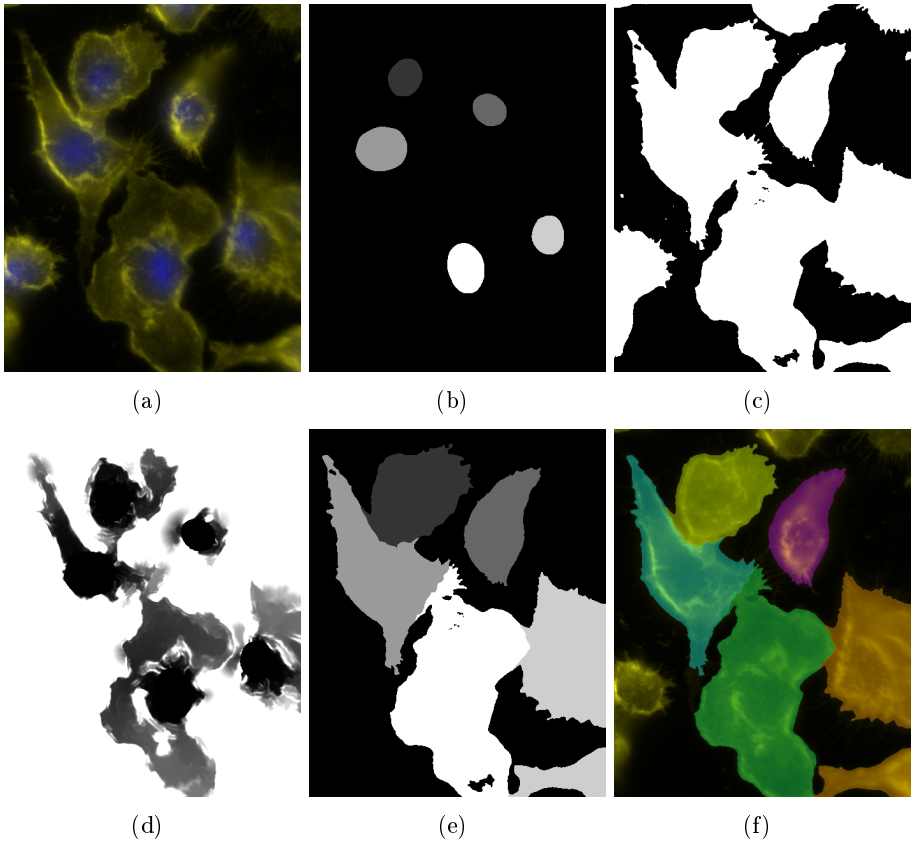
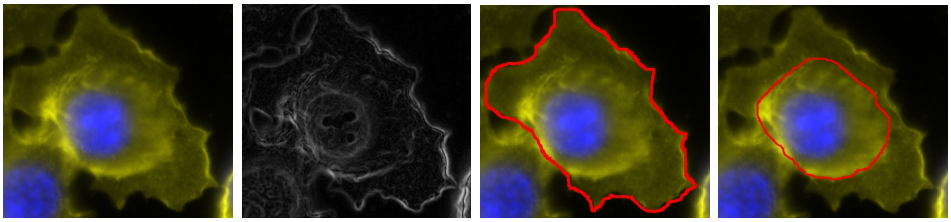


Figure 4.3.: Illustration of the fast marching level set method. Touching macrophages (a) are split incorporating information on position and extension of nuclei (b) and binary fore- and background information (c). Arrival times of the fast marching level set (d) are computed using a gradient magnitude and an image curvature based speed function (for implementation and parameterization of the fast marching level set, see Section 5.4.2, p. 50). The resulting label image (e) and an overlay of the segmentation result (f) show the split macrophages.

pixels. For splitting of touching objects, a unique label can be assigned to each object. Hence, the number of required labels correlates with the number of touching cells, and a so-called multi-label graph cut is required. For both applications, the data energy term and the smoothness energy term must be carefully chosen to provide good results. For a detailed description of the graph cut based threshold selection method, see Section 5.3.4, p. 44. The graph cut based splitting method that uses information on nuclei is described in Section 5.4.3, p. 52.



(a) (b) (c) (d)

Figure 4.4.: Illustration of graph cut shrinking bias: Assume that the macrophage fully contained in (a) shall be separated from the image background based on gradient magnitude (b) using the graph cut based method. Also assume that the regional term forces the nucleus to be assigned to the foreground label and all background pixels to be assigned to the background class. In this case, the contour overlaid in (c) depicts the expected result. However, solution (d) is preferred by the graph cut, as the sum of gradient costs along the object boundary is smaller for (d). The effect that smaller contours are preferred by the graph cut is known as shrinking bias.

The major drawback of the graph cut method is known as shrinking bias, and refers to the fact that small objects are preferred by the graph cut method. The reason for this is the smoothness energy term that sums up costs of all pixels with different neighbor labels. As a result, smoothness energy tends to be smaller for smaller objects. This effect is illustrated in Figure 4.4.

4.2. Calibration of the segmentation pipeline

For automatic calibration of the segmentation pipeline, choice of an appropriate optimization method is crucial. For most optimization problems occurring in this work, testing all possible settings is not applicable, as run time increases exponentially with the number of free parameters. Using such a brute force technique would require a very long time on a standard computer. Instead, more runtime-efficient optimization methods enabling optimization of multi-dimensional parameter spaces are required. As the investigated parameter spaces are currently unknown, ill-conditioned parameter spaces must be considered. Hence, the optimization method must be able to handle local performance maxima and minima.

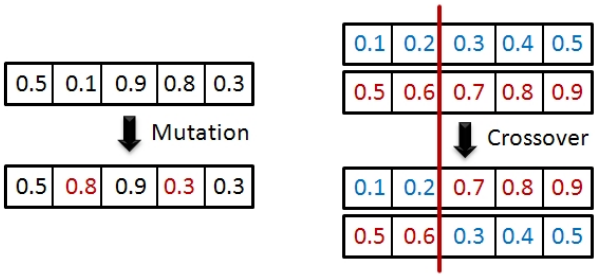


Figure 4.5.: Illustration of a genetic algorithm's crossover and mutation operations for a genome consisting of 5 parameters. In this example, 2 parameters are mutated.

4.2.1. Genetic algorithms

Due to the ability of jumping out of local extrema and efficiently optimizing multi-dimensional parameter spaces, genetic algorithms (Goldberg, 1989) are applied for automated parameter optimization. Using genetic algorithms, parameters to be optimized are regarded as a genome that consists of a set of alleles. Each of the alleles thereby represents a parameter. In this work, discrete parameters defined by a minimum value, a maximum value and a step size are used to reduce size of the parameter space. For initialization of the genome, each parameter is initialized with a random value.

After initialization, the parameter space is explored. Therefore, new individuals defined by their genome are created based on so-called crossover and mutation operations. Depending on encoding of the genome, different mutation and crossover operators can be used. The mutation operator applied for this work creates a new individual based on a parent individual by mutating each of the alleles with a given probability p_{mut}. By mutating an allele, an equally distributed random value is assigned to the corresponding parameter. The crossover operation creates a new individual based on two parent individuals. For this work, the single point crossover operator is used. Thereby, both parent genomes are split at a random location and mixed. Probability that a crossover operation occurs is denoted as p_{cross}. For an illustration of crossover and mutation operations see Figure 4.5. For this work, p_{mut} is adjusted in such way that one to two mutations are expected for each new individual. Assuming that the investigated parameter spaces consist of 5 to 10 parameters, p_{mut} is set to 0.2. Additionally, $p_{\text{cross}} = 0.5$ is used to expect a crossover operation for every second individual. Figure 4.6 shows an example for optimization of a complex parameter space with local

extrema. For implementing the genetic algorithm, the GAlib developed by M. Wall at the MIT is used (Wall, 1996).

For runtime reasons, convergence of the genetic algorithm is assumed if no better individual has been determined for 200 iterations. Using this convergence criterion, in practice about 1000 iterations were performed until convergence for a fixed combination of methods. Using an average reference dataset consisting of 20 images and including tree-fold cross-validation, optimization required about 24 hours for runtime-efficient algorithms. For the less efficient graph cut based cell splitting, parameter optimization required about one week.

4.2.2. Coordinate descent

The coordinate descent optimization method is an approach for minimization of a multidimensional function $f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ that does not require gradient information. The outcome of this method strongly depends on its initialization. Using coordinate descent, each parameter is optimized individually at each iteration by a brute-force approach. Hence, for the first iteration parameter $\mathbf{x}_1$ is varied and optimized. All other parameters remain fixed. The resulting optimal parameter vector is then used for optimization of parameters $\mathbf{x}_2$, ..., $\mathbf{x}_n$. This coordinate-wise optimization of all parameters is then repeated until no new performance maximum is determined. For a 2D example of this search method, see Figure 4.6 (b, c).

Results for this search method strongly depend on initialization, and identification of the global optimum is not guaranteed. For details on convergence of the coordinate descent method, see (Luo and Tseng, 1992).

4.2.3. Cross validation

Automatic parameter optimization requires disjunctive reference and testing datasets to prevent learning by heart and over-fitting. Hence, a k-fold cross validation can be applied. Splitting the dataset into k equally sized subsets, calibration is performed on $k-1$ subsets. The remaining subset is used for testing of the algorithm. This procedure is repeated for each constellation of reference and testing data. Using the k-fold cross validation, a performance measurement is obtained for each sample in the dataset, ensuring that reference and testing data are always disjunctive.

4.3. Evaluation

Comparison of the user-defined hand-labeled data to the automatically generated result is an important aspect of parameter optimization. For this task, different metrics can be used.

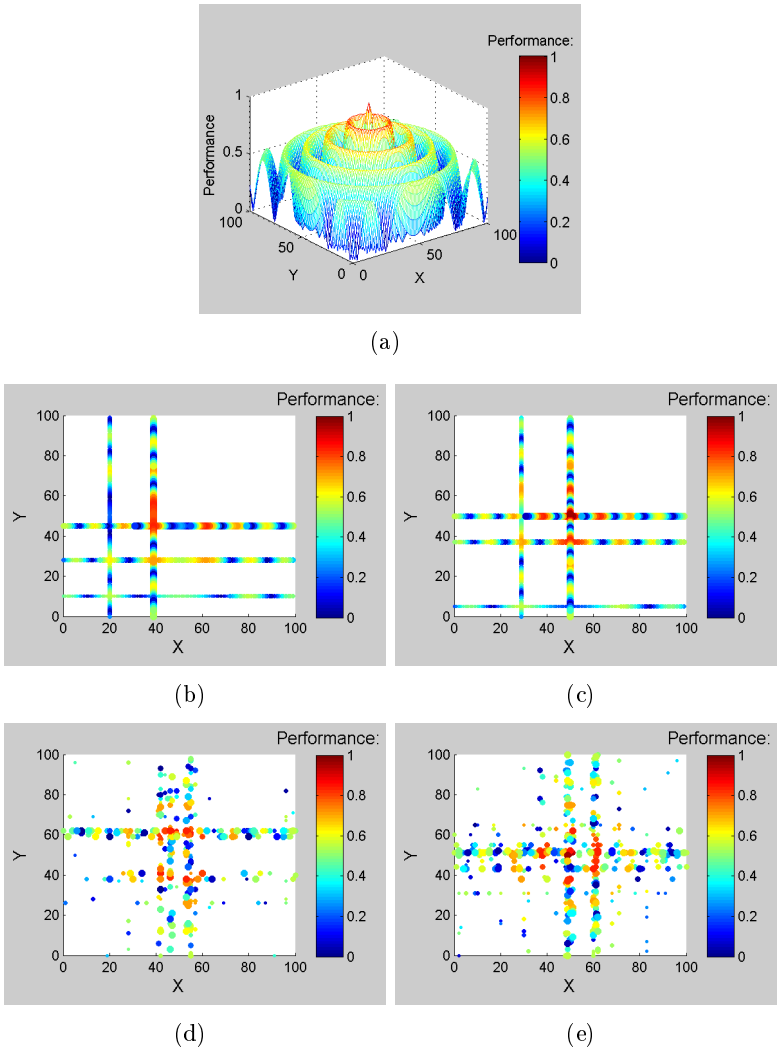


Figure 4.6.: Optimization of a two-dimensional function, with parameters x and y showing several local maxima and minima (a), using coordinate descent optimization (b, c) and genetic algorithms (d, e). Note that size of the dots encodes the sequence of investigated parameter combinations. The smallest point is tested first, whereas the largest point is the last tested one. (b, c) show that output of the coordinate descent method depends on the initialization, whereas output of the genetic algorithm hardly depends on initialization, but rather on chance. Using a bad initialization, the coordinate descent optimization scheme sticks in a local performance optimum (b). In (c), the global maximum is correctly identified. The genetic algorithm can also stick in local maxima as depicted in (d), whereas the global optimum is correctly identified in (e).

4.3.1. Jaccard Overlap

The Jaccard similarity enables pairwise comparison of segmentation results. Denoting the set of pixels representing the hand-labeled data as S_{gt}, and the set of pixels representing the segmentation result as S_{res}, the Jaccard similarity o_{j} is defined as the ratio between intersection and union of S_{gt} and S_{res}:

$$o_{\mathrm{j}} = \frac{|S_{\mathrm{gt}} \bigcap S_{\mathrm{res}}|}{|S_{\mathrm{gt}} \bigcup S_{\mathrm{res}}|}.$$

This results in $o_{\mathrm{j}} = 1$ for a perfect match, and $o_{\mathrm{j}} = 0$ if hand labeling and segmentation results are disjunctive.

4.3.2. Hausdorff distance

In contrast to the Jaccard similarity, the Hausdorff distance considers object boundaries instead of overlaps. Denoting the set of boundary pixels of the hand labeling as B_{gt} and the boundary pixels of the segmentation results as B_{res}, the distance of each pixel contained in B_{gt} to the closest pixel in B_{res} is determined. In this work, for distance measurement, the Euclidean distance was utilized. The Hausdorff distance is then determined by computing the maximum of all minimum distances:

$$o_{\mathrm{hd}}(B_{\mathrm{gt}}, B_{\mathrm{res}}) = \max_{p \in B_{\mathrm{gt}}} \min_{q \in B_{\mathrm{res}}} d(p, q),$$

where $d(p, q)$ denotes the Euclidean distance of boundary pixels p and q.

As the Hausdorff distance is not symmetric, $o_{\mathrm{hd}}(B_{\mathrm{gt}}, B_{\mathrm{res}}) = o_{\mathrm{hd}}(B_{\mathrm{res}}, B_{\mathrm{gt}})$ does not always hold. That is why the following symmetric version of the Hausdorff distance is defined:

$$o_{\mathrm{H}} = \max(o_{\mathrm{hd}}(B_{\mathrm{gt}}, B_{\mathrm{res}}), o_{\mathrm{hd}}(B_{\mathrm{res}}, B_{\mathrm{gt}})).$$

Instead of determining the maximum of all minimum distances for comparison of two segmentation results, which corresponds to measuring only the largest error, the root mean square of minimum distances can be used. This results in the $o_{\mathrm{hd}}^{\mathrm{rms}}$ metric, which takes the complete boundary into account:

$$o_{\mathrm{hd}}^{\mathrm{rms}} = \sqrt{\frac{1}{|B_{\mathrm{res}}|} \sum_{p \in B_{\mathrm{gt}}} (\min_{q \in B_{\mathrm{res}}} d(p, q))^2}.$$

The symmetric root mean square of minimum distances is defined as:

$$o_{\mathrm{H}}^{\mathrm{rms}} = \max(o_{\mathrm{hd}}^{\mathrm{rms}}(B_{\mathrm{gt}}, B_{\mathrm{res}}), o_{\mathrm{hd}}^{\mathrm{rms}}(B_{\mathrm{res}}, B_{\mathrm{gt}})).$$

4.3.3. Accuracy

The Jaccard similarity and the Hausdorff distance only take qualitative overlap or boundary based features into account. Information on the number of correctly identified cells n_{tp}, number of missed cells n_{fn}, and the number of erroneously detected cells n_{fp} is not used for assessment of segmentation performance.

Reduction of n_{fn} and n_{fp} is crucial for many experiments requiring measurement of average cell size or counting objects. In order to evaluate this quantitative information, accuracy o_{a} of the segmentation is defined as:

$$o_{\mathrm{a}} = \frac{n_{\mathrm{tp}}}{n_{\mathrm{tp}} + n_{\mathrm{fp}} + n_{\mathrm{fn}}}.$$

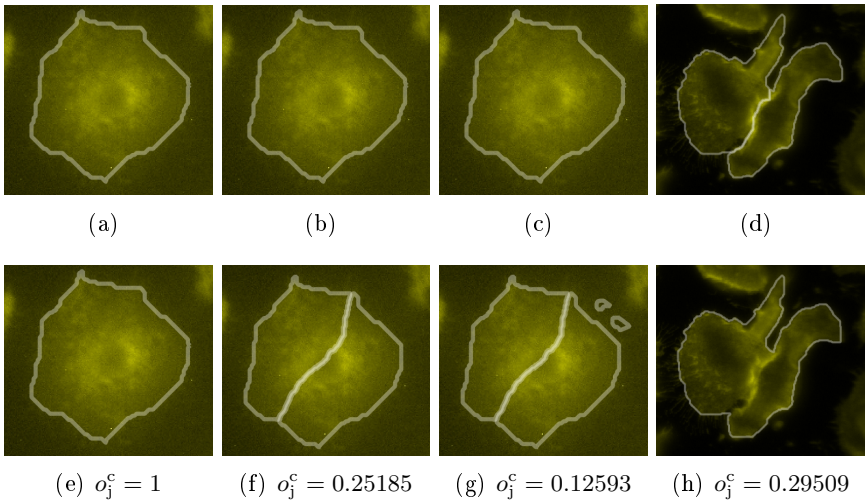
Determining n_{tp}, n_{fp} and n_{fn} requires a mapping between all hand-labeled objects and the segmentation result. For this mapping, each of the segmented result objects is assigned to the best-fitting hand-labeled object. In addition, only a single segmented object can be assigned to each hand-labeled region. Additionally, a rejection class is implemented. This rejection class is used, for instance, for regions that exceed the image boundary and for those that cannot be split unambiguously by the human observer. Objects assigned to the rejection class are excluded from evaluation and do not influence the performance measurement. Hence, a decision rule for determining if a segmented region of interest is assigned to the rejection class is required. For this work, regions that better fit to a rejection class region than to any valid region, in terms of Jaccard similarity, are regarded as objects corresponding to the rejection class and excluded from evaluation.

4.3.4. Combined Jaccard metric

In order to associate quantitative and qualitative performance measurements, the combined Jaccard metric has been defined. Using this metric, a mapping between segmented and hand-labeled regions is performed as described in Section 4.3.3. This also includes incorporation of the rejection class. Based on this mapping, corresponding pairs of hand-labeled regions and segmented regions are obtained. Jaccard similarity of such an optimal pair of corresponding regions is denoted as o_i^* with $i \in 1, 2, ..., n_{\mathrm{tp}}$. Note that each segmentation result can only be assigned to one hand-labeled object, and vice-versa. Based on these measurements, the combined Jaccard metric $o_{\mathrm{j}}^{\mathrm{c}}$ is defined as:

$$o_{\mathrm{j}}^{\mathrm{c}} = \frac{1}{n_{\mathrm{tp}} + n_{\mathrm{fp}} + n_{\mathrm{fn}}} \sum_{i=1}^{n_{\mathrm{tp}}} o_i^*.$$

Examples for some typical constellations of hand-labeling and segmented regions and the corresponding performance values are shown in Figure 4.7.



(a) (b) (c) (d)

(e) $o_j^c = 1$ (f) $o_j^c = 0.25185$ (g) $o_j^c = 0.12593$ (h) $o_j^c = 0.29509$

Figure 4.7.: Illustration for representative scenarios comparing segmentation results to hand-labeled data based on the combined Jaccard similarity measurement. Hand labelings are shown in images (a-d). For segmentations and the resulting performance measurements, compare images (e-h). Note that additional degradation of false positive and false negative detections reduces performance for images (f), (g) and (h).

Note that $o_j^c = 1$ corresponds to a perfect segmentation, whereas $o_j^c = 0$ corresponds to disjunctive hand-labeled data and segmentation results.

4.3.5. Friedman test

In order to determine if the segmentation resulting from different methods is significantly different, a repeated measurement analysis of variance can be performed if the data is normally distributed. For most of the data, the assumption of normal distribution is violated. That is why, for statistical analysis, the non-parametric Friedman test (Friedman, 1937) is used to verify statistically significant differences. In order to carry out this test, the statistical software package R in version 2.13.2 was used.

4.3.6. Wilcox test

The Wilcox signed-rank test Wilcoxon (1945) is a non-parametric statistical hypothesis test that can be used for determining if repeated measurements of a specific experiment are significantly different. In contrast to Student's t-test, no normal distribution of the data is required. For adjustment of P-values for multiple testing, the method of Benjamini and Hochberg (1995)

is applied. In order to carry out this test, the statistical software package R in version 2.13.2 was used.

4.4. Summary

In this chapter, an overview of well-established techniques that constitute a basis for this work was provided. This includes the watershed transform, level set and fast marching level set methods from the field of image processing. From the area of discrete multidimensional optimization, genetic algorithms and the coordinate descent method were briefly described. For evaluation of segmentation performance, different boundary as well as overlap based qualitative and quantitative performance metrics were described, as well as statistical significance tests for non-normally distributed repeated measurements.

5. Design of the segmentation pipeline

The robust segmentation of fluorescence micrographs poses a challenge, as the development of an algorithm applicable for segmentation of a broad range of different fluorescence micrographs is hardly possible. Such an algorithm must be able to handle different cell types, stains and acquisition conditions. In order to support a broad range of application domains, the segmentation pipeline consists of exchangeable image processing methods, each implementing one of the following stages:

1. Preprocessing (noise reduction and shading correction).
2. Figure-ground separation (differentiation between foreground and background pixels).
3. Object splitting (separation of touching, overlapping and overlaying objects).

For an illustration of the segmentation pipeline, see Figure 5.1.

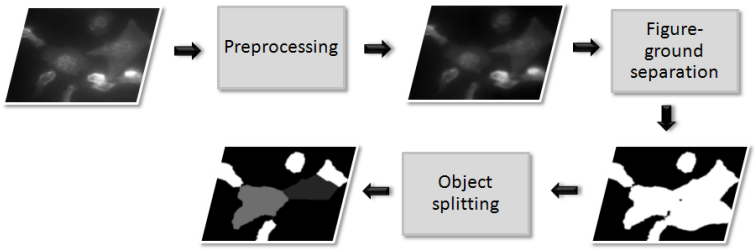


Figure 5.1.: Stages of the segmentation pipeline.

In the following, developed methods are described and discussed for each stage of the segmentation pipeline. In order to decrease runtime for calibration and to improve segmentation for blurred and noisy micrographs and for micrographs showing strong intensity variations, custom methods are developed and included.

As the best performing combination of methods and its parameterization shall be automatically determined on the basis of hand-labeled data, implemented methods must be chosen carefully, considering that, with the number

of supported methods, runtime of the calibration process increases. Hence, only methods are considered that are not too similar. Furthermore, methods with a short runtime are preferred against slower ones. For an overview on implemented methods and corresponding adjustable parameters, see Figure 5.2. Apart from the number of image processing methods, discretization of each method's parameter space strongly affects runtime of the calibration process. That is why discretization of each parameter is performed in such a way that a broad range of fluorescent micrographs can be segmented while keeping the number of discrete values small. For this work, each parameter was discretized by a maximum of 30 values. Due to this limitation, for some parameters, constant step sizes were not appropriate; for them, exponentially increasing step sizes were used. An example for such a parameter is the radius of the minimum filter based preprocessing that is adjustable from 2 to 1024 pixels.

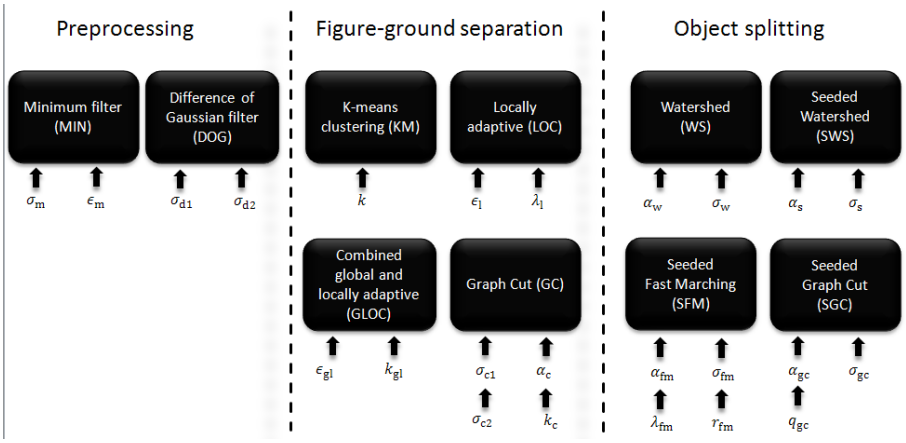


Figure 5.2.: Overview of implemented methods and adjustable parameters for each of the methods.

5.1. Acronyms

A unique acronym was defined for each method of the segmentation pipeline. For an overview on these acronyms, see Figure 5.2. Based on this, a segmentation pipeline can be denoted as M_1-M_2-M_3, where M_1 indicates a preprocessing, M_2 a figure-ground separation, and M_3 an object-splitting method. Thus, a combination of MIN Preprocessing, KM figure-ground separation and WS cell splitting, for instance, is denoted as MIN-KM-WS.

5.2. Preprocessing

Preprocessing methods improve the image in such way that the subsequent figure-ground separation and object splitting tasks are facilitated. Preprocessing methods developed for the segmentation pipeline aim to reduce the noise level in the image and remove shading artifacts caused by uneven illumination or blur that could otherwise harden the segmentation process. Furthermore, tiny details that might mislead figure-ground separation and object splitting are reduced by preprocessing.

In this work, two different preprocessing methods were implemented. Both methods combine a smoothing step for reduction of sensor noise, with a difference imaging based technique for shading correction as well as correction of illumination artifacts. The first method is based on the Difference of Gaussian (DoG) filter, as proposed by Leong et al. (2003). As runtime of the DoG filter increases for large filter kernels, a more runtime-efficient minimum filter based preprocessing method was developed to accelerate calibration of the segmentation pipeline. Due to runtime limitations, edge-preserving methods based on median or bilateral filters were not considered for preprocessing.

5.2.1. Difference of Gaussian based preprocessing (DOG)

The DoG preprocessing method applies a Gaussian smoothing filter with standard deviation $\sigma_{\mathrm{d1}} \in \{1, 2, ..., 20\}$ for noise reduction. Then, a difference imaging based shading correction technique is performed. For shading correction, the estimated background image is subtracted from the smoothed one. Using this preprocessing method, the image background is estimated based on a Gaussian filter with standard deviation σ_{d2} for the removal of background objects. Next, the background corrected image is subtracted from the smoothed image. This is equivalent to applying a difference of Gaussian (DoG) filter (see Figure 5.3). In order to reduce runtime of the Gaussian filter, convolution is implemented using a separable filter kernel. However, runtime of the filter still increases linearly with increasing standard deviations. As a result, runtime is very high for micrographs depicting large objects and for micrographs with a low-frequency image background. In order to prevent the DoG preprocessing method from slowing down the optimization process, σ_{d2} is restricted to $\sigma_{\mathrm{d2}} < 100$. For supporting micrographs that do not require shading correction, no shading correction is performed if $\sigma_{\mathrm{d2}} > 100$. In this case, the smoothed image is returned and no difference imaging operation is performed. These characteristics are encoded by $\sigma_{\mathrm{d2}} \in \{3, 6, ..., 102\}$.

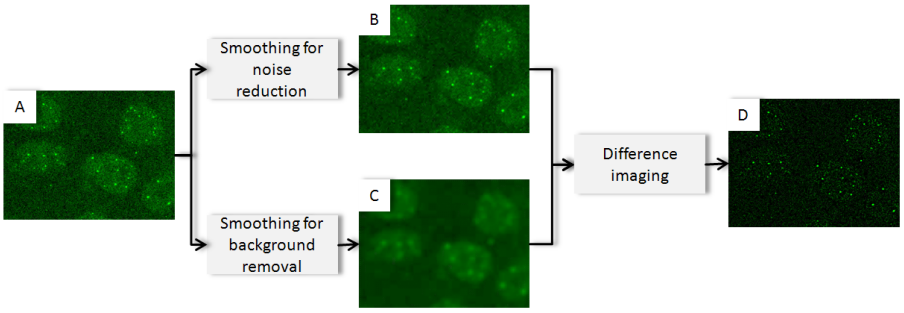


Figure 5.3.: Workflow of the difference imaging based preprocessing method.

5.2.2. Minimum filter based preprocessing (MIN)

As runtime of the Gaussian filter increases for large standard deviations, a more time-efficient preprocessing method was developed. This method also applies a Gaussian smoothing filter with standard deviation $\sigma_{\mathrm{m}} \in \{1, 2, ..., 20\}$ for noise reduction. Image background is estimated based on a minimum filter with a rectangular kernel to remove all background objects. For bandpass-filtering, a difference imaging operation is performed (Figure 5.3). In order to reduce runtime of this method, the minimum filter is implemented according to the method of Van Herk (1992), which results in constant runtime independent from the kernel size O(1). As size of the objects of interest varies from smaller than 10 pixels for spot detection to more than 1000 pixels for the segmentation of large cells, kernel size of the minimum filter based preprocessing method must be adjustable in this range. This leads to a large search space if a fixed step size is used. For more efficient parameterization, exponentially increasing step sizes are applied. Hence, width and height of the filter kernel are set to $2^{\epsilon_{\mathrm{m}}}+1$. Further, adjusting the parameter $\epsilon_{\mathrm{m}} \in \{1, 2, ..., 10\}$ corresponds to a minimum filter radius ranging from 3 to 2049 pixels.

5.3. Figure-ground separation

After reduction of illumination artifacts and noise by preprocessing routines, objects of interest are separated from the image background. In order to solve this task, different figure-ground separation routines are selected, compared and evaluated. For this work a global intensity based k-means clustering method, a locally adaptive threshold selection method and an improved locally and globally combined threshold method, which are closely related to the state of the art, were implemented. In order to improve the segmentation of blurred micrographs and micrographs showing strong variation in intensity, a novel graph cut based figure-ground separation method

was developed.

5.3.1. K-means clustering based figure-ground separation (KM)

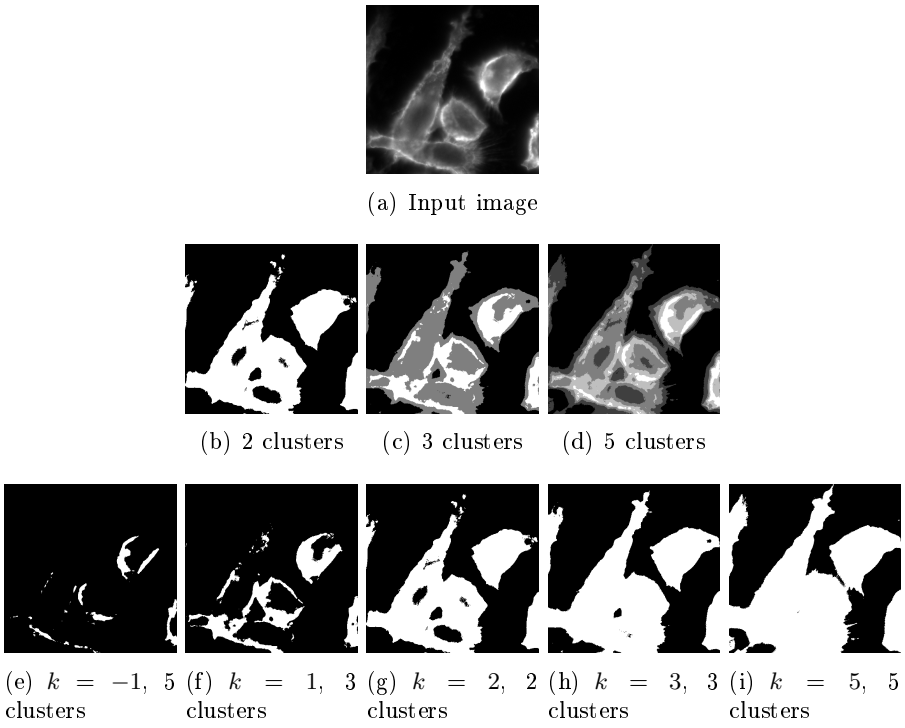
For the separation between fore- and background pixels, a k-means clustering based method (Hartigan and Wong, 1979) was developed. If solely image intensity is used as a feature, k-means clustering can be implemented efficiently using the histogram of the image (Cheng and Rajapakse, 2009). In contrast to alternative routines, this method enables adjustment of the threshold level by changing the number of clusters $k^* \geq 2$. After clustering, pixels belonging to each of the clusters are interpreted as fore- or background.

For simplicity, it is now assumed that the dynamics of the cells are large compared to dynamics of the background. In this case, after clustering with $k^* \geq 2$, the darkest cluster represents the image background; all remaining clusters are interpreted as foreground. This rule for assigning clusters to fore- and background classes is denoted as mode m_1 from here on. Using m_1 for figure-ground separation results in a lower global threshold value if the number of clusters is increased. The drawback of this parameterization is the inability to handle intensity distributions exhibiting more dynamics in the image background than in the image foreground. Using mode m_1, clustering with $k^* \geq 2$ results in an over-segmentation of the image. In order to enable robust figure-ground separation for unexpected image data, a second mode m_2 that interprets only the brightest cluster as image foreground and remaining clusters as image background is offered. Using m_2, increasing the number of clusters results in higher threshold values.

In order to segment data from different fluorescence imaging domains, parameterization of k-means clustering must be adjustable in such way that both modes for assigning clusters to fore- and background are supported. This could be implemented by introducing a second variable switching between the two modes. In this case, the user of this method is required to adjust two parameters, while low threshold values are obtained using a high number of clusters if mode m_1 is applied, or using a low number of clusters for mode m_2.

To improve the usability of this method, parameterization was improved by reparameterization based on a single variable k. Depending on the value of k, the number of clusters used for k-means clustering k^* is defined. Increasing k will then correspond to increasing size of the segmented objects, which results in lower threshold values. That is why, for $k \geq 2$, clustering can be performed with $k^* = k$ clusters while interpreting only the darkest cluster as image background; this corresponds to mode m_1. In order to support m_2 and to guarantee that increasing values for k result in lower threshold values, k^* must decrease if k increases. Hence, the number of clusters could be set to $k^* = -k$, if $k \leq 2$ while interpreting only the brightest cluster as image foreground. This would correspond to m_2.

The drawback of the described approach is that clustering results are the same for $k = 2$ and $k = -2$, and that output is not defined for $k \in \{-1, 0, 1\}$. In order to overcome these drawbacks and to ensure that increasing the number of clusters corresponds to increasing size of foreground objects, the number of clusters is set to $k^* = 4 - k$ if $k < 2$. For an illustration of segmentation results using varying values for k, see Figure 5.4. Using the described method, the k-means clustering based figure-ground separation is parameterizable by a single parameter $k \in \{-8, -7, ..., 12\}$.



(a) Input image

(b) 2 clusters (c) 3 clusters (d) 5 clusters

(e) $k = -1$, 5 clusters (f) $k = 1$, 3 clusters (g) $k = 2$, 2 clusters (h) $k = 3$, 3 clusters (i) $k = 5$, 5 clusters

Figure 5.4.: Illustration of the KM figure-ground separation method based on a representative micrograph (a). Images (b - d) show the output of intensity based k-means clustering for different numbers of clusters. For images (e, f), after clustering, the brightest cluster is assumed to represent the objects of interest, which is parameterized by $k < 2$. For these parameterizations, clustering is performed with k clusters. For images (g - i), all clusters except the darkest one represent the object of interest, which is parameterized by $k \geq 2$, and clustering is performed with $4 - k$ clusters to avoid redundancy in the parameterization.

5.3.2. Locally adaptive figure-ground separation (LOC)

The k-means clustering based threshold selection method, which does not take local image features into account, is prone to shading and illumination artifacts that can only partially be removed by preprocessing. As Harder et al. (2006, 2007) described the locally adaptive threshold selection method as superior to Otsu's method, this approach is considered as an alternative to the k-means clustering based method. Using locally adaptive thresholds, intensity distribution in a local rectangular neighborhood around each pixel is modeled to form a decision rule determining if the current pixel represents an object of interest or noise. Assuming that noise in this local neighborhood is normally distributed, it can be modeled based on the average intensity $\mu(x, y)$ and the standard deviation of the intensity $\sigma(x, y)$. The investigated pixel $I(x, y)$ is then considered as foreground if its intensity deviates from the neighborhood's mean and standard deviation:

$$I(x, y) \geq \mu(x, y) + \lambda_{\mathrm{l}}\sigma(x, y),$$

where λ_{l} adjusts the dissimilarity required to interpret the current pixel as foreground. For time-efficient optimization, width and height of the local neighborhood are set to $2^{\epsilon_{\mathrm{l}}}+1$ to support different object sizes. Based on this, parameter space is discretized by $\epsilon_{\mathrm{l}} \in \{1, 2, ..., 20\}$ and $\lambda_{\mathrm{l}} \in \{-5, -4.5, ..., 5\}$.

5.3.3. Combined global and locally adaptive figure-ground separation (GLOC)

A drawback of the locally adaptive figure-ground separation method (LOC) are outliers produced in background regions. These outliers are caused by the local neighborhood containing only few foreground pixels. For such regions, average and standard deviation of intensities are very low. Utilizing locally adaptive thresholds, this corresponds to low threshold values, which often lead to misinterpreting background pixels as foreground ones. In order to suppress these outliers, the locally adaptive method is combined with a global threshold selection method, as described by Harder et al. (2007). Pixels are then only considered as foreground if both the locally adaptive as well as the global threshold selection method identify a pixel as foreground. Harder et al. (2007) therefore combine the locally adaptive figure-ground separation method with Otsu's global threshold selection method (Otsu, 1975). As previously described, the k-means clustering based threshold selection method is used instead of Otsu's approach due to the possibility of conveniently adjusting threshold values by changing the number of clusters. Combining the global k-means clustering based threshold selection method with the locally adaptive method requires optimization of parameter spaces of both methods. Owing to this, a different parameterization of this figure-ground separation method was developed for this work. Therefore, local

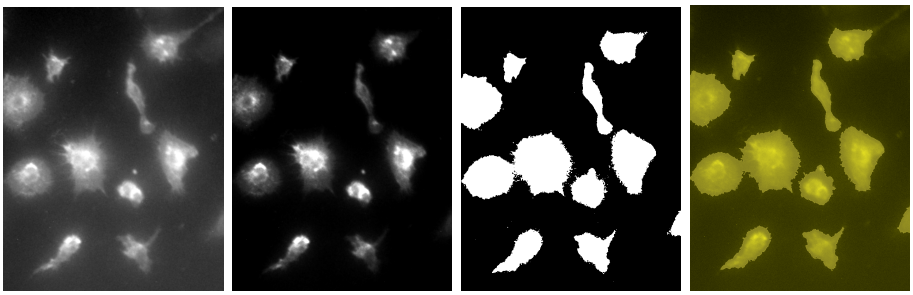
dissimilarity $\lambda_{\mathrm{gl}}(x, y)$ based on local measurements of mean $\mu_{\mathrm{gl}}(x, y)$ and standard deviation $\sigma_{\mathrm{gl}}(x, y)$ in a rectangular neighborhood of width and height $2^{\epsilon_{\mathrm{gl}}} + 1$ is defined as:

$$\lambda_{\mathrm{gl}}(x, y) = \frac{I(x, y) - \mu_{\mathrm{gl}}(x, y)}{\sigma_{\mathrm{gl}}(x, y)}.$$

The local dissimilarity is then used as a weight map for the input image $I(x, y)$ to obtain the locally adaptive weighted image $I_{\mathrm{gl}}(x, y)$ (for an example, see Figure 5.5(b)):

$$I_{\mathrm{gl}}(x, y) = \lambda_{\mathrm{gl}}(x, y) I(x, y).$$

Next, $I_{\mathrm{gl}}(x, y)$ is used as input for the k-means clustering based threshold selection technique. As a result, the described adaptive thresholding scheme enables combining local and global intensity features requiring only two parameters: the number of clusters k_{gl} and the parameter adjusting width and height of the rectangular neighborhood r_{gl}. Figure 5.5 illustrates the workflow of this combined threshold selection method. In order to enable time-efficient adaption to a broad range of fluorescence micrographs, the number of clusters is restricted to a range $k_{\mathrm{gl}} \in \{-8, -7, ..., 12\}$ and size of the neighborhood is discretized by $\epsilon_{\mathrm{gl}} \in \{1, 2, ..., 10\}$.



(a) (b) (c) (d)

Figure 5.5.: Illustration of the combined global and locally adaptive figure-ground separation method. Based on a noisy image with shading artifacts (a), local dissimilarity between each pixel and a local neighborhood is computed (b). Using this method, global k-means clustering based threshold selection is performed, and a binary image is obtained (c, d).

5.3.4. Graph cut based figure-ground separation (GC)

The precise identification of object boundaries for blurred fluorescence micrographs poses a challenge, as object boundaries for some data show a low

contrast to the image background. This task is even more complicated if object intensities strongly vary in a micrograph. In order to improve boundary extraction under such challenging conditions, a novel graph cut based figure-ground separation technique for fluorescent micrographs was developed and included in the segmentation pipeline. This novel algorithm has also been published and evaluated based on publicly available data in (Held et al., 2013b).

The graph cut method (for details, see Chapter 4.1.3, p. 26) interprets the image as a graph and globally minimizes an energy function. Applying the graph cut for figure-ground separation, a two-label energy function classifying pixels as fore- or background is used. This requires definition of the data and the smoothness energy terms $E_{\text{data}}(A)$ and $E_{\text{smooth}}(A)$, with both depending on a labeling A of the image. Both energy terms are defined based on penalty terms $B\{p,q\}$ and $R_p(A)$. In the following, a workflow for estimating both terms based on image data is described that aims to improve robust figure-ground separation for blurred and noisy data.

For this task, foreground objects are separated from the image background at the location of strong gradients. That is why the smoothness energy term depends on the image gradient $\nabla I(x,y)$, which is additionally normalized to a range of $[0,1]$ and denoted as $\nabla I(x,y)^{[0,1]}$. Including an exponential weighting factor $\alpha_{\text{c}} > 0$, which adjusts strength of the gradient magnitude component, the smoothness energy is defined as:

$$B\{p,q\} = (1 - \nabla I(x,y)^{[0,1]})^{\alpha_{\text{c}}}.$$

The ∇ operator is implemented using the non-parametric Sobel operator, as appropriate smoothing has been carried out in the preprocessing step.

The graph cut data term $R_p(A)$ evaluates how well a specific labeling fits to a given data model. In this application, $R_p(A)$ is defined by a set of fore- and background pixels that are very likely to be part of fore- or background objects. A method for identifying fore- and background seeds is described in the following sections. Based on these seeds, the penalty $R_p(A)$ is set to 1 if pixel p is a fore- or background seed, but assigned to the wrong class by the graph cut. Otherwise, penalty $R_p(A)$ is set to 0.

Based on seeds that are known to represent fore- or background structures, remaining pixels are classified by minimization of the graph cut energy and a binary segmentation of the image is obtained. An overview on the workflow of the graph cut based figure-ground separation routine is provided in Figure 5.6.

Identification of background seeds for the graph cut data term

The identification of background seeds is a rather simple task because it is sufficient to use pixels that are very likely to represent the background due to the shrinking bias (for an illustration, see Figure 5.7). Such pixels can be

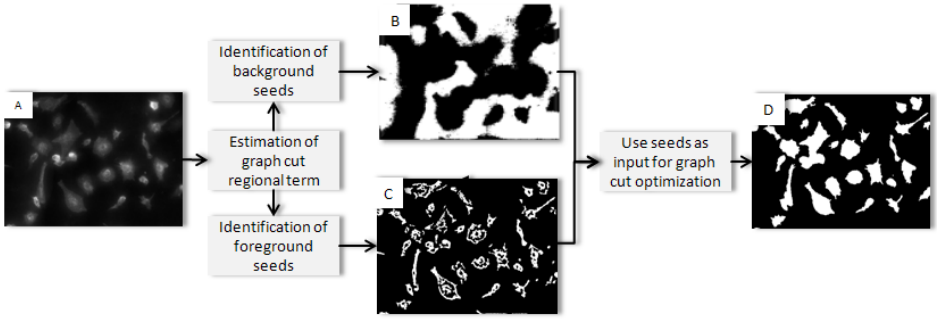


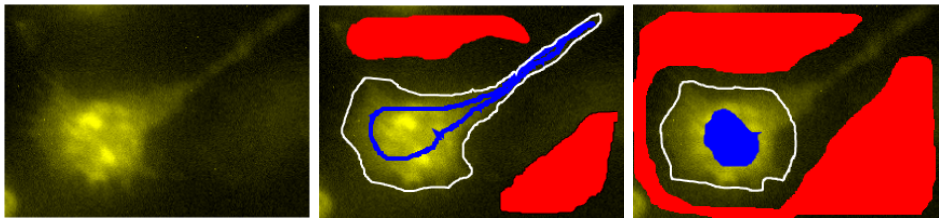
Figure 5.6.: Illustration of the workflow for the graph cut based figure-ground separation routine. First, foreground and background seeds are determined. Foreground seeds are identified by a k-means clustering technique (B) and background seeds by a combination of DoG filters, dual threshold selection and the removal of small objects (C). These seeds are then used to define the graph cut method's energy function. By minimization of this energy function, a segmentation of the image (D) is obtained.

identified by analyzing global image intensity distribution, as illumination artifacts have already been reduced by the preprocessing method. Hence, for identification of background seeds, a very conservative threshold selection method is applied. This conservative method is implemented using a k-means clustering based threshold selection method with a high number of clusters k_c (see Section 5.3.1, p. 41), which interprets only the darkest cluster as image background.

Identification of foreground seeds for the graph cut data term

The identification of foreground seeds poses a challenge, mainly due to the shrinking bias of the graph cut. Hence, it is of utter importance that foreground seeds are contained in each tiny extensions of an object, otherwise small extensions of cells are often not correctly separated from the background and cut off (see Figure 5.7). Furthermore, intensity distribution can vary from cell to cell. That is why global threshold selection methods fail with the robust identification of foreground seeds.

Instead of utilizing global intensity features, pixels are used that are relatively bright compared to the image background. For estimation of the foreground seeds, the input image $I(x, y)$ is smoothened with a strong Gaussian filter function G, with standard deviation σ_{c2}. In order to select relatively bright objects and to reduce liability to noise, this difference imaging scheme can be extended to a DoG based band pass filter by:



(a) (b) (c)

Figure 5.7.: Illustration of the graph cut shrinking bias based on a micrograph depicting a macrophage (a). For the graph cut based figure-ground separation, precise location of foreground seeds (blue) is usually more important than precise location of background seeds (red). Despite imprecise background seeds, segmentation is carried out correctly (white line in image (b)). In contrast, cell extensions are cut off in image (c) if foreground seeds are missing in the cell extensions. The effect, that smaller objects are preferred by the graph cut, is known as shrinking bias and is caused by the graph cut energy function.

$$I_{\mathrm{bp}}(x,y) = (I(x,y) * G(\sigma_{\mathrm{c1}})) - (I(x,y) * G(\sigma_{\mathrm{c2}})), \text{ with } \sigma_{\mathrm{c1}} < \sigma_{\mathrm{c2}},$$

where $I_{\mathrm{bp}}(x,y)$ denotes the band pass filtered image and σ_{c1} and σ_{c2} denote the standard deviation of Gaussian filter kernels. By removing pixels with a negative filter response, $I^{+}_{\mathrm{bp}}(x,y)$ is obtained. Note that parameters of this step are usually adjusted in such way that only the boundary of each object is recognized as foreground to increase the probability that seeds are also contained in small extensions. Figure 5.8(b) shows an example for such a DoG filtered image. For this image, foreground structures must be separated next from noise and artifacts.

The robust separation of foreground structures from noise and artifacts poses a challenge, as cell extensions often show a poor contrast to the image background. This results in image intensity values that are very similar to background structures. Many techniques, including global and locally adaptive threshold selection ones, have been applied to solve this task. This includes all threshold selection techniques described in this work (i.e., k-means clustering based figure-ground separation, locally adaptive thresholds and combined global and locally adaptive thresholds), Isodata threshold (Velasco, 1980), Otsu's threshold selection method (Otsu, 1975), entropy based thresholds (Kapur et al., 1985) and dual thresholds (Buie et al., 2007). However, none of these methods allows robust separation between cell extensions and background noise. Instead, a non-parametric method based

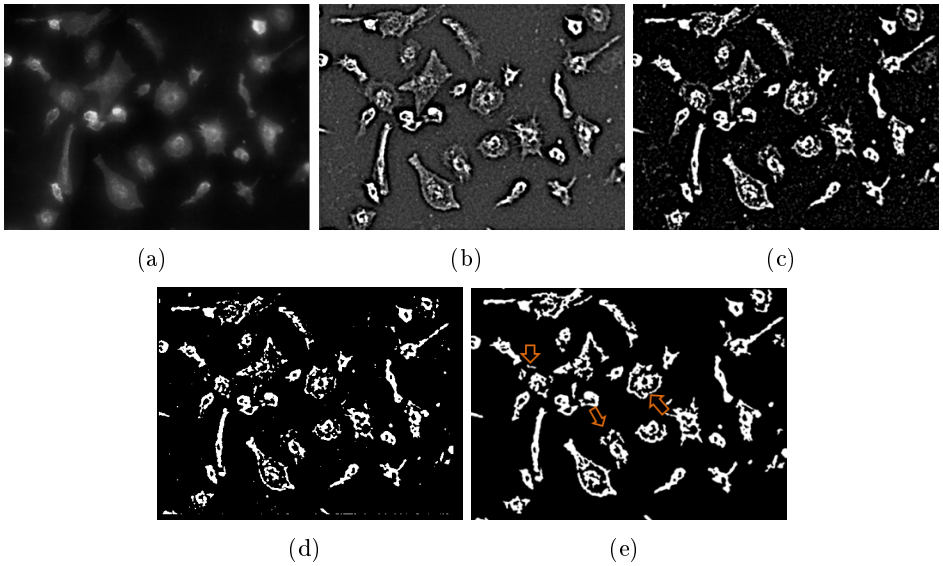


Figure 5.8.: Workflow for the identification of foreground seeds. Input image (a) is filtered with a DoG filter (b). Pixels with a negative filter response are eliminated (c). In the following step, it is important that seeds representing the image background or noise are eliminated. Image (d) is obtained by optimal manual threshold selection. Note that by applying the proposed non-parametric scheme using dual thresholds and removing small objects, high quality seeds preserving small extensions and weak boundaries of each cell are obtained (d). Arrows indicate important differences.

on dual thresholds was developed. This method automatically adapts to the noise level and size of objects represented in the image, based on information on the background seeds. This method showed good results across datasets from different application domains, without requiring adjustment of parameters. In the following, this method will be described and illustrated.

Application of the dual threshold technique requires a lower and a higher threshold value. As background seeds that are known to represent background pixels have already been determined, this information can be used for estimation of both threshold values. Therefore, intensity distribution is analyzed inside and outside the background seeds. The lower threshold is then chosen so that $p_{\mathrm{gc}} = 98\%$ of the background seeds exhibit a lower intensity compared to the threshold value. In contrast to this, for estimation of the higher threshold value only non-background seeds are considered. The higher threshold value is empirically set to the median intensity of the foreground pixels. For both estimations, pixels that hold $I^{+}_{\mathrm{bp}} = 0$ are ignored.

After applying the dual threshold for identification of potential foreground seeds, small objects that are likely to correspond to noise or background structures are removed. Therefore, minimum size of the objects of interest is estimated. As the lower threshold value is chosen in such way that some background or noise pixels are still contained in the image, objects with a size smaller than the average object size can be removed. This results in the foreground seeds required to estimate $R_p(A)$. An overview on this workflow is provided in Figure 5.8.

The method for identification of foreground seeds can additionally be adjusted by replacing routines for computation of lower threshold value, higher threshold value and the size estimation method. Due to the possibility of changing the noise level in the image by the preprocessing step, usage of fixed parameters showed good results across several fluorescence imaging data. Hence, using this method with the described fixed parameters is proposed to reduce size of the parameter space.

5.4. Object splitting

After figure-ground separation, a binary image separating the image into fore- and background pixels is obtained. In this image, touching, overlapping or overlaying objects are not yet separated from each other. This section describes methods for the splitting of touching objects. Most of these approaches exploit characteristics of the cell shape or intensity distribution. For example cell nuclei are often separated from each other incorporating information on their convex shape. If cell morphology and intensity distribution are more complex, these methods become prone to errors and tend to split cells erroneously. In order to improve the segmentation of such complex cells, information on the cell nuclei is incorporated. In this section, methods for cell splitting based on seeds like cell nuclei as well as non-seeded methods are described and investigated. The most commonly used method that can be used with or without seeds is the watershed transform. As an alternative to the seeded watershed transform method, both an approach based on fast marching level sets and another based on graph cuts were developed.

5.4.1. Cell splitting based on watershed transform and seeded watershed transform (WS and SWS)

The watershed transform requires appropriate input data that shows increasing intensities from the core of each object to the object boundary. For fluorescence microscopy image data, the most commonly used input is a distance transformed and inverted binary image, a gradient magnitude image, or a combination of both images, named hybrid image. Applying the distance transform to the binary image allows splitting of the objects of interest based on shape. After distance transform and inversion, it is assumed

that each local intensity minimum represents an individual object. Using this method, the convex shape of some cell types or cell nuclei is exploited (for an example, see Figure 4.2, p. 25). Using the gradient magnitude based approach, it is assumed that gradient magnitude is 0 at the core of each object, and that touching objects can be split at the location of strongest gradient magnitude. In contrast to the distance transform based method, the gradient magnitude based approach is misled by objects showing several local intensity minima or maxima. In order to combine and benefit from both methods, the hybrid approach was implemented in this work as described by (Malpica et al., 1997). Denoting the gradient magnitude based image as $\nabla I(x, y)$ and the distance transformed binary image as $I_{\mathrm{D}}(x, y)$, the input for the hybrid watershed transform $I_{\mathrm{H}}(x, y)$ is defined as:

$$I_{\mathrm{H}}(x, y) = (1 - \alpha_{\mathrm{w}}) I_{\mathrm{D}}(x, y) + \alpha_{\mathrm{w}} \nabla I(x, y),$$

where $\alpha_{\mathrm{w}} \in [0, 1]$ is a weighting factor that allows balancing between the impact of the gradient magnitude based and the distance transform based components.

Application of the watershed transform in various scenarios has shown that it tends to result in an over-segmentation of the image. In order to reduce over-segmentation, a merging step is commonly applied using, for example, information on cell shape or maximizing gradient magnitude at the cell boundary. As an alternative to this approach, a Gaussian smoothing filter with standard deviation σ_{w} can be used for smoothing of $I_{\mathrm{H}}(x, y)$ and thus reducing over-segmentation artifacts. For this application, the letter approach is implemented, as no prior knowledge on cell size, shape or intensity distribution is required.

In contrast to the watershed transform, the seeded watershed transform does not regard local intensity extrema as seeds. Instead, cell nuclei that have been previously segmented in a different channel are used as seeds. The seeded watershed transform works analogously to the watershed transform, and uses parameters α_{s} for balancing of gradient and distance information and σ_{s} for smoothing of the hybrid image.

5.4.2. Cell splitting using seeded fast marching level sets (SFM)

For micrographs with available information on nuclei, a fast marching level set method can be used as alternative to the seeded watershed transform. For a brief description of the level set and the fast marching level set method, see Section 4.1.2, p. 25. Preliminary results for this approach have been published in (Held et al., 2011). Using the fast marching level sets for splitting of touching cells can be illustrated by several wavefronts released at the location of each nucleus. Each wavefront expands until it touches another wavefront or the image boundary. Motion of the front is controlled by a

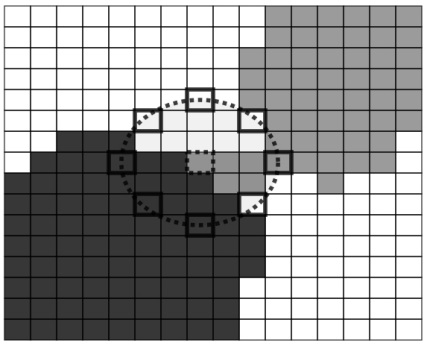
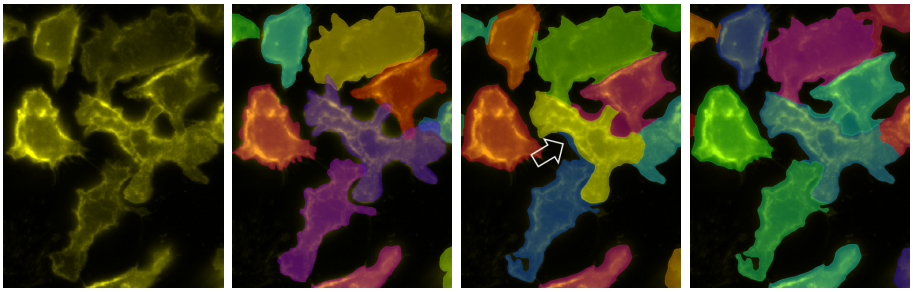


Figure 5.9.: Estimation of curvature $\kappa(x, y) \in [-0.5, 0.5]$ based on a label image used by the fast marching level set method. A circle with n (in this example 8) equally spaced points (bold pixels) is investigated, and the number of pixels that share the label of the center point is counted. In this example, 1 out of 8 points shares the label of the center point, which results in a curvature $\kappa(x, y) = \frac{1}{8} - 0.5 = -0.375$. Based on this information, the front of the level set can be slowed down for this pixel in order to reduce leaking.

speed function $F(x, y)$. In order to increase probability that touching objects are split at a location of strong gradient magnitude, the wavefronts are slowed down at locations of large gradient values $\nabla I(x, y)$. The ∇ operator is implemented using a DoG filter with standard deviation σ_{fm}. Furthermore, an exponential weighting factor α_{fm} determining impact of the gradient magnitude information is included. $F(x, y)$ is additionally chosen so that the fast marching level set method is prevented from running into the image background. Therefore, information on background pixels is used. Note that a binary image $I_{\mathrm{B}}(x, y)$, that holds $I_{\mathrm{B}}(x, y) = 0$ if pixel (x, y) is a background pixel and $I_{\mathrm{B}}(x, y) = 1$ otherwise, is available from the figure-ground separation step. As the fast marching level set method has a known tendency to leaking, a curvature term $\kappa(x, y)$ is incorporated as proposed by Nilsson and Heyden (2003). For this work, a slightly modified version of this term was used, which was described in (Held et al., 2011). Modification is required, because in the present application, the image contains multiple fronts starting at locations of different nuclei. Owing to this, motion of each front is tracked in an additional label image $I_{\mathrm{L}}(x, y)$. For estimation of the curvature $\kappa(x, y)$, n equally spaced points $\Delta(x, y)_i, i = 1, ..., n$ located on a discrete circle with radius r_{fm} are analyzed. Curvature is then estimated based on counting the number of points that share the label of the current pixel:



(a) (b) (c) (d)

Figure 5.10.: Macrophages (a) and corresponding manual annotation (b). The splitting resulting from fast marching level sets (c) shows leaking (white arrow). Including the curvature based term, leaking is reduced(d).

$$\kappa(x,y) = \frac{1}{n}\#\{i : I_{\mathrm{L}}(x,y) = I_{\mathrm{L}}((x,y) + \Delta(x,y)_i)\} - \frac{1}{2}.$$

For illustration of this term, see Figure 5.9. In order to increase or decrease influence of the curvature term, a weighting factor $\lambda_{\mathrm{fm}} \in \{0, 0.2, ..., 1.8\}$ is included in the speed function. For an example showing that including this curvature based term can reduce leaking, see Figure 5.10. Combining information on seeds and image background, gradient magnitude and curvature, the speed function is defined as:

$$F(x,y) = \frac{1 + \lambda_{\mathrm{fm}}\kappa(x,y)}{(\nabla I(x,y))^{\alpha_{\mathrm{fm}}}} I_{\mathrm{B}}(x,y).$$

To allow time-efficient optimization of parameters of the seeded fast marching level set method, parameter space is discretized by $r_{\mathrm{fm}} \in \{3, 6, ..., 17\}$, $\lambda_{\mathrm{fm}} \in \{0, 0.2, ..., 1.8\}$, $\sigma_{\mathrm{fm}} \in \{1, 2, ..., 9\}$ and $\alpha_{\mathrm{fm}} \in \{1, 2, ..., 10\}$.

5.4.3. Cell splitting with seeded graph cuts (SGC)

In contrast to the fast marching level set method that determines a local optimum of an energy function for splitting of touching cells, the graph cut algorithm is able to globally optimize an energy function. That is why the graph cut approach is considered as an alternative to both the seeded fast marching level set method and the seeded watershed transform. Instead of the binary graph cut used for figure-ground separation (Section 5.3.4, p. 44), a multi-label graph cut is applied for the splitting of touching objects. Using the multi-label graph cut, each seed is represented by a unique label. As described in Section 4.1.3 (p. 26), the graph cut method minimizes an energy

function $E(A)$ based on sum of smoothness energy $E_{\text{smooth}}(A)$ and a data based regional energy term $E_{\text{data}}(A)$:

$$E(A) = E_{\text{data}}(A) + E_{\text{smooth}}(A).$$

In order to enable fair comparison of splitting capabilities of the graph cut and the fast marching level set method, $E_{\text{smooth}}(A)$ is defined such that cells are split at the location with strongest gradient magnitude. For balancing of the influence of gradient magnitude values $\nabla I(x, y)$, an exponential weighting factor $\alpha_{\text{gc}} \in \{1, 2, ..., 10\}$ is included.

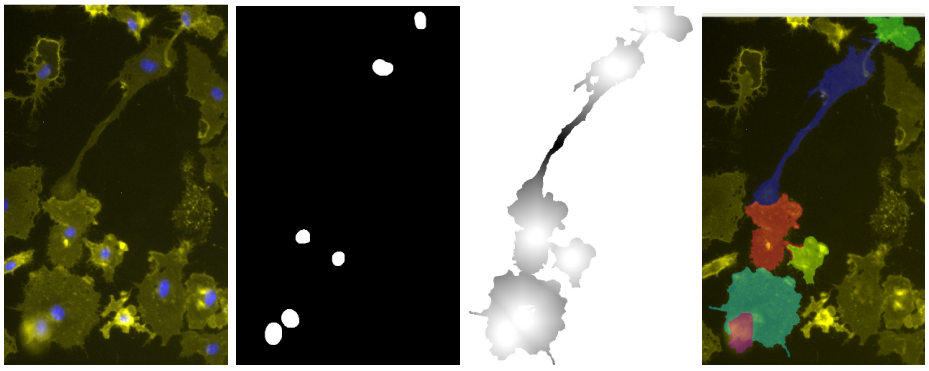
For consistence with the level set method, the ∇ operator is implemented by a differential of Gaussian filter with standard deviation $\sigma_{\text{gc}} \in \{1, 2, ..., 9\}$. One of the major problems of the graph cut method is the well-known shrinking bias (Boykov and Funka-Lea, 2006). As a result of this, initialization of each cell with the contour of its nucleus often leads to erroneous segmentations. For an illustration of the shrinking bias, see Figure 5.8, p. 48. In order to reduce this shrinking bias, an additional distance prior based weight map $I_{\text{D}}(x, y)$ is included.

Assuming that a binary segmentation partitioning the image into fore- and background is known from the previous figure-ground separation stage and denoted as $I_{\text{B}}(x, y)$, the distance based weight map can be determined by computing the distance of each pixel to the closest nucleus. The distance of pixels assigned to the image background as well as the distance inside the cell nuclei is set to 0. This image is then inverted and normalized to the range $[1, q_{\text{gc}}]$, with $q_{\text{gc}} \in \{0, 0.2, ..., 1\}$, and denoted as $I_{\text{D}}^{[1,\text{q}_{\text{gc}}]}(x, y)$. An example for the resulting distance image is shown in Figure 5.11 (c). Combining information on the distance prior and gradient magnitude, the smoothness energy term is obtained:

$$E_{\text{smooth}}(A) = (1 - I_{\text{D}}^{[1,\text{q}_{\text{gc}}]}(x, y))(1 - \nabla I(x, y))^{\alpha_{\text{gc}}} \delta(A_p, A_q),$$

where $\delta(A_p, A_q)$ is 0 if $A_{\text{p}} = A_{\text{q}}$, and 1 if otherwise.

For the data based energy term, information on figure-ground separation is combined with information on the cell nuclei. Assuming that each cell contains only one nucleus, an object containing n nuclei is split into n cells. By assigning a unique label to each nucleus and an additional label to the image background, $n + 1$ labels are required for definition of the data term. For pixels belonging to the image background, the data term is set to ∞ for all labels, except the background label. For pixels representing a nucleus, the data term is set to ∞ for all labels, except its unique label. For pixels that are not part of the image background and do not represent a nucleus, the regional term is set to ∞ for the background label only. All remaining combinations of target label and label assigned by the graph cut are plausible. Hence, the corresponding data term is set to 0. An example for application of the described graph cut based cell splitting scheme is provided in Figure 5.11(d).



(a) (b) (c) (d)

Figure 5.11.: Example for splitting of touching macrophage cells (a) by means of a graph cut based splitting scheme, which uses information on corresponding nuclei (b). Cells touch each other after figure-ground separation. For splitting of these cells, a multi-label graph cut method is used. For the smoothness energy term, a distance based weight map (c) is incorporated, providing information on the splitting of the cell nuclei if no gradient information is available. Shrinking bias is thus reduced. The smoothness cost term combines the distance function with information on the gradient magnitude. Together with information on size and extension of the cell nuclei, information on the image background and the smoothness energy term, cells are separated from each other (d).

5.5. Parameterization of the segmentation pipeline

Runtime of the calibration strongly depends on discretization of the parameters of each method. Thus, discretization of each parameter is performed in such way that a broad range of fluorescent micrographs can be segmented while keeping the number of discrete values small. For this work, each parameter was discretized by a maximum of 30 values. For an overview of parameterization of all methods implemented for the segmentation pipeline, see Table 5.1.

5.6. Brute force calibration of the segmentation pipeline

Using the described modular segmentation pipeline (see Chapter 5, p. 37) allows segmentation of a broad range of fluorescent image data. In order

Method	Parameter	Range
MIN	σ_{m}	{1, 2, ..., 20}
	ϵ_{m}	{1, 2, ..., 11}
DOG	σ_{d1}	{1, 2, ..., 10}
	σ_{d2}	{3, 6, ..., 102}
KM	k	{-8, -7, ..., 12}
LOC	ϵ_{l}	{1, 2, ..., 20}
	λ_{l}	{-5, -4, ..., 5}
GLOC	ϵ_{gl}	{1, 2, ..., 20}
	k_{gl}	{-9, -7, ..., 12}
GC	σ_{c1}	{1, 2, ..., 9}
	σ_{c2}	{5, 7, ..., 19}
	α_{c}	{1, 2, ..., 10}
	k_{c}	{3, 6, ..., 21}
WS	α_{w}	{0, 0.1, ..., 1}
	σ_{w}	{1, 2, ..., 10}
SWS	α_{s}	{0, 0.1, ..., 1}
	σ_{s}	{1, 2, ..., 10}
SFM	r_{fm}	{3, 6, ..., 18}
	λ_{fm}	{0, 0.2, ..., 1.8}
	σ_{fm}	{1, 2, ..., 9}
	α_{fm}	{1, 2, ..., 10}
SGC	σ_{gc}	{1, 2, ..., 9}
	α_{gc}	{1, 2, ..., 10}
	q_{gc}	{0, 0.2, ..., 1}

Table 5.1.: Overview of discretization of all methods.

to calibrate the segmentation pipeline based on hand-labeled data, a dual optimization problem must be solved (see Figure 5.12). Hence, using brute force optimization, all combinations of methods and all combinations of free parameters must be tested.

Using a brute force approach for calibrating the segmentation pipeline is very time-consuming. For the described segmentation pipeline consisting of two preprocessing, four figure-ground separation and four splitting methods, 32 combinations of methods must be tested to identify the optimal combination thereof. As the average number of different parameterizations for each method corresponds to approximately 1500, about $1500^3 \approx 3.4 \cdot 10^9$ combinations of parameters must be tested for each combination of methods. This results in $3.4 \cdot 10^9 \cdot 32 \approx 10^{11}$ different parameterizations of the segmentation pipeline. Assuming that only a single image is used for calibration of the segmentation pipeline and application of the segmentation pipeline requires $1s$, this would result in a runtime of 3170 years.

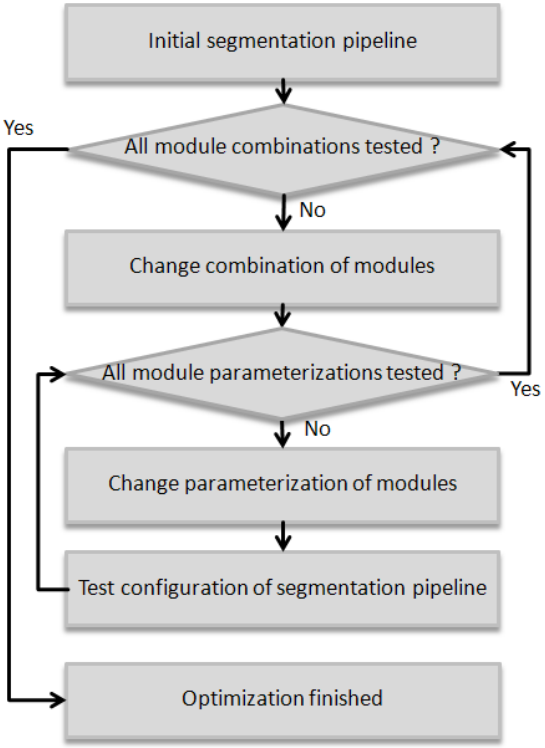


Figure 5.12.: Illustration of the workflow to solve the dual optimization problem. In order to determine the best configuration of the segmentation pipeline, the optimal combination of methods and all method parameters must be automatically adjusted.

The question of how to speed up the calibration process by utilizing more time-efficient segmentation methods is addressed in the following chapter.

5.7. Summary

In this chapter, design of the segmentation pipeline consisting of the stages of preprocessing, figure-ground separation and object splitting was described. This included the development of exchangeable parameterized methods for each of the stages, as well as discretization of each method. All methods were developed with a focus on the segmentation of fluorescent micrographs depicted in this work (compare Figures 6.1-6.7, pp. 60-66), but can potentially be used for segmentation of similar data. Range and discretization of each parameter were carefully chosen so that a broad range of data can be segmented, while keeping search spaces small.

Most of the described methods are mainly based on the state of the art. An exception to this is the graph cut based figure-ground separation that was developed for this work. The k-means clustering based and the combined global and locally adaptive figure-ground separation methods are based on state of the art, and were improved to decrease runtime of the calibration.

6. Experiments and Results

Automatic calibration of the segmentation pipeline based on hand-labeled data requires both determining an optimal combination of segmentation methods and optimizing each method's parameters. As optimal parameterization of each method depends on the combination of methods, and vice-versa; this task is called dual optimization problem. Using a brute force approach, runtime of the optimization increases exponentially with the number of methods and parameters. For the described segmentation pipeline, brute force optimization would take several years and is hence not applicable (for details, compare Section 5.6, p. 54). Thus, more efficient strategies to solve the dual optimization problem are investigated.

In this chapter, an overview of the data on which this work is based on is provided (Section 6.1). Based on this, different objective functions required for calibration of the segmentation pipeline are compared (Section 6.2). Furthermore, applicability of each method for segmentation of data from various fluorescence imaging domains is investigated (Section 6.3). As data from all application domains shows high image quality, micrographs were overlaid with simulated shading and noise artifacts (Section 6.4) to explore the limits of the presented segmentation methods. Next, a detailed investigation of method parameter spaces was performed to allow selection of a time-efficient parameter optimization method (Section 6.5). Based on this, different optimization methods are compared (Section 6.6). Finally, strategies for providing high-quality hand-labeled data in a short time are compared and discussed (Section 6.7).

6.1. Materials

For validation and improvement of the segmentation pipeline, micrographs from different application domains are required. In the following sections, a brief overview is provided on representative image data and the hand-labeled data used for evaluation.

6.1.1. Fluorescence image data used in this thesis

In order to provide an insight into various fluorescence datasets, for each application domain one of the simplest and one of the most challenging images are depicted in Figures 6.1 to 6.7.

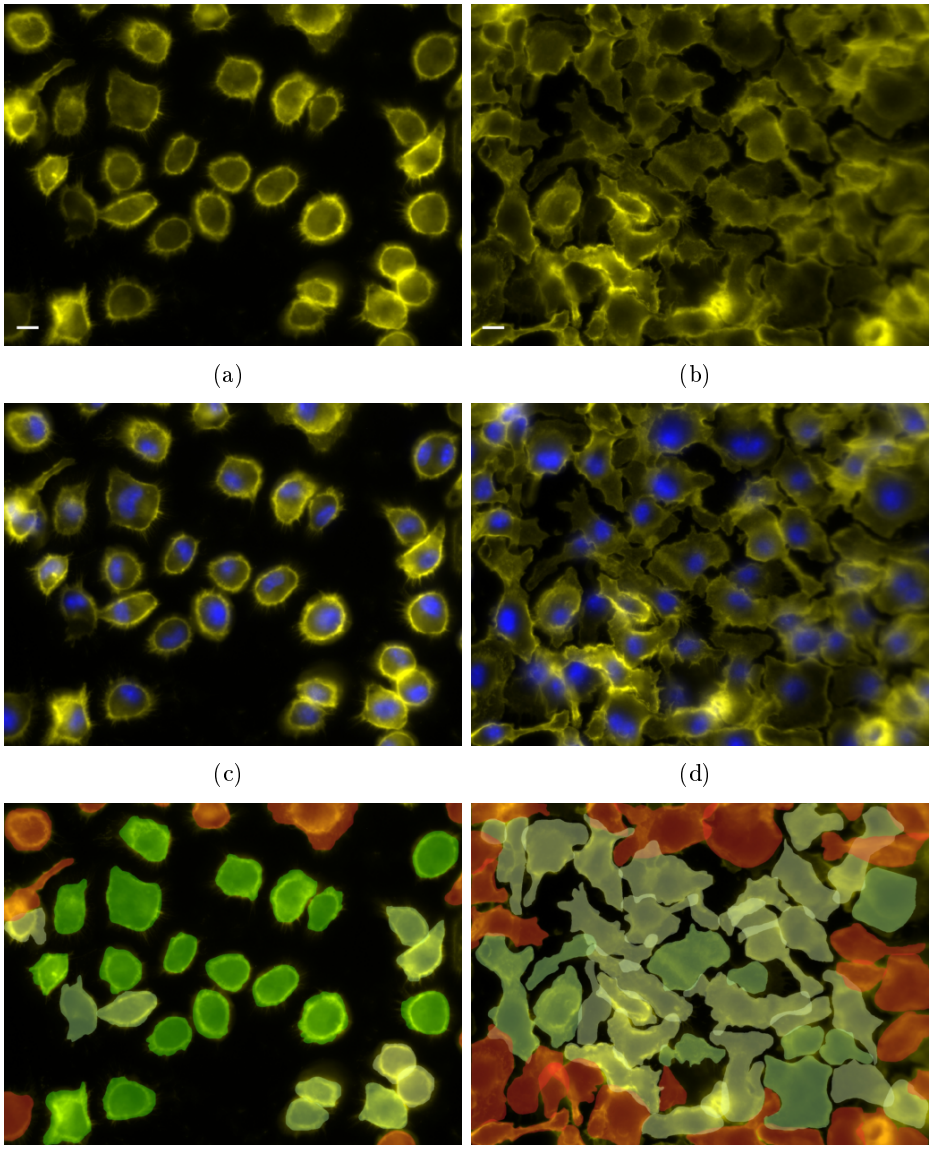
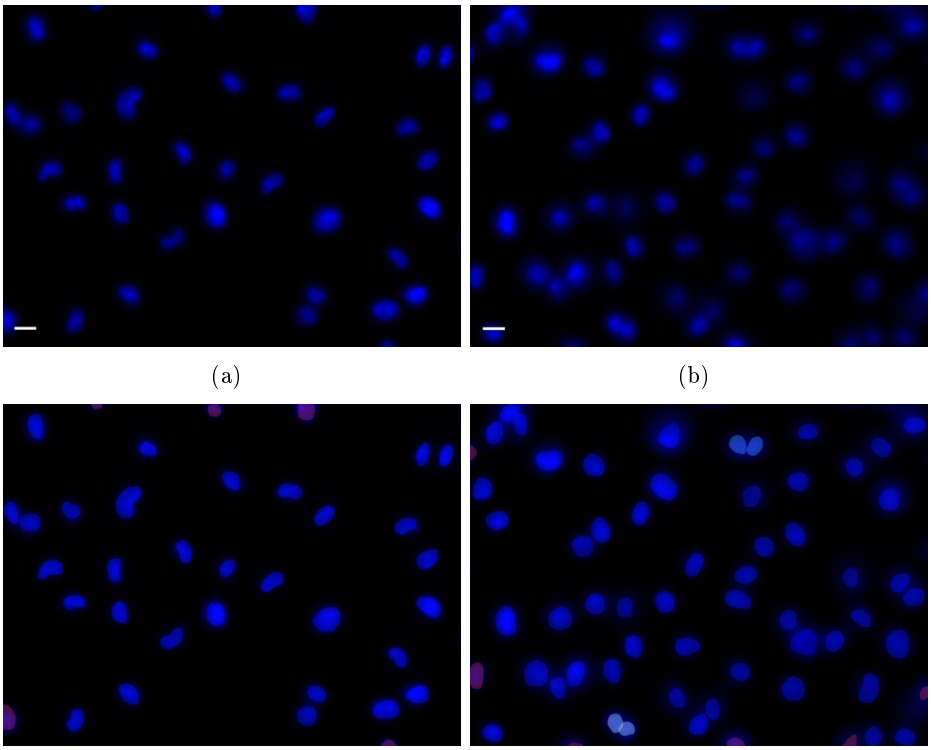


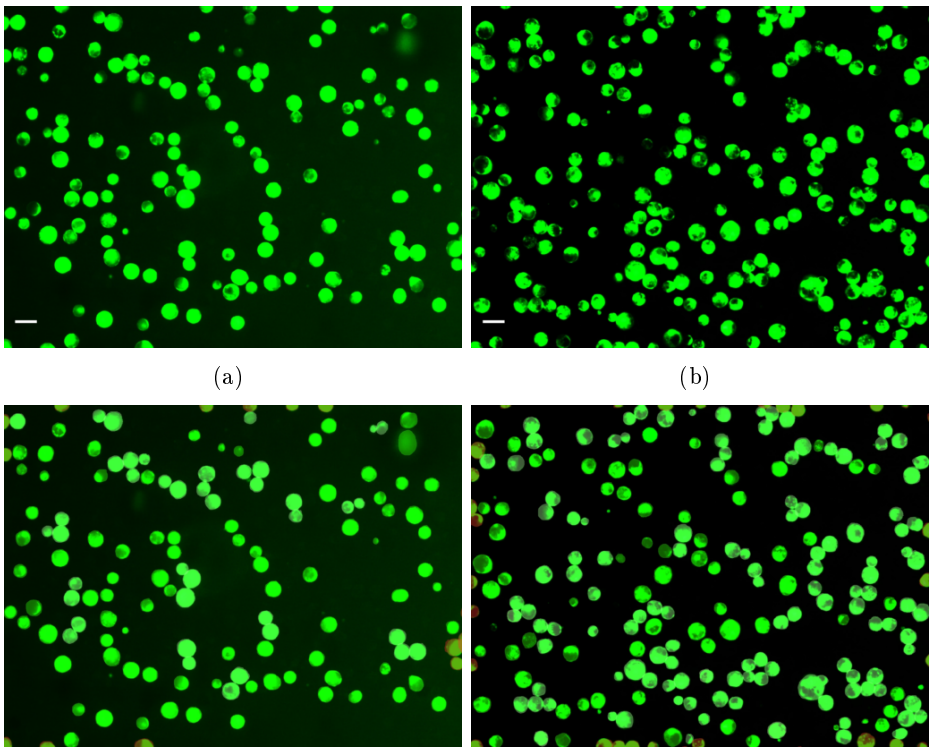
Figure 6.1.: **D1: CD11b/APC-stained macrophages:** Corresponding images of simplest (a, c, e) and most challenging micrograph (b, d, f) representing bone marrow-derived macrophages from $C57BL/6$ mice that were stimulated with LPS. For visualization of the macrophages, a $CD11b/APC$ staining was applied (a, b). Nuclei were visualized with DAPI staining (overlay images c, d). Images with a size of 1388×1040 pixels were captured using a Zeiss Axiovert microscope with a 20x objective (scale bar corresponds to 10 μm). Note that cells exceeding the image boundary and cells whose boundaries could not be recognized by the human evaluator were assigned to a rejection class (red cells in images e, f). For details on the different classes see Section 6.1.2.



(a) (b)

(c) (d)

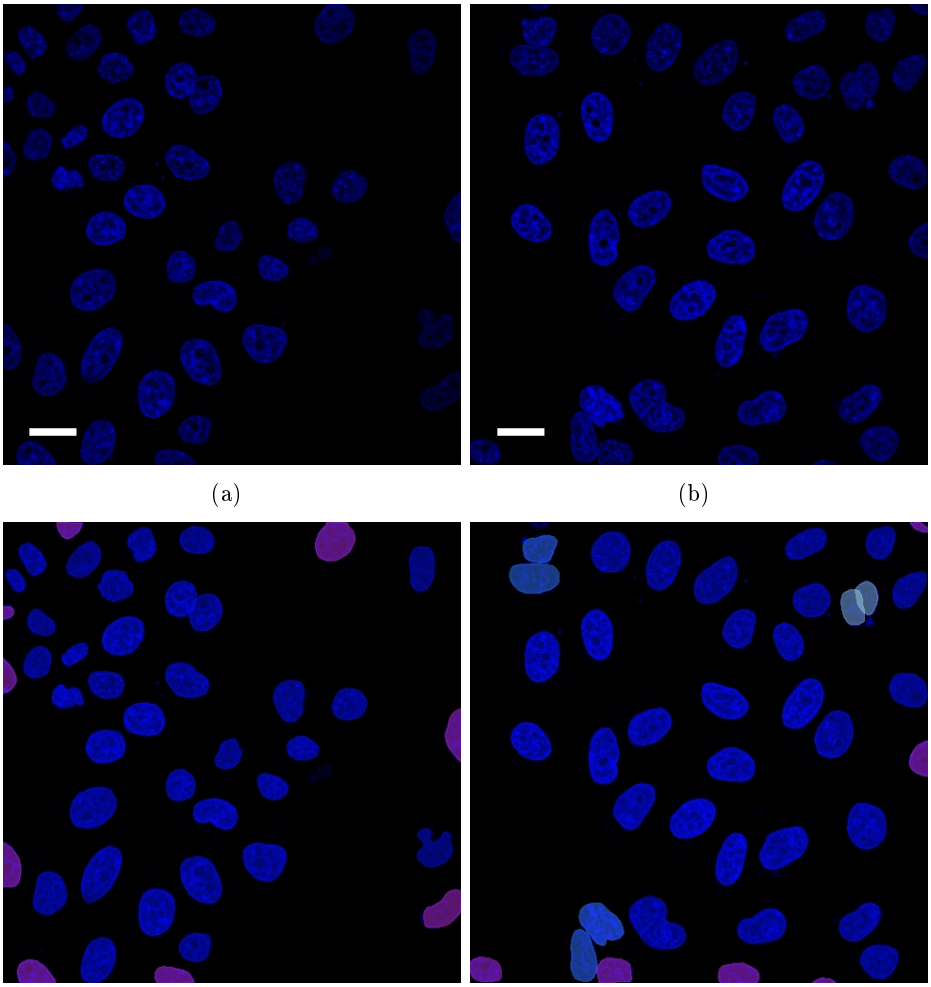
Figure 6.2.: **D2: DAPI-stained nuclei:** Corresponding images depicting simplest (a, c) and most challenging micrograph (b, d) consisting of cell nuclei of bone marrow-derived macrophages from $C57BL/6$ mice. For visualization of the nuclei, a DAPI staining was applied (images a, b). Images with a size of 1388×1040 pixels were captured using a Zeiss Axiovert microscope with a 20x objective (scale bar corresponds to 10 μm). Images (c, d) show overlays from hand-labeled nuclei with colors encoding feasibility classes.



(a) (b)

(c) (d)

Figure 6.3.: **D3: FDA-stained chloroplasts:** Corresponding images showing simplest (a, c) and most challenging micrograph (b, d) showing chlorophyll inside chloroplasts after Fluorescein diacetate (FDA) staining (a, b). Images with a resolution of 1392×1040 pixels were acquired with an Olympus BX60 microscope using a $10\times$ objective and a Leica DFC300 Fx camera. Images (c, d) show the hand-labeled cells with colors encoding feasibility classes. Scale bar corresponds to 10 μm.



(a) (b)

(c) (d)

Figure 6.4.: **D4: Confocal DAPI-stained nuclei:** Corresponding images showing simplest (a, c) and most challenging (b, d) DAPI-stained cell nuclei of HeLa cells (a, b). Images (c, d) show the hand-labeled cells with colors encoding feasibility classes. Scale bar corresponds to 20 μm.

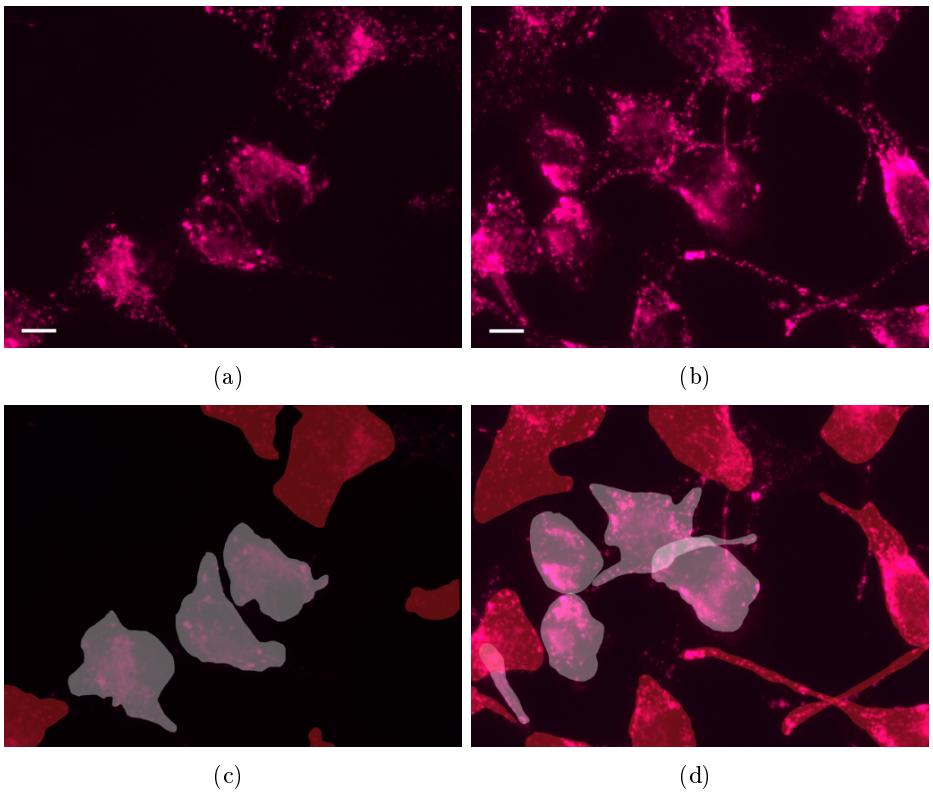


Figure 6.5.: **D5: Alexa 568-stained HeLa cells:** Corresponding images of simplest (a, c) and most challenging (b, d) micrographs showing the actin cytoskeleton of HeLa cells after Phalloidin Alexa 568 staining (a, b). A Zeiss Axiovert microscope with 63x oil immersion objective was used for image acquisition. The image size is 1388×1040 pixels. Images (c, d) show corresponding hand labelings, with colors encoding feasibility classes. Scale bar corresponds to 10 μm.

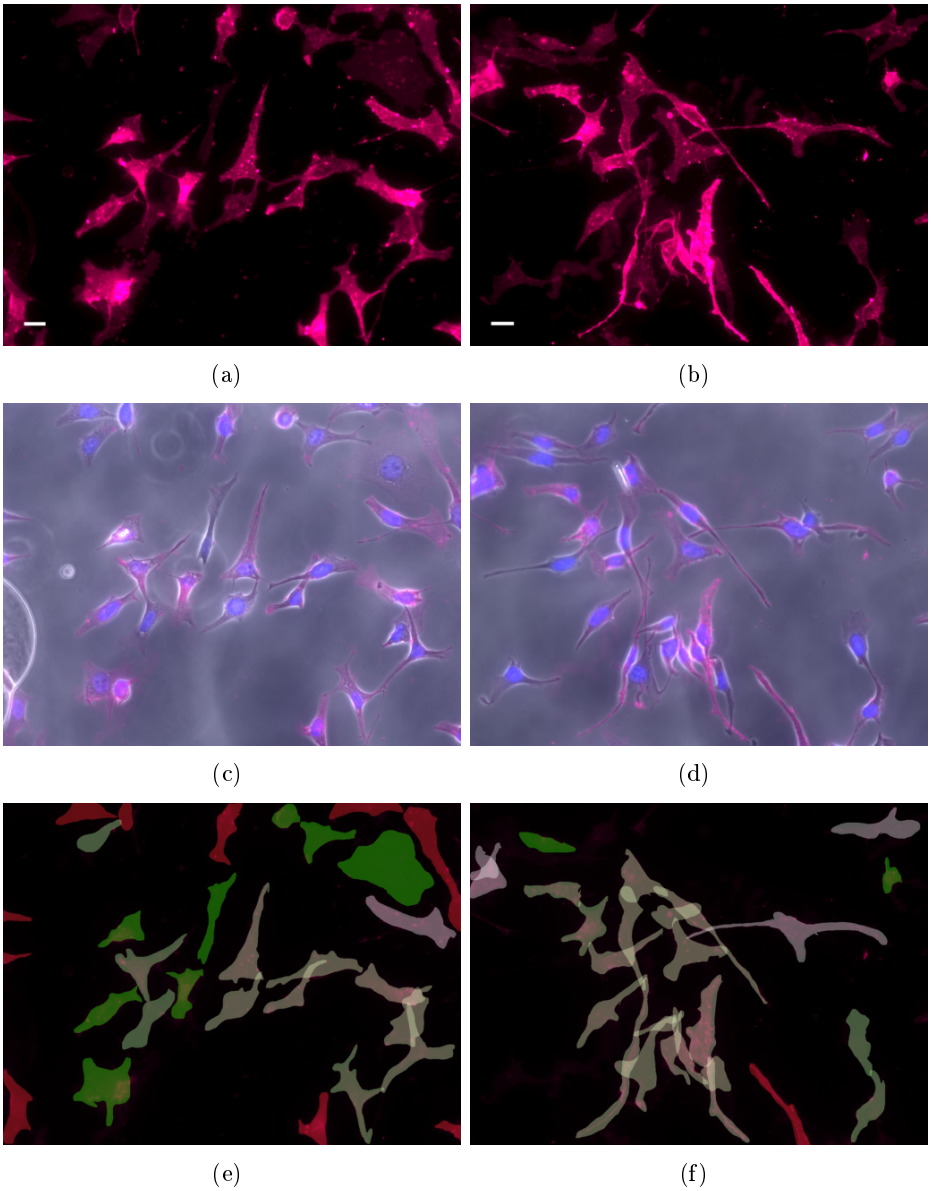


Figure 6.6.: **D6: DiD-stained HeLa cells:** The cytoplasmic membrane of HeLa cells was labeled with DID ™ (a, b). Corresponding images (a, c, e) depict the simplest and (b, d, f) the most challenging micrograph. For image acquisition, a 20x oil immersion objective was mounted onto a Zeiss Axiovert microscope. Images were captured with a size of 1388×1040 pixels. For this micrograph, additional DAPI and phase contrast images were captured, as can be seen in merged images (c, d). An overlay of hand labelings is shown in (e, f). Scale bar corresponds to 20 μm.

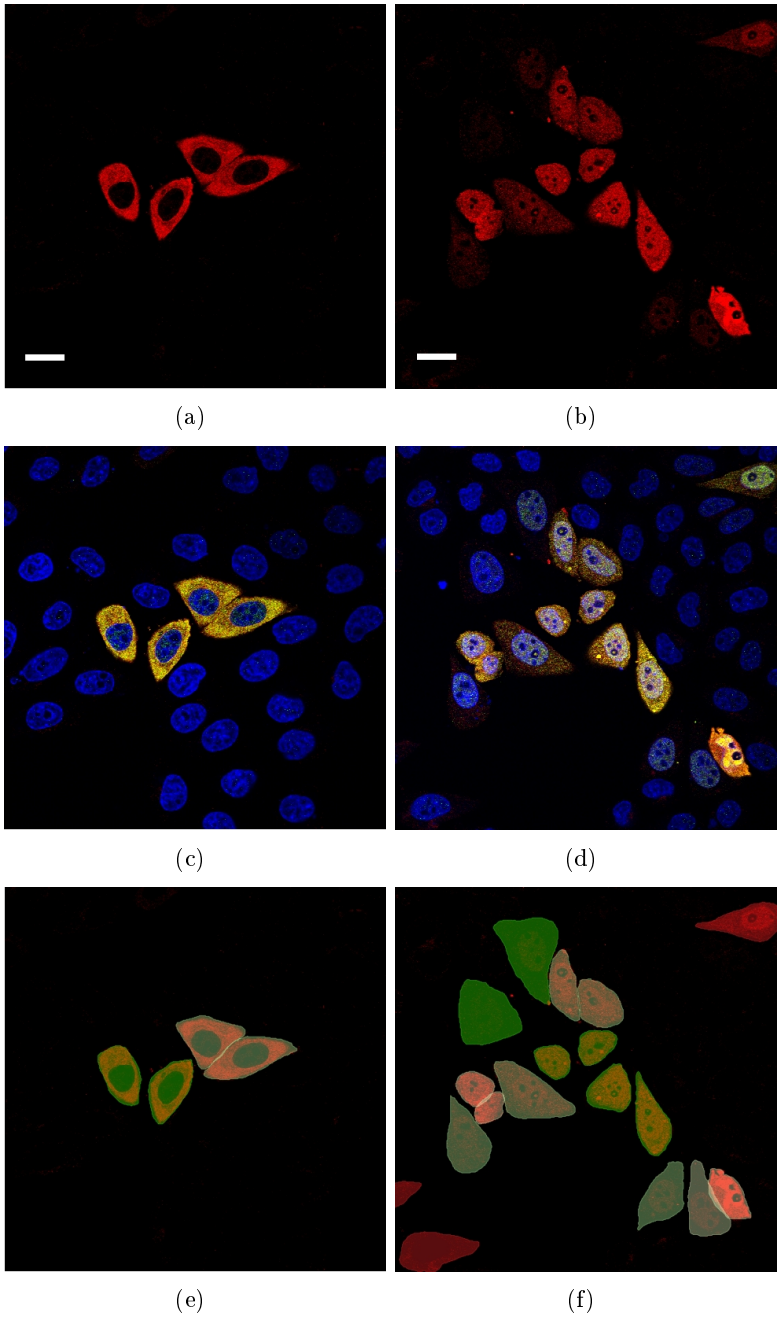


Figure 6.7.: **D7: pUL97-stained HeLa cells:** HeLa cells were stained with pUL97. Corresponding images (a, c, e) show the simplest and (b, d, f) the most challenging micrographs. For visualization of the cell nuclei, DAPI staining was applied. Images (c, d) show an overlay of both channels, including an additional green GFP channel. Images with a resolution of 512×512 pixels were captured using a confocal Leica TCS SP5 microscope. Images (e, f) depict the hand-labeled HeLa cells with colors encoding feasibility classes. Scale bar corresponds to 20 μm.

6.1.2. Generation of hand-labeled data

In order to enable comparison between hand-labeled and automatically segmented data, the solution of the segmentation problem must be defined. Therefore, a trained and experienced user provides a high-accuracy annotation of each dataset by delineating each object manually. To obtain high-quality hand labelings in a short time, a Wacom Cintiq 21UX^{TM} digital drawing board (Wacom Europe, Krefeld, Germany) was used for labeling of the data.

As each sample contains cells of different difficulty, feasibility classes are assigned to each cell or nucleus. This information is required, for instance, to investigate if usage of simple or challenging cells for calibration results in the best parameterization of the segmentation pipeline.

Objective assessment of the difficulty in segmenting an object is hardly possible, particularly if objects from different datasets are compared. Thus, objects are differentiated based on whether they are isolated, touching or overlapping neighbor cells.

For the segmentation pipeline, cells exceeding the image boundary pose a particular challenge. As only part of the cell is visible, it is often hard to correctly split touching or overlapping cells near the image boundary. Additionally, information on nuclei is missing for some of these cells. In order to exclude such cells from the evaluation and to be able to differentiate between cells of different complexity, one of the following feasibility classes is assigned to each object:

- Isolated object
- Touching object
- Overlapping or overlaying object
- Invalid (Object that exceeds image or boundaries and cannot be recognized by the human observer, e.g. due to blur or very strong overlap)

A detailed overview on the composition of data used in this work is provided in Table 6.1. This table shows that both datasets showing cell nuclei rarely contain overlapping and touching objects, whereas feasibility classes strongly vary for micrographs depicting cells. This supports the hypothesis that segmentation of cells usually poses a stronger challenge than the segmentation of the corresponding nuclei.

6.2. Experiment I: Comparison of objective functions

Results from automated calibration of the segmentation pipeline are strongly determined by the objective function used for comparison between segmen-

Application Domain	# images	# GT	I	T	O	Inv
D1: CD11b/APC-stained macrophages	20	588	116	169	**273**	320
D2: DAPI-stained nuclei	20	710	**688**	18	4	110
D3: FDA-stained chloroplasts	5	2287	**1150**	1075	62	437
D7: pUL97-stained HeLa cells	46	303	**214**	77	12	36
D4: Confocal DAPI-stained nuclei	46	1672	**1656**	14	2	444
D5: Alexa 568-stained HeLa cells	64	294	**177**	90	27	509
D6: DiD-stained HeLa cells	20	717	215	125	**377**	263

Table 6.1.: Composition of the datasets depicting the number of captured images (# images) and the number of manually annotated hand-labeled cells (# GT) for each application domain. Additionally, the number of isolated (I), touching (T), overlapping (O) and invalid cells (Inv) is provided.

tation result and hand-labeled data. For identification of an appropriate metric, the overlap based Jaccard similarity metric o_{j} (see Section 4.3.1, p. 32), the boundary based Hausdorff metric o_{H}, and the root mean squared Hausdorff distance $o_{\mathrm{H}}^{\mathrm{rms}}$ (see Section 4.3.2, p. 32) were compared. These measurements were selected as the Jaccard similarity represents a pure overlap based measurement, whereas the Hausdorff metric solely depends on the object boundaries. By determining similarity between segmentation result and hand labeling, a performance measurement is obtained for each segmented cell. Respectively, a vector of performance measurements $\mathbf{x}$ is obtained evaluating segmentation performance for a set of fluorescent micrographs. For evaluating segmentation quality for the complete dataset, the L^p norm of $\mathbf{x}$ is used:

$$||\mathbf{x}||_{\mathbf{p}} = \left(\sum_{\mathbf{i=1}}^{\mathbf{n}} |\mathbf{x_i}|^{\mathbf{p}} \right)^{\mathbf{1/p}},$$

where n denotes the number of elements in $\mathbf{x}$.

Experimental setup

In the following, different metrics using boundary or overlap based measurements are compared considering the L^1, $\mathrm{L}^2$and the L^∞ norms. For determining a good objective function, all parameters are automatically calibrated based on the hand labelings. For this comparison, a fixed image segmentation pipeline consisting of DoG based preprocessing, k-means clustering based figure-ground separation and fast marching cell splitting (DOG-KMSFM) is applied, as this pipeline enables time-efficient calibration of the segmentation pipeline. Parameters of all methods are adapted towards the hand-labeled data by using genetic algorithms. The reference images were

chosen in such way as to include cells with varying shapes and intensities. Note that the reference images are also used for testing, as ability to adapt to the hand labelings was investigated for varying objective functions. In order to minimize random effects of the optimization procedure, this experiment was carried out three times.

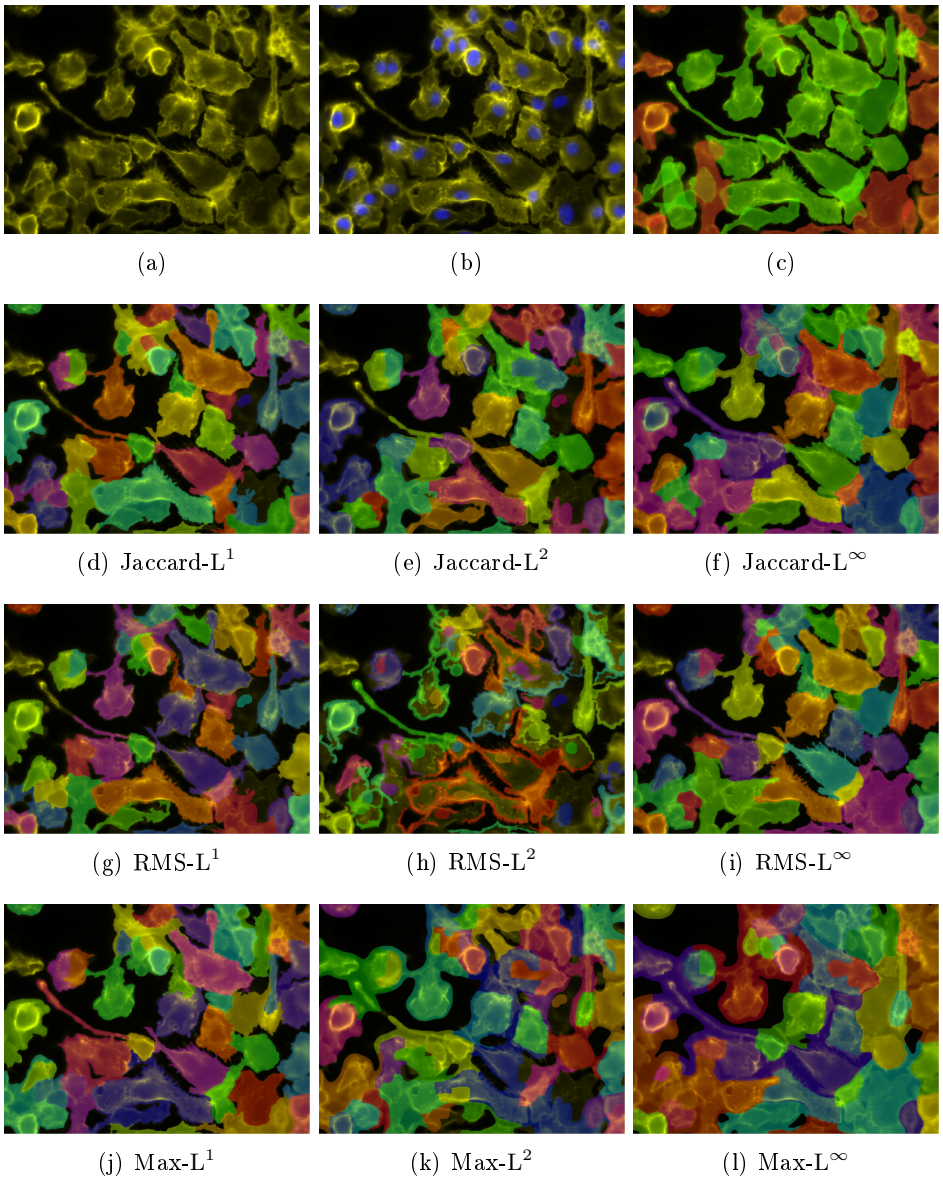
Results and discussion

Reference image, as well as the results from automated calibration of the segmentation pipeline, are shown in Figures 6.8 to 6.10. The experimental data clearly shows that the overlap based Jaccard similarity provides the most stable results. Using the boundary based Hausdorff metric can lead to exclusive segmentation of the cell boundary, which can result in a segmentation excluding the inside of the cells (Figures 6.8(h), 6.9(k) and 6.10(h)). Furthermore, quality of results varies much more strongly for the Hausdorff based metric than for the Jaccard similarity based one. This indicates that the Jaccard similarity is more suitable for automated parameter optimization.

Comparison of the L^1, L^2 and the L^∞ norms indicates that using the L^1 norm results in the best segmentation. This can be explained by the larger impact of well-segmented cells when using the L^2 or the L^∞ norm. Hence, a parameter setting may be accepted that allows a very good segmentation of some cells while disregarding other, erroneously segmented, cells. Applying the L^∞ norm, only the best segmented cell is considered.

Conclusion

This experiment clearly demonstrates that using the Jaccard similarity and the L^1 norm outperforms Hausdorff based measurements for calibration of the segmentation pipeline. In order to improve robustness with outliers, the median Jaccard similarity can be used instead of the L^1 norm.



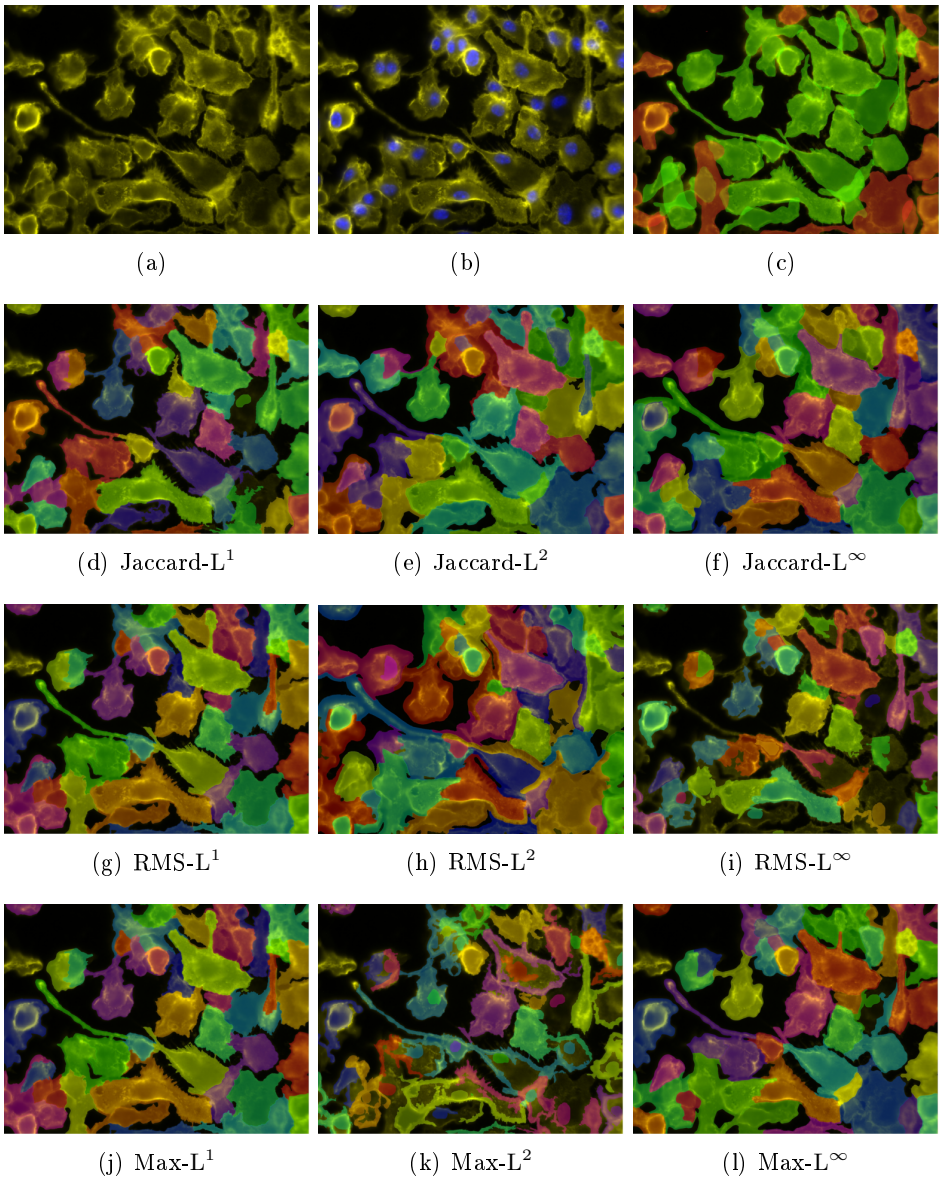
(a) (b) (c)

(d) Jaccard-L^1 (e) Jaccard-L^2 (f) Jaccard-L^∞

(g) RMS-L^1 (h) RMS-L^2 (i) RMS-L^∞

(j) Max-L^1 (k) Max-L^2 (l) Max-L^∞

Figure 6.8.: Calibration of the segmentation pipeline for corresponding images (a) and (b), based on hand-labeled data (c), using the Jaccard similarity, root mean squared of the Hausdorff distance (RMS), the maximum Hausdorff distance (Max) and L^1, L^2, L^∞norms (d-l). Red cells in (c) are excluded due to exceeding the image boundary, which might lead to errors. In general, the Jaccard similarity metric and the L^1 norm provide the best results, whereas use of Hausdorff RMS based measurements can lead to exclusive segmentation of the cell boundary (h).



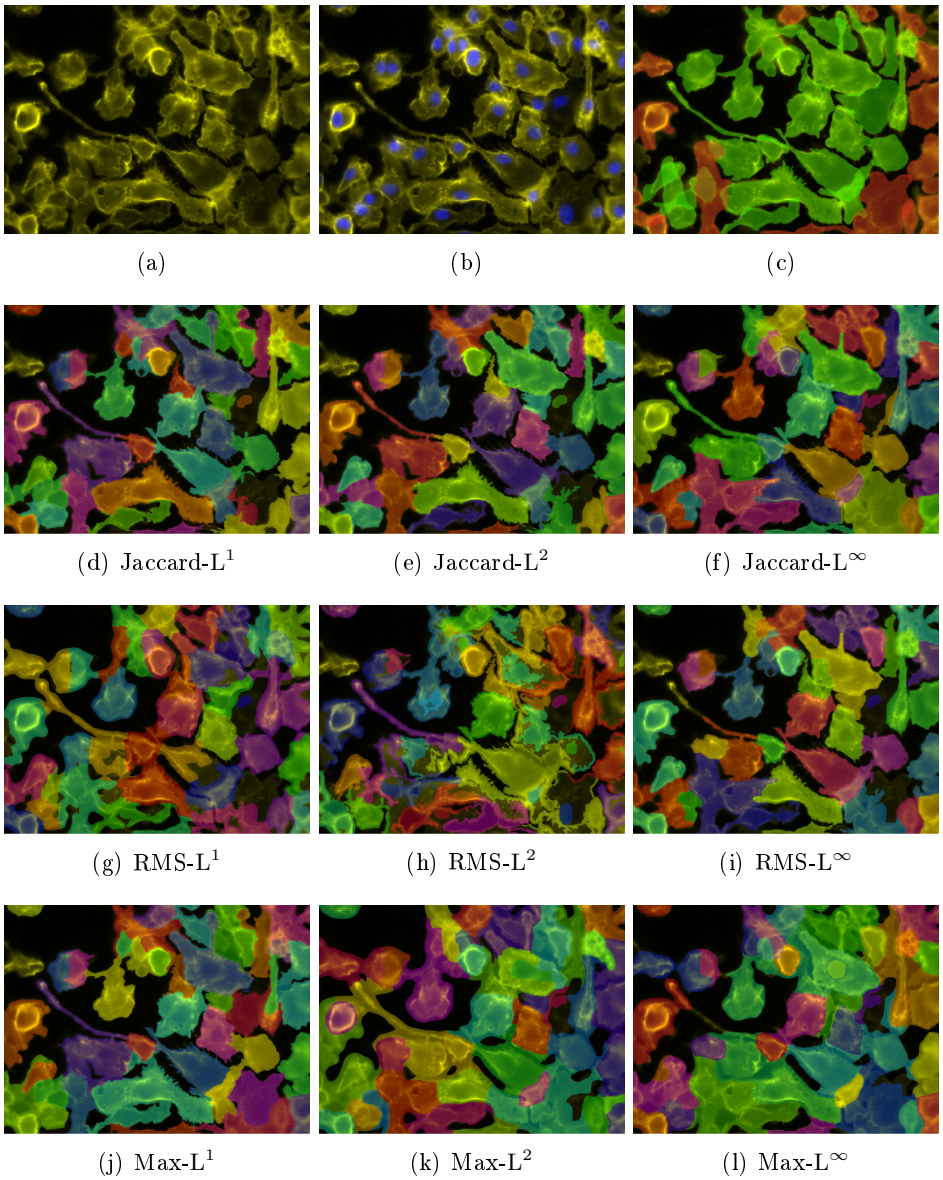
(a) (b) (c)

(d) Jaccard-L^1 (e) Jaccard-L^2 (f) Jaccard-L^∞

(g) RMS-L^1 (h) RMS-L^2 (i) RMS-L^∞

(j) Max-L^1 (k) Max-L^2 (l) Max-L^∞

Figure 6.9.: Repeated calibration of the segmentation pipeline for corresponding images (a) and (b), based on hand-labeled data (c) for different objective functions and norms (d-l). Carrying out this experiment for a second time for reduction of random effects confirmed that the Jaccard similarity metric and the L^1 norm result in a good calibration and that usage of Hausdorff based measurements can lead to exclusive segmentation of the cell boundary (k).



(a) (b) (c)

(d) Jaccard-L^1 (e) Jaccard-L^2 (f) Jaccard-L^∞

(g) RMS-L^1 (h) RMS-L^2 (i) RMS-L^∞

(j) Max-L^1 (k) Max-L^2 (l) Max-L^∞

Figure 6.10.: Repeated calibration of the segmentation pipeline for corresponding images (a) and (b), based on hand-labeled data (c) for different objective functions and norms (d-l). Carrying out this experiment for a second time for reduction of random effects confirmed that the Jaccard similarity metric and the L^1 norm result in a good calibration.

6.3. Experiment II: Comparison of segmentation performance for different application domains

For testing each method of the segmentation pipeline, applicability for segmentation of fluorescence image data from different application domains was investigated. State of the art algorithms are mostly developed and evaluated for a single application domain. Comparing different state of the art as well as new methods based on a broad range of fluorescence image data constitutes one of the contributions of this work.

Based on this experiment, it could be concluded which kind of fluorescence microscopy data can be handled by each segmentation method. This also enables the removal of methods that are outperformed by alternative ones, for every dataset, to reduce the amount of computational time required for calibration of the segmentation pipeline. For comparison of different methods and combinations thereof, all free parameters were optimized towards the hand-labeled data using genetic algorithms. Additionally, a three-fold cross validation was included in the calibration process to avoid overfitting and to allow fair comparison of different combinations of methods. Convergence of the genetic algorithm was assumed if no new performance maximum was identified for 200 iterations. Using the proposed setup, calibration for most datasets was performed in about 24 hours, using a given combination of image processing methods. An exception to this was the graph cut based splitting scheme, which requires up to one week for calibration due to its longer runtime.

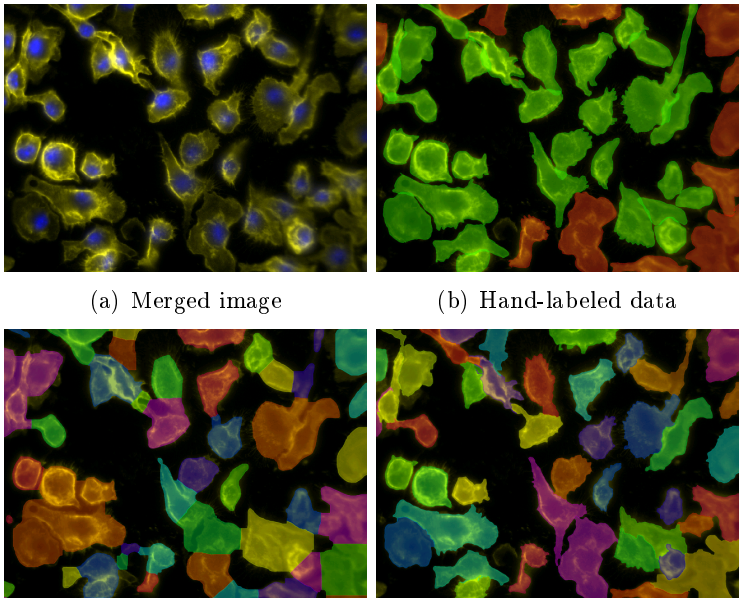
6.3.1. Evaluation of segmentation performance for different combinations of methods

For comparison of different preprocessing, figure-ground separation or object splitting methods for a specific application domain, the segmentation pipeline is calibrated for each combination of methods. Comparing segmentation results to the hand-labeled data, a Jaccard performance measurement is obtained for each cell in the dataset. Based on this, segmentation performance for each dataset can be represented by the median Jaccard similarity. In order to evaluate applicability of an image processing method for segmentation of a specific dataset, evaluation of performance for all method combinations making use of this specific method is necessary. Hence, the average of median Jaccard similarity values is used for comparison of these methods. This is best illustrated by an example. Assuming that average of median Jaccard similarity shall be analyzed for MIN preprocessing, then the average value of median Jaccard similarities is determined for segmentation pipelines MIN-M_2-M_3, where M_2-M_3 denotes all possible combinations of figure-ground separation and object splitting methods.

In order to investigate if performance of a method is significantly better

than performance of an alternative method, all plausible combinations of methods are tested for significance. Assuming for example that differences in segmentation performance are analyzed for DOG preprocessing and MIN preprocessing, then significance is tested for all pairs of method combinations MIN-M_2-M_3 and DOG-M_2-M_3. By counting the number of method combinations that show statistically significant differences, a conclusion on statistical significance can be drawn.

6.3.2. Application domain D1: CD11b/APC-stained macrophages



(a) Merged image (b) Hand-labeled data

(c) Segmentation without using information on nuclei (d) Segmentation using information on nuclei

Figure 6.11.: Macrophages stained with CD11b/APC and DAPI (merged image (a)) and corresponding hand labelings (b). Note that invalid hand-labeled cells in (b) exceed the image and are shown in red. These cells are assigned to a rejection class and excluded from calibration and evaluation. Results from the calibrated segmentation pipeline without and with use of information on the corresponding nuclei are shown in (c) and (d).

The CD11b/APC-stained macrophage dataset (Figure 6.1, p. 60) is used for comparison of segmentation performance because of its overlapping cells that are partially out of focus and exhibit only low contrast to the image background. Furthermore, the number of cells strongly varies inside the dataset which poses a challenge for most figure-ground separation methods. For comparison of preprocessing, figure-ground separation or object splitting methods, segmentation performance was evaluated for all 32 combinations of methods. The resulting accuracy, median Jaccard similarity and combined performance values are summarized in Figure 6.12. Additionally, representative result images based on the calibrated segmentation pipeline are depicted in Figure 6.11.

Pre	Fg	Split	Accuracy	Jaccard	Combined
		WS	0.51	0.49	0.254
	KM	SWS	0.94	0.73	0.689
		SFM	0.94	0.73	0.684
		SGC	0.94	0.73	0.686
		WS	0.51	0.43	0.221
	LOC	SWS	0.93	0.71	0.653
		SFM	0.94	0.69	0.651
		SGC	0.93	0.67	0.626
		WS	0.52	0.44	0.228
MIN	GLOC	SWS	0.84	0.56	0.471
		SFM	0.89	0.59	0.529
		SGC	0.78	0.59	0.461
		WS	0.55	0.62	0.339
	GC	SWS	0.93	**0.80**	0.752
		SFM	**0.95**	0.79	0.752
		SGC	0.94	**0.80**	**0.757**
		WS	0.53	0.48	0.253
	KM	SWS	0.94	0.73	0.688
		SFM	0.94	0.73	0.685
		SGC	0.94	0.73	0.692
		WS	0.43	0.45	0.195
DOG	LOC	SWS	0.91	0.65	0.596
		SFM	0.94	0.70	0.659
		SGC	0.93	0.68	0.631
		WS	0.40	0.35	0.141
	GLOC	SWS	0.86	0.58	0.499
		SFM	0.94	0.70	0.659
		SGC	0.93	0.68	0.631
		WS	0.54	0.63	0.334
	GC	SWS	**0.95**	**0.80**	0.753
		SFM	**0.95**	0.79	0.753
		SGC	0.94	**0.80**	0.749
Interobserver			0.94	0.83	0.784
Intraobserver			0.89	0.83	0.742

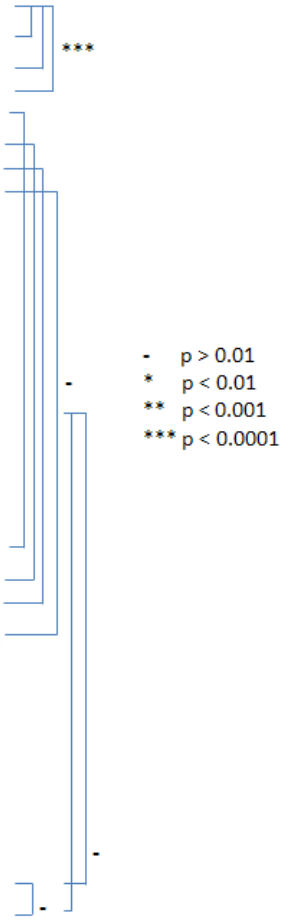


Figure 6.12.: Accuracy, median Jaccard coefficient and combined performance for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods applied for the segmentation of $CD11b/APC$ stained macrophages. Additionally, some significant or non-significant differences are highlighted, which are identified by a Friedman test followed by pairwise Wilcox post-hoc test.

Comparison between manual and automated segmentation

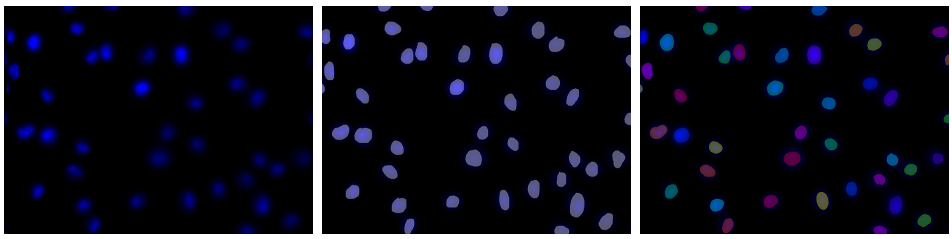
Using the described experimental setup, a combination of minimum filter based preprocessing (MIN), graph cut based figure-ground separation (GC) and seeded graph cut (SGC) (MIN-GC-SGC) showed the highest combined performance measurement, $p_c = 0.757$, a Jaccard similarity $p_j = 0.80$, and an accuracy of $p_a = 0.94$ (see Figure 6.12). Comparing this to the inter- and intraobserver combined Jaccard similarity of $p_c = 0.784$ and $p_c = 0.742$ indicates that, for this dataset, the proposed segmentation pipeline can compete with the human if the combined segmentation performance is taken into account. This could be confirmed by a Friedman test, followed by a pairwise Wilcox post-hoc test, showing that resulting Jaccard similarities for interobserver, intraobserver and the optimal segmentation pipeline are not statistically significant. Comparison of segmentation accuracy using MIN-GC-SGC ($p_a = 0.94$) to accuracy of inter- and intraobserver ($p_a = 0.94$ and $p_a = 0.89$) demonstrates that accurate splitting of cells is an ambiguous task for human experts, as the intraobserver measurement is outperformed by the automated approach if information on position and extension of the cell nuclei is incorporated.

Comparison of image processing methods

Comparison of the applied image processing methods shows only a slight difference between MIN and DOG preprocessing applied for segmentation of $CD11b/APC$-stained macrophages. Comparison of the average combined performance for all method combinations using DOG preprocessing ($p_c^{av} = 0.559$) to the average combined performance of methods using MIN preprocessing ($p_c^{av} = 0.547$) indicates that the MIN preprocessing method and the DOG preprocessing methods result in a very similar segmentation. A pairwise Wilcox test confirms that segmentations are significantly different based on the Jaccard similarity metric for 3 of 16 method combinations (MIN-M_2-M_3, DOG-M_2-M_3).
Comparing methods for figure-ground separation, KM results in an average combined segmentation performance of $p_c^{av} = 0.579$. Usage of LOC, GLOC and GC results in the values $p_c^{av} = 0.529$, $p_c^{av} = 0.452$ and $p_c^{av} = 0.648$. The deduced hypothesis that the GC figure-ground separation method significantly outperforms alternative figure-ground separation ones was confirmed by the pairwise Wilcox test showing significant differences between Jaccard similarities for all method combinations, comparing graph cut based figure-ground separation to alternative figure-ground separation methods.
For comparison of object splitting methods, the average of median Jaccard similarities was determined for WS ($p_c^{av} = 0.246$), SWS ($p_c^{av} = 0.638$), SFM ($p_c^{av} = 0.672$) and SGC ($p_c^{av} = 0.654$). Based on statistical analysis using a Friedman test, followed by a pairwise Wilcox test, the conclusion may

be drawn that incorporating information on the cell nuclei significantly increases the median Jaccard similarity for the investigated data ($p < 0.0001$). Comparison of different seeded cell splitting methods only shows tiny differences, whereof about 50% of the observed Jaccard similarities are significantly different. This indicates that identification of the optimal seeded scheme for splitting of the cells can be neglected when accepting a small decrease in segmentation performance for the analyzed data.

6.3.3. Application domain D2: DAPI-stained nuclei



(a) Input image (b) Hand-labeled data (c) Segmentation without using information on nuclei

Figure 6.13.: Results from the calibrated segmentation pipeline applied for segmentation of DAPI stained nuclei of macrophages (a-c).

An overview of performance for different combinations of image processing methods applied for segmentation of fluorescently stained cell nuclei is provided in Figure 6.14. Calibration of the segmentatino pipeline was performed based on hand-labeled data by a GA including three-fold cross validation. For a representative example image and the corresponding hand labelings see Figure 6.13.

Pre	Fg	Split	Accuracy	Jaccard	Combined
	KM	WS	0.94	0.81	0.756
MIN	LOC	WS	0.94	0.79	0.749
	GLOC	WS	0.93	0.80	0.743
	GC	WS	0.92	0.80	0.738
	KM	WS	**0.96**	0.81	0.782
DOG	LOC	WS	**0.96**	**0.83**	**0.794**
	GLOC	WS	0.95	0.82	0.778
	GC	WS	**0.96**	0.82	0.786
Interobserver			0.97	0.75	0.721
Intraobserver			0.98	0.80	0.783

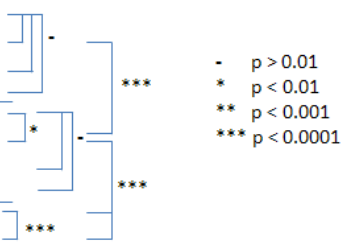


Figure 6.14.: Accuracy, median Jaccard coefficient and combined performance for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods. For the DAPI-stained nuclei data, performance of all methods is comparable and can compete with human performance in terms of accuracy and Jaccard similarity.

Comparison between manual and optimal automated segmentation

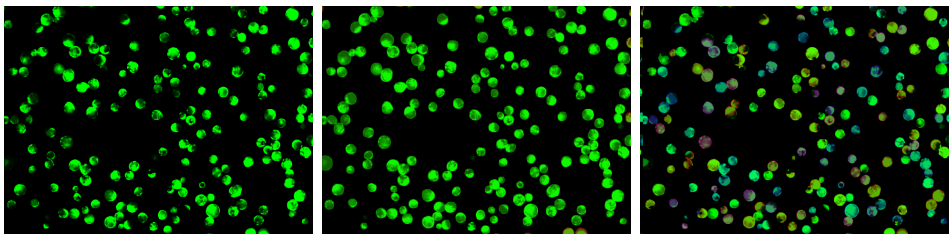
Comparison of combinations of segmentation methods for DAPI-stained nuclei (see Figure 6.14) shows that a segmentation pipeline DOG-LOC-WS

enables segmentation with highest combined performance ($p_c = 0.794$), accuracy ($p_a = 0.96$) and median Jaccard similarity ($p_j = 0.83$). For this dataset, a combined performance of $p_c = 0.721$ was obtained for interobserver analysis, and $p_c = 0.783$ for intraobserver analysis. Focusing on combined performance, the human observer is slightly outperformed by automated image analysis. Detailed analysis showed that the human part achieves higher accuracy ($p_a = 0.97$ and $p_a = 0.98$ vs. $p_a = 0.96$), but median Jaccard similarity was slightly better for automated segmentation. The obtained inter- and intraobserver measurements nicely demonstrate that, for this simple dataset, automated image analysis can outperform the human element in terms of objectivity and reproducibility.

Comparison of different methods

Comparison of different combinations of methods indicates that DOG preprocessing outperforms MIN preprocessing for the DAPI-stained nuclei. This is supported by an average combined performance $p_c^{av} = 0.785$ for the DOG and $p_c^{av} = 0.747$ for the MIN preprocessing method. Statistical analysis of the obtained data with a Friedman and a pairwise Wilcox post-hoc test shows that observed Jaccard similarity measurements for pairs of observations (MIN-LOC-WS, DOG-LOC-WS) and (MIN-GC-WS, DOG-LOC-WS) are significantly different ($p < 0.01$), whereas measurements (MIN-KM-WS, DOG-KM-WS) and (MIN-GLOC-WS, DOG-GLOC-WS) are not significantly different ($p > 0.01$).
Comparing different methods for figure-ground separation demonstrates that all of them can robustly handle the DAPI-stained nuclei data. This is confirmed by the resulting average combined Jaccard similarity measurements $p_c^{av} = 0.769$ for KM, $p_c^{av} = 0.771$ for LOC, $p_c^{av} = 0.761$ for GLOC and $p_c^{av} = 0.762$ for GC figure-ground separation. This hypothesis is further strengthened by the pairwise Wilcox test showing significantly different measurements for only three out of 12 pairs of measurements ($p < 0.01$).

6.3.4. Application domain D3: FDA-stained chloroplasts



(a) Input image (b) Hand-labeled data (c) Segmentation without using information on nuclei

Figure 6.15.: Corresponding images (a, b, c) depict results from use of the calibrated segmentation pipeline for FDA-stained chloroplasts.

An overview of performance for different combinations of image processing methods for segmentation of fluorescently stained chloroplasts is provided in Figure 6.16. Calibration of the segmentation pipeline was automatically performed based on hand-labeled data using a GA including three-fold cross validation. For a representative exemplary image, see Figure 6.15.

Pre	Fg	Split	Accuracy	Jaccard	Combined
	KM	WS	0.94	0.89	0.813
MIN	LOC	WS	0.91	0.89	0.787
	GLOC	WS	0.92	0.89	0.772
	GC	WS	0.93	**0.90**	0.779
	KM	WS	0.93	0.89	**0.820**
DOG	LOC	WS	0.93	0.88	0.810
	GLOC	WS	0.94	0.89	0.789
	GC	WS	0.93	0.84	0.813
Interobserver			0.98	0.87	0.850
Intraobserver			0.99	0.90	0.888

* p > 0.01
* p < 0.01
** p < 0.001
*** p < 0.0001

Figure 6.16.: Median Jaccard coefficients and resulting combined performance measurements for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods. For the FDA-stained chloroplast data, information on cell nuclei is not available.

Comparison between manual and optimal automated segmentation

Configuring the segmentation pipeline for segmentation of FDA-stained chloroplasts (see Figure 6.16), the pipeline DOG-KM-WS resulted in the highest performance ($p_c = 0.820$, $p_a = 0.93$, $p_j = 0.89$). Analysis of interobserver variability lead to $p_c = 0.850$, $p_a = 0.98$, $p_j = 0.87$, and intraobserver

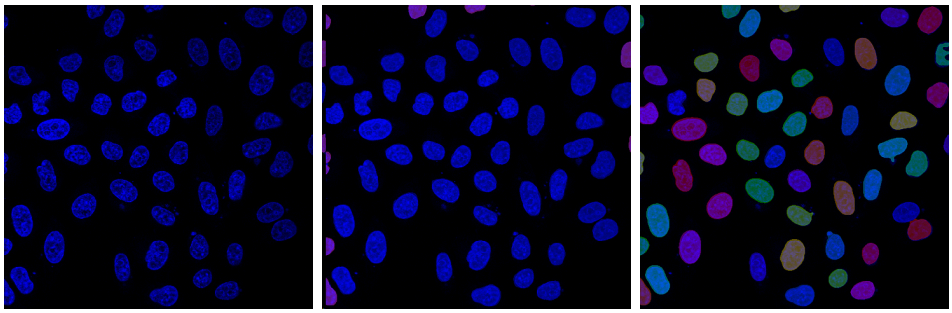
variability to $p_c = 0.888$, $p_a = 0.99$, $p_j = 0.90$. These measurements demonstrate that segmentation of the analyzed chloroplasts is an unambiguous and well-reproducible task for the human observer ($p_a = 0.99$, $p_a = 0.98$). For automated image analysis, robust splitting poses a challenge, as some chloroplasts cannot be completely separated from the image background in the figure-ground separation step. These incomplete cells often mislead shape-based splitting methods. Comparing Jaccard similarity for the intraobserver measurements based on a pairwise Wilcox test, automated methods are significantly outperformed with $p < 0.0001$ by man, though for the interobserver analysis differences are not statistically significant ($p > 0.01$). This data nicely demonstrates that quality of figure-ground separation using automated methods is often comparable to manual annotation, but splitting capabilities of the automated approaches cannot compete with the human element for this data.

Comparison of different methods

Detailed analysis of different combinations of methods showed a small performance advantage using DOG preprocessing (average combined performance $p_c^{av} = 0.808$, compared to $p_c^{av} = 0.790$ for MIN preprocessing). A pairwise Wilcox test confirmed that observed differences are small, as only differences for pairs of observations (MIN-KM-WS, DOG-KM-WS) and (MIN-GC-WS, DOG-GC-WS) are statistically significant.

Comparison of KM, LOC, GLOC and GC for figure-ground separation yielded the average combined performance values $p_c^{av} = 0.817$, $p_c^{av} = 0.798$, $p_c^{av} = 0.785$ and $p_c^{av} = 0.796$. This demonstrates that k-means clustering based threshold selection allows the best segmentation for the FDA-stained chloroplast data. This hypothesis is supported by a pairwise Wilcox test showing that results are significantly different for eight out of 12 combinations of methods based on KM, compared to combinations of methods not using KM.

6.3.5. Application domain D4: Confocal DAPI-stained nuclei



(a) Input image (b) Hand-labeled data (c) Segmentation without using information on nuclei

Figure 6.17.: Representative micrographs depicting DAPI-stained nuclei captured with a confocal microscope (a), corresponding hand labelings (b) and result of the calibrated segmentation pipeline (c).

Comparison of segmentation performance for different combinations of image processing methods applied for segmentation of cell nuclei captured with a confocal microscope is provided in Figure 6.18. Configuration of the segmentation pipeline was performed based on hand-labeled data by applying a GA including three-fold cross validation. Representative segmentation results are depicted in Figure 6.17.

Comparison between manual and automated segmentation

Calibration of the segmentation pipeline for confocal DAPI-stained nuclei (see Figure 6.18) resulted in a pipeline MIN-KM-WS with $p_c = 0.848$, $p_a = 0.95$ and $p_j = 0.89$. Based on interobserver measurements $p_c = 0.881$, $p_a = 0.97$, $p_j = 0.91$ and intraobserver measurements $p_c = 0.914$, $p_a = 0.99$, $p_j = 0.92$, it could be concluded that the proposed segmentation pipeline, for this dataset, can identify cell boundaries nearly as well as the human observer. However, for splitting of touching cells some additional errors are implemented for ambiguous cases. Statistical analysis using a pairwise Wilcox test showed that difference between observed Jaccard similarities are highly statistically significant, with $p < 0.0001$ for all pairs of interobserver, intraobserver and automated segmentation, but also for intra- and interobserver measurements.

Pre	Fg	Split	Accuracy	Jaccard	Combined
	KM	WS	**0.95**	0.89	**0.848**
MIN	LOC	WS	**0.95**	0.89	**0.848**
	GLOC	WS	0.94	0.89	0.833
	GC	WS	0.94	**0.90**	0.847
	KM	WS	0.93	0.89	0.824
DOG	LOC	WS	**0.95**	0.88	0.837
	GLOC	WS	0.94	0.89	0.835
	GC	WS	0.94	**0.90**	0.845
Interobserver			0.97	0.91	0.881
Intraobserver			0.99	0.92	0.914

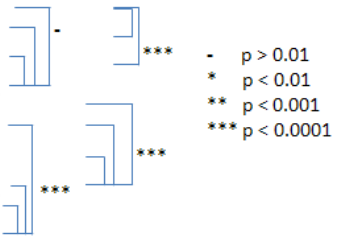
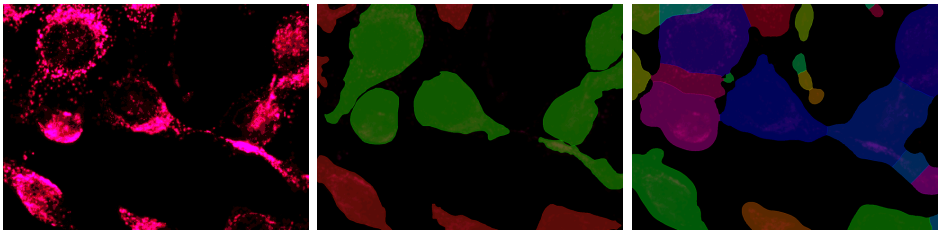


Figure 6.18.: Median Jaccard coefficient and accuracy for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods. For the DAPI-stained confocal cell nuclei, human handling is outperformed by machine in terms of Jaccard similarity. However, human is superior in splitting of touching cells, which is concluded from an observed higher accuracy. Analyzing combined performance, machine is slightly outperformed by man.

Comparison of different methods

In this section, different methods from the segmentation pipeline are compared to each other. For confocal DAPI-stained nuclei, MIN preprocessing outperformed that with DOG as an average Jaccard median of $p_c^{av} = 0.844$ was obtained for segmentation pipelines using MIN and $p_c^{av} = 0.835$ for pipelines using DOG preprocessing. For figure-ground separation, the GC method slightly outperformed alternative methods. This was confirmed by analyzing average combined Jaccard similarity for KM ($p_c^{av} = 0.836$), LOC ($p_c^{av} = 0.842$), GLOC ($p_c^{av} = 0.834$) and GC ($p_c^{av} = 0.846$). A Friedman test followed by a Wilcox post-hoc test comparing Jaccard coefficients for pairs of observations (M_1-GC-M_3, M_1-M_2-M_3) showed that all these combinations generated significantly different results. Hence, it could be concluded that GC figure-ground separation outperforms competing figure-ground separation methods for the confocal nuclei data.

6.3.6. Application domain D5: Alexa 568-stained HeLa cells



(a) Input image (contrast enhanced) (b) Hand-labeled data (c) Segmentation without using information on nuclei

Figure 6.19.: Corresponding images (a, b, c) depicting merged input image (a), hand labelings (b) and results for the calibrated segmentation pipeline (c) for segmentation of Alexa568-stained HeLa cells.

A detailed overview of different combinations of methods applied for segmentation of Alexa568-stained HeLa cells is provided in Figure 6.20. Optimal configuration of the segmentation pipeline was automatically determined based on hand-labeled data by a GA including three-fold cross validation. Representative segmentation results are provided in Figure 6.19.

Pre	Fg	Split	Accuracy	Jaccard	Combined
	KM	WS	0.31	**0.61**	**0.187**
MIN	LOC	WS	0.28	0.55	0.155
	GLOC	WS	0.30	0.34	0.100
	GC	WS	**0.33**	0.45	0.170
	KM	WS	0.14	**0.67**	0.096
DOG	LOC	WS	0.19	0.50	0.094
	GLOC	WS	0.13	0.30	0.039
	GC	WS	0.16	0.52	0.073
Interobserver			0.96	0.81	0.780
Intraobserver			0.85	0.79	0.668

* ***
***
\-

\- p > 0.01
* p < 0.01
** p < 0.001
*** p < 0.0001

Figure 6.20.: Accuracy, median Jaccard coefficient and combined performance for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods applied to Alexa568-stained HeLa cells.

Comparison between manual and optimal automated segmentation

For this very challenging dataset containing overlapping cells that can only be recognized by dot patterns, the optimal segmentation pipeline (MIN-KM-WS) achieved a combined performance $p_c = 0.187$, median Jaccard coeffi-

cient $p_j = 0.61$ and accuracy $p_a = 0.31$. When this dataset was annotated twice by two independent observers, a combined performance of $p_c = 0.780$ with $p_j = 0.81$ and $p_a = 0.96$ was obtained. Intraobserver measurements resulted in $p_c = 0.668$ with $p_j = 0.79$ and $p_a = 0.85$. For this dataset, human observation clearly outperformed the proposed algorithms, particularly in terms of accuracy, as parameters cannot be adjusted in such way that cells are automatically split correctly. A Friedman test followed by a pairwise Wilcox post-hoc test confirmed that ability to recognize cell boundaries is significantly greater for inter- and intraobserver measurements.

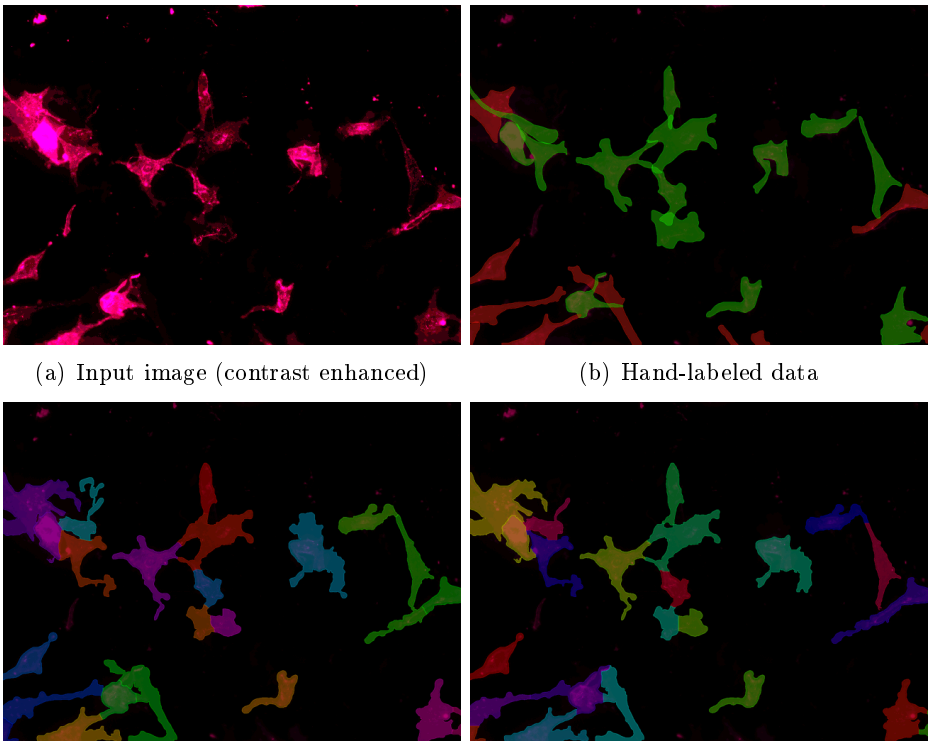
In order to improve segmentation for this challenging data, information on cell nuclei could be included.

Comparison of different methods

Comparison of the utilized image processing methods demonstrates that average combined performance of MIN preprocessing ($p_c^{av} = 0.153$) is superior to that of DOG preprocessing ($p_c^{av} = 0.076$). Differences between both methods are mainly caused by accuracy, not by observed median Jaccard measurements (see Figure 6.20). This is supported by the pairwise Wilcox test showing statistically significant differences, with $p < 0.01$ for two out of four pairs of observations (MIN-KM-WS, DOG-KM-WS) and (MIN-GLOC-WS, DOG-GLOC-WS).

Comparing combined performance for figure-ground separation methods, KM achieves $p_c^{av} = 0.142$, LOC $p_c^{av} = 0.124$, GLOC $p_c^{av} = 0.070$ and GC $p_c^{av} = 0.122$. This suggests that the KM figure-ground separation method outperforms LOC, GLOC and GC for this dataset. Comparison of Jaccard similarities confirms this hypothesis, as these are statistically significant for all combinations of preprocessing and figure-ground separation methods.

6.3.7. Application domain D6: DiD-stained HeLa cells



(a) Input image (contrast enhanced)

(b) Hand-labeled data

(c) Segmentation without using information on nuclei

(d) Segmentation using information on nuclei

Figure 6.21.: Corresponding images (a-d) illustrating segmentation performance of calibrated segmentation pipeline without (c) and with use of information on nuclei (d) for the segmentation of DiD-stained HeLa cells.

Results for calibrating the segmentation pipeline based on hand labelings of DiD-stained HeLa cells are provided in Figure 6.22. For this experiment, calibration was carried out using a GA and including three-fold cross validation. For resulting images, see Figure 6.21.

Comparison between manual and optimal automated segmentation

Calibrating the segmentation pipeline, the combination of methods MIN-GC-SGC performs best for DiD-stained HeLa cells, and results in a combined segmentation performance of $p_c = 0.649$, $p_j = 0.70$ and $p_a = 0.93$. Interobserver measurements $p_c = 0.705$, $p_j = 0.75$ and $p_a = 0.94$ as well as intraobserver measurements $p_c = 0.666$, $p_j = 0.77$ and $p_a = 0.87$ demonstrate that automatically segmented data shows slightly lower quality than

Pre	Fg	Split	Accuracy	Jaccard	Combined
		WS	0.63	0.50	0.317
	KM	SWS	**0.93**	0.64	0.595
		SFM	**0.93**	0.64	0.592
		SGC	0.91	0.64	0.578
		WS	0.57	0.47	0.265
	LOC	SWS	**0.93**	0.68	0.635
		SFM	**0.93**	0.69	0.641
		SGC	**0.93**	0.69	0.638
		WS	0.56	0.49	0.272
MIN	GLOC	SWS	0.92	0.61	0.569
		FM	**0.93**	0.62	0.570
		SGC	0.91	0.59	0.540
		WS	0.62	0.60	0.377
	GC	SWS	**0.93**	**0.70**	0.647
		SFM	**0.93**	0.68	0.633
		SGC	**0.93**	**0.70**	**0.649**
		WS	0.46	0.49	0.226
	KM	SWS	0.92	0.67	0.612
		SFM	0.92	0.66	0.606
		SGC	0.91	0.68	0.617
		WS	0.48	0.52	0.245
DOG	LOC	SWS	**0.93**	0.67	0.624
		SFM	0.92	0.67	0.618
		SGC	0.90	0.68	0.613
		WS	0.48	0.48	0.227
	GLOC	SWS	**0.93**	0.61	0.568
		SFM	0.92	0.61	0.567
		SGC	0.91	0.59	0.540
		WS	0.48	0.51	0.244
	GC	SWS	0.92	0.66	0.611
		SFM	0.92	0.67	0.621
		SGC	**0.93**	0.69	0.639
Interobserver			0.94	0.75	0.705
Intraobserver			0.87	0.77	0.666

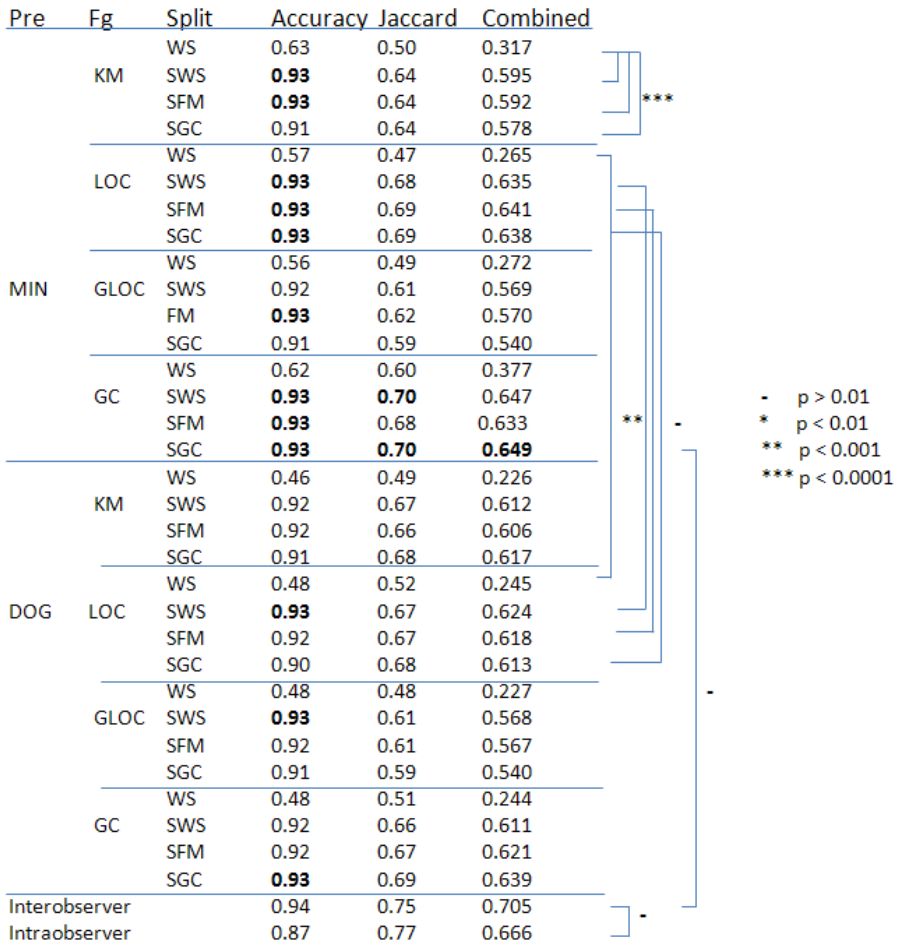


Figure 6.22.: Median Jaccard coefficient and accuracy for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods applied for segmentation of DiD-stained HeLa cells. Due to information on cell nuclei, algorithms can compete with the human handling in terms of accuracy. However, analysis of median Jaccard measurement shows that the calibrated segmentation pipeline is nearly able to resolve cell boundaries as well as humans.

hand-labeled data. Note that, in terms of accuracy, measurements from automated methods outperform the intraobserver measurements. Statistical analysis using a Friedman test, followed by a pairwise Wilcox post-hoc test, shows that Jaccard similarities are not significantly different for the calibrated segmentation pipeline and the interobserver measurements. This demonstrates that automated image segmentation can compete with manual annotation for DiD-stained HeLa cells.

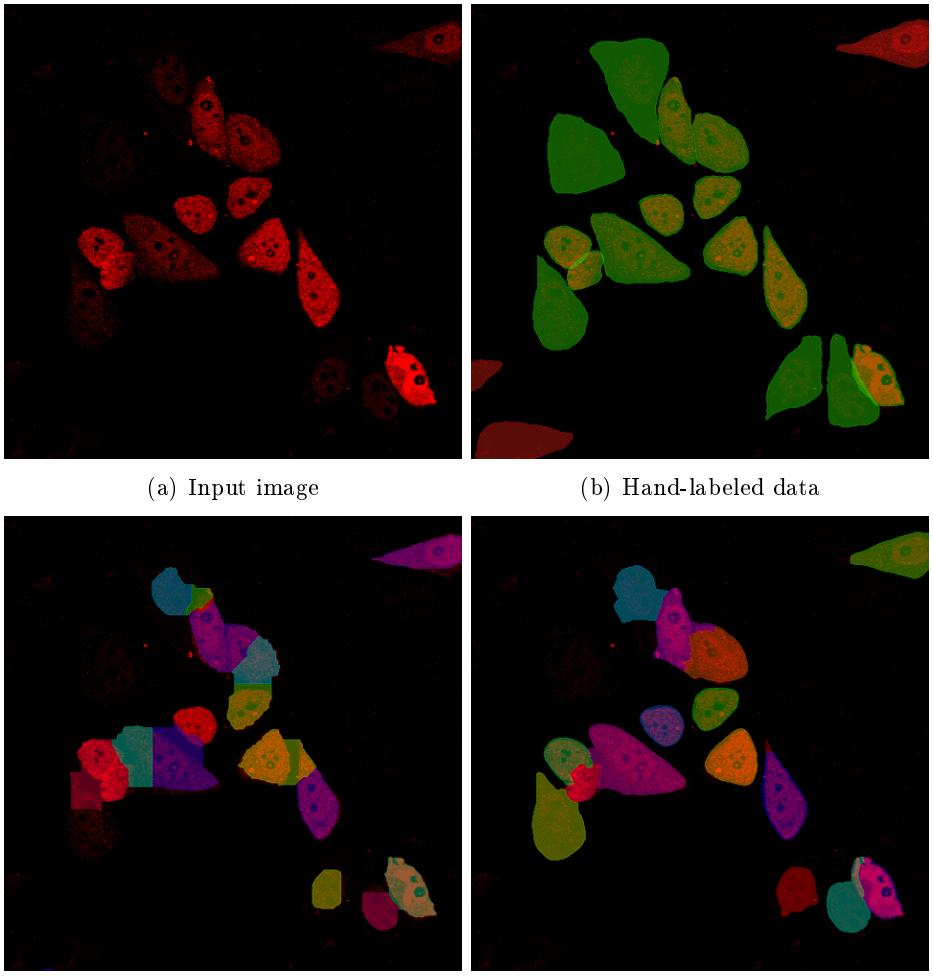
Comparison of different methods

Comparing MIN to DOG preprocessing, average combined Jaccard measurements of $p_c^{av} = 0.543$ for MIN and $p_c^{av} = 0.511$ for DOG preprocessing are obtained. A pairwise Wilcox test shows that Jaccard similarities for two out of 16 method combinations comparing pairs of observations (MIN-M_2-M_3, DOG-M_2-M_3) are statistically significant. Hence, using MIN or DOG preprocessing has only little influence on segmentation quality for DiD-stained HeLa cells.

For figure-ground separation, $p_c^{av} = 0.518$ is obtained for KM, $p_c^{av} = 0.535$ for LOC, $p_c^{av} = 0.481$ for GLOC and $p_c^{av} = 0.553$ for GC based segmentation pipelines. This indicates that the GC figure ground separation method outperforms competing approaches for the presented data. This is also strongly supported by the pairwise Wilcox test showing statistically significant differences for 32 out of 40 pairs of segmentation pipelines (M_1-GC-M_3, M_1-M_2-M_3, with $M_2 \neq$ GC).

Analyzing average combined performance for WS ($p_c^{av} = 0.272$), SWS ($p_c^{av} = 0.608$), SFM ($p_c^{av} = 606$), and SGC ($p_c^{av} = 0.602$) object splitting shows that information on nuclei strongly supports the cell splitting. However, performance hardly differs for different cell splitting methods using information on nuclei. To substantiate this hypothesis with statistical data, a pairwise Wilcox test was performed comparing the averagely best splitting method (SWS) to alternative seeded splitting methods. This test confirmed that none of the differences for all pairs of observed Jaccard measurements (M_1-M_2-SWS, M_1-M_2-M_3) with ($M_3 \neq$ SWS) is statistically significant.

6.3.8. Application domain D7: pUL97-stained HeLa cells



(a) Input image

(b) Hand-labeled data

(c) Segmentation without using information on nuclei

(d) Segmentation using information on nuclei

Figure 6.23.: Merged input image (a), corresponding hand-labeled data (b) and results from the calibrated segmentation pipeline using non-seeded (c) and seeded (d) splitting methods for the segmentation of pUL97-stained HeLa cells. These example images nicely demonstrate that the challenging task of splitting touching cells can be facilitated when information on cell nuclei is included.

An overview of performance for different combinations of methods applied for the segmentation of pUL97-stained HeLa cells is provided in Figure 6.24. For the pUL97-stained HeLa dataset, information on combined performance

and accuracy was not determined for seeded methods, as for this dataset many cells do not express any signal (see Figure 6.7, p. 66). This leads to numerous false positive detections if information on position and extension of the nuclei is used for initialization. Removal of these false positive detections is easily possible by adapting the splitting methods (e.g., cells can be removed that are only few pixels larger than the corresponding nuclei). However, such postprocessing routines are specific for the presented dataset, and therefore not incorporated into this general-purpose approach. For examples of resulting segmentations, see Figure 6.23.

Comparison between manual and optimal automated segmentation

Due to many false positive detections caused by nuclei with missing corresponding cell signals in the pUL97 channel, the median Jaccard similarity was used for comparison of segmentation methods. The best combination of methods (MIN-KM-FM) resulted in a median Jaccard similarity of $p_{\mathrm{j}} = 0.88$. Median Jaccard similarity for interobserver measurements was $p_{\mathrm{j}} = 0.87$, and $p_{\mathrm{j}} = 0.91$ for intraobserver measurements. A statistical pairwise Wilcox test showed that differences between inter- and intraobserver Jaccard similarities are significant with $p < 0.0001$. The described measurements nicely demonstrate that automated figure-ground separation can compete with manual figure-ground separation.

Comparison of different methods

Comparing average median Jaccard similarity of MIN preprocessing ($P_{\mathrm{c}}^{\mathrm{av}} = 0.81$) to that of DOG ($P_{\mathrm{c}}^{\mathrm{av}} = 0.81$) indicates that both methods show identical performance for the pUL97-stained HeLa cells. This could be confirmed by the pairwise Wilcox test, which showed that none of the pairs of observations (MIN-M_2-M_3, DOG-M_2-M_3) are significantly different.

Next, average median Jaccard similarities were compared for KM ($P_{\mathrm{c}}^{\mathrm{av}} = 0.80$), LOC ($P_{\mathrm{c}}^{\mathrm{av}} = 0.81$), GLOC ($P_{\mathrm{c}}^{\mathrm{av}} = 0.78$) and GC ($P_{\mathrm{c}}^{\mathrm{av}} = 0.83$) figure-ground separation. Observed measurements and the pairwise Wilcox test showed that median Jaccard similarity using GC figure-ground separation was significantly better than performance of KM, LOC or GLOC for most method combinations.

Evaluating median Jaccard similarity for different object splitting methods resulted in $P_{\mathrm{c}}^{\mathrm{av}} = 0.67$ for WS, $P_{\mathrm{c}}^{\mathrm{av}} = 0.85$ for SWS, $P_{\mathrm{c}}^{\mathrm{av}} = 0.86$ for SFM and $P_{\mathrm{c}}^{\mathrm{av}} = 0.85$ for SGC based object splitting. The pairwise Wilcox test confirmed that observed measurements were hardly significantly different. In detail, only three out of 21 pairs of observations (M_1-M_2-M_3,M_1-M_2-M_3^*) with $M_3^* \neq M_3$ are significantly different with $p < 0.01$.

Pre	Fg	Split	Accuracy	Jaccard	Combined
		WS	0.63	0.58	0.368
	KM	SWS	-	0.86	-
		SFM	-	**0.88**	-
		SGC	-	0.87	-
		WS	0.63	0.65	0.412
	LOC	SWS	-	0.87	-
		SFM	-	**0.88**	-
		SGC	-	0.87	-
		WS	0.56	0.66	0.372
Min	GLOC	SWS	-	0.82	-
		SFM	-	0.85	-
		SGC	-	0.85	-
		WS	0.58	0.77	0.499
	GC	SWS	-	0.86	-
		SFM	-	0.87	-
		SGC	-	0.81	-
		WS	0.60	0.64	0.382
	KM	SWS	-	0.86	-
		SFM	-	0.87	-
		SGC	-	0.86	-
		WS	0.59	0.61	0.360
DoG	LOC	SWS	-	0.88	-
		SFM	-	0.87	-
		SGC	-	0.87	-
		WS	0.62	0.62	0.379
	GLOC	SWS	-	0.80	-
		SFM	-	0.84	-
		SGC	-	0.84	-
		WS	**0.67**	0.80	0.539
	GC	SWS	-	0.87	-
		SFM	-	0.86	-
		SGC	-	0.83	-
Interobserver			0.94	0.87	0.810
Intraobserver			0.98	0.91	0.894

*** ** - - p > 0.01 * p < 0.01 ** p < 0.001 *** p < 0.0001

Figure 6.24.: Accuracy, median Jaccard coefficient and combined segmentation performance for different combinations of preprocessing (Pre), figure-ground separation (Fg) and object splitting (Split) methods. For the pUL97-stained HeLa data, information on combined performance and accuracy is not determined for methods requiring information on the cell nuclei, as many cells do not express any signal located in the cytoplasm (compare to Figure 6.23). Otherwise, this would result in numerous false positive detections.

6.3.9. Summary and Discussion

In order to draw conclusions on applicability of a specific image processing method for a given task, this work compared different state of the art as well as new methods, based on a broad range of fluorescence microscopy data. The resulting measurements of average combined performance for each segmentation and image processing method are shown in Table 6.2.

Method	D1	D2	D3	D4	D5	D6	D7
MIN	0.55	**0.75**	0.79	**0.84**	**0.15**	**0.53**	**0.75**
DOG	**0.56**	0.79	**0.81**	0.84	0.08	0.51	0.74
KM	0.58	0.77	**0.82**	0.84	**0.14**	0.52	0.75
LOC	0.53	**0.77**	0.80	0.84	0.12	0.54	0.75
GLOC	0.45	0.76	0.79	0.83	0.07	0.48	0.70
GC	**0.65**	0.76	0.80	**0.85**	0.12	**0.55**	**0.78**
WS	0.25	0.75	0.80	0.84	0.11	0.27	0.52
SWS	0.64	-	-	-	-	**0.61**	0.85
SFM	**0.67**	-	-	-	-	0.61	**0.86**
SGC	0.65	-	-	-	-	0.60	0.75
Manual	0.76	0.75	0.87	0.90	0.72	0.69	0.89

Table 6.2.: Average of median combined performance measurements for each image processing method and each dataset (D1-D7). Missing values "-" could not be determined as these methods require information on nuclei. Note that for D7, Jaccard similarities are provided instead of combined Jacccard similarities (for details see Figure 6.24). In order to assess performance of manual annotation, average combined performance of inter- and intra-observer measurements are provided.

Measurements of segmentation performance showed that MIN preprocessing outperforms DOG preprocessing for five out of seven application domains. For some of the remaining data, DoG preprocessing shows significantly improved Jaccard similarity measurements. Hence, it can be concluded that applicability of both preprocessing methods is specific for each dataset.

Comparing methods for figure-ground separation, the novel GC based method shows best performance for four out of seven application domains. Particularly for data showing strong variation in intensity inside each image, GC outperforms KM based figure-ground separation. This is consistent with the results published in (Held et al., 2013b), where the GC approach was shown to outperform the KM method and the CellProfiler for several publicly available fluorescence datasets. Comparison of global methods (KM and GC) to locally adaptive ones (LOC and GLOC) shows that the latter can compete for data with a constant confluence. If confluence varies strongly inside the

dataset, area occupied by cells also varies. As a result, estimated average and standard deviation of intensity distribution also change. This reduces performance for LOC and GLOC, as can be seen for data from application domain D1: CD11b/APC-stained macrophages.

Comparing methods for splitting of cells clearly demonstrates that additional information on seeds (e.g. cell nuclei) significantly improves the segmentation process. Comparing seeded SWS, SFM and SGC for splitting indicates that all seeded methods achieve comparable results. Surprisingly, the SGC based method is slightly outperformed by SFM based ones, despite ability of SGC to enable global minimization of the corresponding energy function, whereas SFM performs a local optimization of a very similar energy function. This is consistent with findings from Hodneland et al. (2009), who compared a watershed transform based method to a level set based approach for segmentation of fluorescence micrographs.

6.3.10. Conclusion

Based on the described results, it can be concluded that the novel GC figure-ground separation method outperforms alternative approaches for several application domains. In addition, it was shown, based on a broad range of fluorescence microscopy data, that cell splitting methods based on SWS, SFM and SGC result in similar segmentation quality.

6.4. Experiment III: Comparison between KM and GC figure-ground separation using partially simulated micrographs

Fluorescence image data used in this work consist of several micrographs with varying cell types and staining. Most of these micrographs show high image quality with large contrast between object and background. Hence, the effect of noise and shading artifacts cannot be studied in detail using this data. In order to analyze how such artifacts influence segmentation performance, micrographs were overlaid with simulated artifacts. Based on these images, performance of KM can be compared to the novel GC figure-ground separation method, which has been developed to improve segmentation of blurred and noisy micrographs.

6.4.1. Simulation of noise and illumination artifacts

For simulation of degraded reference and testing data, images $I(x, y)$ are overlaid with simulated additive noise artifacts $A(x, y)$. Additionally, a multiplicative term $M(x, y)$ is included for simulation of illumination artifacts. Based on this, a simulated micrograph $S(x, y)$ is defined as:

$$S(x, y) = M(x, y)I(x, y) + A(x, y).$$

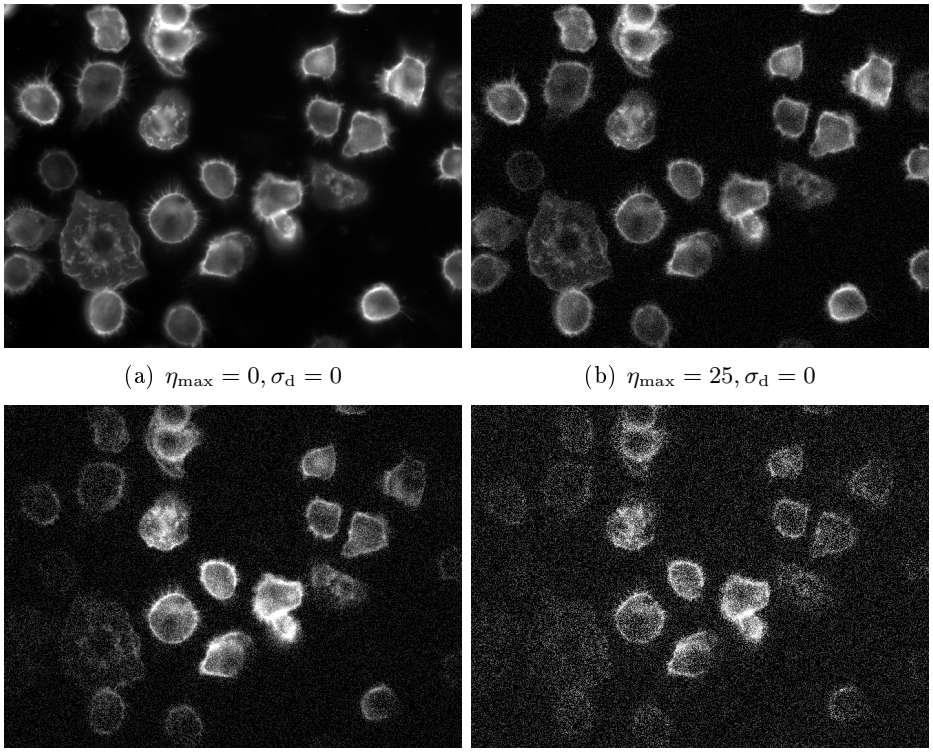
For simulation of the additive noise component, an equally distributed random number $A(x, y) \in [-\eta_{\max}, \eta_{\max}]$ is applied. For the multiplicative noise component, intensity is assumed to decrease from the center of the image to the boundary. Let $d(x, y)$ denote the distance of pixel (x, y) to the image center using the Euclidean norm, and $d_{\max}$ be the distance from the center to an image corner. Then, the multiplicative noise component $M(x, y)$ is defined as:

$$M(x, y) = 2^{-\left(\frac{d(x,y)}{\sigma_{\mathrm{d}}}\right)^2 d_{\max}}.$$

Additionally, values of the resulting image are cropped to a range $[0, 255]$ by setting values > 255 to 255 and values < 0 to 0. Resulting images for varying parameters σ_{d} and $\eta_{\max}$ are depicted in Figure 6.25.

6.4.2. KM vs. GC figure-ground separation

Using the described methods for simulation of shading artifacts and noise, performance of KM based figure-ground separation (Figure 6.26) is compared to performance of GC figure-ground separation (Figure 6.27). Therefore, noise and shading artifacts with different parameterization are added to the reference and the testing image. For each reference image, calibration



(a) $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (b) $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$

(c) $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (d) $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

Figure 6.25.: Examples for different values of additive noise component, parameterized by $\eta_{\max}$, and shading artifacts, parameterized by σ_{d}, added to the reference image (a - d).

of the segmentation pipeline is performed using KM and GC based figure-ground separation. For reduction of the number of configurations and for reduction of runtime, SFM object splitting is applied, as it showed good results for segmentation of the CD11b/APC-stained macrophages (compare Figure 6.12, p. 76) and is much more time-efficient than SGC object splitting. Optimal parameterization for all combinations of methods is then determined based on hand-labeled data utilizing genetic algorithms and the combined Jaccard similarity metric as objective function. Convergence of the genetic algorithm is assumed if no new performance maximum is determined for 200 iterations. The resulting calibrated segmentation pipeline is then applied for segmentation of different test images. Corresponding test images with different strengths of noise and shading artifacts using KM figure-ground separation are shown in Figure 6.26. Results from GC figure-ground separation are provided in Figure 6.27. Enlarged images enabling better comparison between KM and DOG figure-ground separation are pro-

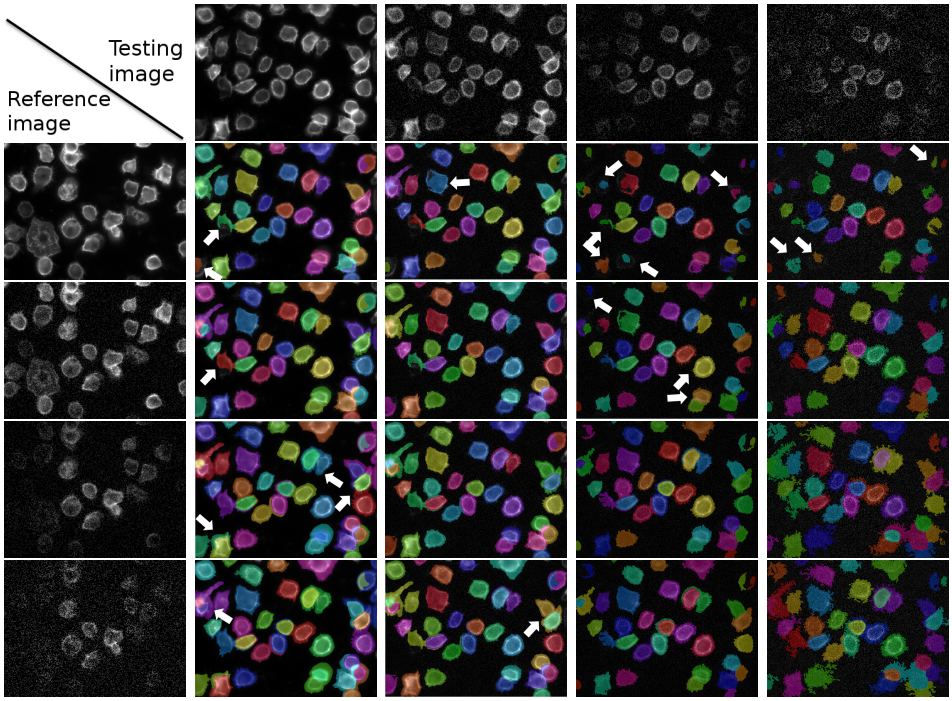


Figure 6.26.: Examples for calibrating the segmentation pipeline based on KM figure-ground separation and SFM object splitting for segmentation of partially simulated micrographs. For enlarged images, see Figures A.2 - A.5, pp. 136 - 139. Note that most severe segmentation errors were highlighted by white arrows.

vided in Appendix A, pp. 135 ff.

For quantitative comparison, median Jaccard similarities enabling comparison of KM and GC for different constellations of reference and testing data are summarized in Figure 6.28.

6.4.3. Discussion

In general, highest segmentation performance is achieved if strength of shading and illumination artifacts is equal for reference and testing data (Figure 6.28). This shows that both methods are not able to adapt parameters independently of noise and illumination artifacts. Hence, it can be concluded that algorithms are able to adapt to a constant level of noise or illumination artifacts. However, if strength of these artifacts varies inside the dataset, segmentation performance usually decreases.

Analyzing variation in performance for a fixed reference image, for testing images with varying noise and illumination artifacts, clearly shows that such

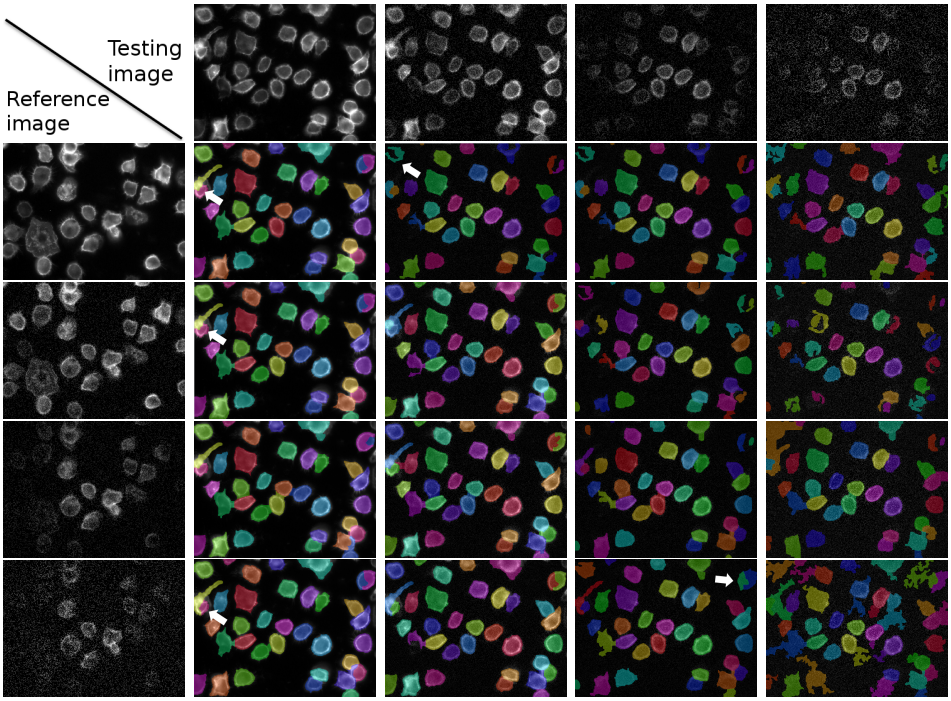


Figure 6.27.: Examples for calibrating the segmentation pipeline based on GC figure-ground separation and SFM object splitting for segmentation of partially simulated micrographs. For enlarged images see Figures A.2 - A.5, pp. 136 - 139. Note that the most severe segmentation errors were highlighted with white arrows.

variations are less strong for GC than for KM. This nicely demonstrates that parameterization of GC based figure-ground separation is less affected by noise and illumination artifacts than parameterization of KM figure-ground separation. This supports the hypothesis that GC based figure-ground selection enables improved segmentation for images showing large variation in intensities, noise, as well as illumination artifacts. This hypothesis is supported by visual inspection of the image data (Figures 6.26 and 6.27 as well as Figures A.2-A.5, pp. 136-139) since GC is often able to recognize cells that could not be recognized by KM. Furthermore, GC enables improved recognition of cell boundaries and often shows less leaking. This could be confirmed by qualitative median Jaccard measurements, depicted in Figure 6.28. If measurements for calibration with $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$ in Figure 6.28(c) are interpreted as an outlier, median Jaccard similarity of GC is always larger than median Jaccard similarity of KM figure-ground separation.

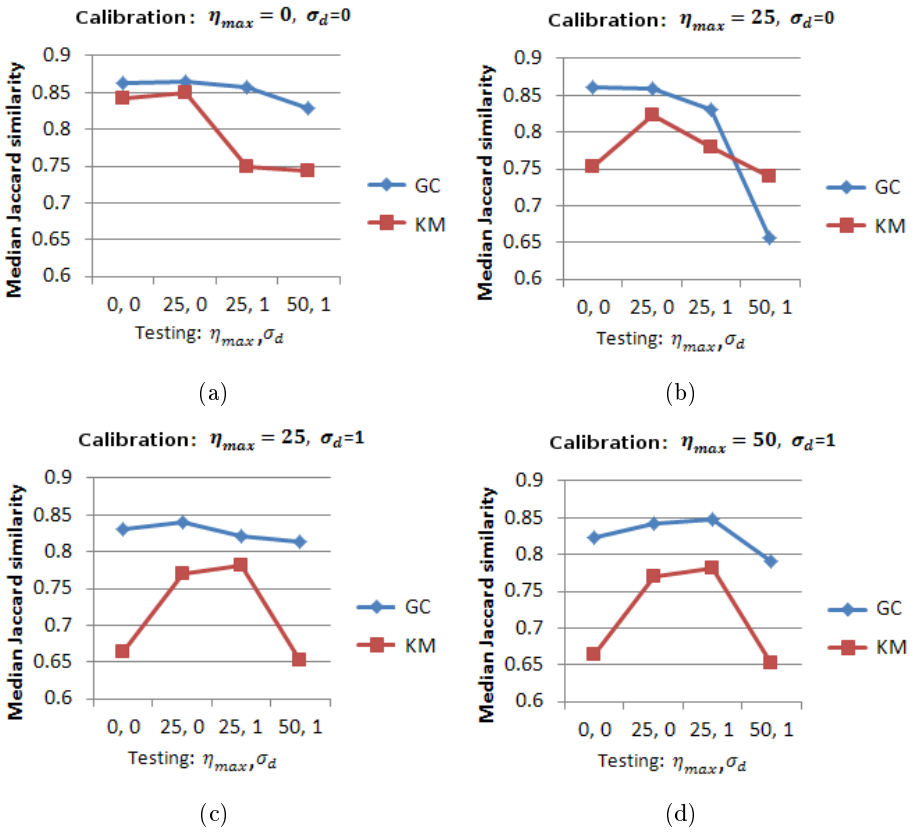


Figure 6.28.: Median Jaccard similarities for various constellations of simulated artifacts, for reference and testing data of GC and KM. Calibration based on original data $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (a), noisy data $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (b), noisy data with illumination artifacts $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (c) and very noisy data with illumination artifacts $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$ (d).

6.4.4. Conclusion

The described data confirms that the novel GC figure-ground separation method outperforms the state of the art KM method, particularly for the segmentation of blurred and noisy micrographs.

6.5. Experiment IV: Exploratory parameter space analysis

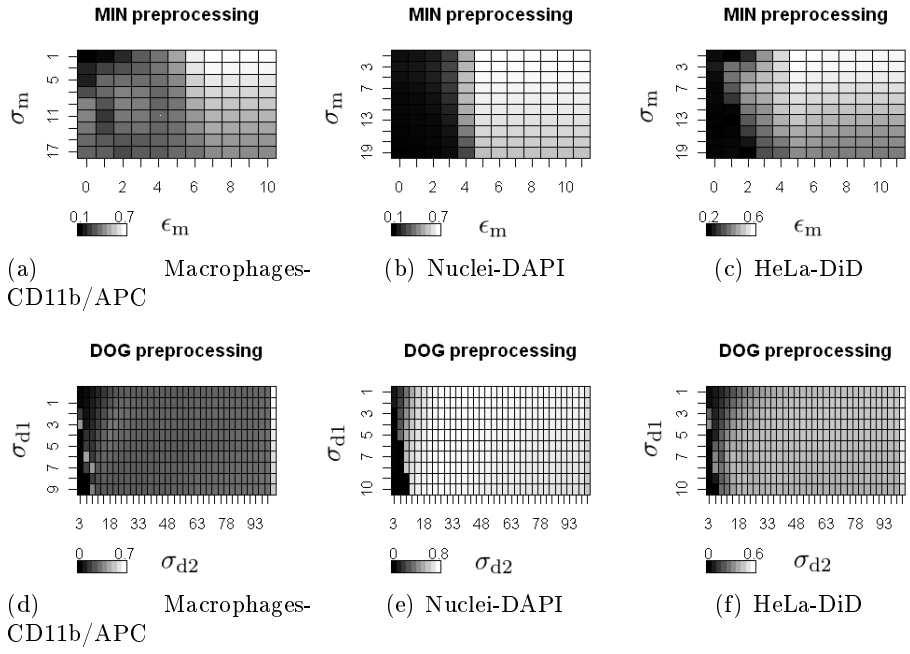
Investigation of the parameter spaces for each of the utilized image processing methods is essential for choice of an automated parameter optimization approach. As this topic is hardly covered in literature, an exploratory parameter space analysis was carried out based on a broad range of fluorescence microscopy image data.

As the segmentation pipeline uses different methods for preprocessing, figure-ground separation and object splitting, analysis of each parameter for all combinations of methods is hardly possible for all datasets. In order to reduce runtime of this experiment, the number of data and combinations of methods, and hence the number of investigated parameter spaces, was reduced. For this reason, only CD11b/APC-stained macrophages, D2: DAPI-stained nuclei and D6: DiD-stained HeLa cells were considered for detailed analysis of parameter search spaces. D1: CD11b/APC-stained macrophages were selected, as they represent a challenging dataset, depicting cells with varying intensities as well as overlapping and overlaying cells. D6: DiD-stained HeLa cells were analyzed due to the multipolar nature of these cells, showing strong overlaps, as well as D2: DAPI-stained nuclei, for including a simple nuclei dataset. Besides restriction to three application domains, the number of investigated combinations of methods was also restricted. Therefore, the optimal combination of image processing methods was determined for each application domain by a genetic algorithm including three-fold cross validation. Note that optimal corresponding combinations of methods for each dataset have already been determined in Section 6.3. Parameter spaces of all methods were then investigated by replacing a single method in this pipeline while keeping the remaining ones. Denoting the optimal combination of methods as (M_1^*, M_2^*, M_3^*), the combinations of methods (M_1-M_2^*-M_3^*) are considered for preprocessing, (M_1^*-M_2-M_3^*) for figure-ground separation and (M_1^*-M_2^*-M_3) for the splitting of objects. For this experiment, a genetic algorithm without cross-validation was applied for calibration of the segmentation pipeline. Based on this, all pairs of parameters were varied around the performance maximum, so as to provide an insight into the parameter space for each dataset. Examples and visualizations of the resulting parameter spaces are shown in the following sections. Note that parameter spaces for segmentation pipelines MIN-KM-WS and MIN-KM-SWS using the macrophage dataset have already been published in (Held et al., 2013a).

6.5.1. Parameter spaces for preprocessing

For CD11b/APC-stained HeLa cells, the optimal combination of image processing methods used GC figure-ground separation methods and SGC for

splitting of the cells. As a result of this, parameter spaces of the preprocessing methods were explored using pipelines MIN-GC-SGC and DOG-GC-SGC. For the DAPI-stained nuclei, pipelines MIN-LOC-WS, DOG-LOC-WS were explored, whereas for DiD-stained HeLa, pipelines MIN-GC-SGC and DOG-GC-SGC were used. Parameters of all methods were calibrated in such way that optimal performance could be obtained by a genetic algorithm. Then, pairs of parameters were varied for each preprocessing method. The resulting visualizations of parameter spaces are shown in Figure 6.29.



(a) Macrophages-CD11b/APC (b) Nuclei-DAPI (c) HeLa-DiD

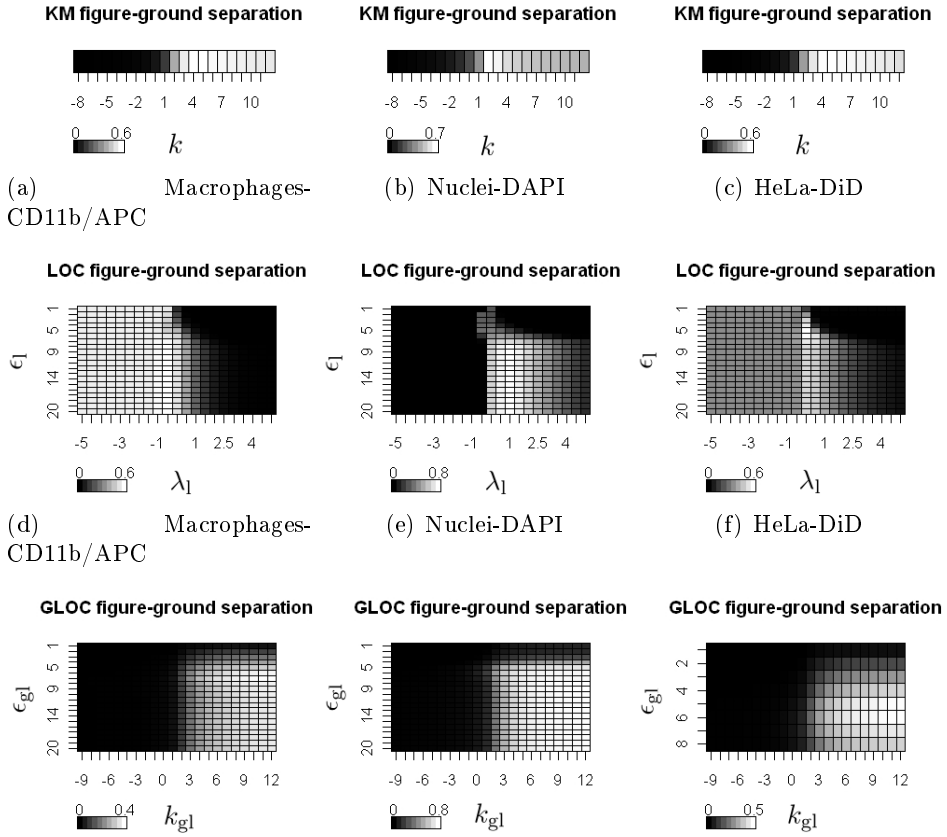
(d) Macrophages-CD11b/APC (e) Nuclei-DAPI (f) HeLa-DiD

Figure 6.29.: Exploration of parameter spaces based on different data for preprocessing (a-f) with intensities encoding combined Jaccard similarity for varying parameterization of MIN (a-c) and DOG (d-f) preprocessing methods.

6.5.2. Parameter spaces for figure-ground separation

For investigation of parameter spaces for figure-ground separation, methods KM, LOC, GLOC and GC were investigated based on data from CD11b/APC-stained macrophages. Therefore, parameters of segmentation pipelines MIN-KM-SGC, MIN-LOC-SGC, MIN-GLOC-SGC and MIN-GC-SGC were configured using genetic algorithms and no cross-validation. Parameter spaces were then investigated by varying all pairs of parameters around the performance maximum.

Accordingly, combinations of image processing methods DOG-KM-WS, DOG-LOC-WS, DOG-GLOC-WS and DOG-GC-WS were optimized and investigated for DAPI-stained nuclei. In order to investigate parameter spaces for DiD-stained HeLa cells, combinations of methods MIN-KM-SGC, MIN-LOC-SGC, MIN-GLOC-SGC and MIN-GC-SGC were analyzed. The resulting parameter spaces for KM, LOC and GLOC can be visualized in Figure 6.30 as well as Figures 6.31 to 6.32 for the GC figure-ground separation method.

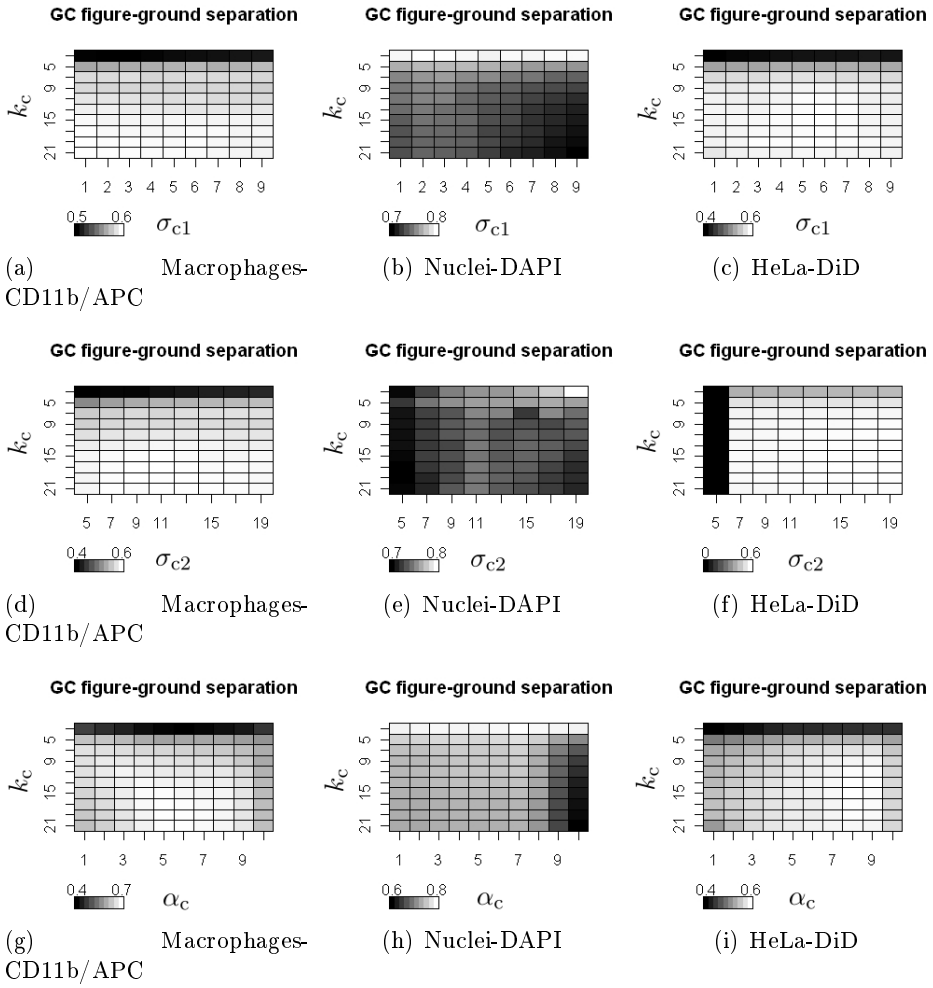


(a) Macrophages-CD11b/APC (b) Nuclei-DAPI (c) HeLa-DiD

(d) Macrophages-CD11b/APC (e) Nuclei-DAPI (f) HeLa-DiD

(g) Macrophages-CD11b/APC (h) Nuclei-DAPI (i) HeLa-DiD

Figure 6.30.: Parameter spaces for KM (a-c), LOC (d-f) and GLOC (g-i) figure-ground separation for different datasets (a-i), with intensities encoding combined Jaccard similarity for varying parameterization of the methods. Parameter space for GC figure-ground separation is given in Figures 6.31 and 6.32.

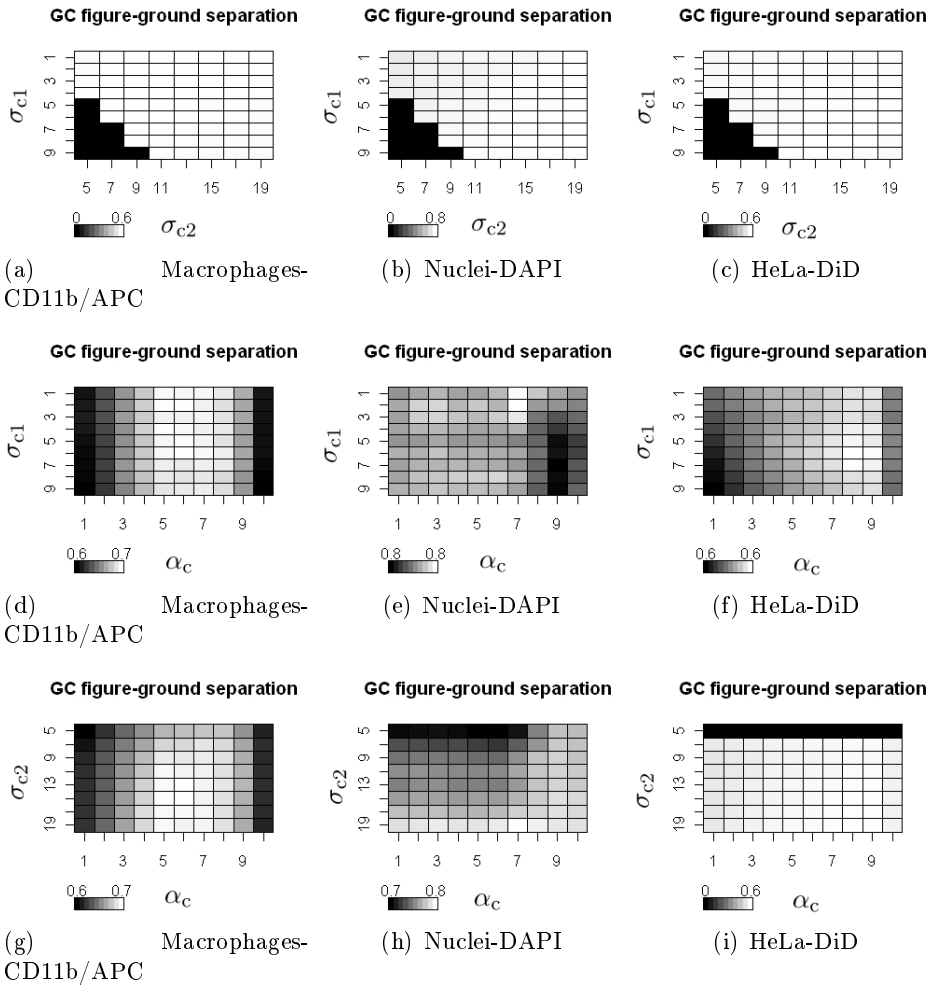


(a) Macrophages-CD11b/APC (b) Nuclei-DAPI (c) HeLa-DiD

(d) Macrophages-CD11b/APC (e) Nuclei-DAPI (f) HeLa-DiD

(g) Macrophages-CD11b/APC (h) Nuclei-DAPI (i) HeLa-DiD

Figure 6.31.: Parameter space of GC figure-ground separation (a-i), part 1 of 2, with intensities encoding combined Jaccard similarity for varying parameterization of the methods. Further visualizations are given in Figure 6.32.



(a) Macrophages-CD11b/APC (b) Nuclei-DAPI (c) HeLa-DiD

(d) Macrophages-CD11b/APC (e) Nuclei-DAPI (f) HeLa-DiD

(g) Macrophages-CD11b/APC (h) Nuclei-DAPI (i) HeLa-DiD

Figure 6.32.: Parameter space of GC figure-ground separation (a-i), part 2 of 2, with intensities encoding combined Jaccard similarity for varying parameterizations.

6.5.3. Parameter spaces for object splitting

In order to investigate parameter spaces for object splitting, segmentation pipelines MIN-GC-WS, MIN-GC-SWS, MIN-GC-SFM and MIN-GC-SGC were investigated for CD11b/APC-stained macrophages. For DAPI-stained nuclei, only DOG-LOC-WS were investigated due to the lack of information on nuclei. For DiD-stained HeLa cells, optimal method combinations MIN-GC-WS, MIN-GC-SWS, MIN-GC-SFM and MIN-GC-SGC were analyzed. In order to explore parameter spaces for object splitting methods, all parameters were automatically optimized towards the hand labelings using genetic algorithms. Then, pairs of parameters were varied around the optimum. The resulting parameter spaces are available in Figures 6.33 and 6.34.

6.5.4. Results and discussion

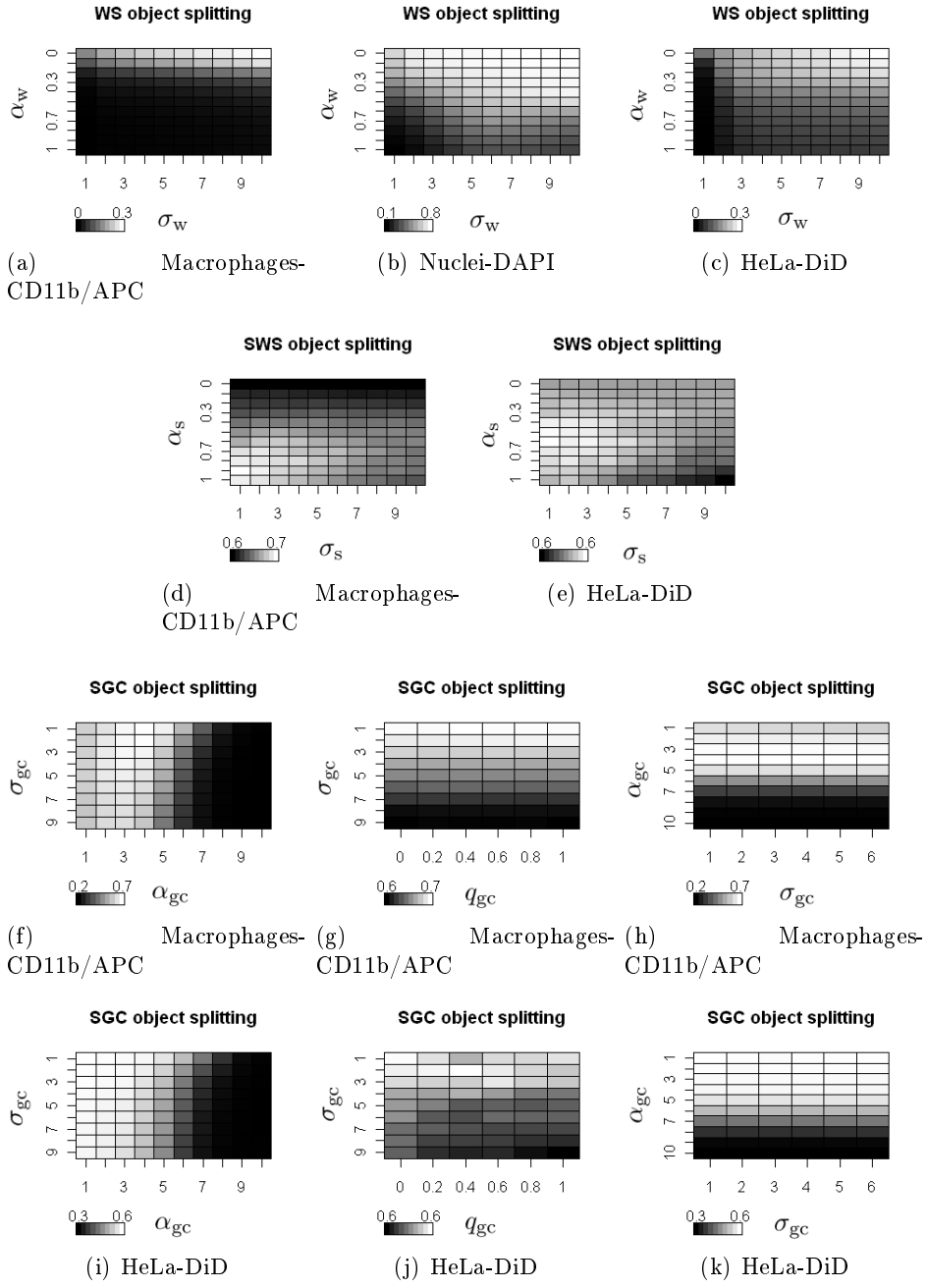
Detailed analysis of parameter spaces for preprocessing (Figure 6.29, p. 101) showed that combined Jaccard performance strongly depends on optimization of the parameter space for preprocessing methods. Hence, sticking to a local performance maximum could lead to a strong decrease in segmentation performance. However, probability to stick at a local performance optimum is low if adequate optimization techniques are applied, as corresponding parameter spaces are quite monotonous and show only few local performance maxima.

Exploring parameter spaces for different figure-ground separation methods (Figures 6.30-6.32, pp. 103-105) indicates that, as for the preprocessing methods, identification of the global performance maximum is crucial, otherwise segmentation performance could strongly decrease. Most of the corresponding parameter spaces are quite monotonous, showing only few local performance maxima. Solely for the graph cut based method, pairs of parameters (σ_{c2}, k_c) and (α_c, σ_{c1}) showed low monotony and several local performance maxima. Fortunately, segmentation performance only slightly depends on optimization of these parameters.

For object splitting methods, most parameter spaces are very monotonous. Moreover, parameter space of WS is more monotonous than that of SWS (Figures 6.33 and 6.34, pp. 107, 108). This is consistent with the findings published in (Held et al., 2013a).

6.5.5. Conclusion

Taken together, exploration of method parameter spaces showed that many approaches exhibit local performance maxima. Performance of the global optimum is thereby often much higher than performance of local optima. Consequently, sticking at a local performance optimum must be avoided. This demonstrates that optimization schemes able to jump out of local performance optima are required. In practice, this poses a challenge for manual



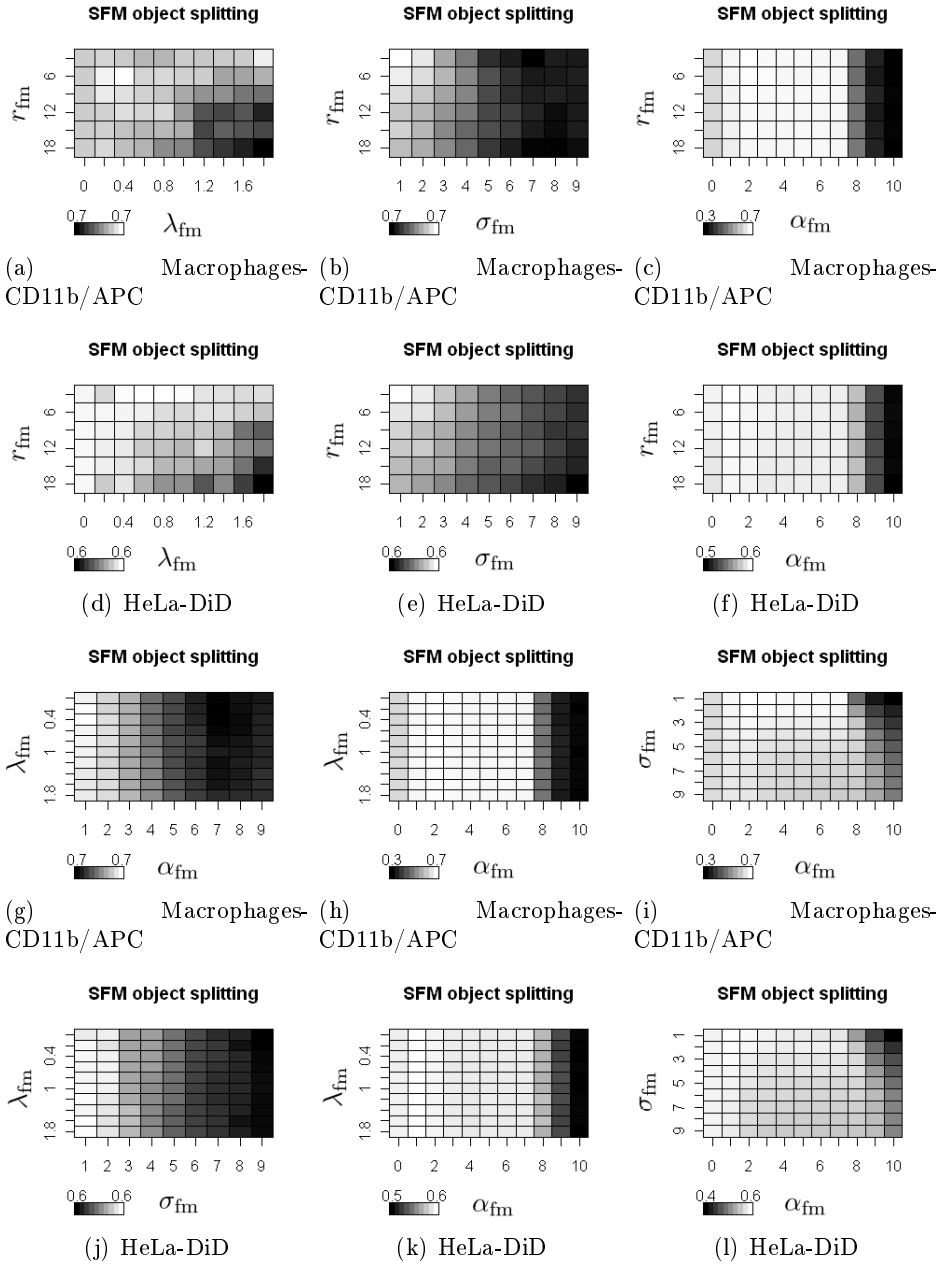
(a) Macrophages-CD11b/APC (b) Nuclei-DAPI (c) HeLa-DiD

(d) Macrophages-CD11b/APC (e) HeLa-DiD

(f) Macrophages-CD11b/APC (g) Macrophages-CD11b/APC (h) Macrophages-CD11b/APC

(i) HeLa-DiD (j) HeLa-DiD (k) HeLa-DiD

Figure 6.33.: Investigation of parameter spaces for WS (a-c), SWS (d, e) and SGC (f-k) object splitting methods, with intensities encoding combined Jaccard similarity for varying parameterizations. Note that parameter space for SFM is shown in Figure 6.34 .



(a) Macrophages-CD11b/APC (b) Macrophages-CD11b/APC (c) Macrophages-CD11b/APC

(d) HeLa-DiD (e) HeLa-DiD (f) HeLa-DiD

(g) Macrophages-CD11b/APC (h) Macrophages-CD11b/APC (i) Macrophages-CD11b/APC

(j) HeLa-DiD (k) HeLa-DiD (l) HeLa-DiD

Figure 6.34.: Investigation of parameter space for SFM object splitting with intensities encoding combined Jaccard similarity for varying parameterizations and application domains.

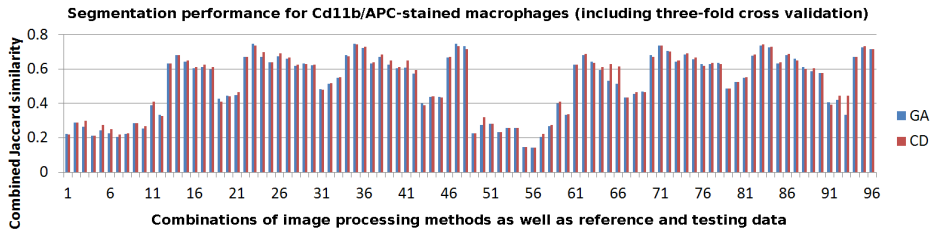
calibration of a segmentation pipeline, as local performance maxima are often accepted.

6.6. Experiment V: Comparison of parameter optimization techniques

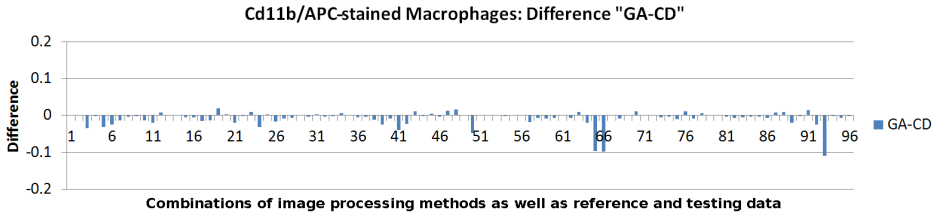
Exploratory parameter space analysis for different data showed that automated optimization algorithms must be able to handle parameter spaces showing several local extrema. This confirms that genetic algorithms (GA) are a good choice for optimization of the depicted parameter spaces. For a further speedup of the calibration process, more runtime-efficient optimization techniques are required. As no comparison on parameter optimization techniques for fluorescence microscopy image data are available, GA was compared to the coordinate descent (CD) method in this experiment. The CD method is considered an alternative due to its ability to jump out of local performance optima as well as to its capability to efficiently optimize high dimensional parameter spaces. As the GA and the CD method are not always able to identify a globally optimal parameter setup, using a brute force technique for identification of the global performance optimum is desirable. Based on this, errors resulting from the GA or CD could be assessed. Unfortunately, runtime of brute force exceeds several years, even for a single image, and is hence not applicable. In the following, convergence as well as quality of the parameterization are compared for GA and CD.

6.6.1. Comparison of maximum performance

For this experiment, combined Jaccard similarity was compared for GA and CD based parameter optimization. In this context, parameters were optimized for all combinations of image processing methods using GA and CD. Note that this experiment was only carried out for the application domains D1: CD11b/APC-stained macrophages - D2: DAPI-stained nuclei and D6: DiD-stained HeLa cells, in order to reduce the amount of required computational time. In this experiment, the segmentation pipeline was calibrated based on hand-labeled data for all combinations of image processing methods. The GA as well as the CD method both include a three-fold cross validation for separation between reference and testing data. Figure 6.35 shows the resulting segmentation performance measurements for both optimization techniques for CD11b/APC-stained macrophages. Results for DAPI-stained nuclei are provided in Figure 6.36. Measurements based on DiD-stained HeLa cells are available in Figure 6.37.



(a)



(b)

Figure 6.35.: Comparison between segmentation performance of genetic algorithm (GA) and coordinate descent (CD) parameter optimization for different combinations of image processing methods as well as reference and testing data (1-96) based on CD11b/APC-stained macrophages (a). Additionally, three-fold cross validation is included to separate reference from testing data. The corresponding difference plot is provided in (b).

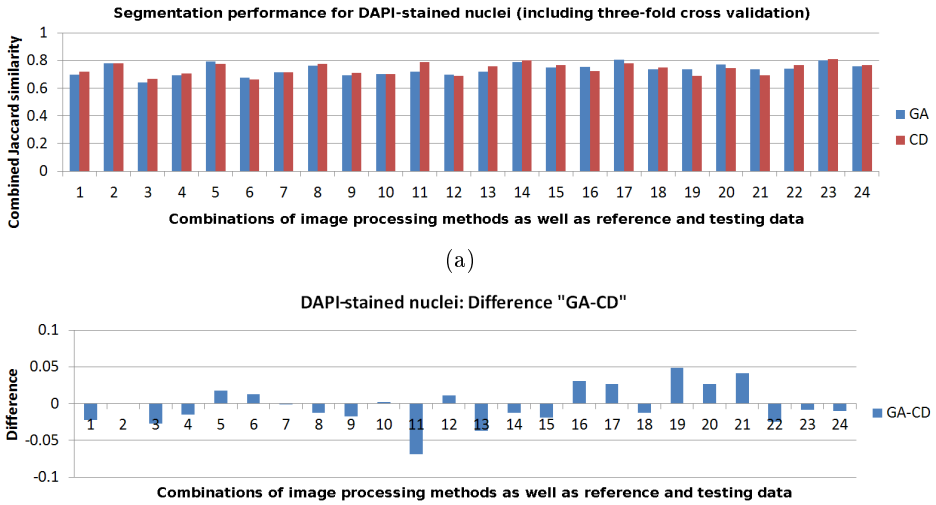
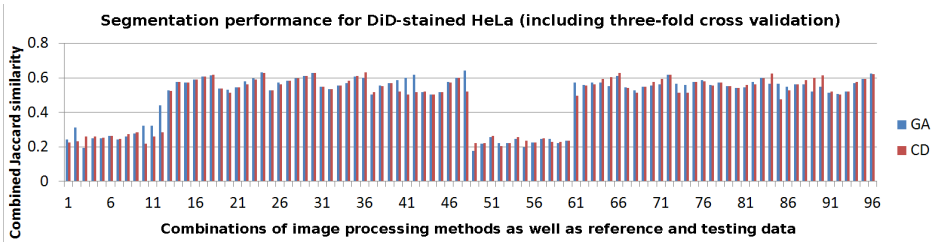


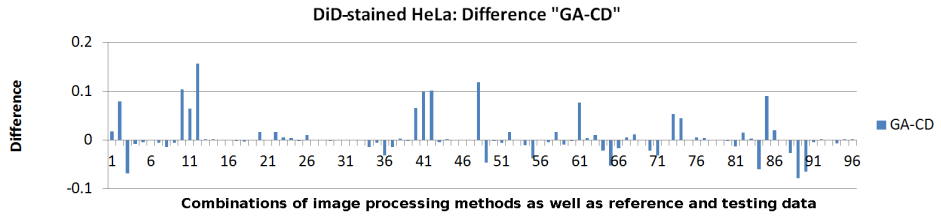
Figure 6.36.: Comparison between segmentation performance of GA and CD parameter optimization based on DAPI-stained nuclei. Three-fold cross validation is included to separate reference from testing data. The corresponding difference plot is shown in (b).

6.6.2. Comparison of convergence

Apart from ability to determine a parameterization close to the global optimum, convergence rate is crucial for this application. For fair comparison, convergence is analyzed for combinations of methods and data where GA and CD result in the most different or most similar parameterization. This corresponds to the combinations of methods where GA is most outperformed by CD, and vice-versa. Resulting convergence plots are shown in Figure 6.38. Additionally, convergence is investigated for combinations of image processing methods where GA and CD achieve most similar results (see Figure 6.39).

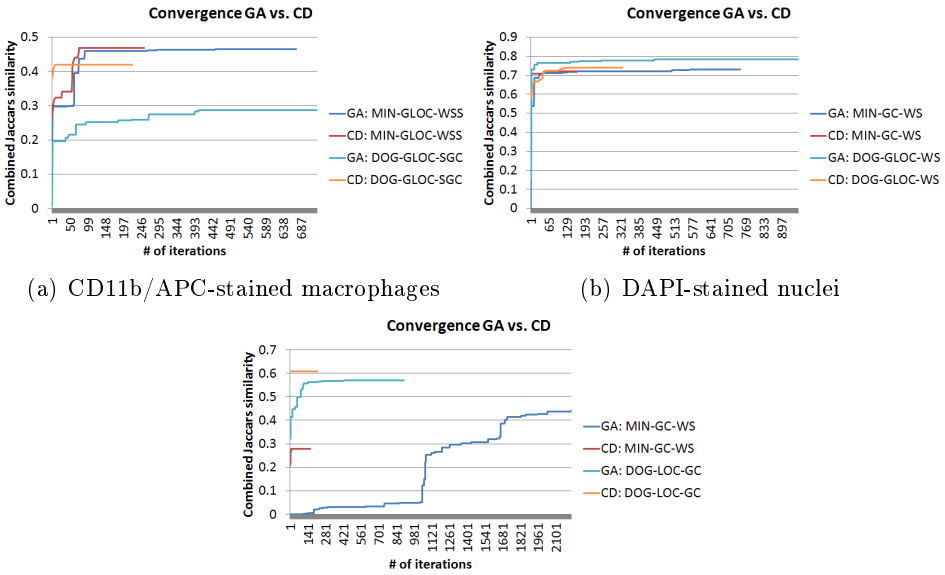


(a)



(b)

Figure 6.37.: Comparison between segmentation performance of GA and CD parameter optimization for different combinations of image processing methods as well as reference and testing data (1-96) based on DiD-stained HeLa cells. Note that three-fold cross validation is included. The corresponding difference plot is provided in (b).

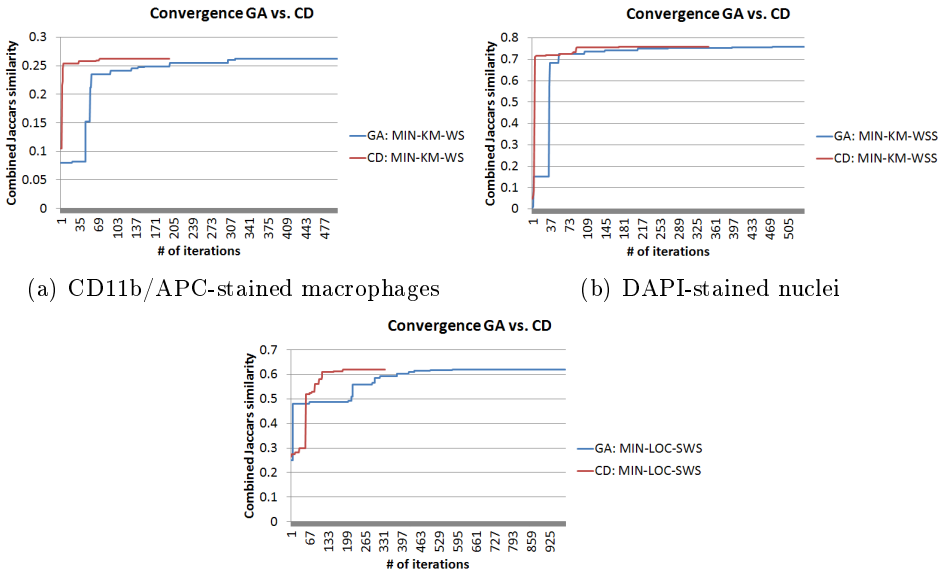


(a) CD11b/APC-stained macrophages

(b) DAPI-stained nuclei

(c) DiD-stained HeLa

Figure 6.38.: Comparison of convergence for GA (blue) and CD (red) optimization techniques for different application domains (a, b, c). For this plot, only combinations of methods where GA and CD result in the most different parameterization are considered. This corresponds to scenarios where CD most outperforms GA and vice versa.



(a) CD11b/APC-stained macrophages

(b) DAPI-stained nuclei

(c) DiD-stained HeLa

Figure 6.39.: Convergence using GA (blue) and CD (red) optimization for segmentation of different data (a, b, c). For this plot, combinations of image processing methods are selected where GA and CD result in the most similar segmentation performance.

6.6.3. Discussion

In this section, GA and CD parameter optimization methods were compared for calibration of the segmentation pipeline for data from a broad range of application domains. For the CD11b/APC-stained macrophage data, CD resulted in a slightly better parameterization than GA (compare Figure 6.35(b)). For DAPI-stained nuclei of macrophages, both methods showed comparable results (see Figure 6.36(b)), whereas for DiD-stained HeLa cells, GA tended to outperform CD (compare Figure 6.37(b)). Based on these experiments, it was concluded that both methods are comparable for calibration of the segmentation pipeline.

Comparing convergence of both methods indicates that convergence of CD is more time-efficient than GA, as a lower number of iterations is required until convergence. This effect is mainly caused by the convergence criterion of the GA, assuming convergence if no new performance maximum is identified for 200 iterations. Convergence plots clearly demonstrate that time efficiency of the GA could be improved by modification of the convergence criterion. For example, convergence could be assumed if performance increased for less than δ in the past 200 iterations. Optimal choice of δ would depend on the applied performance metric and the image data. Hence, adjusting δ might pose a challenge.

A further question that must be addressed is whether GA or CD parameter optimization could replace parameter optimization by a greedy brute force technique. Comparison of results from GA or CD to those from brute force search is not possible due to the long runtime of the automated methods. Hence, data from the comparison between GA and CD is employed to discuss this question. Comparing performance of the calibrated segmentation pipeline using GA and CD shows a difference in combined Jaccard similarity up to 0.1. This difference is quite large and poses a serious drawback for non-greedy parameter optimization methods. Fortunately, large errors are relatively rare and differences between Jaccard similarity for GA and CD are usually below 0.025. Additionally, for solving the dual optimization problem, only the best combination of methods is used. Hence, probability is low that a configuration of the segmentation pipeline that is far below the possible optimum is selected, as such a configuration is usually outperformed by alternative configurations. This hypothesis is supported by the performance plots (Figures 6.35-6.37) showing similar performance values for different combinations of methods.

6.6.4. Conclusion

Based on this experiment, it could be concluded that CD slightly outperforms GA, as both methods result in similar segmentation performance, though convergence of CD is often faster. It was further shown that errors

implemented by using CD or GA are usually small (below 0.025 combined Jaccard similarity) and can be accepted.

6.7. Experiment VI: Recommended input to the segmentation pipeline

For most applications in research and industry, measurement of segmentation performance is usually not included in the experimental setup. Hand-labeled data is hence only used for calibration of the segmentation pipeline. As a result, a small number of hand-labeled cells shall be used for calibration. In the following, research carried out on the influence of the number of hand-labeled cells on the resulting segmentation performance is reported. Additionally, different strategies for selecting cells for manual annotation are compared. Note that experiments in this chapter were carried out solely for application domains D1: CD11b/APC-stained macrophages, D2: DAPI-stained nuclei and D6: DiD-stained HeLa cells, due to limited availability of hand labelings from experts. These three datasets have been selected because they range from simple nuclei data to challenging touching and overlapping cells.

6.7.1. Required number of cells

In order to investigate the influence of the number of hand labelings on the resulting segmentation performance, each dataset was split into a reference and a testing subset for each application domain. Size of the reference set was chosen in such way that as to consist of about 100 valid cells. Hence, each testing dataset still contained more than 400 cells. For each reference dataset, 1, 5, 10, 25, 50 and 100 cells were randomly selected and manually annotated. Remaining cells were assigned to a rejection class. Based on these different amounts of reference cells, calibration of the segmentation pipeline was performed using genetic algorithms and the combined Jaccard similarity metric. Additionally, a fixed segmentation pipeline was used to reduce the amount of required computational time. Therefore, a combination of image processing methods was selected that had shown good segmentation performance in previous experiments (see Chapter 6.3). In order to draw a conclusion on the required number of hand labelings, calibration was performed for varying amounts of hand labelings. The calibrated segmentation pipeline was applied to the testing subset of data. Resulting combined Jaccard similarity performance values for varying numbers of reference cells are shown in Figure 6.40.

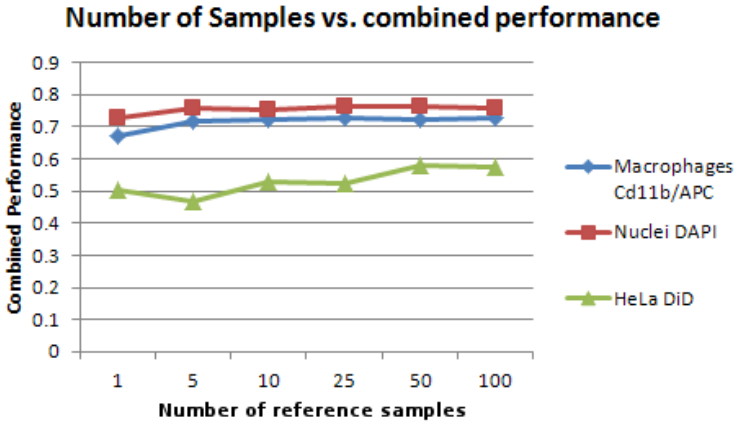


Figure 6.40.: Measurement of combined performance for calibration of the segmentation pipeline based on varying numbers of randomly selected hand-labeled reference samples which are not included in the testing dataset.

6.7.2. Selection of cells for hand labeling

In order to reduce of the amount of required hand labelings, recommendations for the selection of cells for this purpose were required. Hence, the question was addressed of whether calibration based on isolated, touching or overlapping cells would result in the best segmentation performance.

For this experiment, the dataset was split into equally sized reference and testing subsets of data. Calibration of the segmentation pipeline was then performed by using genetic algorithms without cross-validation on the reference dataset, whereas only isolated (respectively touching or overlapping) cells were used. All remaining cells were assigned to the rejection class. Performance evaluation was then carried out using all cells in the testing dataset (excluding only invalid cells). Carrying out this experiment for CD11b/APC-stained macrophages yielded combined Jaccard similarity measurements $p_c = 0.749$ for calibration with only isolated macrophages, $p_c = 0.736$ for calibration with only touching macrophages, and $p_c = 0.726$ for calibration using only overlapping cells.

6.7.3. Discussion

Comparison of combined Jaccard similarity for calibration based on varying numbers of hand labelings indicated that segmentation performance tendentiously increased with the number of hand labelings. For most datasets, usage of 10 hand-labeled cells resulted in good segmentation. Only for DiD-stained HeLa cell data segmentation performance increased if more than 50

cells were used for calibration. This indicates that the selection of representative cells is more important than quantity of the annotated cells.

In the second part of this experiment, only isolated, touching or overlapping cells were used for calibration of the segmentation pipeline. The results indicated that usage of isolated cells results in the best parameterization of the segmentation pipeline. However, this piece of information must be interpreted very carefully, as isolated cells can touch close neighbor cells in case of an over-segmentation. As a result, calibrating parameters for splitting of touching cells may also be possible based on isolated cells.

From a theoretical point of view, using only isolated cells for calibration results in well-adjusted parameters for preprocessing and figure-ground separation. However, parameters for object splitting are mainly determined by chance. In contrast, all parameters can be adjusted appropriately if touching cells are used for calibration. Using overlapping cells for calibration is only useful if the segmentation pipeline is able to perform an overlapping segmentation of these cells. As the described segmentation pipeline as well as state of the art methods can only perform a segmentation of touching cells, using these cells for calibration results in a bias towards small objects when applying a combined Jaccard similarity based objective function. The reason for this bias is the larger impact of small cells. An example as well as a detailed explanation are provided in Figure 6.41. Apart from this, parameters of cell splitting cannot be calibrated correctly if the reference data consists of only overlapping cells.

As a result of the described measurements and the bias towards correct segmentation of smaller objects, touching cells are regarded as optimal input for the proposed segmentation pipeline, as long as segmentation of overlapping and overlaying cells is not supported.

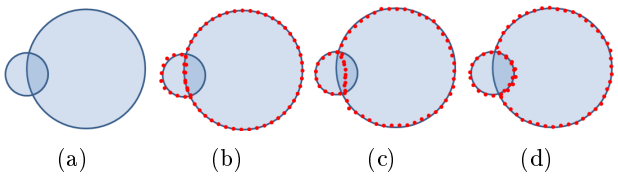


Figure 6.41.: Calibration of the segmentation pipeline with overlapping cells (a - d) can result in biased parameter optimization when using the combined Jaccard similarity metric. This bias leads to an improved splitting for smaller objects, whereas splitting performance for larger objects decreases. Hence, segmentation (d) results in the best combined performance. More specifically, in (b) a low Jaccard similarity ≈ 0.5 is obtained when evaluating segmentation performance for the small cell and $p_{\mathrm{j}} \approx 1$ for the large cell. This results in a combined performance of ≈ 0.75. In (c) Jaccard similarity is mainly degraded by the small cell with $p_{\mathrm{j}} \approx 0.75$. Combining this with a Jaccard similarity $p_{\mathrm{j}} \approx 1$ for the large cell leads to a combined segmentation performance of ≈ 0.88. Splitting the small cell correctly and neglecting splitting of the large cell (d) leads to the highest combined segmentation performance of ≈ 0.95, as $p_{\mathrm{j}} = 1$ for the small cell and $p_{\mathrm{j}} \approx 0.9$ for the larger cell.

6.8. Summary of experiments

In this chapter, research was carried out towards improving runtime, efficiency and segmentation performance for the calibration of a segmentation pipeline based on hand-labeled data. As this topic is not considered in the state of the art, the described experiments constitute a contribution.

In order to compare quality of the segmentation for different configurations of the segmentation pipeline to the hand labelings, an objective function is required. Based on fluorescence micrographs of macrophage cells, segmentation pipeline was calibrated using different objective functions. This experiment clearly showed that the overlap based combined Jaccard similarity outperforms Hausdorff based measurements, as the latter occasionally resulted in an exclusive segmentation of the cell boundary. This experiment further demonstrated that using the L^1norm should be preferred to alternative norms, as segmentation performance of all cells should have an equal impact.

After identification of an objective function, performance of different segmentation methods was compared based on a broad range of fluorescence image data. A three-fold cross validation was applied for the separation of reference and testing data, and genetic algorithms were used for optimizing parameters of all method combinations. This experiment clearly demonstrated that performance differences between preprocessing methods are relatively small. For figure-ground separation, locally adaptive methods were outperformed by alternative ones for most datasets. In addition, the novel graph cut based figure-ground separation outperformed alternative approaches for most datasets. For cell splitting, the seeded fast marching level set method slightly outperformed alternative approaches, particularly the graph cut based splitting method. Comparing results from automated image segmentation to inter- and intraobserver results showed that automated methods can compete with the human performance for simple image data. In this context, most nuclei and cell data with convex shape, objects with homogeneous intensity distribution as well as data that does not contain touching or overlapping cells can be considered as simple. Experimental data confirmed that information on the cell nuclei allows separation of touching cells for most data. However, the segmentation of overlapping cells still remains a challenge. Particularly for overlapping cells, manual annotations remain more accurate than automated segmentation results. As these evaluations were carried out on a broad range of fluorescence micrographs, using hand-labeled data and automated calibration for a fair comparison of the methods, this experiment represents an extension to the state of the art.

So far, the area of application for the segmentation pipeline was demonstrated based on high-quality image data. However, several types of artifacts can hamper image segmentation in practice, even after careful sample preparation and image acquisition. Cells expressing different amounts of staining

can lead to varying intensities of cells in a micrograph, as well as to a high noise level. In order to study the influence of such effects on segmentation performance, micrographs were overlaid with simulated noise and shading artifacts. The resulting data nicely demonstrated that shading artifacts, which are very similar to cells of varying intensity, can decrease segmentation performance. Nevertheless, particularly for these data, the novel graph cut based figure-ground separation algorithm was shown to be more robust against noise and artifacts than the k-means clustering based state of the art approach.

In order to reduce runtime for calibration of the segmentation pipeline, more runtime-efficient optimization methods are required. For identification of applicable methods, an exploratory parameter space analysis was performed. As this experiment revealed that parameter spaces were mostly monotonous, but showed several local performance maxima, usage of very time-efficient optimization methods (e.g. the hill-walking method) often results in a local performance optimum. Hence, optimization methods must be able to jump out of local performance optima.

Based on this result, genetic algorithms and the coordinate descent method were considered for optimization and compared to each other. Comparing Jaccard similarity for the optimal parameter settings identified by GA or CD showed comparable results for both methods. Comparing the number of iterations, required until convergence, the coordinate descent method outperformed genetic algorithms for most data. Comparing GA to CD also enabled estimation of the error caused by not using a brute force approach for parameter optimization. It was shown that median Jaccard similarity using GA or CD optimization should rarely decrease by more than 0.025 compared to brute force optimization.

The last experiment carried out in this section addresses the question of how many and which kind of data is best to be used for calibration of the segmentation pipeline. Based on experimental data, it could be shown that a small amount of about 10 hand-labeled samples often results in a good calibration of the segmentation pipeline. In addition, it was argued that representative, ideally touching, but not overlapping data should be used for calibration.

7. Summary: Design of efficient segmentation pipeline(s) - a trade-off between performance and runtime

Based on experimental data from the previous chapter, two very efficient segmentation pipelines were developed. Efficient means that high-quality calibration of the segmentation pipeline is performed in a short runtime. The first version of this segmentation pipeline performs a trade-off between a high segmentation quality and runtime, whereas the second version focuses on the reduction of runtime. Compared to the first segmentation pipeline, the second version can result in a decreased segmentation performance. However, due to its short runtime, the second variant can be used in an iterative workflow that facilitates the selection of representative reference data.

7.1. Choice of objective function and optimization strategies

Choice of the objective function is essential for an efficient calibration of the segmentation pipeline. Based on quantitative comparison of different objective functions (Section 6.2, p. 67), the combined Jaccard similarity is used for calibration of the segmentation pipeline, as Hausdorff distance based measurements can lead to exclusive segmentation of the object boundary.

Runtime of the calibration is strongly influenced by choice of the optimization technique. Based on an exploratory parameter space analysis, it can be concluded that optimization techniques must be able to handle local performance extrema. Comparison of genetic algorithms (GA) to coordinate descent (CD) optimization techniques (Section 6.6, p. 110) demonstrates that both methods enable identification of parameter settings with comparable segmentation quality, but convergence of CD is faster. As a result, CD is used for calibration of the highly efficient segmentation pipeline.

A further issue that must be addressed is the convergence criterion used by CD. As this algorithm converges if no new performance maximum is detected in an iteration, the required number of iterations is unknown for each new optimization task. Hence, time required for optimization is not predictable. Analysis of convergence for different combinations of methods

and application domains shows that the CD method usually converges after less than 5 iterations (Section 6.6.2, p. 110). Based on this information, runtime of the CD method can be restricted by limiting the number of maximum iterations.

7.2. Focusing on performance and runtime

The first segmentation pipeline performs a trade-off between segmentation quality and runtime of the calibration process. For the development of this segmentation pipeline, methods that have not outperformed alternative methods on any dataset are removed. Hence, runtime of the calibration is reduced, while the resulting segmentation performance is not influenced. Additionally, parameter spaces of all image processing methods are redesigned to improve the trade-off between runtime and segmentation performance.

Based on comparison of image processing methods for various application domains (Section 6.3.9, p. 93), MIN as well as DOG based preprocessing are considered for the segmentation pipeline. For figure-ground separation, KM, LOC and GC based methods are utilized, whereas the GLOC method is removed. Comparison between seeded object splitting methods shows that choice of the optimal method only leads to small improvements in segmentation performance as long as information on nuclei is incorporated. As SFM object splitting showed best performance for two out of three datasets that contain information on nuclei, SFM is applied for splitting of cells if information on nuclei is available. Otherwise, WS is used for splitting of the cells. As a result, SGC and SWS are removed from this improved segmentation pipeline.

Apart from removing image processing methods that cannot compete in practical applications, parameter spaces of the remaining methods are redesigned to reduce runtime of the parameter optimization. Adjustment of range and step size of each parameter must be done carefully to preserve ability to adapt to a broad range of fluorescence image data.

Based on the initial discretization (see Table 5.1, p. 55) and the exploratory parameter space analysis, range and step size are adjusted for all parameters. A summary on the resulting ranges of parameters and corresponding numbers of discrete values is provided in Table 7.1.

To ensure that calibration of the segmentation pipeline can be performed over-night, maximum number of iterations of the CD method must be restricted. Analysis of convergence for CD demonstrates that, in practice, a maximum of five iterations is required until convergence. In the following, number of iterations that can be carried out in 8 hours are estimated based on a typical image of 1024×1024 pixels. For such an image, segmentation usually takes about one second using a single-core implementation on an Intel Core 2 Duo with 2.66 GHz. A single iteration of CD testing

Method	Parameter	Discretization	Number of steps
MIN	σ_{m}	$\{1, 2, 4, ..., 32\}$	6
	ϵ_{m}	$\{1, 2, 3, ..., 11\}$	11
DOG	σ_{d1}	$\{1, 2, 4, ..., 32\}$	6
	σ_{d2}	$\{4, 8, 16, ..., 256\}$	7
KM	k	$\{0, 1, 2, ..., 10\}$	11
LOC	ϵ_{l}	$\{2, 4, 6, ..., 20\}$	10
	λ_{l}	$\{-5, -4, -3, ..., 5\}$	11
GC	σ_{c1}	$\{1, 2, 4, 8\}$	4
	σ_{c2}	$\{4, 8, 16, ..., 128\}$	6
	α_{c}	$\{1, 2, 3, ..., 10\}$	10
	k_{c}	$\{2, 4, 8, 16, 32\}$	5
WS	α_{w}	$\{0, 0.2, 0.4, ..., 1\}$	6
	σ_{w}	$\{1, 2, 4, ..., 64\}$	7
SWS	α_{s}	$\{0, 0.2, 0.4, ..., 1\}$	6
	σ_{s}	$\{1, 2, 4, ..., 64\}$	7
SFM	r_{fm}	$\{2, 4, 8, 16\}$	4
	λ_{fm}	$\{0, 0.45, 0.9, 1.35, 1.8\}$	5
	σ_{fm}	$\{1, 2, 4, 8, 16\}$	5
	α_{fm}	$\{1, 2, 4, 8, 16\}$	5

Table 7.1.: Overview on discretization and number of parameter values for the different image processing methods from the improved segmentation pipeline.

all combinations of methods requires testing 284 combinations of parameters, if information on the cell nuclei is excluded for segmentation, and 320 combinations of parameters if this information is incorporated. Hence, for calibration based on a multi-channel image, $284+320 = 604$ parameter combinations must be tested. This results in a maximum number of $\frac{3600s \cdot 8}{2s \cdot 604} \approx 24$ iterations if a runtime of $2s$ per image is assumed. Hence, maximum number of iterations for CD is restricted to 8 to enable calibration based on three reference images in less than 8 hours for multichannel image data.

7.3. Focusing on runtime

The second segmentation pipeline focuses on the reduction of runtime. Hence, losing a small amount of segmentation performance can be tolerated. For identification of image processing methods used for this highly time-efficient segmentation engine, runtime of all methods was compared. As runtime of each method depends on its parameterization, parameters were varied manually according to Table 7.1 to estimate the runtime of each method.

Table 7.2 summarizes runtime measurements for a representative micrograph from the macrophage dataset.

Method	Min runtime (ms)	Max runtime (ms)
MIN	**78**	**281**
DOG	200	1373
KM	**16**	**32**
LOC	125	140
GLOC	187	1230
GC	900	1100
WS	**811**	**1590**
SWS	**796**	**1060**
SFM	2409	2686
SGC	1124	25000

Table 7.2.: Runtime estimations for all methods based on a macrophage image of size 1388×1040 pixels. Parameters of each method are varied for estimation of minimum and maximum runtime.

These measurements demonstrate that segmentation pipelines MIN-KM-WS and MIN-KM-SWS allow the most time-efficient segmentation. Using MIN-KM-WS, a runtime between $0.90s$ and $1.90s$ can be expected for a 1.44 megapixel image. For MIN-KM-SWS, a runtime between $0.89s$ and $1.37s$ is measured. Using these segmentation pipelines, one iteration of the CD method requires testing 41 parameter combinations for each of the method combinations. Hence, a single iteration of the CD method is estimated to take $\frac{0.90s+1.90s}{2}41 \approx 57s$ for MIN-KM-WS and $\frac{0.89s+1.37s}{2}41 \approx 46s$ for MIN-KM-SWS. Setting the maximum number of iterations for CD to five, a runtime of less than 9 minutes is achieved using a single core from an Intel Core 2 Duo with 2.66 GHz.

Focusing solely on runtime, parameterization is optimized for a single combination of methods. This can result in a decreased segmentation performance, particularly for blurred micrographs and micrographs showing strong variations in image intensity, as the graph cut based figure-ground separation method is not incorporated. Furthermore, errors caused by the optimization method sticking in local performance optimum cannot be leveled if only a single combination of methods is investigated. The effect of these errors can be estimated based on performance measurements in Section 6.3. The data indicates that errors are between 0 and 7% for the datasets used in this work. Note that, for these measurements, genetic algorithms were used for parameter optimization. However, as comparison between GA and CD showed that both methods result in a similar segmentation performance, usage of CD is assumed to result in similar estimations.

7.4. Facilitating the selection of representative reference samples by an iterative workflow

Segmentation quality of both segmentation pipelines strongly depends on selection of the reference samples. This task poses a challenge, particularly for users with little knowledge of image processing. In order to overcome this problem, an interactive workflow is proposed. A further advantage of this concept is the very small amount of initial hand-labeled data required for calibration of the segmentation pipeline. For this workflow, starting with a single cell is usually sufficient. Based on this hand-labeled reference data, a calibration of the segmentation pipeline is performed.

By visualizing preliminary segmentation results, information on challenging cells is provided to the user, as such cells show the largest errors in segmentation performance. By correcting such cells using interactive software tools, the hand-labeled reference data is improved. In turn, using this improved data usually results in a better segmentation of the cells. An illustration of this workflow is depicted in Figure 7.1.

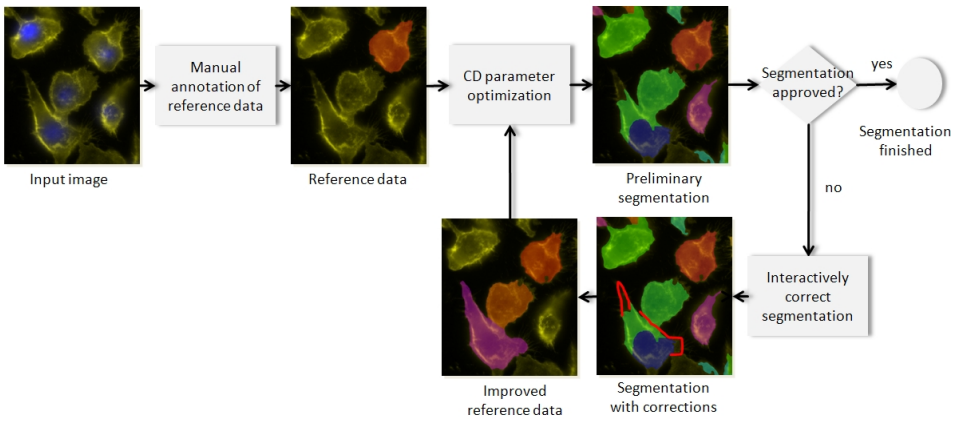


Figure 7.1.: Illustration of the interactive calibration process. Based on a very small amount of reference samples (usually a single cell is sufficient), parameter optimization is performed by the CD method. During calibration, a preliminary segmentation is shown to the user, presenting the performance for the currently best configuration of the segmentation pipeline. During calibration, the user can approve the segmentation. In this case, CD is terminated. Alternatively, the user can correct cells interactively. This user input is used to create improved reference data, and ultimately calibration.

7.5. Discussion

Using a large number of methods for the segmentation pipeline usually results in a good segmentation performance, but runtime of the calibration also increases. That is why, in this work, two efficient segmentation pipelines using only a subset of the described image processing methods were presented that can be optimized more time-efficiently. For the first version of the segmentation pipeline, methods GLOC, SWS and SGC were removed, as they were outperformed by alternative methods in most application domains. Additionally, all parameter spaces were redesigned to enable more time-efficient calibration (see Table 7.1, p. 125). In the second version, only methods with the shortest runtimes were considered. This resulted in the segmentation pipeline MIN-KM-WS for single-channel data, and MIN-KM-SWS if information on nuclei is available. Experimental data showed that using the segmentation pipeline can lead to a decrease in segmentation performance of up to 7% (combined Jaccard similarity), but allows using an iterative workflow, avoiding the task of selecting representative sample data.

8. Conclusion and outlook

As existing software solutions require the user to select a combination of segmentation methods and to adjust all parameters manually, research was carried out on a novel image analysis concept enabling automated selection of segmentation methods and automated optimization of all parameters based on hand-labeled data. For implementation of this concept, two different modular segmentation pipelines were developed, consisting of the stages of preprocessing, figure-ground separation and object splitting. Both segmentation pipelines allow the segmentation of a broad range of fluorescence micrographs. The first segmentation pipeline was developed in such way that both segmentation performance and runtime are optimized, whereas a strong focus was put on a short runtime in the second version of the segmentation pipeline. Based on experimental data, it was shown that ten reference samples often result in a good parameterization of the segmentation pipeline. Assuming that representative reference samples have been selected, segmentation results are often at least comparable to those from manually calibrated segmentation pipelines. The reason for this is the large number of different parameterizations that are tested systematically by the coordinate descent optimization method. Furthermore, runtime for calibration of the described segmentation pipelines was shown to be below 8 hours for three manually annotated reference samples. If a stronger focus is put on runtime, calibration for a single image takes less than 9 minutes. In addition, an iterative workflow was proposed for the more runtime-efficient segmentation pipeline, which facilitates the challenging task of selecting representative reference data.

The aim of this work was the improvement of efficiency and automation in fluorescence image analysis. Due to the low amount of hand-labeled data required for calibration of the segmentation pipeline, as well as to the low runtime of the calibration process, it could be concluded that the amount of time researchers have to spend in the evaluation of their data could strongly reduced with the proposed image analysis concept. Hence, efficiency and automation in fluorescence microscopy image analysis could be improved.

A second topic in this work was carrying out research on image processing methods that are applicable for segmentation of a broad range of fluorescence micrographs. Based on this, the area of application and objectivity for automated image analysis software will be described. In addition, reduction of runtime for the calibration based on parallel implementation of the segmentation pipeline will be discussed and an outlook of this work will be

provided.

Area of application for automated image analysis software

In the introduction of this work, fluorescence micrographs were classified according to the features irregularity, intensity distribution and overlaps. Based on these features, the current area of application for automated image analysis software will be described below.

Figure 8.1 provides an overview of segmentation performance for different combinations of the named features based on the experiments carried out in this work. This illustration shows that the developed segmentation pipeline is able to perform a robust segmentation of cells with irregular boundaries or complex intensity distribution, as long as cells do not touch or overlap each other. In this case, the results of automated segmentation routines are often comparable to manual annotation. This demonstrates that, for the presented data, figure-ground separation can usually be performed with high accuracy. An exception to this are micrographs showing strong variations in intensity, blur or shading artifacts. In order to extend the application area of automated image analysis, a novel graph cut based figure-ground separation algorithm was developed for improving the segmentation for such micrographs.

Touching and overlapping cells can be considered as the greatest challenge for the segmentation of fluorescence micrographs. For touching or overlapping cells, segmentation performance usually strongly decreases if no additional information on nuclei is available, particularly for cells with irregular cell boundaries or non-homogeneous intensity distribution. If information on nuclei is available, touching cells can often be handled robustly, but overlapping cells still pose a challenge, as overlaps are not resolved in the cell splitting step. This is due to the lack of algorithms enabling robust splitting of overlapping micrographs for different application domains. Consequently, using automated segmentation methods for these micrographs results in an erroneous segmentation. Depending on the application, such errors can impede automated analysis and evaluation of an experiment. In order to improve applicability of automated image analysis for challenging objects, two different workflows are plausible, whereas only the first is applicable for high-throughput scenarios:

1. Prevention of overlapping cells: During slide preparation, cells can be seeded less confluently, so that touching or overlapping cells appear less frequently. Alternatively, only images showing isolated or touching cells can be captured or selected by the experimenter.

2. Manual correction: Based on fully automatic segmentation, erroneously segmented objects can be corrected manually by interactive software tools. For this purpose, application tools for splitting and merging of

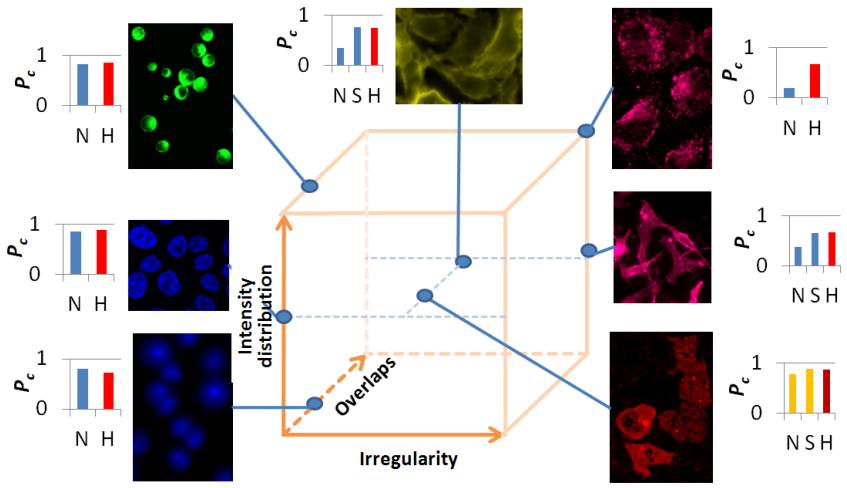


Figure 8.1.: Illustration of combined segmentation performance p_c using the calibrated segmentation pipeline. Measurements denoted with (S) make use of information on cell nuclei or seeds, whereas measurements denoted by (N) do not use this information. Minimum of intra- and interobserver measurements is denoted by (H). Note that for the pUL97-stained HeLa cells, Jaccard similarity is shown (orange bars) instead of combined performance.

cells, and tools for interactively adding regions to objects, or removing regions from objects, were implemented.

Objectivity of automated image analysis for fluorescent micrographs

In general, evaluations carried out by using automated image analysis software are reproducible and objective. However, a bias can already be implemented at the image acquisition stage. Usually, each captured image represents only a tiny portion of the complete slide; hence, only a small part of the cell population is analyzed. Selection of regions of interest that are captured and used for automated image analysis is thereby done by the experimenter. As a result, it can happen that only regions showing objects that confirm the experimental hypothesis are selected. This results in a biased evaluation that can be avoided by acquisition and analysis of the complete object slide. Segmentation of whole slide image data based on the described segmentation pipeline could result in a long runtime or huge memory requirements, particularly if runtime or memory consumption of an algorithm increases more than linearly with the number of pixels, which is the case for all described object splitting routines. This drawback can be eliminated by tiling the whole slide image into several smaller sub-images. By tiling, calibration of the segmentation remains possible over-night for whole slide image data.

Reducing runtime by parallelization

Today, most computers contain multi-core processors as well as multi-core graphics hardware that can be used for parallel computing. The CPUs hexa-core (e.g. AMD Phenom II X6) or octa-core (e.g. Intel Xeon E7-2820), consisting of 6 to 8 processors, can be regarded as state of the art. Modern graphics cards consist of several hundreds of processors (e.g., a NVIDIA GeForce GTX 560 TI consists of 448 cores). By using such hardware, runtime for calibrating the segmentation pipeline can be reduced.

For parallelization of the calibration, the coordinate descent optimization strategy as well as each of the image processing methods can be considered. Parallelization of the coordinate descent is easily achievable, as several performance values must be determined for optimization of each coordinate. Computation of these performance values can be executed in parallel because sequential dependencies do not exist. As computation of each performance measurement requires application of the segmentation pipeline with a different configuration, memory requirement increases nearly linearly, while runtime decreases nearly linearly with the number of processors. Hence, parallelizing coordinate descent parameter optimization can efficiently be done by using a multi-core CPU.

Apart from parallelization of the optimization strategy, each method of the segmentation pipeline can also be considered for parallelization. The drawback of this strategy is that most of the methods can hardly be parallelized. Examples for this are the watershed transform, fast marching level sets and the graph cut method. For such algorithms, parallel implementations require lots of effort and time, and performance decreases far from linearly with the number of multiprocessors. As a result, parallelizing the coordinate descent optimization strategy is proposed. Afterwards, GPU acceleration can be considered for a further speedup of specific algorithms.

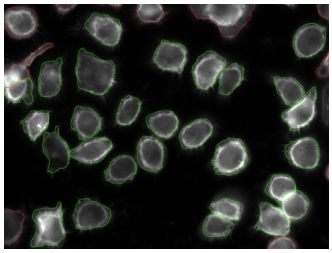
Future Research

In this work, applicability of the concept of calibrating a segmentation pipeline using hand-labeled data was demonstrated for the segmentation of fluorescence imaging data. However, all methods proposed in this work focus on segmentation of 2D image data. Thus, future work may include the development of methods for segmentation of multi-dimensional image data, including segmentation of 3D, 4D or whole-slide image data, as well as cell tracking.

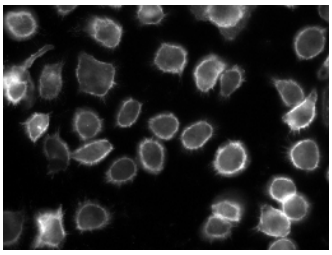
Instead of focusing on data with a higher dimension, research might be carried out to develop segmentation methods that take several image channels into account. State of the art methods as well as algorithms proposed in this thesis incorporate information on size and extension of the cell nuclei during segmentation of the cells. Similarly, information on phase contrast or further staining could be included to improve segmentation.

A further still unsolved topic is the segmentation of overlapping or overlaying objects with irregular boundaries or complex intensity distribution. In this context, independent level sets using information on position and extension of the nuclei did not show satisfying results (Yan et al., 2008). Hence, future research could include the development of methods applicable to segmentation of irregularly overlapping fluorescent micrographs.

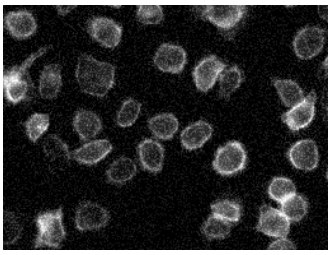
A. Appendix: KM vs. GC figure-ground separation



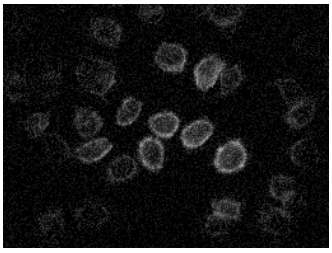
(a) Hand-labeled data



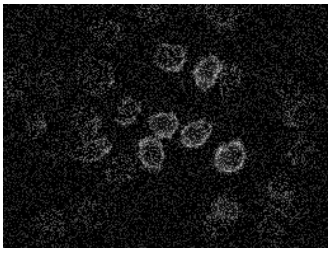
(b) $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$



(c) $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$

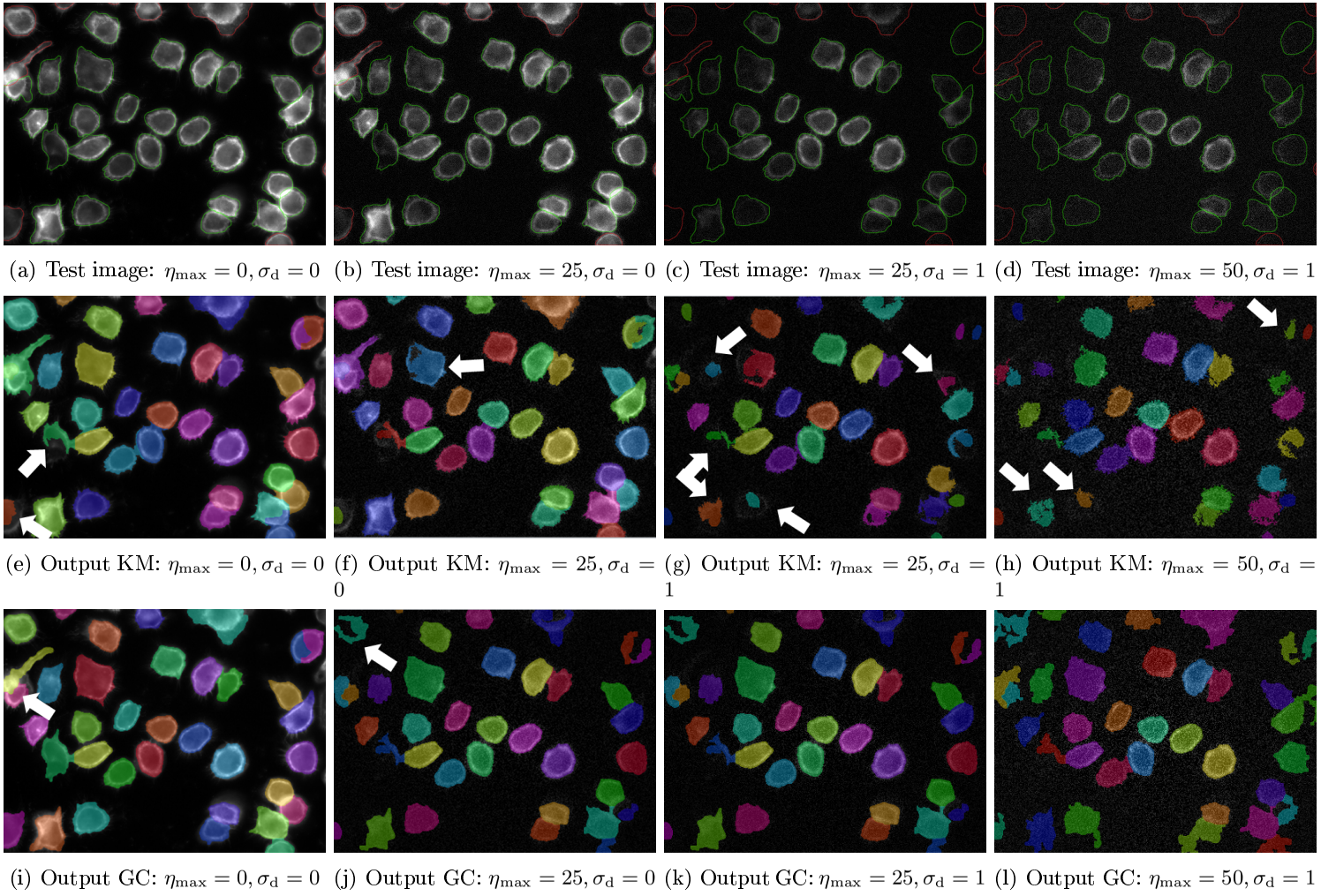


(d) $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$



(e) $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

Figure A.1.: Examples for different strengths of additive noise component, parameterized by $\eta_{\max}$, and shading artifacts, parameterized by σ_{d}, for the test image.

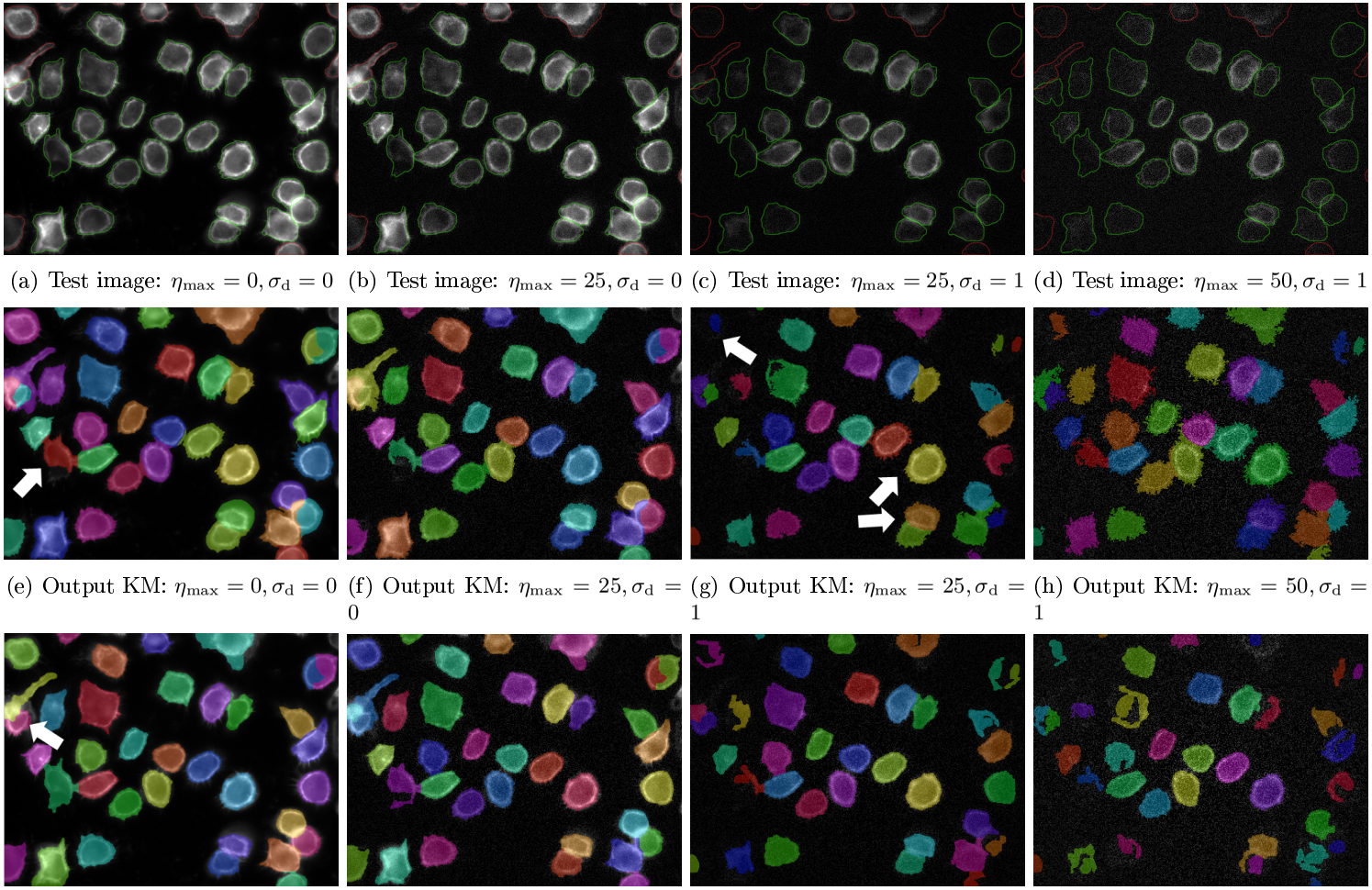


(a) Test image: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (b) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (c) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (d) Test image: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(e) Output KM: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (f) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (g) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (h) Output KM: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(i) Output GC: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (j) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (k) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (l) Output GC: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

Figure A.2.: Test images (a-d), output using KM (e-h) and output using GC (i-l) figure-ground separation. For calibration of the segmentation pipeline, the reference image with parameters $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ was used. Some important differences between KM and GC are highlighted with white arrows.

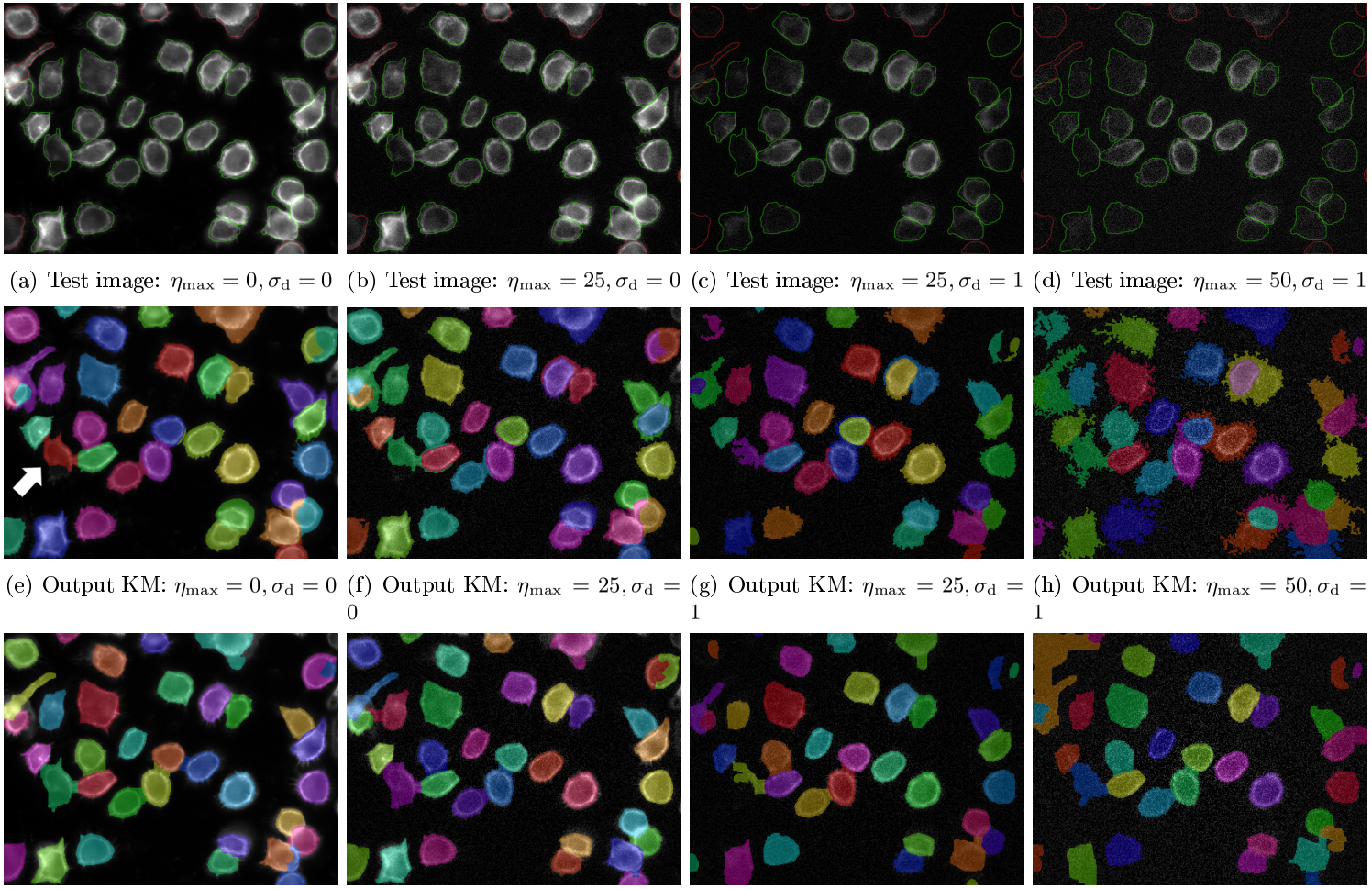


(a) Test image: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (b) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (c) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (d) Test image: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(e) Output KM: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (f) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (g) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (h) Output KM: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(i) Output GC: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (j) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (k) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (l) Output GC: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

Figure A.3.: Test images (a-d), output using KM (e-h) and output using GC (i-l) figure-ground separation. For calibration of the segmentation pipeline, the reference image with parameters $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ was used. Some important differences between KM and GC are highlighted with white arrows.

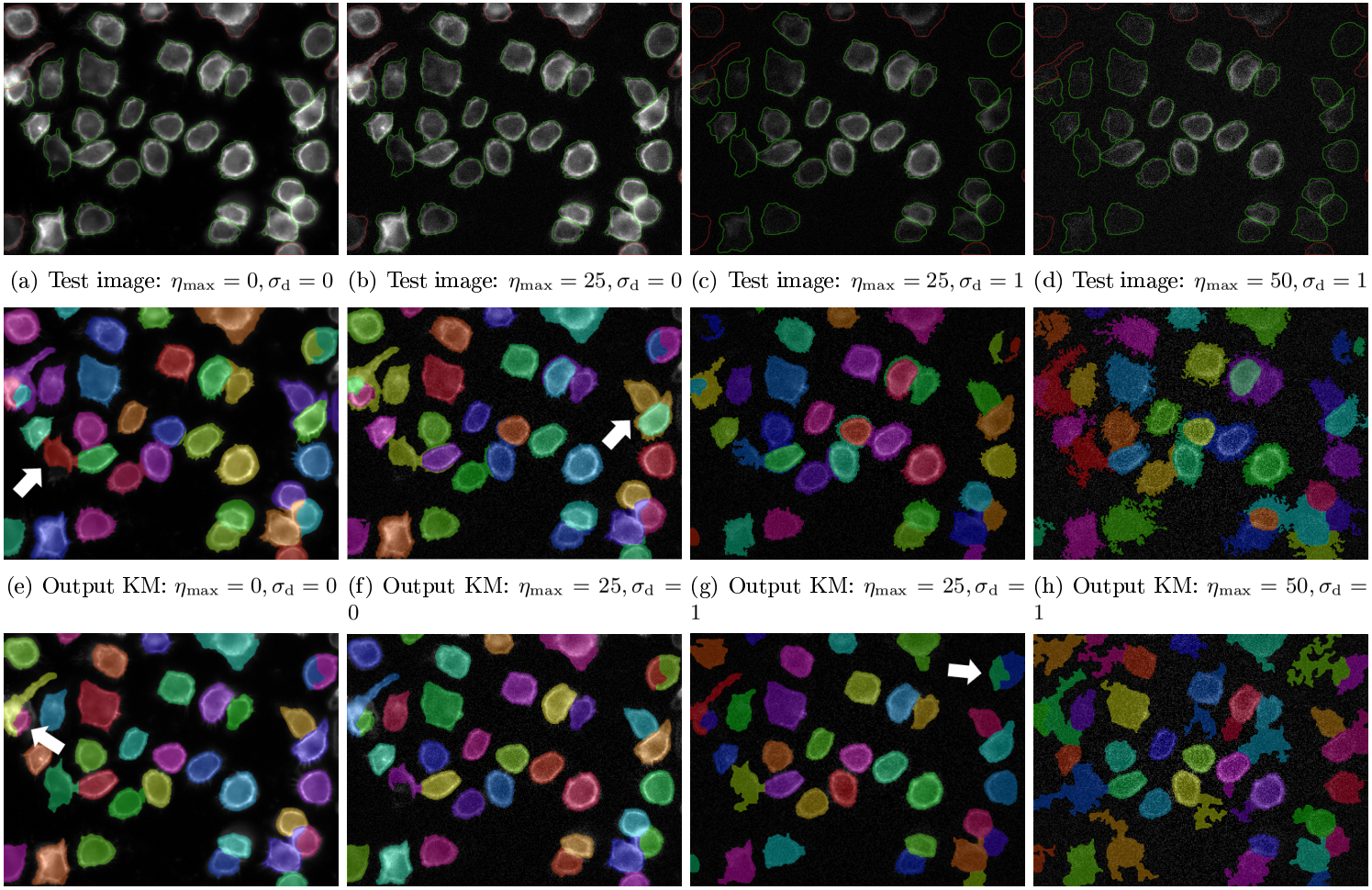


(a) Test image: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (b) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (c) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (d) Test image: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(e) Output KM: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (f) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (g) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (h) Output KM: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(i) Output GC: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (j) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (k) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (l) Output GC: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

Figure A.4.: Test images (a-d), output using KM (e-h) and output using GC (i-l) figure-ground separation. For calibration of the segmentation pipeline, the reference image with parameters $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ was used. Some important differences between KM and GC are highlighted with white arrows.



(a) Test image: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (b) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (c) Test image: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (d) Test image: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(e) Output KM: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (f) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (g) Output KM: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (h) Output KM: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

(i) Output GC: $\eta_{\max} = 0, \sigma_{\mathrm{d}} = 0$ (j) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 0$ (k) Output GC: $\eta_{\max} = 25, \sigma_{\mathrm{d}} = 1$ (l) Output GC: $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$

Figure A.5.: Test images (a-d), output using KM (e-h) and output using GC (i-l) figure-ground separation. For calibration of the segmentation pipeline, the reference image with parameters $\eta_{\max} = 50, \sigma_{\mathrm{d}} = 1$ was used. Some important differences between KM and GC are highlighted with white arrows.

Bibliography

M. Abràmoff, P. Magalhães, and S. Ram. Image processing with ImageJ. *Biophotonics international*, 11(7):36–42, 2004.

A. Allalou and C. Wählby. BlobFinder, a tool for fluorescence microscopy image cytometry. *Computer Methods and Programs in Biomedicine*, 94 (1):58–65, 2009.

D. Baggett, M. Nakaya, M. McAuliffe, T. P. Yamaguchi, and S. Lockett. Whole cell segmentation in solid tissue sections. *Cytometry Part A*, 67A (2):137–143, 2005.

Bengtsson, E. and C. Wahlby. Robust cell image segmentation methods. *Pattern Recognition and Image Analysis*, 14(2):157–167, 2004.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

S. Blanchoud, Y. Budirahardja, F. Naef, and P. Gönczy. ASSET: a robust algorithm for the automated segmentation and standardization of early caenorhabditis elegans embryos. *Developmental Dynamics*, 239(12):3285–3296, 2010.

Y. Boykov and G. Funka-Lea. Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, volume 2 of *ICCV '03*, page 26–33, Washington, DC, USA, 2003.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *EEE Transactions on Pattern Analysis and Machine Intelligence*, volume 1, page 377–384, 1999.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

L. Bradbury and J. W. L. Wan. A spectral k-means approach to bright-field cell image segmentation. *Conf Proc IEEE Eng Med Biol Soc*, 2010: 4748–4751, 2010.

D. D. Brüllmann, A. Pabst, K. M. Lehmann, T. Ziebart, M. O. Klein, and B. d'Hoedt. Counting touching cell nuclei using fast ellipse detection to assess in vitro cell characteristics: a feasibility study. *Clinical Oral Investigations*, 16(1):33–38, 2012.

H. Buie, G. Campbell, R. Klinck, J. MacNeil, and S. Boyd. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone*, 41(4): 505–515, 2007.

F. Bunyak, K. Palaniappan, S. K. Nath, T. L. Baskin, and G. Dong. Quantitative cell motility for in vitro wound healing using level set-based active contour tracking. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*, pages 1040–1043. IEEE, 2006.

A. Carpenter, T. Jones, M. Lamprecht, C. Clarke, I. Kang, O. Friman, D. Guertin, J. Chang, R. Lindquist, J. Moffat, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):R100, 2006.

X. Chen, X. Zhou, and S. T. Wong. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Transactions on Biomedical Engineering*, 53(4):762–766, 2006.

J. Cheng and J. Rajapakse. Segmentation of clustered nuclei with shape markers and marking function. *IEEE Transactions on Biomedical Engineering*, 56(3):741–748, 2009.

R. Chinta and M. Wasser. Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live drosophila embryos. *Cytometry. Part A: The Journal of the International Society for Analytical Cytology*, 81(1):52–64, 2011.

D. M. Chudakov, M. V. Matz, S. Lukyanov, and K. A. Lukyanov. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological reviews*, 90(3):1103–1163, 2010.

T. Collins et al. ImageJ for microscopy. *Biotechniques*, 43(1):25–30, 2007.

A. H. Coons, H. J. Creech, and R. N. Jones. Immunological properties of an antibody containing a fluorescent group. In *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, NY)*, volume 47, page 200–202, 1941.

H. Daims, S. Lücker, and M. Wagner. Daime, a novel image analysis program for microbial ecology and biofilm research. *Environmental microbiology*, 8 (2):200–213, 2006.

O. Danek, P. Matula, C. Ortiz-de Solorzano, A. Muoz-Barrutia, M. Maska, and M. Kozubek. Segmentation of touching cell nuclei using a two-stage graph cut model. In A.-B. Salberg, J. Y. Hardeberg, and R. Jenssen, editors, *Image Analysis*, volume 5575, pages 410–419. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

R. N. Day and M. W. Davidson. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.*, 38(10):2887–2921, 2009.

F. de Chaumont, S. Dallongeville, and J. C. Olivo-Marin. ICY: a new open-source community image processing software. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 234–237. IEEE, 2011.

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

O. Dzyubachyk, J. Essers, W. A. v. Cappellen, C. Baldeyron, A. Inagaki, W. J. Niessen, and E. Meijering. Automated analysis of time-lapse fluorescence microscopy images: from live cell images to intracellular foci. *Bioinformatics*, 26(19):2424 –2430, 2010a.

O. Dzyubachyk, W. A. van Cappellen, J. Essers, W. J. Niessen, and E. Meijering. Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *IEEE Transactions on Medical Imaging*, 29(3):852–867, 2010b.

I. Ersoy, F. Bunyak, V. Chagin, M. C. Cardoso, and K. Palaniappan. Segmentation and classification of cell cycle phases in fluorescence imaging. *Lecture notes in computer science*, 5762:617–624, 2009.

M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937.

I. C. Ghiran. Introduction to fluorescence microscopy. In H. Chiarini-Garcia and R. C. N. Melo, editors, *Light Microscopy*, volume 689, pages 93–136. Humana Press, Totowa, NJ, 2011.

D. E. Goldberg. Genetic algorithms in search, optimization, and machine learning. 1989.

I. Goldberg. Open microscopy environment. In *Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts. IEEE*, page 380, 2005.

D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society Series B Methodological*, 51(2):271–279, 1989.

N. Harder, F. Mora-Bermúdez, W. J. Godinez, J. Ellenberg, R. Eils, and K. Rohr. Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences. *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, 9(1):840–848, 2006.

N. Harder, F. Mora-Bermudez, W. J. Godinez, J. Ellenberg, R. Eils, and K. Rohr. Determination of mitotic delays in 3D fluorescence microscopy images of human cells using an error-correcting finite state machine. In *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1044–1047. IEEE, 2007.

J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *JR Stat. Soc., Ser. C*, 28:100–108, 1979.

C. Held, J. Wenzel, R. Webel, M. Marschall, R. Lang, R. Palmisano, and T. Wittenberg. Using multimodal information for the segmentation of fluorescent micrographs with application to virology and microbiology. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011:6487–6490, 2011.

C. Held, J. Wenzel, R. Lang, R. Palmisano, and T. Wittenberg. Comparison of methods for splitting of touching and overlapping macrophages in fluorescent micrographs. In *Proceedings of the 9th international conference on Image Analysis and Recognition - Volume Part II*, ICIAR'12, page 456–464, Berlin, Heidelberg, 2012. Springer-Verlag.

C. Held, T. Nattkemper, R. Palmisano, and T. Wittenberg. Approaches to automatic parameter fitting in a microscopy image segmentation pipeline: An exploratory parameter space analysis. *Journal of Pathology Informatics*, 4(2):5, 2013a.

C. Held, J. Wenzel, V. Wiesmann, R. Palmisano, R. Lang, and T. Wittenberg. Enhancing automated micrograph-based evaluation of LPS-stimulated macrophage spreading. *Cytometry Part A*, 83(4):409–418, Apr. 2013b.

E. Hodneland, N. V. Bukoreshtliev, T. W. Eichler, X.-C. Tai, S. Gurke, A. Lundervold, and H. H. Gerdes. A unified framework for automated 3-d segmentation of surface-stained living cells and a comprehensive segmentation evaluation. *IEEE Transactions on Medical Imaging*, 28(5):720–738, 2009.

M. R. Jeong, B. C. Ko, and J. Y. Nam. Overlapping nuclei segmentation based on bayesian networks and stepwise merging strategy. *Journal of Microscopy*, 235(2):188–198, 2009.

J. N. Kapur, P. K. Sahoo, and A. K. C. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing*, 29(3):273–285, 1985.

P. S. Karvelis, D. I. Fotiadis, I. Georgiou, and M. Syrrou. A watershed based segmentation method for multispectral chromosome images classification. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006. EMBS '06*, pages 3009–3012, 2006.

K. Kvilekval, D. Fedorov, B. Obara, A. Singh, and B. S. Manjunath. Bisque: a platform for bioimage analysis and management. *Bioinformatics*, 26(4): 544–552, 2009.

M. Lamprecht, D. Sabatini, and A. Carpenter. CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques*, 42(1): 71–75, 2007.

F. Leong, M. Brady, and J. McGee. Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of clinical pathology*, 56 (8):619, 2003.

G. Lin, U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roysam. A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry Part A*, 56A(1):23–36, 2003.

J. Lindblad, C. Wählby, E. Bengtsson, and A. Zaltsman. Image analysis for automatic segmentation of cytoplasms and classification of rac1 activation. *Cytometry Part A*, 57A(1):22–33, 2004.

C. Loyek, N. Rajpoot, M. Khan, and T. Nattkemper. BioIMAX: a web 2.0 approach for easy exploratory and collaborative access to multivariate bioimage data. *BMC bioinformatics*, 12(1):297, 2011.

Z. Q. Luo and P. Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72(1):7–35, 1992.

N. Malpica, C. O. de Solórzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. García-Sagredo, and F. del Pozo. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, 28(4):289–297, 1997.

F. Mech, A. Thywißen, R. Guthke, A. A. Brakhage, and M. T. Figge. Automated image analysis of the host-pathogen interaction between phagocytes and aspergillus fumigatus. *PLoS ONE*, 6(5):e19591, 2011.

S. G. Megason and S. E. Fraser. Imaging in systems biology. *Cell*, 130(5): 784–795, Sept. 2007.

V. Metzler, H. Bienert, T. Lehmann, K. Mottaghy, and K. Spitzer. A novel method for quantifying shape deformation applied to biocompatibility testing. *ASAIO J.*, 45(4):264–271, 1999.

V. Metzler, T. Lehmann, H. Bienert, K. Mottaghy, and K. Spitzer. Scale-independent shape analysis for quantitative cytology using mathematical morphology. *Computers in Biology and Medicine*, 30(3):135–151, 2000.

B. Möller, O. Greß, N. Stöhr, S. Hüttelmaier, and S. Posch. Adaptive segmentation of cells and particles in fluorescent microscope images. In P. Richard and J. Braz, editors, *VISAPP (2)*, pages 97–106. INSTICC Press, 2010.

K. Nandy, P. R. Gudla, K. J. Meaburn, T. Misteli, and S. J. Lockett. Automatic nuclei segmentation and spatial FISH analysis for cancer detection. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009*, pages 6718–6721. IEEE, 2009.

Nattkemper. A neural network architecture for automatic segmentation of fluorescence micrographs. *Neurocomputing*, 48(1-4):357–367, 2002.

T. W. Nattkemper, H. Wersing, W. Schubert, and H. Ritter. Fluorescence micrograph segmentation by gestalt-based feature binding. In *IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000*, volume 1, pages 348–353 vol.1. IEEE, 2000.

B. Nilsson and A. Heyden. A fast algorithm for level set-like active contours. *Pattern Recognition Letters*, 24(9-10):1331–1337, 2003.

C. Ortiz De Solorzano, R. Malladi, S. A. Lelièvre, and S. J. Lockett. Segmentation of nuclei and cells using membrane related protein markers. *Journal of Microscopy*, 201(3):404–415, 2001.

N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11:285–296, 1975.

M. Palmieri, C. J. Nowell, M. Condron, J. Gardiner, A. B. Holmes, J. Desai, A. W. Burgess, and B. Catimel. Analysis of cellular phosphatidylinositol (3,4,5)-trisphosphate levels and distribution using confocal fluorescent microscopy. *Analytical Biochemistry*, 406(1):41–50, 2010.

H. Peng. Bioimage informatics: A new area of engineering biology. *Bioinformatics*, 24(17):1827–1836, 2008.

T. D. Pham, D. I. Crane, T. H. Tran, and T. H. Nguyen. Extraction of fluorescent cell puncta by adaptive fuzzy segmentation. *Bioinformatics*, 20(14):2189 –2196, 2004.

A. J. Pretorius, M.-A. Bray, A. E. Carpenter, and R. A. Ruddle. Visualization of parameter space for image analysis. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2402–2411, 2011.

X. Qian, C. Peng, X. Wang, and D. Ye. Self-organizing map-based multithresholding on neural stem cells images. *Medical & Biological Engineering & Computing*, 47:801–808, 2009.

L. Qu, F. Long, X. Liu, S. Kim, E. Myers, and H. Peng. Simultaneous recognition and segmentation of cells: application in c.elegans. *Bioinformatics (Oxford, England)*, 27(20):2895–2902, 2011.

H. Robenek. *Mikroskopie in Forschung und Praxis*. Git Verlag, Nov. 1999.

J. Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Mathematical morphology*, 41(1-2): 187, 2000.

F. W. D. Rost. *Fluorescence Microscopy*. Cambridge University Press, July 1995. ISBN 9780521410885.

J. A. Sethian. *Level set methods and fast marching methods*. Cambridge University Press, 1999.

G. Srinivasa, M. Fickus, and J. Kovaevi. Active contour-based multiresolution transforms for the segmentation of fluorescence microscope images. In *Wavelets XII*. SPIE, 2007.

G. Srinivasa, M. Fickus, and J. Kovačević. Multiresolution multiscale active mask segmentation of fluorescence microscope images. *Proceedings - Society of Photo-Optical Instrumentation Engineers*, 7446(744603):744603, 2009.

D. Tomaževič, B. Likar, and F. Pernuš. Comparative evaluation of retrospective shading correction methods. *Journal of Microscopy*, 208(3):212–223, 2002.

M. Van Herk. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. *Pattern Recognition Letters*, 13(7): 517–521, 1992.

F. R. Velasco. Thresholding using the ISODATA clustering algorithm. *IEEE Transactions on Systems Man and Cybernetics*, 10:771–774, 1980.

M. Wall. *GAlib: A C++ library of genetic algorithm components.* 1996.

M. Wang, X. Zhou, F. Li, J. Huckins, R. W. King, and S. T. Wong. Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics*, 24(1):94 –101, 2008.

Q. Wang, J. Niemi, C. M. Tan, L. You, and M. West. Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry Part A*, 77A(1):101–110, 2010a.

X. Wang, B. Zheng, R. R. Zhang, S. Li, X. Chen, J. J. Mulvihill, X. Lu, H. Pang, and H. Liu. Automated analysis of fluorescent in situ hybridization (FISH) labeled genetic biomarkers in assisting cervical cancer diagnosis. *Technology in cancer research & treatment*, 9(3):231–242, 2010b.

C. Wählby, J. Lindblad, M. Vondrus, E. Bengtsson, and L. Björkesten. Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Analytical Cellular Pathology: The Journal of the European Society for Analytical Cellular Pathology*, 24(2-3):101–111, 2002.

C. Wählby, I. m. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *Journal of Microscopy*, 215(1):67–76, 2004.

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

T. Wittenberg, F. Becher, M. Hensel, and D. G. Steckhan. Image segmentation of cell nuclei based on classification in the color space. *IFMBE Proceedings*, 22(7):613–616, 2009.

P. Yan, X. Zhou, M. Shah, and S. T. C. Wong. Automatic segmentation of high-throughput RNAi fluorescent cellular images. *IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society*, 12(1):109–117, 2008.

W. Yu, H. K. Lee, S. Hariharan, W. Bu, and S. Ahmed. Quantitative neurite outgrowth measurement based on image segmentation with topological dependence. *Cytometry. Part A: The Journal of the International Society for Analytical Cytology*, 75(4):289–297, 2009.

W. Yu, H. K. Lee, S. Hariharan, W. Bu, and S. Ahmed. Evolving generalized voronoi diagrams for accurate cellular image segmentation. *Cytometry Part A*, 77A(4):379–386, 2010.

A. Zaritsky, S. Natan, J. Horev, I. Hecht, L. Wolf, E. Ben-Jacob, and I. Tsarfaty. Cell motility dynamics: a novel segmentation algorithm to quantify multi-cellular bright field microscopy images. *PloS One*, 6(11):e27593, 2011.

G. Zhang, X. Jia, T. D. Pham, and D. I. Crane. Multistage spatial property based segmentation for quantification of fluorescence distribution in cells. *AIP Conference Proceedings*, 1210(1):3–12, 2010a.

Z. Zhang, Y. Ge, D. Zhang, and X. Zhou. High-content analysis in monastrol suppressor screens. *Methods of Information in Medicine*, 50:265–272, 2010b.

D. Zhu, S. Jarmin, A. Ribeiro, F. Prin, S. Q. Xie, K. Sullivan, J. Briscoe, A. P. Gould, F. M. Marelli-Berg, and Y. Gu. Applying an adaptive watershed to the tissue cell quantification during t-cell migration and embryonic development. *Methods in Molecular Biology (Clifton, N.J.)*, 616:207–228, 2010.